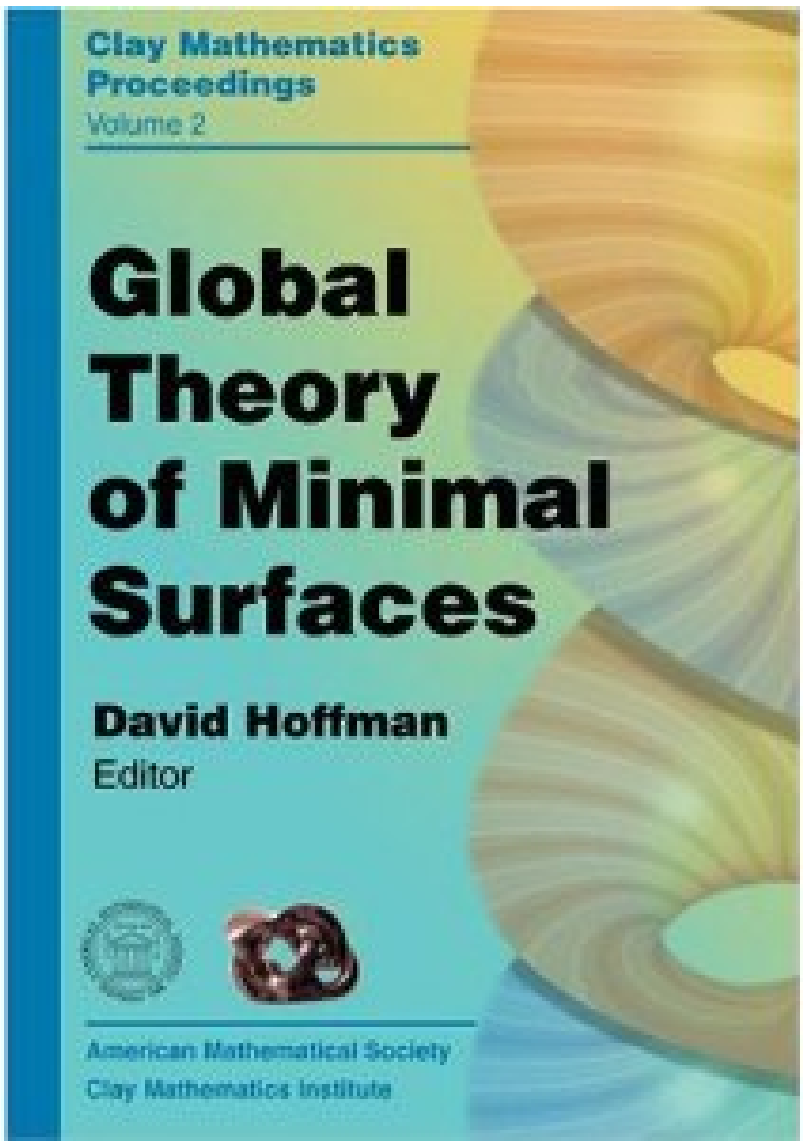

Clay Mathematics Proceedings
Volume 2
Global Theory of Minimal Surfaces
David Hoffman
Editor
American Mathematical Society
Clay Mathematics Institute

Clay Mathematics Proceedings
Volume 2, 2004

Contents

Preface, *David Hoffman* vii

Geometric Measure Theory and the Proof of the Double Bubble Conjecture, *Frank Morgan and Manuel Ritoré* 1

Classical Minimal Surfaces in Euclidean Space by Examples: Geometric and Computational Aspects of the Weierstrass Representation, *Matthias Weber* 19

Computational Aspects of Discrete Minimal Surfaces, *Konrad Polthier* 65

Mean Curvature in Riemannian Geometry and General Relativity, *Richard Schoen* 113

Introduction to Conjugate Plateau Constructions, *Hermann Karcher* 137

Parabolicity and Minimal Surfaces, *Joaquín Pérez* 163

The Isoperimetric Problem, *Antonio Ros* 175

Flat Structures, Teichmüller Theory and Handle Addition for Minimal Surfaces, *Michael Wolf* 211

The Genus-One Helicoid as a Limit of Screw-Motion Invariant Helicoids with Handles, *Matthias Weber* 243

Computing Minimal Surfaces, *David Hoffman* 259

Geometric Aspects of the Theory of Fully Nonlinear Elliptic Equations, *Joel Spruck* 283

Hyperbolic Surfaces of Constant Mean Curvature One with Compact Fundamental Domains, *Hermann Karcher* 311

Isoperimetric Inequalities of Minimal Submanifolds, *Jaigyoung Choe* 325

Complete Nonorientable Minimal Surfaces in $\mathbb{R}^3$, *Francisco Martín* 371

Some Picard-Type Results for Properly Immersed Minimal Surfaces in $\mathbb{R}^3$, *Francisco J. López* 381

©2004 Clay Mathematics Institute

Optimal Isoperimetric Inequalities for Three-Dimensional Cartan–Hadamard Manifolds, *Manuel Ritoré* 395

Embedded Minimal Disks, *Tobias H. Colding and William P. Minicozzi II* 405

Construction of Minimal Surfaces by Gluing Weierstrass Representations, *Martin Traizet* 439

Global Problems in Classical Minimal Surface Theory, *William H. Meeks III* 453

Minimal Surfaces of Finite Topology, *William H. Meeks III and Harold Rosenberg* 471

Constructions of Minimal Surfaces by Gluing Minimal Immersions, *Nikolaos Kapouleas* 489

The Conformal Theory of Alexandrov Embedded Constant Mean Curvature Surfaces in R^3, *Rafe Mazzeo, Frank Pacard and Daniel Pollack* 525

Constructing Mean Curvature 1 Surfaces in H^3 with Irregular Ends, *Wayne Rossman, Masaaki Umehara and Kotaro Yamada* 561

Conformal Structures and Necksizes of Embedded Constant Mean Curvature Surfaces, *Rob Kusner* 585

Uniqueness of the Riemann Minimal Surfaces, *Joaquín Pérez* 597

The Mathematical Protein Folding Problem, *Yi Fang* 611

Minimal and CMC Surfaces Obtained by Ribaucour Transformations, *Keti Tenenblat* 623

Meromorphic Data for Surfaces of Mean Curvature One in Hyperbolic Space, II, *Ricardo Sa Earp and Eric Toubiana* 635

Special Lagrangian Submanifolds, *Richard Schoen* 655

Lectures on Special Lagrangian Geometry, *Dominic Joyce* 667

Variational Problems in Lagrangian Geometry: $\mathbb{Z}_2$-Currents, *Jon Wolfson* 697

Minimal Surfaces and the Topology of Three-Manifolds, *Joel Hass* 705

Minimal Surfaces in Geometric 3-Manifolds, *J. Hyam Rubinstein* 725

Cousins of Constant Mean Curvature Surfaces, *Karsten Große-Brauckmann* 747

An Approach to the Willmore Conjecture, *Peter Topping* 769

Minimal Surfaces and Harmonic Maps into Singular Geometry, *Chikako Mese* 773

Shortest Networks in 2 and 3 Dimensions, *J. Hyam Rubinstein* 783

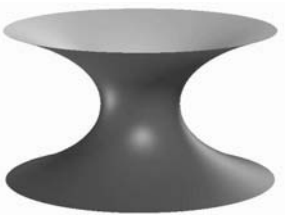
List of Participants 791

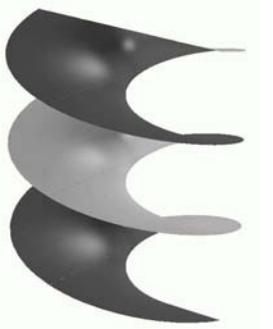
Clay Mathematics Proceedings
Volume 2, 2004

Preface

David Hoffman

For five weeks in the summer of 2001, the Clay Mathematics Institute organized a summer school on the Global Theory of Minimal Surfaces at the Mathematical Sciences Research Institute. MSRI became the world center for the study of minimal surfaces, with 150 mathematicians — undergraduates, young researchers, and most of the world's experts — participating in what was probably the most extensive meeting ever held on the subject in its 250-year history.





In 1744, Euler posed and solved the problem of finding the surfaces of rotation that minimize area. The only such surface is the catenoid. Some eleven years later, in a series of letters to Euler, Lagrange, at age 19, discussed the problem of finding a graph over a region in the plane, with prescribed boundary values, that was a critical point for the area function. He wrote down what would now be called the Euler–Lagrange equation for the solution (a second-order, nonlinear and elliptic equation), but he did not provide any new solutions. In 1776, Meusnier showed that the helicoid was also a solution. Equally important, he gave a geometric interpretation of the Euler–Lagrange equation in terms of the vanishing of the average of the principal curvatures of the surface, a quantity now known, after a suggestion of Sophie Germain, as the mean curvature. (This short outline of the earliest history of the subject is based on the introduction to J. C. C. Nitsche's *Lectures on Minimal Surfaces*, published by Cambridge University Press.)

A *minimal surface* is one whose mean curvature is everywhere zero. It has the defining property that every sufficiently small piece of it (small enough, say, to be a graph over some plane) is the surface of least area among all surfaces with the same boundary. The simplest minimal surface is the flat plane, but other minimal

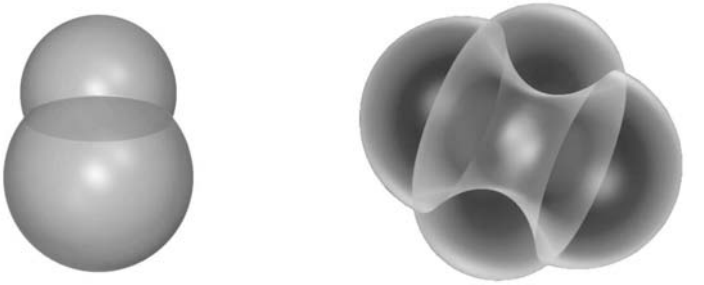
©2004 Clay Mathematics Institute

surfaces are far from simple. They are realized in the physical world by soap films spanning closed curves, and they appear as interfaces where the pressure is the same on either side. Finding a surface of least area spanning a given contour is known as the Plateau Problem, after the nineteenth-century Belgian physicist Felix Plateau, who posed the problem of finding a mathematical description of these solutions (and, by implication, a proof that they exist).

Closely related to minimal surfaces are *CMC-surfaces*: those whose mean curvature is constant but not zero. A CMC-surface minimizes area subject to a volume constraint. The most famous example is the round sphere, which minimizes surface area subject to the constraint of enclosing a fixed volume. They appear as soap bubbles and, more generally, as interfaces when there is a pressure difference from one side to the other.

The Clay Summer School, of which I was the Director, was organized by a committee of researchers whose expertise spans the subject and its applications: Joel Hass, UC Davis; Arthur Jaffe (ex officio), Harvard and CMI; Rick Schoen, Stanford; Antonio Ros, Granada; Harold Rosenberg, Paris VII; Mike Wolf, Rice. The first two weeks of the Summer School also functioned as MSRI's Summer Graduate Program on Minimal Surfaces, coordinated by Joel Hass. Over forty students from MSRI's sponsoring institutions joined sixty Clay Summer School participants, lecturers and mentors for these two weeks. It was an advanced-graduate introduction to the theory of minimal surfaces. The Clay student participants were chosen competitively from candidates nominated by senior mathematicians. Joining them were the lecturers, session assistants — all working researchers in the subject — and other senior mathematicians who attended the lecture series, served as mentors and gave special colloquia.

The featured event of the first two weeks of the meeting was Frank Morgan's lecture course on *Geometric Measure Theory and the Proof of the Double Bubble Conjecture*. The conjecture says that the least-area way to enclose and separate two given volumes of air is the familiar standard double soap bubble shown on



the left, rather than some more exotic version such as the one on the right. Several years ago, Joel Hass and Roger Schlafly proved the conjecture when the two volumes are assumed to be the same. They used some important theoretical observations, including a key idea of Michael Hutchings, followed by a novel proof that employed interval programming. Hutchings, Morgan, Manuel Ritore and Antonio Ros proved the Double Bubble Conjecture for arbitrary volumes. Geometric measure theory was key to their work, as one is forced to consider surfaces with singular sets of an unpredictable nature.

Morgan's lectures (page 1) discuss modern, measure-theoretic definitions of surface, compactness of spaces of surfaces, and finally the proof of the Double Bubble Conjecture. The notes were prepared by Gary Lawler

Complementing Morgan's geometric measure theory were the innovative lectures of Matthias Weber (page 19), who spoke on the computational aspects of minimal surface theory. The course doubled as a high-level introduction to the classical representation of minimal surfaces using complex function theory and conformal mapping. The software he used was a combination of his own well developed Mathematica notebooks and a new Java version of MESH, a software suite developed by Jim Hoffman of MSRI. Weber's course and computing assignments were complemented by lectures by John McCuan, who covered the basics of the subject. Special topics such as the conjugate surface construction, Schwarz reflection, associate families and periodic minimal surfaces were discussed in companion lectures by Hermann Karcher (p. 137), Harold Rosenberg (p. 471), Bill Meeks (p. 453) and myself (p. 259).

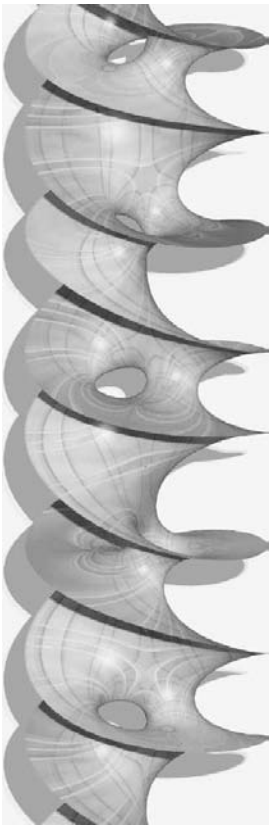
Rick Schoen (p. 113) gave a three-lecture course on the applications of mean curvature to general relativity and Riemannian geometry, including the Penrose Inequality and a description of the Huisken–Ilmanen and Bray method of solution. Konrad Polthier (U. Berlin) lectured about discrete minimal surfaces (p. 65). The idea of a discrete minimal surface (Pinkall and Polthier) makes sense of what a harmonic map should be when the domain of definition is a polygonal surface.

The most striking thing about the first two weeks of the CMI Summer School was the enormous amount of interaction between established mathematicians, graduate students and post docs. In addition to scheduled help sessions, voluntary group meetings were held after the day's lectures were done. Frequently, two or more groups of five to ten students met at the blackboards on the second and third floors with two or three senior mathematicians in attendance. These meetings often lasted until 6:30PM or later. Hermann Karcher had a series of meetings with a small group of graduate students that took place in various locations of MSRI. One of them, held on the MSRI patio, was so intense that the members were oblivious to call for participation in the group photograph, which had to be retaken after we found out that this fanatic study group had missed it. The response from students was extremely positive. Several students stayed longer than they had originally planned because of the excitement of the meeting and the intensity of the interaction with senior mathematicians.

The research component of the School ran from July 9 to July 27. In the first week, Antonio Ros gave a series of lectures (p. 175) about the isoperimetric problem, a subject intimately tied the study of constant-mean curvature in Riemannian manifolds. The emphasis was on methods: symmetrization, suitable in spaces with many symmetries; stability, most useful in negatively curved spaces; and comparison with model situations, including novel comparisons to Gaussian measure models. Ritoré gave the last lecture of this series (p. 395), on optimal inequalities in Hadamard–Cartan manifolds. In that week, Joel Spruck and Bo Guan gave a series of lectures (p. 283) about the geometry of fully nonlinear elliptic equations, the prototype of which is the Monge-Ampere equation.

Prominently discussed were new methods of constructing minimal surfaces, both from the point of view of PDEs and the perspective of the Weierstrass representation and complex function theory. Nicos Kapouleas (p. 489) and (separately)

Rafe Mazzeo and Frank Pacard (p. 525) outlined their work on what could be termed gluing together embedded minimal surfaces to construct new examples, and desingularizing the intersection of embedded minimal surfaces of finite total curvature. Pascal Collin and Karsten Große-Brauckmann (p. 747) described various existence and structure results for constant mean curvature surface in Euclidean and hyperbolic space.



Martin Traizet (p. 439) described an elegant theory of deformation families of embedded minimal surfaces of finite total curvature in flat three manifolds and their singular limit configurations. He used this technique to produce a complete embedded minimal surface of finite total curvature in $\mathbb{R}^3$ with no nontrivial symmetries. All previous examples, ever since Euler discovered the catenoid in 1744, had been found by assuming symmetry in order to simplify the analysis. Traizet's announcement of this result was received with a round of applause.

Matthias Weber and Mike Wolf presented their theory of flat structures (p. 211), which allows handle addition to minimal surfaces. Other methods from Teichmüller theory are used to develop a description of the moduli space of examples and to solve the period problem for minimal surfaces in a geometric fashion. A strong application of these new techniques is the proof of embeddedness of a genus one helicoid as the limit of a deformation family of screw-motion-invariant surfaces (such as the one shown here). This surface is the first properly embedded minimal surface with infinite total curvature to be found since Meusnier showed that the helicoid was minimal in 1776.

Bill Minicozzi gave five lectures (p. 405) in which he developed his theory of embedded minimal surfaces of finite topology in a ball of $\mathbb{R}^3$. This represents joint research with Toby Colding. Primarily, he discussed the case when M is a disk. The key analysis concerns behavior of M near points where the curvature becomes large; a multi-graph forms near such points, very much like a piece of a helicoid near its axis. Minicozzi explained the different scales on which one sees this multigraph, and from this perspective there is emerging a global picture of the geometry of M. The main tools are the stability equation and the conformal geometry of minimal surfaces. This work leads to their theory of limit minimal laminations for blow-downs of uniformly locally simply connected minimal surfaces.

The Summer School showed that the study of minimal surfaces is in a period of dramatic progress. Example: Harold Rosenberg and Bill Meeks proved in their lectures (p. 471) that the helicoid is the only non-planar properly embedded and simply connected minimal surface. No assumption is made about the conformal type of the surface or the behavior of the Gauss normal map at infinity. The proof involves a marvelous combination of geometric and analytic techniques; a key step makes essential use of the new analytic results of Colding and Minicozzi.

The lectures of Joyce (p. 667), Schoen (p. 655), and Wolfson (p. 697) discussed the mathematics of lagrangian and special lagrangian geometry. The variational theory of the volume functional among lagrangian competitors was presented as an approach to constructing minimal lagrangian and special lagrangian cycles in Kähler–Einstein manifolds.

Joel Hass (p. 705) and Hyam Rubinstein (p. 725) lectured on the many applications of minimal surfaces to the study of three-manifolds. After Hass gave a survey of some of the spectacular achievements in this area over the past 25 years, Rubinstein discussed some recent investigations into Heegaard surfaces and hyperbolic manifolds using the study of minimal surfaces in these manifolds.

On July 22, we boarded buses in the morning for an all-day trip to Monterey Bay, which included a stop for lunch and a swim in Carmel, a walking tour of Point Lobos State Preserve, and dinner in the Monterey Bay Aquarium, which we had all to ourselves. Our dining room was in a room one of whose walls (perhaps 80 by 30 feet) was the transparent side of an enormous tank filled with the denizens of Monterey Bay, including turtles, tuna, and many large sharks.

All who attended are grateful to the Clay Mathematics Institute for organizing this Summer School, and to MSRI and its staff for hosting it. It was a memorable meeting that will have a lasting and positive effect on the subject. The organizers of the meeting who are the editors of these proceedings hope that this volume conveys the content and excitement of the Summer School.

The creation of these proceedings was greatly aided by the editorial and TeX talents of Silvio Levy, MSRI's librarian and editor. We are grateful for his aid and good advice.

Berkeley, Summer 2003
David Hoffman,
On behalf of the organizing committee:

Joel Hass, UC Davis
David Hoffman, MSRI
Arthur Jaffe, Harvard University and CMI
Antonio Ros, Universidade de Granada
Harold Rosenberg, Université Paris VII
Rick Schoen, Stanford University
Mike Wolf, Rice University

Note: The lectures for the entire Summer School can be seen on the Web, at
`www.msri.org/publications/video/index02.html`.
The software JMEsh can be found at
`www.msri.org/publications/sgp/jim/software/jmesh/indexc.html`
and an archive of pre-computed minimal surfaces is located at
`www.msri.org/publications/sgp/jim/software/jmesh/surfacearchive/indexc.html`.
The Clay Mathematics Institute website is: `www.claymath.org`.

Clay Mathematics Proceedings
Volume 2, 2004

Geometric Measure Theory and the Proof of the Double Bubble Conjecture

Frank Morgan and Manuel Ritoré

Abstract. These notes by Ritoré are based on a nine-hour course given by Morgan in June–July 2001 during the Clay Mathematics Institute Summer School at MSRI. They follow closely the text [**30**] and can be read as an introduction to it. These notes provide a very brief overview of geometric measure theory, with a few sketches of proofs and references to the appropriate sources.

1. Introduction

Geometric measure theory was developed in the second half of the twentieth century to treat existence and regularity questions in the calculus of variations.

The lack of compactness properties of the spaces of submanifolds made it necessary to consider "generalized" submanifolds (currents, varifolds, finite perimeter sets, and so on), with appropriate topologies, to apply direct methods from the calculus of variations. To find a minimum of some functional, take a sequence approaching the infimum. If this sequence is compact in some topology, extract a convergent subsequence to a generalized submanifold realizing the minimum. Then prove a regularity result to recover a classical solution. In some sense, this theory is similar to Sobolev space theory, in which one enlarges the space of smooth functions to find a solution to a differential equation, and then proves regularity to recover a classical solution.

These notes provide a brief overview of geometric measure theory. A few sketches of proofs will be given, with references to appropriate sources. Recommended texts include Morgan [**30**], Simon [**41**], Giusti [**18**], Massari–Miranda [**26**], Ziemer [**47**] on finite perimeter sets and functions of bounded variation, and of course Federer's treatise [**12**] titled *Geometric Measure Theory.*

2. Rectifiable Sets

2.1. Lebesgue and Hausdorff measures [**30**, Ch. 2]**.** Lebesgue measure in $\mathbb{R}^n$ will be denoted by $\mathcal{L}^n$. An important outer measure in $\mathbb{R}^n$, defined for any nonnegative integer m, is Hausdorff measure $\mathcal{H}^m$. Let $\alpha_m = \mathcal{L}^m(\mathbf{B}(0,1))$, and

©2004 Clay Mathematics Institute

$A \subset \mathbb{R}^n$. Fix $\delta > 0$ and consider

$$\mathcal{H}^m_\delta(A) = \inf\left\{ \sum_j \alpha_m r_j^m : A \subset \bigcup_j S_j,\ \operatorname{diam} S_j = 2r_j \leqslant \delta \right\}.$$

Then define the m-dimensional Hausdorff measure of A as

$$\mathcal{H}^m(A) = \lim_{\delta \to 0} \mathcal{H}^m_\delta(A).$$

Note that this limit always exists, although it could be ∞. We recall that Hausdorff measure is just an outer measure, but restricted to the σ-algebra of measurable sets is a Borel-regular measure by Carathéodory's criterion. In $\mathbb{R}^n$, we have $\mathcal{L}^n = \mathcal{H}^n = \mathcal{H}^n_\delta$, for any $\delta > 0$. One may extend the definition of Hausdorff measure to any real dimension, by replacing α_m by the corresponding expression in terms of gamma functions ($\alpha_m = \pi^{m/2}/\Gamma(m/2+1)$). $\mathcal{H}^0$ is just counting measure.

If $m < p$, then $\mathcal{H}^m(A) < \infty$ implies $\mathcal{H}^p(A) = 0$. If $p < m$, then $\mathcal{H}^m(A) > 0$ implies $\mathcal{H}^p(A) = \infty$. The Hausdorff dimension of $A \subset \mathbb{R}^n$ is defined as the real number

$$\begin{aligned}\inf\{p : \mathcal{H}^p(A) < \infty\} &= \inf\{p : \mathcal{H}^p(A) = 0\} \\ &= \sup\{p : \mathcal{H}^p(A) > 0\} = \sup\{p : \mathcal{H}^p(A) = \infty\}.\end{aligned}$$

The m-dimensional Hausdorff measure of an m-dimensional submanifold of $\mathbb{R}^n$ coincides with the Riemannian volume of the submanifold.

2.2. Densities [30, 2.5]. If $A \subset \mathbb{R}^n$, one can define the m-dimensional density of A at a by

$$\Theta(A, a) = \lim_{r \to 0} \frac{\mathcal{H}^m(A \cap \mathbf{B}^n(a, r))}{\alpha_m r^m},$$

when this limit exists. In the formula, $\mathbf{B}^n(a, r)$ is the closed ball of center a and radius $r > 0$ in $\mathbb{R}^n$. Similarly, one can define the density of A with respect to a measure μ in $\mathbb{R}^n$ by replacing $\mathcal{H}^m$ by μ in the previous formula.

The m-dimensional density of any smooth m-dimensional submanifold of $\mathbb{R}^n$ equals 1 at any of its points. If A is the boundary of a cube in $\mathbb{R}^3$, the 2-dimensional density is 1 at every point except at the vertices, where it is $\frac{3}{4}$.

2.3. Lipschitz functions [30, Ch. 3]. A Lipschitz function f between metric spaces is one which satisfies the inequality $d(f(x), f(y)) \leqslant C\, d(x, y)$ for some constant $C > 0$. The smallest such C is called the Lipschitz constant of f. We can say that Lipschitz functions are those with controlled slope. A Lipschitz function between open sets of Euclidean spaces is differentiable almost everywhere by Rademacher's Theorem [30, 3.2]. Hence the Jacobian of a Lipschitz function can be integrated and we have the following important formulae.

THEOREM 2.1 (Area formula). *Consider a Lipschitz function $f : \mathbb{R}^m \to \mathbb{R}^n$ for $m \leqslant n$.*

(i) *If $A \subset \mathbb{R}^m$ is an $\mathcal{L}^m$-measurable set, then*

$$\int_A J(f)\, d\mathcal{H}^m = \int_{\mathbb{R}^n} \mathcal{H}^0(A \cap f^{-1}(y))\, d\mathcal{H}^m(y).$$

(ii) *If $u : A \to \mathbb{R}$ is an $\mathcal{L}^m$-integrable function, then*

$$\int_A u\, J(f)\, d\mathcal{H}^m = \int_{\mathbb{R}^n} \sum_{x \in f^{-1}(y)} u(x)\, d\mathcal{H}^m(y).$$

We have denoted the Jacobian of f by $J(f)$. This formula shows that Hausdorff measure is compatible with classical mapping area.

THEOREM 2.2 (Coarea formula for Lipschitz maps). *Consider a Lipschitz function $f : \mathbb{R}^m \to \mathbb{R}^n$ with $m > n$. If $A \subset \mathbb{R}^m$ is an $\mathcal{L}^m$-measurable set, then*

$$\int_A J(f)\, d\mathcal{H}^m = \int_{\mathbb{R}^n} \mathcal{H}^{m-n}(A \cap f^{-1}(y))\, d\mathcal{H}^n(y).$$

The coarea formula is very useful in geometric analysis since it allows slicing to recover important geometric quantities. As we shall show, it can be generalized to maps between rectifiable sets.

2.4. Tangent cones [30, 3.9]. For a set $A \subset \mathbb{R}^n$, one may consider the tangent cone of A at a, $\mathrm{Tan}(A, a)$, as the cone over the subset S of the unit sphere given by

$$S = \bigcap_{\varepsilon > 0} \mathrm{Cl}\left\{ \frac{x-a}{|x-a|} : x \in A, 0 < |x-a| < \varepsilon \right\}.$$

For a C^1 m-dimensional submanifold the tangent cone is precisely the tangent m-plane. We define a smaller approximate tangent cone by

$$\mathrm{Tan}^m(A, a) = \bigcap \{ \mathrm{Tan}(S, a) : \Theta^m(A - S, a) = 0 \},$$

which omits parts of $\mathrm{Tan}(A, a)$ with dimension less than m.

2.5. Rectifiable sets [30, 3.10], [41, Ch. 3]. A set $A \subset \mathbb{R}^n$ is called $(\mathcal{H}^m, m)$-rectifiable set if $\mathcal{H}^m(A)$ is finite and $\mathcal{H}^m$-almost all of A is the union of the images of countably many Lipschitz maps from $\mathbb{R}^m$ to $\mathbb{R}^n$. We shall often refer to such sets as m-rectifiable sets.

In some sense, a rectifiable set behaves like a C^1 submanifold of $\mathbb{R}^n$ since it has an approximate tangent plane almost everywhere.

PROPOSITION 2.3. *If A is an m-rectifiable set in $\mathbb{R}^n$, then for $\mathcal{H}^m$-almost all points $a \in A$, we have $\Theta(A, a) = 1$ and $\mathrm{Tan}^m(A, a)$ is an m-dimensional plane.*

On rectifiable sets, one can define the notion of the gradient of a function, the divergence of a vector field, and the Jacobian of a map. Hence there is a generalization of the coarea formula to such sets.

THEOREM 2.4 (General coarea formula). *Consider an m-dimensional rectifiable set $A \subset \mathbb{R}^n$, a μ-dimensional rectifiable set $B \subset \mathbb{R}^\nu$, $m \geqslant n \geqslant 1$, and a Lipschitz function $f : A \to B$. Then*

$$\int_A J(f)\, d\mathcal{H}^m = \int_B \mathcal{H}^{m-\mu}(f^{-1}\{z\})\, d\mathcal{H}^\mu(z).$$

More generally, for any $\mathcal{H}^m$-integrable function g on A,

$$\int_A g\, J(f)\, d\mathcal{H}^m = \int_z \int_{f^{-1}\{z\}} g\, d\mathcal{H}^{m-\mu}\, d\mathcal{H}^\mu(z).$$

In the above result, $J(f)$ is the Jacobian of the Lipschitz map f, which is defined almost everywhere. An important consequence of the coarea formula is Crofton's formula:

COROLLARY 2.5 (Crofton's formula). *If $A \subset \mathbb{R}^n$ is an m-rectifiable set, then*

$$\mathcal{H}^m(A) = \frac{1}{\beta(n,m)} \int_{P \in \mathcal{P}} \#\{P \cap A\},$$

where $\beta(n,m)$ is a constant depending only on n and m, and $\mathcal{P}$ is the set of $(n-m)$ planes.

Finally we state the celebrated Federer Structure Theorem, which allows the decomposition of any set in $\mathbb{R}^n$ as the union of a rectifiable set and a set with negligible projection on almost all m-dimensional subspaces.

THEOREM 2.6 (Federer's structure theorem). *Let $A \subset \mathbb{R}^n$ be an arbitrary subset with $\mathcal{H}^m(A) < \infty$. Then A can be decomposed as the union of two disjoint sets $A = B \cup C$, where B is $(\mathcal{H}^m, m)$-rectifiable and $\mathcal{J}^m(C) = 0$.*

In this theorem $\mathcal{J}^m$ is the m-dimensional integral-geometric measure. Equality $\mathcal{J}^m(C) = 0$ implies that the m-dimensional measure of the orthogonal projection of C onto an m-dimensional linear subspace is zero for almost all subspaces.

One can define an orientation on a rectifiable set, just by choosing a measurable orientation in the set of tangent planes. Recall that there is a tangent plane for almost all the points of the set. Of course one can define an uncountable number of such orientations.

3. Currents

3.1. Definition and properties [30, Ch. 4]. The concept of current was introduced by Federer and Fleming [**14**] in 1960. Consider the space $\mathcal{D}^m$ of smooth differential m-forms with compact support in $\mathbb{R}^n$, endowed with the topology of C^∞ convergence on compact subsets. The space of continuous linear maps $T : \mathcal{D}^m \to \mathbb{R}$ will be denoted by $\mathcal{D}_m$. A *current* is an element of $\mathcal{D}_m$.

An oriented m-rectifiable set $A \subset \mathbb{R}^n$ defines a current $T_A \in \mathcal{D}_m$ in the following way. For any $a \in A$ for which there is a tangent plane, let $\vec{S}(a)$ denote the m-vector associated to the oriented tangent plane, and let μ be an integer-valued function in $L^1(A)$. Then

$$T_A(\varphi) = \int_A \mu \left\langle \vec{S}, \varphi \right\rangle d\mathcal{H}^m.$$

The function $\mu(a)$ measures the multiplicity of the rectifiable set in the point a. A current induced by a *compact* oriented m-rectifiable set will be called a *rectifiable current*.

The *support* of a current T is the smallest closed set C such that, for any $\varphi \in \mathcal{D}^m$ with $\operatorname{spt}(\varphi) \cap C = \varnothing$, one has $T(\varphi) = 0$.

One can define the *boundary* of a current in the following way: if $T \in \mathcal{D}_m$ and $\varphi \in \mathcal{D}^{m-1}$, then

$$\partial T(\varphi) = T(d\varphi).$$

Property $d^2 = 0$ of the exterior derivative of forms implies that ∂T has no boundary. The boundary of a rectifiable current T need not be rectifiable. If this is the case, then we say that T is an *integral current*.

The *mass* of a current T is

$$\mathbf{M}(T) = \sup\{T(\varphi) : |\varphi| \leqslant 1\},$$

where $|\varphi|$ is the supremum norm of the differential form φ. If $T \in \mathcal{D}_m$ is a rectifiable current associated to a rectifiable set A with multiplicity $\mu \equiv 1$, then $\mathbf{M}(T) = \mathcal{H}^m(A)$.

A natural topology in the space of currents is the weak topology. We say that $T_i \to T$ weakly in $\mathcal{D}_m$ if and only if $T_i(\varphi) \to T(\varphi)$ for any differential form $\varphi \in \mathcal{D}^m$. If T_i converges weakly to T, then one obtains easily

$$\mathbf{M}(T) \leqslant \liminf_i \mathbf{M}(T_i).$$

Another seminorm in the space of currents is the *flat norm*. It is defined as

$$\mathcal{F}(T) = \inf\{\mathbf{M}(A) + \mathbf{M}(B) : T = A + \partial B, A \text{ } m\text{-rect.}, B \text{ } (m+1)\text{-rect.}\},$$

with $T \in \mathcal{D}_m$. If one has two m-currents T_1 and T_2, then saying that they are close in the flat norm is equivalent to saying that there is an $(m-1)$-current of small mass such that its boundary is T_1 plus T_2 plus another m-current of small mass.

The flat norm is weaker in general than the mass norm, and stronger than the weak topology.

Another classical definition is that of a *normal current*, which is a current T with compact support and $\mathbf{M}(T) + \mathbf{M}(\partial T) < \infty$.

3.2. The deformation theorem and consequences [30, Ch. 5]. One of the most interesting results concerning integral currents is the deformation theorem, which deforms a given integral current $T \in \mathcal{D}_m$ to a current contained in a given m-dimensional grid of prescribed mesh.

THEOREM 3.1 (Deformation theorem [**30**, 5.1]). *Let $T \in \mathcal{D}_m$ be an integral current in $\mathbb{R}^n$. Then there are integral currents $Q \in \mathcal{D}_m$, $S \in \mathcal{D}_{m+1}$, and a current $P \in \mathcal{D}_m$ such that*

$$T = P + Q + \partial S,$$

so that $\mathbf{M}(P) \leqslant \gamma\,\mathbf{M}(T)$, $\mathbf{M}(Q) \leqslant \varepsilon\,\mathbf{M}(\partial T)$ and $\mathbf{M}(S) \leqslant \varepsilon\gamma\,\mathbf{M}(T)$, where $\gamma = 2n^{2m+2}$, and hence $\mathcal{F}(T-P) \leqslant \varepsilon\gamma\,(\mathbf{M}(T) + \mathbf{M}(\partial T))$.

Moreover $\operatorname{spt}(T)$ is contained in a given m-dimensional 2ε grid, and $\operatorname{spt}(\partial T)$ is contained in the $(m-1)$-dimensional 2ε grid.

An important first consequence of the deformation theorem is

THEOREM 3.2 (Isoperimetric inequality [**30**, 5.3]). *An m-dimensional cycle T ($\partial T = 0$) in $\mathbb{R}^n$ bounds an $(m+1)$-dimensional integral current S with*

$$\mathbf{M}(S)^{m/(m+1)} \leqslant \gamma\,\mathbf{M}(T),$$

with $\gamma = 2n^{2m+2}$.

Almgren [**5**] improved this theorem by showing that the best constant γ one can place in the above inequality corresponds to the classical isoperimetric inequality in $\mathbb{R}^{m+1}$.

Another consequence of the deformation theorem is

THEOREM 3.3. *The set of integral currents T in $\mathbf{B}(0,R)$ such that $\mathbf{M}(T) \leqslant C$ and $\mathbf{M}(\partial T) \leqslant C$ is totally bounded in the flat norm topology.*

A consequence of this result is the equivalence of the weak topology and the flat norm topology on the subset of integral currents T in $\mathbf{B}(0,R)$ such that $\mathbf{M}(T) \leqslant C$, $\mathbf{M}(\partial T) \leqslant C$, for any constant $C > 0$.

The previous theorem is the first ingredient in the proof of

THEOREM 3.4 (Compactness theorem). *Given $R > 0$, $C > 0$, the set of integral currents with support in $\mathbf{B}(0,R)$ such that $\mathbf{M}(T) \leqslant C$ and $\mathbf{M}(\partial T) \leqslant C$ is compact in the flat norm topology.*

Now we are in position to apply these results to prove existence of area-minimizing surfaces for the Plateau problem.

COROLLARY 3.5. *Let B be an $(m-1)$-rectifiable current in $\mathbb{R}^n$ with $\partial B = 0$. Then there is an m-dimensional area-minimizing rectifiable current S with $\partial S = B$.*

PROOF. The proof is given in several steps.

(i) First show that B bounds some surface by taking a cone over B.
(ii) Take a sequence of rectifiable currents S_i with $\partial S_i = B$ approaching the infimum of the area.
(iii) Project the S_i into a ball containing B. Since the projection to the boundary of the ball is distance nonincreasing, the mass is reduced.
(iv) By the Compactness Theorem, get a limit S.
(v) By lower semicontinuity $\mathbf{M}(S) \leqslant \liminf_i \mathbf{M}(S_i)$, which shows that S is minimizing.

The limit we have obtained in this process is an integral current. □

The existence of this limit is not obvious since it is not clear that there is an area-minimizing surface among the infinitely many candidates.

An example of an area-minimizing surface is a minimal graph over a convex region. This follows from projection over the right cylinder over the convex region and a calibration argument.

3.3. Regularity theorems [30, Ch. 8]. We are interested now in seeing how regular is the solution to the Plateau problem we have obtained in the previous subsection.

In this direction, the first regularity result was proved by Fleming in 1962, who showed [**15**] that an area-minimizing rectifiable current in $\mathbb{R}^3$ is an embedded surface in the interior. This regularity theorem was generalized to three-dimensional surfaces in $\mathbb{R}^4$ by Almgren [**2**] in 1966, and to m-dimensional surfaces in $\mathbb{R}^{m+1}$, $m \leqslant 6$, by Simons [**42**] in 1968. In 1969, Bombieri, De Giorgi and Giusti [**7**] gave an example of a seven-dimensional, area-minimizing rectifiable current in $\mathbb{R}^8$ which is not smooth at the origin (the cone over $\mathbb{S}^3(1/\sqrt{2}) \times \mathbb{S}^3(1/\sqrt{2})$ in $\mathbb{R}^8$). The complete regularity result was given by Federer.

THEOREM 3.6. [**13**] *An $(n-1)$-dimensional, area-minimizing rectifiable current in $\mathbb{R}^n$ is a smooth, embedded manifold on the interior except for a singular set of Hausdorff dimension at most $n-8$.*

Similarly in a smooth Riemannian manifold M^n, for $n \leqslant 7$, an area-minimizing rectifiable current is a smooth embedded manifold and, for $n \geqslant 8$, the dimension of the singular set is less than or equal to $n-8$. The same result is true with a volume constraint (almost minimizing is enough). This was proved by Gonzalez,

Massari and Tamanini [**19**] for finite perimeter sets. We state the equivalent result for currents.

THEOREM 3.7. [**19**] *Let T be an $(n-1)$-dimensional rectifiable current of least area in the unit ball $B \subset \mathbb{R}^n$, with prescribed boundary in ∂B, bounding an oriented region of prescribed volume.*

Then T is a smooth hypersurface of constant mean curvature in its interior, except for a singular set of dimension at most $n-8$.

The result also holds in Riemannian manifolds (see [**31**]).

Regularity in higher codimension is much weaker. It was already proved by Federer [**12**, 5.3.16] that the set of regular points is dense in the interior. On the other hand, it was known that complex analytic varieties are area-minimizing, with singular sets of codimension 2. The most conclusive result was proved by Almgren.

THEOREM 3.8. [**4**],[**6**] *An m-dimensional, area-minimizing rectifiable current in $\mathbb{R}^n$ is a smooth, embedded manifold on the interior except for a singular set of Hausdorff dimension at most $m-2$.*

In 1988, Chang [**9**] showed that there are only isolated singularities for 2-dimensional currents in $\mathbb{R}^n$.

The only known examples in $\mathbb{R}^4$ are complex curves. It is known that the tangent cone must consist of complex planes (with multiplicities). The only known examples of 3-dimensional surfaces in Euclidean 5-space are the 2-dimensional examples in 4-space cross a line.

Why are regularity results stronger in codimension 1? First of all one can reduce to multiplicity 1, since a nesting lemma yields decomposition into nested, multiplicity 1 area-minimizing surfaces. One then proves that each one is regular. If they touch, they must coincide by the maximum principle. All the ingredients also hold in Riemannian manifolds.

There are several open questions concerning singularities. A first one is about the structure of singularities. A second one is if area-minimizing surfaces are stratified manifolds, or could they have fractional dimensional singular sets? Also, what are the possible tangent cones?

Concerning boundary regularity, the definitive result is the one by Hardt and Simon.

THEOREM 3.9. [**20**] *Let T be an $(n-1)$-dimensional, area-minimizing rectifiable current in $\mathbb{R}^n$, bounded by a C^2 oriented submanifold with multiplicity 1. Then at every boundary point,* $\operatorname{spt} T$ *is a C^1 embedded manifold with boundary.*

In a smooth Riemannian manifold M^3 with nonempty boundary ∂M, area-minimizing surfaces S in homology classes of M (mod. ∂M) are also regular along the boundary.

The main tool of regularity theory is the monotonicity in r of (mass inside r-ball)$/r^m$ [**30**, Ch. 9]. More precisely, for an m-dimensional rectifiable current T, we define the mass ratio $M(r)$ as $\mathbf{M}(T \llcorner \mathbf{B}(a,r))/\alpha_m r^m$, which is an increasing function. It is immediate that the density $\Theta(T,a)$ equals $\lim_{r\to 0} M(r)$. Then we have

THEOREM 3.10. *Let T be an area-minimizing locally rectifiable current, and let $a \in \operatorname{spt}(T)$. Then, for $0 < r < \operatorname{dist}(a, \operatorname{spt}(T))$, the mass ratio $M(r)$ is an increasing function of r.*

PROOF. As $f(r) = \mathbf{M}(T \llcorner \mathbf{B}(a,r))$ is increasing, it is differentiable almost everywhere. By slicing we have $\mathbf{M}(\partial(T \llcorner \mathbf{B}(a,r))) \leqslant f'(r)$. If C_r is the cone over $\partial(T \llcorner \mathbf{B}(a,r))$ with vertex a, we get $\mathbf{M}(C_r) = (r/m)\,\mathbf{M}(\partial(T \llcorner \mathbf{B}(a,r)))$ by the coarea formula. Hence

$$f(r) \leqslant \mathbf{M}(C_r) \leqslant \frac{r}{m}\, f'(r),$$

and

$$\alpha_m\, M'(r) = \left(\alpha_m\, \frac{f(r)}{r^m}\right)' = \frac{m\alpha_m}{r^{m+1}} \left[\frac{r}{m}\, f'(r) - f(r)\right] \geqslant 0,$$

and one easily concludes that $M(r)$ is increasing. □

It is not difficult to show that equality holds if and only if T is a cone. A first corollary of the above result is that

$$\mathbf{M}(T \llcorner \mathbf{B}(a,r)) \geqslant \Theta(T,a)\, \alpha_m r^m.$$

As $\Theta(T,a) = 1$ for almost all points of a rectifiable current T, we easily conclude that $\Theta(T,a) \geqslant 1$ everywhere on $\operatorname{spt} T - \operatorname{spt} \partial T$. In particular, if T is a 2-dimensional, area-minimizing surface in $\mathbb{R}^3$, then $\mathbf{M}(T,a) \geqslant \pi r^2$.

Another interesting consequence of monotonicity is the existence of a tangent cone (limit of homothetic expansions) at any point of an area-minimizing rectifiable current [**30**, 9.8].

When proving the regularity of an area-minimizing surface there are two steps: the first one is to show that the tangent cone is a plane. The second one is to conclude that the existence of the tangent plane implies that the surface is regular.

Allard [**1**] generalized monotonicity to stationary surfaces (mean curvature 0) and proved that for bounded mean curvature surfaces in a Riemannian manifold

$$M(r)\, e^{Cr} r^{-m}$$

is monotone, where $M(r)$ is the mass inside a ball of radius r. This result was proved in the very general context of varifolds. It is not valid for surfaces stationary for general integrands Φ, although for Φ-minimizers we still have

$$M(r) \geqslant C\, r^m.$$

4. Other Types of Surfaces

The Belgian physicist Plateau [**33**] (see also [**45**]) observed around 1870 two kinds of singularities in soap films: either three sheets meet along a seam at 120°, or four such seams meet at a point at about 109°. The latter phenomenon corresponds to the triangles obtained from any given edge of a regular tetrahedron and from the segments which join the barycenter with the vertexes of the edge. This behavior of soap films cannot be modeled by currents since the singular set is very large and contradicts the regularity results of the previous section. Hence other definitions of generalized surfaces are necessary to treat this problem.

For instance, to treat nonorientable surfaces one can study rectifiable currents modulo 2. Two rectifiable currents T_1 and T_2 are congruent modulo 2 if there is some rectifiable current Q such that $T_1 - T_2 = 2Q$. In general one can consider rectifiable currents modulo any integer number ν. However, these new objects are not appropriate to model the tetrahedral soap film; see the discussion in [**30**, pp. 107 ff.].

We discuss now several different types of generalized surfaces.

4.1. Varifolds [30, 11.2]. A varifold is a measure concentrated on a set in space and certain "tangent planes". The notion of varifold allows multiplicity, but neither cancellation nor obvious definition of orientation. Also tangent planes need not be associated with the set. More precisely, an m-dimensional *varifold* is a Radon measure on $\mathbb{R}^n \times G_m\mathbb{R}^n$, where $G_m\mathbb{R}^n$ is the Grassmannian of unoriented unit m-planes through the origin in $\mathbb{R}^n$.

An *integral varifold* is one associated to a rectifiable set. Let $S \subset \mathbb{R}^n$ be a rectifiable set, say with multiplicity 1, and m-tangent plane $\vec{S}$. Define

$$\mathbf{v}(S) = \mathcal{H}^m \llcorner \{(x, \vec{S}(x)) : x \in S\}.$$

Instead of the notion of boundary we get the first variation of a varifold. $\delta V(\vec{v})$ is the initial rate of change of area in $\mathbb{R}^n$ under a smooth variation with initial velocity $\vec{v}$. A *stationary varifold* is one for which $\delta V = 0$ for all $\vec{v}$. Stationary integral varifolds include area-minimizing rectifiable currents (modulo ν also), and soap films. For integral varifolds there are a compactness theorem and isoperimetric inequalities. Also the following regularity result holds

THEOREM 4.1 (Allard's regularity [1]). *An integral k-dimensional varifold with mean curvature bounded or in L^p, $p > k$, is a $C^{1,\alpha}$ submanifold at points of density* 1.

An important open question for varifolds is whether a two-dimensional stationary integral varifold in an open subset of $\mathbb{R}^3$ is a smooth embedded manifold almost everywhere.

4.2. Finite perimeter sets and functions of bounded variation. The notion of a set of finite perimeter was introduced by Cacciopoli and De Giorgi [11] and the theory was developed independently for some time. Later it became clear that a finite perimeter set in $\mathbb{R}^n$ is $(\mathcal{H}^n, n)$-rectifiable [18, Ch. 4], [41, §14].

We shall say that $A \subset \mathbb{R}^n$ is a *finite perimeter set* if, for any vector field X with compact support in $\mathbb{R}^n$, we have

$$\int_A \operatorname{div} X \, d\mathcal{H}^n \leqslant c(A)|X|,$$

where $|X|$ is the sup norm of the vector field X, and $c(A) > 0$ is a constant depending on the set A. It can be shown that this property is equivalent to saying that the characteristic function χ_A of the set A is of *bounded variation*. The perimeter $\mathcal{P}(A)$ of the set A is defined as

$$\mathcal{P}(A) = \sup_{|X| \leqslant 1} \int_A \operatorname{div} X \, d\mathcal{H}^n.$$

This definition coincides with $\mathcal{H}^{n-1}(\partial A)$ if ∂A is a smooth hypersurface.

From the definition and Riesz's representation theorem, one can get the existence of an $\mathcal{H}^{n-1}$-measurable vector-valued function $\nu : A \to \mathbb{R}^n$ and a Borel measure μ on $\mathbb{R}^n$ such that

$$\int_A \operatorname{div} X \, d\mathcal{H}^n = \int_{\mathbb{R}^n} \langle \nu, X \rangle \, d\mu.$$

Both the measure μ and the function ν has support contained in the topological boundary of A.

For finite perimeter sets, the natural convergence is in measure. We shall say that $A_i \to A$ if $\chi_{A_i} \to \chi_A$ in L^1. With this notion of convergence the perimeter is lower semicontinuous, i.e., if $A_i \to A$ then

$$\mathcal{P}(A) \leqslant \liminf_i \mathcal{P}(A_i).$$

Finite perimeter sets also admit slicing and isoperimetric inequalities. The interested reader should consult Giusti [**18**], Massari–Miranda [**26**], Simon [**41**], and Ziemer [**47**].

4.3. $(\mathbf{M}, \varepsilon, \delta)$-minimal sets [30, 11.3]. Soap films were modeled by Almgren as $(\mathbf{M}, 0, \delta)$-minimal sets, which satisfy two conditions: they touch the boundary, and tiny portions (of diameter $\leqslant \delta$) cannot be pinched or deformed to make the area go down. To be more precise, $S \subset \mathbb{R}^n - B$ bounded, with $\mathcal{H}^m(S) < \infty$, is $(\mathbf{M}, 0, \delta)$-minimal with respect to a closed set B (the boundary) if, for every Lipschitz deformation φ of $\mathbb{R}^n$ which differs from the identity map only in a δ-ball disjoint from B,

$$\mathcal{H}^m(S) \leqslant \mathcal{H}^m(\varphi(S)).$$

The Lipschitz function φ need not be a diffeomorphism. For more general functions $\varepsilon(r) = C\, r^\alpha$, $\alpha > 0$, the weaker inequality,

$$\mathcal{H}^m(S) \leqslant (1 + \varepsilon(r))\, \mathcal{H}^m(\varphi(S)),$$

defines $(\mathbf{M}, \varepsilon, \delta)$-minimal sets, which include soap bubbles with volume constraints.

These sets were treated by Almgren in his monograph [**3**]. Three basic properties of $(\mathbf{M}, \varepsilon, \delta)$-minimal sets are $(\mathcal{H}^m, m)$-rectifiability [**3**, II.3(9)], monotonicity formula [**44**, II.1], and existence of an $(\mathbf{M}, 0, \delta)$-minimal cone at every point [**44**, II.2]. The regularity of $(\mathbf{M}, \varepsilon, \delta)$-minimal sets in $\mathbb{R}^3$ was studied by Taylor [**44**], who proved

THEOREM 4.2. *$(\mathbf{M}, \varepsilon, \delta)$-minimal sets in $\mathbb{R}^3$ can have two kinds of singularities: the ones predicted by Plateau.*

PROOF. The proof consists in showing that there are only two possible singularities. To prove this, take any singularity and consider its linear approximation, which is a cone determined over a geodesic net in the sphere. This idea goes back to Lamarle (1864). There are precisely 10 such geodesic nets [**30**, pp. 134–135] (Lamarle missed one). One then shows that all but three of them are unstable. One corresponds to a great circle in the sphere, which yields regular portions, and the other two correspond to the singular sets.

Hence, for approximating cones, there are only two possible singularities. The really hard part is to show that the cone is a good enough approximation of the original soap film. Reifenberg's epiperimetric inequality is used to show that the minimizers converge rapidly to an asymptotic cone. □

For singularities of $(\mathbf{M}, 0, \delta)$-minimal sets in $\mathbb{R}^4$ one has the following conjecture.

CONJECTURE 4.3. *A cluster in $\mathbb{R}^4$ has only the following singularities:*

(i) 3 *hypersurfaces meeting along surfaces at* 120°.
(ii) 4 *such surfaces meeting along a curve at almost* 109°.
(iii) 5 *or* 16 *of such curves meeting at a point as in the cone over a regular simplex or the cone over a hypercube.*

Brakke [**8**] proved that these are the only polyhedral minimizing cones (see Sullivan [**43**]). White [**46**] has shown that the link of a nonpolyhedral minimizing cone would have to have positive Euler characteristic.

Many questions remain open.

- Boundary regularity.
- The existence of a least-area $(\mathbf{M}, 0, \delta)$-minimal set in $\mathbb{R}^3$ with a given smooth boundary curve.
- The cone over the regular tetrahedron is the least-area separator of the four faces. Is there a smaller equilibrium surface not separating the four regions?
- It is known that a smooth minimal surface is locally area-minimizing in arbitrary codimension. Is this property still true whenever the tangent cone is area-minimizing?

5. Double Bubbles

5.1. Existence and regularity of area-minimizing clusters. Bubble clusters seek the least-area way to enclose and separate several regions of prescribed volume. But they do *not* always find the absolute least area shape (they could be stable, but not minimizers).

A *cluster* is a collection of disjoint regions $R_1, \ldots, R_m$ (n-dimensional locally integral currents of multiplicity 1 in $\mathbb{R}^{n+1}$), with surface area

$$\frac{1}{2}\Big(\sum_i \mathbf{M}(\partial R_i) + \mathbf{M}(\partial \sum_i R_i)\Big).$$

A region is not assumed to be connected. The existence of an area-minimizing cluster enclosing prescribed volumes is guaranteed by Almgren's work [**3**] in a very general context. A simplified proof was given by Morgan [**29**].

THEOREM 5.1. *In $\mathbb{R}^n$, given volumes $V_1, \ldots, V_m > 0$, there exists an area-minimizing cluster for those volumes.*

The proof of this result needs some lemmas, the first of which is the following observation.

LEMMA 5.2. *Given any cluster, there exists $C > 0$, such that arbitrary small changes in volume may be accomplished inside small balls at a cost*

$$|\Delta A| \leqslant C\,|\Delta V|.$$

Another interesting lemma which follows from the monotonicity formula and the isoperimetric inequality is

LEMMA 5.3. *An area-minimizing cluster is bounded in $\mathbb{R}^n$.*

With these ingredients we are ready to prove Theorem 5.1. The proof of this result is complicated because when taking a minimizing sequence the volume could disappear at infinity. Let us take a minimizing sequence $\mathcal{C}_\alpha$ with the prescribed volumes.

Step 1. Get a convergent subsequence to a nonzero limit $\mathcal{C}$. One can choose points p_α such that

$$\mathrm{vol}(R_{1,\alpha} \cap \mathbf{B}(p_\alpha, c_1)) \geqslant c_2,$$

where $R_{1,\alpha}$ is the first region of $\mathcal{C}_\alpha$. This is a technical result known as the Concentration Lemma. The proof follows easily since we partition $\mathbb{R}^n$ into cubes K_j. By an isoperimetric inequality, we may assume that

$$\operatorname{area}(\partial R_{1,\alpha} \cap K_j) \geqslant \gamma \, (\operatorname{vol}(R_{1,\alpha} \cap K_j))^{(n-1)/n},$$

for some isoperimetric constant γ, and so

$$\operatorname{area}(\partial R_{1,\alpha} \cap K_j) \geqslant \gamma \, \frac{\operatorname{vol}(R_{1,\alpha} \cap K_j)}{\max_i \operatorname{vol}(R_{1,\alpha} \cap K_i)^{1/n}}.$$

Summing over j we get

$$\max_i \operatorname{vol}(R_{1,\alpha} \cap K_i)^{1/n} \geqslant \frac{\gamma}{A} V_1 = c_2,$$

which proves the desired result. (This is a minor correction of the statement in [**30**, 13.7(1)].)

Translate to assume $p_\alpha = 0$.

Step 2. Show that the limit $\mathcal{C}$ is area-minimizing for its volumes. If not, then one can get another (minimizing) sequence with less area than the original one. If no volume is lost, then we are done.

Step 3. If some volume disappears at infinity, then repeat the process with some p'_α.

Then countably many repetitions capture the total volume and yield a solution. This solution is bounded by Lemma 5.3, and so only a finite number of repetitions is needed.

Concerning regularity the first observation is that an area-minimizing cluster is $(\mathbf{M}, \varepsilon, \delta)$-minimal [**30**, 13.8]. The proof of this result relies on Lemma 5.2. Then one can apply Taylor's regularity result in $\mathbb{R}^3$, which implies that an $(\mathbf{M}, \varepsilon, \delta)$-minimal set consists of smooth, constant mean curvature surfaces meeting in threes at 120° along curves, in turn meeting in fours at about 109°.

In higher dimensions, Almgren [**3**] showed that $(\mathbf{M}, \varepsilon, \delta)$-minimal sets in $\mathbb{R}^n$, $n \geqslant 3$, are $C^{1,\alpha}$ almost everywhere. A simple treatment of $(\mathbf{M}, \varepsilon, \delta)$-minimal sets in $\mathbb{R}^2$ can be found in Morgan [**28**]. White [**46**] proved that they consist of constant mean curvature hypersurfaces meeting in threes at 120° along $(n-2)$-dimensional submanifolds, which meet in fours along smooth $(n-3)$-dimensional surfaces, which meet in an $(n-4)$-dimensional set.

5.2. Characterization of area-minimizing clusters. For an account of results on single bubbles the interested reader can consult [**30**, 13.2].

For double bubbles the only known optimal shapes are those of the double bubble in $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^4$. For the remaining volumes and dimensions, the problem is still open.

The *standard double bubble* in $\mathbb{R}^n$ is composed of three $(n-1)$-spherical caps meeting along an $(n-2)$-sphere at 120°. In case we consider two equal volumes one of the spherical caps degenerates to a flat disc. The whole configuration is obtained by rotating three circles (or two circles and a segment) in a plane meeting at 120° around an $(n-2)$-dimensional subspace.

The double bubble conjecture states that the standard double bubble provides the least perimeter way to enclose and separate two regions of given volumes in $\mathbb{R}^n$.

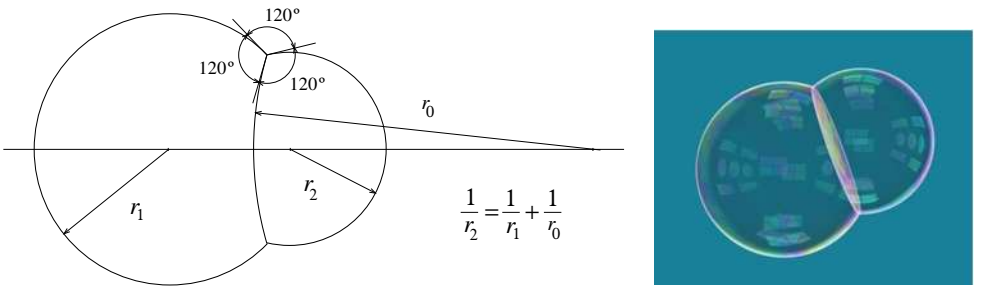


FIGURE 1. The standard double bubble.

This conjecture appeared in Foisy's undergraduate thesis [**16**]. The *planar* double bubble conjecture was proved in 1990 by Foisy, Alfaro, Brock, Hodges and Zimba [**17**]. The double bubble conjecture in $\mathbb{S}^2$ was solved by Masters [**27**] in 1994.

The study of double bubbles is complicated since we do not know anything about their structure. The regions can be disconnected with any number of components, and also the exterior of the bubble can be disconnected, with precisely one noncompact component, and possibly more compact ones which will be called *empty spaces* or *empty chambers*.

The first results towards proving the double bubble conjecture in Euclidean 3-space were

(i) The bubble is a surface of revolution around some line in $\mathbb{R}^3$, White, Foisy [**16**], Hutchings [**23**].

(ii) The isoperimetric profile $A(v_1, v_2)$, which assigns to every pair of volumes (v_1, v_2) the perimeter of the area-minimizing double bubble of these volumes is a *concave* function. This implies that A is increasing in each variable, which shows that there are no empty spaces inside, Hutchings [**23**].

It was also proved in Foisy [**16**] and Hutchings [**23**] that the bubble must touch the axis of revolution. In order to prove this one considers the transformations $r \mapsto (r^2 - \varepsilon)^{1/2}$, where r is the distance to the axis of revolution. This deformation decreases area and keeps constant the volume enclosed. For two equal volumes Hutchings also proved that each region is connected. By using this result Hass and Schlaffly proved, by making use of computer techniques, that

THEOREM 5.4. [**22**] *The standard double bubble is the least perimeter way to enclose and separate two equal volumes in* $\mathbb{R}^3$.

In the general case it was proved by Hutchings, Morgan, Ritoré and Ros [**24**], [**25**] that there are at most three components (one of the regions can be disconnected and has at most two connected components). Finally they proved

THEOREM 5.5. [**25**] *The standard double bubble is the least perimeter way to enclose and separate two arbitrary given volumes in* $\mathbb{R}^3$.

We already know that an area-minimizing bubble must be a surface of revolution, that there are no empty chambers, and that the number of components is less than or equal to three. This is enough to reduce the number of candidates to

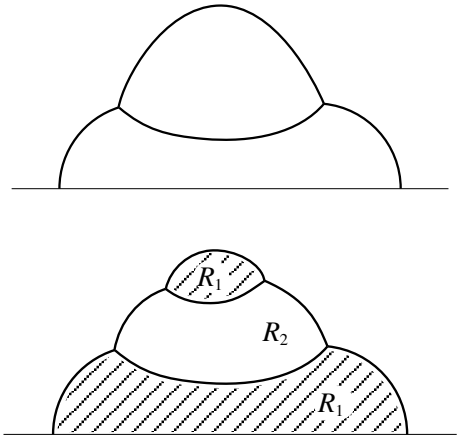


FIGURE 2. Candidate graphs to produce area-minimizing bubbles when rotated.

two possible configurations which are depicted in Figure 2. Both can be obtained by rotating a graph Γ in a half-plane about the boundary line.

The main ingredient in the proof is an instability argument, based on Courant's Nodal Domain Theorem. It essentially says that a Killing field cannot induce too many nodal regions in the bubble. If X is a Killing field, the one induced by a one-parameter group of isometries, a nodal region as a connected component of the function $\langle X, N\rangle$, where N is a unit vector field normal to the sheets of the bubble. This argument has been previously used to study stability and isoperimetric questions in [**35**], [**37**], [**36**] and [**32**]. From this principle one can prove the following proposition.

PROPOSITION 5.6. *Consider a double bubble of revolution $\Sigma \subset \mathbb{R}^n$, $n \geqslant 3$, with axis L, as in Figure 2. Assume there is a finite number of nonsingular points $p_1, \cdots, p_k$, with $x = f(p_1) = \cdots = f(p_k)$, which separate the generating graph Γ.*

Then Σ is unstable.

In the above proposition, $f(p)$ is the intersection of the normal line to p with the axis of revolution L. Of course $f(p)$ can be ∞. This proposition reduces the proof of the double bubble conjecture to elementary planar geometry modulo some properties of Delaunay hypersurfaces (hypersurfaces of revolution with constant mean curvature). For instance, we can easily discard the two components case. Consider Figure 3. Let L' be the line equidistant from a and b. Assume that this line meets the axis of revolution L at some point x. In each one of the curves joining a and b there is at least a point at maximum or minimum distance from x. Call them p_1 and p_2. Then p_1 and p_2 separate the generating graph Γ, which implies that the bubble is unstable.

The three-component case is more involved, and some special properties of hypersurfaces of revolution with constant mean curvature are used. We refer the interested reader to [**25**] for the discussion of this case.

By using the arguments in [**25**], Reichardt, Heilmann, Lai and Spielman [**34**] were able to prove the double bubble conjecture in $\mathbb{R}^4$ and in some particular cases

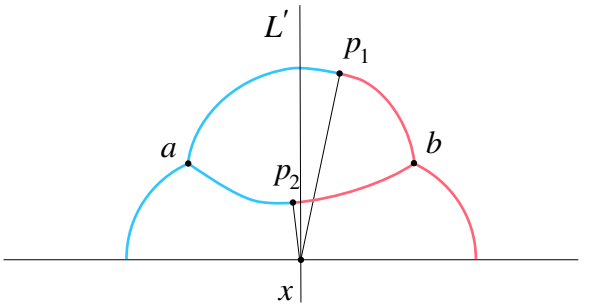


FIGURE 3. The partition method.

in $\mathbb{R}^n$. Cotton and Freeman [**10**] obtained partial results on the double bubble conjecture in the three-dimensional sphere and hyperbolic space.

6. Integrands and Isoperimetric Inequalities

6.1. Integrands. An *integrand* is a map $\Phi : \mathbb{R}^n \times \Lambda_m \mathbb{R}^n \to \mathbb{R}$ which is homogeneous and continuous. Sometimes it is required that Φ be smooth except at the origin, positive, even ($\Phi(x,\xi) = \Phi(x,-\xi)$) or convex ($\Phi(x,\xi_1+\xi_2) \leqslant \Phi(x,\xi_1)+\Phi(x,\xi_2)$). An integrand defines a functional on the space of m-dimensional rectifiable sets by

$$\Phi(S) = \int_S \Phi(x, \vec{S}(x))\, d\mathcal{H}^m(x).$$

For instance, the integrand $\Phi(x,\xi) = |\xi|$ defines the area functional.

Existence of Φ-minimizers follows from the compactness theorem. Interior regularity of a hypersurface minimizing a smooth elliptic integrand Φ in $\mathbb{R}^n$ or in a Riemannian manifold M^n except for a singular set of $\mathcal{H}^{n-3}$-measure 0 was proved by Schoen, Simon, and Almgren [**39**]. In higher codimension, regularity on an open dense set was proved by Almgren and Federer [**12**, 5.3.16–17]. Boundary regularity is known just if the surface has density 1/2, Hardt [**21**].

The Φ-isoperimetric problem for curves in $\mathbb{R}^n$ seeks curves of least Φ-energy bounding a fixed least-area. In $\mathbb{R}^2$ the optimal shape is the Wulff shape (the unit ball in the dual norm).

6.2. Isoperimetric inequalities. It has been known since ancient times that the circle in $\mathbb{R}^2$ encloses the most area for fixed length. The first rigorous proof of this fact was given by Weierstrass in the late 1800s. That a round sphere in $\mathbb{R}^3$ encloses the most volume for fixed area was proved by Schwarz [**40**]. Later Schmidt [**38**] proved that a geodesic ball in Euclidean, hyperbolic or spherical space encloses a fixed volume with the least possible perimeter.

In higher codimension there are many submanifolds with a given boundary, so that large area need not imply large perimeter. However isoperimetric inequalities can be obtained in $\mathbb{R}^n$ if we require that the spanning surface be area-minimizing

THEOREM 6.1 (Almgren [**5**]). *An m-dimensional area-minimizing integral current in $\mathbb{R}^n$ $(2 \leqslant m \leqslant n)$ has no more area than a round disc of the same perimeter, with equality only for the round disc itself.*

Let us give the main ideas of the proof for surfaces in $\mathbb{R}^3$ (although this case was known earlier). Amongst area-minimizing surfaces in $\mathbb{R}^3$ with area π there is one Q with boundary C of least length L. We assume regularity to simplify the arguments. Of course, by direct comparison with the planar disc of area π, we get

$$L \leqslant 2\pi.$$

Let us see first that the curvature κ of the curve satisfies the inequality $\kappa \leqslant 1$. Otherwise we could find a local variation of C (around some point where $\kappa > 1$) so that

$$\frac{dL}{dA} > 1.$$

We use this variation to reduce the length of C, reducing at the same time the area of Q. On the other hand, scaling up to restore the area we get

$$\frac{dL}{dA} = \frac{L}{2A} \leqslant 1,$$

so that the combination of both transformations keeps the area constant but reduces boundary length, a contradiction to the minimality of the length of C.

Finally we look at the Gauss map of the boundary B of the convex hull of C (Q lies inside by the convex hull property for minimal surfaces). The total curvature of B is 4π and the only contribution to this total curvature comes from $B \cap C$ since $B - C$ is a flat surface. Geometrical considerations [**30**, p. 183] imply that this contribution to the total curvature equals $(2/\pi)\,\kappa\alpha$, where $\alpha \in (0, \pi)$ is an angle. Hence

$$4\pi = \frac{2}{\pi} \int_C \kappa\, \alpha\, ds \leqslant \frac{2}{\pi} 1\, \pi\, L = 2L.$$

This shows that $L = 2\pi$. If equality holds, then $\kappa \equiv 1$ and $\alpha \equiv \pi$ and this is enough to show that Q is a flat disc bounded by the circle C.

To make this argument completely rigorous, Almgren [**5**] considered the equidistants (sets at fixed distance) from the boundary B of the convex hull of B.

Finally we state some open questions concerning isoperimetric inequalities

(i) We know that amongst area-minimizing surfaces the disc of area π is the largest minimal surface bounded by a system of curves of total length 2π. Is this still true for all minimal surfaces? Is it true for stable minimal surfaces?

(ii) Does any isoperimetric inequality hold for Φ-minimal surfaces?

References

1. William K. Allard, *On the first variation of a varifold*, Ann. of Math. (2) **95** (1972), 417–491. MR 46 #6136
2. F. J. Almgren, Jr., *Some interior regularity theorems for minimal surfaces and an extension of Bernstein's theorem*, Ann. of Math. (2) **84** (1966), 277–292. MR 34 #702
3. ———, *Existence and regularity almost everywhere of solutions to elliptic variational problems with constraints*, Mem. Amer. Math. Soc. **4** (1976), no. 165, viii+199. MR 54 #8420
4. ———, *Q valued functions minimizing Dirichlet's integral and the regularity of area minimizing rectifiable currents up to codimension two*, Bull. Amer. Math. Soc. (N.S.) **8** (1983), no. 2, 327–328. MR **84b:**49052
5. ———, *Optimal isoperimetric inequalities*, Indiana Univ. Math. J. **35** (1986), no. 3, 451–547. MR **88c:**49032
6. ———, *Almgren's big regularity paper*, World Sci. Publishing, River Edge, NJ, 2000. MR **2003d:**49001

7. E. Bombieri, E. De Giorgi, and E. Giusti, *Minimal cones and the Bernstein problem*, Invent. Math. **7** (1969), 243–268. MR 40 #3445
8. Kenneth A. Brakke, *Polyhedral minimal cones in* $\mathbb{R}^4$, preprint, 1993.
9. Sheldon Xu-Dong Chang, *Two-dimensional area minimizing integral currents are classical minimal surfaces*, J. Amer. Math. Soc. **1** (1988), no. 4, 699–778. MR **89i:**49028
10. A. Cotton and D. Freeman, Int. J. Math. Math. Sci. **32** (2002), no. 11, 641–699. MR 1 949 693
11. Ennio De Giorgi, *Frontiere orientate di misura minima*, Editrice Tecnico Scientifica, Pisa, 1961. MR 31 #3897
12. Herbert Federer, *Geometric Measure Theory*, Springer-Verlag New York Inc., New York, 1969, Die Grundlehren der mathematischen Wissenschaften, Band 153. MR 41 #1976
13. ———, *The singular sets of area minimizing rectifiable currents with codimension one and of area minimizing flat chains modulo two with arbitrary codimension*, Bull. Amer. Math. Soc. **76** (1970), 767–771. MR 41 #5601
14. Herbert Federer and Wendell H. Fleming, *Normal and integral currents*, Ann. of Math. (2) **72** (1960), 458–520. MR 23 #A588
15. Wendell H. Fleming, *On the oriented Plateau problem*, Rend. Circ. Mat. Palermo (2) **11** (1962), 69–90. MR 28 #499
16. Joel Foisy, *Soap Bubble Clusters in* $\mathbb{R}^2$ *and in* $\mathbb{R}^3$, Undergraduate thesis, Williams College, 1991.
17. Joel Foisy, Manuel Alfaro, Jeffrey Brock, Nickelous Hodges, and Jason Zimba, *The standard double soap bubble in* $\mathbb{R}^2$ *uniquely minimizes perimeter*, Pacific J. Math. **159** (1993), no. 1, 47–59. MR **94b:**53019
18. Enrico Giusti, *Minimal surfaces and functions of bounded variation*, Birkhäuser Verlag, Basel, 1984. MR **87a:**58041
19. E. Gonzalez, U. Massari, and I. Tamanini, *On the regularity of boundaries of sets minimizing perimeter with a volume constraint*, Indiana Univ. Math. J. **32** (1983), no. 1, 25–37. MR **84d:**49043
20. Robert Hardt and Leon Simon, *Boundary regularity and embedded solutions for the oriented Plateau problem*, Ann. of Math. (2) **110** (1979), no. 3, 439–486. MR **81i:**49031
21. Robert M. Hardt, *On boundary regularity for integral currents or flat chains modulo two minimizing the integral of an elliptic integrand*, Comm. Partial Differential Equations **2** (1977), no. 12, 1163–1232. MR 58 #23954
22. Joel Hass and Roger Schlafly, *Double bubbles minimize*, Ann. of Math. (2) **151** (2000), no. 2, 459–515. MR 1 765 704
23. Michael Hutchings, *The structure of area-minimizing double bubbles*, J. Geom. Anal. **7** (1997), no. 2, 285–304. MR **99j:**53010
24. Michael Hutchings, Frank Morgan, Manuel Ritore, and Antonio Ros, *Proof of the double bubble conjecture*, Electron. Res. Announc. Amer. Math. Soc. **6** (2000), 45–49 (electronic). MR 1 777 854
25. Michael Hutchings, Frank Morgan, Manuel Ritoré, and Antonio Ros, *Proof of the double bubble conjecture*, Ann. of Math. (2) **155** (2002), 459–489. MR **2003c:**53013 Avalaible at `http://www.ugr.es/~ritore/bubble/bubble.htm`, 2000
26. Umberto Massari and Mario Miranda, *Minimal surfaces of codimension one*, North-Holland Publishing Co., Amsterdam, 1984, Notas de Matemática [Mathematical Notes], 95. MR **87f:**49058
27. Joseph D. Masters, *The perimeter-minimizing enclosure of two areas in* S^2, Real Anal. Exchange **22** (1996/97), no. 2, 645–654. MR **99a:**52010
28. Frank Morgan, $(\mathbf{M}, \epsilon, \delta)$*-minimal curve regularity*, Proc. Amer. Math. Soc. **120** (1994), no. 3, 677–686. MR **94e:**49018
29. ———, *Clusters minimizing area plus length of singular curves*, Math. Ann. **299** (1994), no. 4, 697–714. MR **95g:**49083
30. ———, *Geometric Measure Theory: A Beginner's Guide*, third ed., Academic Press Inc., San Diego, CA, 2000. MR 1 775 760
31. ———, *Regularity of isoperimetric hypersurfaces in riemannian manifolds*, Trans. Amer. Math. Soc. **355** (2003), 5041–5052.

32. Renato H. L. Pedrosa and Manuel Ritoré, *Isoperimetric domains in the Riemannian product of a circle with a simply connected space form and applications to free boundary problems*, Indiana Univ. Math. J. **48** (1999), no. 4, 1357–1394. MR 1 757 077
33. Joseph Plateau, *Statique expérimentale et théorique des liquides soumis aux seules forces moléculaires*, Gauthier-Villars, Paris, 1873
34. Ben W. Reichardt, Cory Heilmann, Yuan Y. Lai, and Anita Spielmann, *Proof of the double bubble conjecture in $\mathbb{R}^4$ and certain higher dimensional cases*, Pacific J. Math. **208** (2003), 347-366
35. Manuel Ritoré and Antonio Ros, *Stable constant mean curvature tori and the isoperimetric problem in three space forms*, Comment. Math. Helv. **67** (1992), no. 2, 293–305. MR **93a:**53055
36. Antonio Ros and Rabah Souam, *On stability of capillary surfaces in a ball*, Pacific J. Math. **178** (1997), no. 2, 345–361. MR **98c:**58029
37. Antonio Ros and Enaldo Vergasta, *Stability for hypersurfaces of constant mean curvature with free boundary*, Geom. Dedicata **56** (1995), no. 1, 19–33. MR **96h:**53013
38. Erhardt Schmidt, *Beweis der isoperimetrischen Eigenschaft der Kugel im hyperbolischen und spharischen Raum jeder Dimensionenzahl*, Math. Z. **49** (1943), 1–109.
39. Richard Schoen, Leon Simon, and Frederick J. Almgren, Jr., *Regularity and singularity estimates on hypersurfaces minimizing parametric elliptic variational integrals. I, II*, Acta Math. **139** (1977), no. 3-4, 217–265. MR 57 #7333
40. Hermann Amadeus Schwarz, *Beweis des Satzes, dass die Kugel kleinere Oberfläche besitzt, als jeder andere Körper gleichen Volumens*, Nachrichten von der Königlichen Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen (1884), 1–13.
41. Leon Simon, *Lectures on Geometric Measure Theory*, Australian National University Centre for Mathematical Analysis, Canberra, 1983. MR **87a:**49001
42. James Simons, *Minimal varieties in riemannian manifolds*, Ann. of Math. (2) **88** (1968), 62–105. MR 38 #1617
43. John M. Sullivan, *Convex deltatopes in all dimensions, and polyhedral soap films*, preprint, 1994.
44. Jean E. Taylor, *The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces*, Ann. of Math. (2) **103** (1976), no. 3, 489–539. MR 55 #1208a
45. D'Arcy Wentworth Thompson, *On Growth and Form: the complete revised edition*, Dover Publications, Inc., New York, 1992.
46. Brian White, in preparation.
47. William P. Ziemer, *Weakly Differentiable Functions*, Springer-Verlag, New York, 1989. MR **91e:**46046

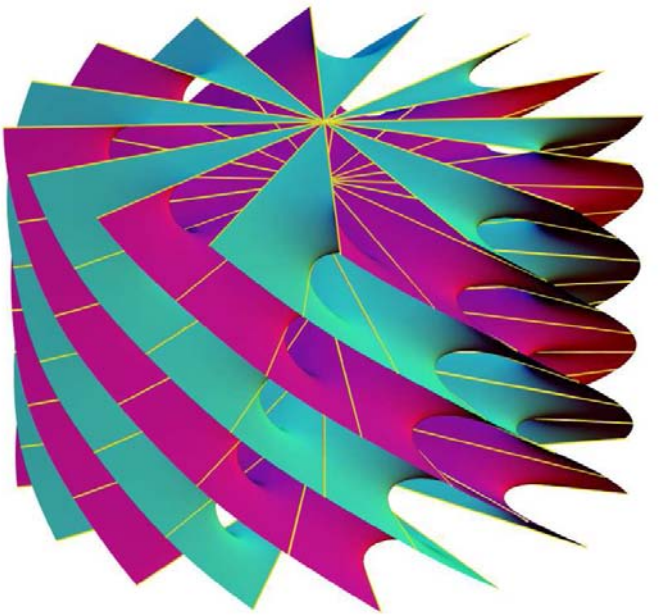
Department of Mathematics, Williams College, Williamstown, MA 01267
E-mail address: Frank.Morgan@williams.edu

Departamento de Geometría y Topología, Universidad de Granada, E–18071 Granada, España
E-mail address: ritore@ugr.es

Clay Mathematics Proceedings
Volume 2, 2004

Classical Minimal Surfaces in Euclidean Space by Examples: Geometric and Computational Aspects of the Weierstrass Representation

Matthias Weber



Contents

1. Introduction 20
2. Basic Differential Geometry of Surfaces 20
3. The Weierstrass Representation 23
4. Minimal Surfaces on Punctured Spheres 29
5. The Scherk Surfaces 33
6. Minimal Surfaces Defined on Punctured Tori 40
7. Higher Genus Minimal Surfaces 53
References 62

©2004 Clay Mathematics Institute

1. Introduction

These notes are an enriched version of the lecture notes I distributed for the participants of the Clay Mathematical Summer School at MSRI, Berkeley, in the summer 2001.

The purpose of the five lectures covered by these notes was to introduce beginning graduate students to minimal surfaces so that they would understand the close interaction between the surface geometry and their complex analytic description.

To achieve this in a fully satisfactory way, I would have needed many prerequisites from differential geometry, partial differential equations, complex variables, Riemann surfaces and topology.

Since I was not allowed to assume all this, I have chosen an informal and concrete way to present the subject. Instead of proving theorems, I construct examples, and the more complicated ones only numerically.

The workshop was accompanied by computer sessions where I used my Mathematica notebooks and Jim Hoffman's MESH program to illustrate the examples. The Mathematica notebooks are available on the web. However, they make use of an add-on to Mathematica, which is not available anymore, the OpenGL Explorer. However, as Mathematica 5 still has no acceptable support for high quality 3D graphics, I am currently migrating the notebooks to an open source system, and the code will be made available as soon as it is ready.

Acknowledgments. I thank the referee for the enormous number of helpful comments, some of which can only be taken care of in a largely expanded version of these notes.

I have borrowed from the sources [**DHKW92, HK97, Kar89, Oss86**], which are recommended for further reading. I hope that these notes make these sources more accessible.

2. Basic Differential Geometry of Surfaces

We review basic facts about surfaces in euclidean space. Any textbook on elementary differential geometry will do as a reference.

2A. Local surface parameterizations. Denote the standard scalar product of the euclidean space $\mathbb{R}^3$ by $\langle \cdot, \cdot \rangle$. A surface in euclidean space is given by a smooth map from a parameter domain $\Omega \subset \mathbb{R}^2$,

$$\phi : \Omega \to \mathbb{R}^3.$$

We require the map to be nonsingular. This means that the differential $d\phi$ has maximal rank 2. Associated to such a surface parametrization are various data which are intended to describe geometric quantities on the surface. Of course, the most interesting data (like curvatures) will be independent of the parametrization. In such a case, it is best to imagine these data as being attached to the surface in $\mathbb{R}^3$ but to compute with these data by representing them in the domain Ω of the surface.

DEFINITION 2.1. The first fundamental form or Riemannian metric of a surface $\phi : \Omega \to \mathbb{R}^3$ is defined by

$$g(V, W) = \langle d\phi \cdot V,\ d\phi \cdot W \rangle,$$

where V and W are tangent vectors at some point $p \in \Omega$. It can be used to measure angles and length of tangent vectors. For instance:

DEFINITION 2.2. Let $c : [a, b] \to \Omega$ be a smooth curve. Then the length of $\tilde{c} = \phi \circ c$ is

$$\text{length}(\tilde{c}) = \int_a^b g(c'(t), c'(t))^{1/2} dt.$$

Using the length of curves on the surface, one can define an intrinsic metric (in the usual sense) by declaring the distance between two points as the infimum of the lengths of curves on the surface connecting these two points. There is a useful criterion when a surface, equipped with this metric, is complete. We use this criterion as a definition:

DEFINITION 2.3. A parametrized surface $\phi : \Omega \to \mathbb{R}^3$ is complete if and only if every smooth curve $c : [0, \infty) \to \Omega$ with $c(t) \to \partial\Omega$ for $t \to \infty$ has infinite length.

We will mainly be concerned with complete surfaces.

Of particular importance to us is the following special kind of parametrization:

DEFINITION 2.4. A parametrization is conformal or isothermal if it preserves the angles between tangent vectors:

$$\langle d\phi \cdot V,\, d\phi \cdot W \rangle = \lambda^2 \langle V,\, W \rangle.$$

We call the factor λ the stretch factor of ϕ.

In other words, the Riemannian metric is pointwise proportional to the euclidean scalar product of the domain Ω.

If one draws a surface by mapping a square grid to space using an isothermal parametrization, the image quadrilaterals will be close to squares. This makes conformal parametrizations attractive for visualizing surfaces.

2B. The stereographic projection. An important example of a conformal parametrization is the inverse of the stereographic projection. Recall that the stereographic projection maps the unit sphere to the (extended) complex plane. It can be defined as follows: Take a sphere and a plane in $\mathbb{R}^3$ tangent to the sphere at P. Now, for any point $Q \neq P$ on the sphere there is a straight line l through P and Q. This line intersects the plane in some point Q'. We call the image of P itself the infinite point ∞. The map $Q \mapsto Q'$ is the stereographic projection. It is one-to-one, since for every point Q' in the plane, the line through P and Q' will intersect the sphere in a second point Q.

There are different formulas for this map, depending on the various choices we made. We will use the following formula for the inverse of the stereographic projection. It maps the $z = x + iy$-plane onto the unit sphere in $\mathbb{R}^3$. In complex coordinates,

$$\sigma^{-1}(z) = \left(\frac{2\,\text{Re}(z)}{1 + |z|^2}, \frac{2\,\text{Im}(z)}{1 + |z|^2}, 1 - \frac{2}{1 + |z|^2} \right). \tag{2–1}$$

This choice of a stereographic projection maps the north pole to ∞, the south pole to 0, and the equator to the unit circle.

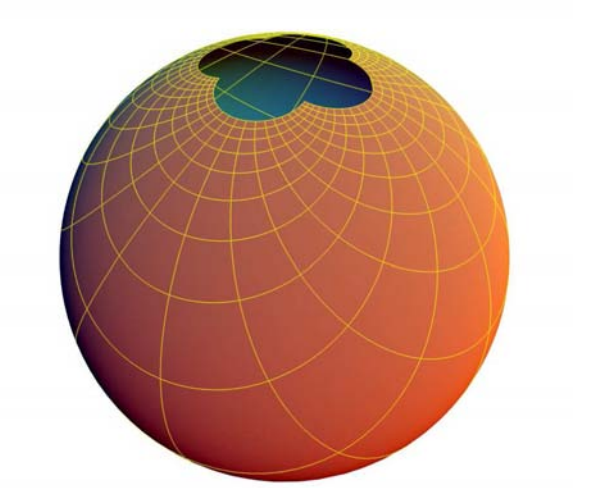


FIGURE 1. Image of a square grid in the plane under the inverse stereographic projection.

MATHEMATICA NOTE 2.5. The notebook Stereographic.nb teaches you how to use Mathematica to compute the differential of a parametrization and check whether it is conformal. In particular, it proves that the stereographic projection is conformal.

2C. The normal vector or Gauss map. Consider a surface parametrization $\phi : \Omega \to \mathbb{R}^3$. For each point $p \in \Omega$ and two coordinate vectors $U = \frac{\partial}{\partial u}, V = \frac{\partial}{\partial v}$, the vectors $d\phi(p) \cdot U, d\phi(p) \cdot V$ span a linear subspace of $\mathbb{R}^3$, the tangent space of Σ at $\phi(p)$.

Orthogonal to this tangent space is a unit vector in $\mathbb{R}^3$ that is used to define an important map from Ω into the unit sphere:

DEFINITION 2.6. The normal vector to the tangent space is called the Gauss map:

$$N : \Omega \to S^2, \quad N(p) = \left. \frac{d\phi \cdot U \times d\phi \cdot V}{|d\phi \cdot U| \cdot |d\phi \cdot V|} \right|_p .$$

REMARK 2.7. There is an ambiguity here, in that there are two different normal vectors which point in opposite directions. This can be helped by fixing once and for all orientations of the domain $\Omega \subset \mathbb{R}^2$ and $\mathbb{R}^3$. Then the normal vector is chosen so that for an oriented basis U, V of the tangent space to some point p of Ω at some point, the vectors $d\phi U$, $d\phi V$, N form an oriented basis of the tangent $\mathbb{R}^3$ at $\phi(p)$.

So far we have only considered first order derivatives of the surface parametrization (to write down tangent vectors). The second order derivatives lead to important geometric data, called curvatures. Here is what we need:

DEFINITION 2.8. The Weingarten map or shape operator S is defined as

$$d\phi \cdot S \cdot V = dN \cdot V.$$

In this definition, we have identified the tangent space of the surface at $\phi(p)$ with the tangent space of the unit sphere at $N(p)$ by a parallel translation.

The *tensor* S maps tangent vectors in Ω to tangent vectors in Ω and is thus solely defined in the domain of definition of the surface $\phi : \Omega \to \mathbb{R}^3$.

PROPOSITION 2.9. *The Weingarten map is a symmetric endomorphism of the tangent spaces of* Ω *with respect to the Riemannian metric* g.

DEFINITION 2.10. The eigenvalues of the Weingarten map are called principal curvatures. The eigendirections indicate where the surface bends most and least. The principal curvatures are combined into mean curvature

$$H = \operatorname{trace} S$$

and Gauss curvature

$$K = \det S.$$

3. The Weierstrass Representation

Our goal will be to understand the interaction:

Riemann Surface Meromorphic Function G Holomorphic Differential dh	$\Longleftrightarrow$	Geometry

3A. The Gauss map. A minimal surface in $\mathbb{R}^3$ is a surface which locally minimizes area. In this sense it is a two-dimensional generalization of a straight line, a path which locally minimizes arc length. To make this precise and achieve minimal completeness of exposition, we give the definition of area even though we won't need it later on:

DEFINITION 3.1. The area of a surface Σ, parametrized by $\phi : \Omega \to \mathbb{R}^3$, is defined as

$$\operatorname{area}(\Sigma) = \int_\Omega dA.$$

Here dA is the area form (or density) of the surface. Using coordinates (x_1, x_2) of Ω and coordinate tangent vectors $X_j = \partial/\partial x_j$,

$$dA = (\det g(d\phi X_i, d\phi X_j))^{1/2}\, dx_1 dx_2.$$

Equivalently, for minimal surfaces, the Gauss map is anticonformal. This means: It preserves angles and changes the orientation.

PROOF. The mean curvature being zero implies that S has eigenvalues $\lambda \geq 0$ and $-\lambda$. Thus the square of S is equal to $\lambda^2\,\mathrm{id}$. Hence

$$\langle dnU,\, dnU\rangle = \langle d\phi SU,\, d\phi SU\rangle = g(SU, SU) = g(S^2U, U) = \lambda^2 g(U, U).$$

This implies that the Gauss map preserves angles. It reverses the orientation, because for a positively oriented basis of curvature directions,

$$\begin{aligned}\det(dNU, dNV, N) &= \det(d\phi SU, d\phi SV, N) = \det(\lambda d\phi U, -\lambda d\phi V, N)\\ &= -\lambda^2 \det(d\phi U, d\phi V, N) \leq 0,\end{aligned}$$

because, we have chosen n such that $d\phi U, d\phi V, N$ are positively oriented in $\mathbb{R}^3$. $\square$

Precomposing the Gauss map with a conformal parametrization of the surface and postcomposing it with an anticonformal stereographic projection, yields a conformal map from the domain of definition of the surface to the Riemann sphere. This meromorphic function is denoted by G and called the Gauss map as well.

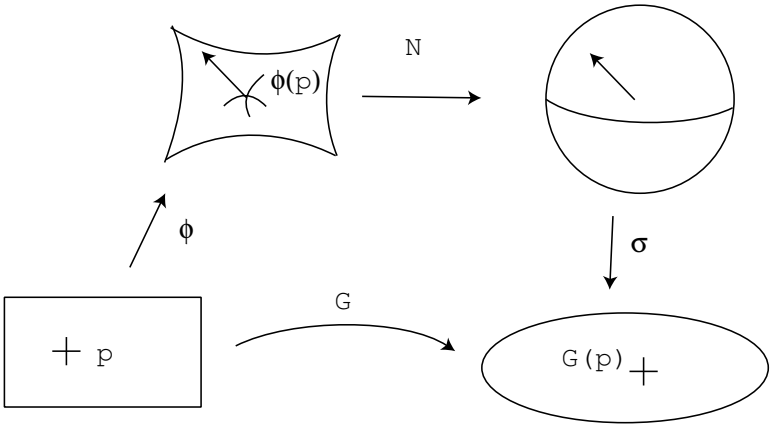


FIGURE 2. The Gauss map.

3B. Harmonic coordinate functions. Consider a conformally parametrized surface

$$\phi : \Omega \to \mathbb{R}^3,$$
$$z = u + iv \mapsto (x_1, x_2, x_3).$$

This surface is minimal if and only if the coordinate functions are harmonic, that is, if they satisfy

$$\Delta x_i = \Big(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\Big) x_i = 0.$$

In the domain Ω of definition, the harmonic coordinate differentials dx_i possess (locally) holomorphic extensions. It is customary to denote the holomorphic extension of the third coordinate differential dx_3 by dh and to call it the height differential. We have

$$dx_3 = \operatorname{Re} dh.$$

3C. Weierstrass data. Now suppose that we are given a conformal minimal surface parametrization $\phi = (\phi_1, \phi_2, \phi_3)$, defined in a domain Ω. Since the ϕ_j are harmonic functions in Ω, we can write them locally as the real parts of holomorphic functions Φ_j. Or, globally in a simply connected domain, as

$$\phi = \operatorname{Re} \int^p (\omega_1, \omega_2, \omega_3),$$

with holomorphic 1-forms ω_j defined in all of Ω. Suppose, on the other hand, we are given such a parametrization. This is conformal (and hence parametrizes a minimal surface) if and only if

$$\omega_1^2 + \omega_2^2 + \omega_3^2 \equiv 0$$

in Ω. The proof of this uses only the Cauchy–Riemann equations and is omitted.

The conformal factor λ^2 is given by

$$\lambda^2 = |\omega_1|^2 + |\omega_2|^2 + |\omega_3|^2. \tag{3–1}$$

Together we have

THEOREM 3.2. *Consider holomorphic forms $\omega_1, \omega_2, \omega_3$ in a simply connected domain Ω such that $\omega_1^2 + \omega_2^2 + \omega_3^2 \equiv 0$ and $|\omega_1|^2 + |\omega_2|^2 + |\omega_3|^2 \neq 0$. Then*

$$\phi(p) = \operatorname{Re} \int^p (\omega_1, \omega_2, \omega_3)$$

defines a conformally parametrized minimal surface in $\mathbb{R}^3$, and every conformal minimal surface parametrization arises this way.

We call the $\int \omega_j$ the holomorphic extensions of the coordinate functions.

We have associated two pieces of holomorphic data to a conformally parametrized minimal surface so far, namely its Gauss map and the holomorphic extensions of the coordinate functions. Because the Gauss map can be expressed in terms of the tangent vectors, it is natural to expect that the meromorphic function G can be written in terms of the holomorphic coordinate differentials.

In fact, using the formula for the stereographic projection (2–1), one obtains

$$G = \frac{-\omega_1 + i\omega_2}{\omega_3}.$$

This allows us to use meromorphic data G and $dh = \omega_3$ defined in a domain Ω to write down the surface parametrization in the so-called Weierstrass representation

$$\phi(z) = \operatorname{Re} \int^z \left(\frac{1}{2}\Big(\frac{1}{G} - G\Big),\ \frac{i}{2}\Big(\frac{1}{G} + G\Big),\ 1 \right) dh. \tag{3–2}$$

This integral may (and frequently will) depend on the path of integration if the domain Ω is not simply connected.

As explained, G depends on the choice of the stereographic projection, and dh depends on how we rotate the surface in space. However, a different choice changes G and dh only by a fractional linear transformation.

We will treat G and dh as the two basic holomorphic objects associated to a minimal surface. One of our main goals will be to understand how the geometry of a minimal surface is related to complex-analytic properties of G and dh.

3D. First examples.

EXAMPLE 3.3 (Catenoid and Helicoid). The catenoid is the only minimal surface of revolution worth mentioning. It is defined by

$$G(z) = z, \qquad dh = \frac{1}{z} dz.$$

We have to be careful with the choice of the domain Ω here. Depending on our knowledge, we can choose the safe method and take as Ω the simply connected plane slit along the negative real axis. Or, we could choose as Ω the universal cover of $\mathbb{C}^*$. No matter how we proceed, integrating these data gives the parametrization

$$\phi(z) = \operatorname{Re} \left(-\frac{1}{2}\Big(\frac{1}{G} + G\Big),\ \frac{i}{2}\Big(G - \frac{1}{G}\Big),\ \log z \right)$$

in $\Omega = \mathbb{C}^*$. Observe that while $\log z$ is not well-defined in $\mathbb{C}^*$, its real part is.

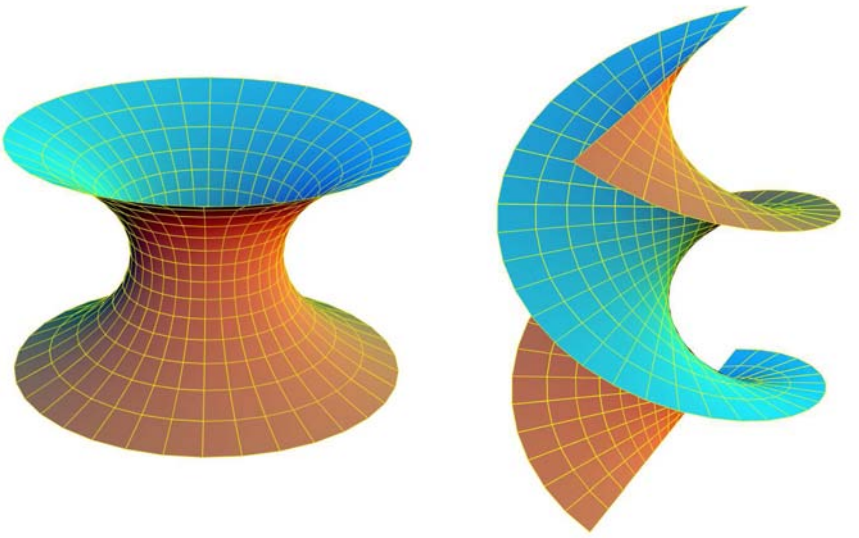


FIGURE 3. The catenoid and the helicoid.

The helicoid is the only ruled minimal surface worth mentioning:

$$G(z) = z, \qquad dh = \frac{i}{z}dz.$$

Integrating these data according to (3–2) gives the parametrization

$$\phi(z) = \operatorname{Re}\left(-\frac{i}{2}(\frac{1}{G} + G), -\frac{1}{2}(G - \frac{1}{G}), i\log z\right).$$

Here we face for the first time a problem which will become more and more prominent: While the data G and dh are well-defined in all of $\Omega = \mathbb{C}^*$, the integral $\operatorname{Re}\int^z \omega_3$ depends on the homotopy class of the path of integration. The effect is visible in the figure: The images of the same point in Ω reached by different paths differ by a vertical translation which is a multiple of 2π.

Technically, there are three ways out of this problem: We can restrict the domain Ω to a simply connected subset of Ω, we can use the universal cover of Ω, or we can consider surfaces into space forms $\mathbb{R}^3/\Gamma$.

EXAMPLE 3.4 (The Enneper surface). Taking

$$G(z) = z, \quad dh = z\,dz \tag{3–3}$$

in the complex plane defines the Enneper surface. Integrating the formula (3–2) gives the surface parametrization

$$\phi(z) = \operatorname{Re}\left(\frac{1}{2}\Big(z - \frac{z^3}{3}\Big),\ \frac{i}{2}\Big(z + \frac{z^3}{3}\Big),\ \frac{z^2}{2}\right).$$

The surface has two horizontal straight lines on which are the horizontal diagonals $y = \pm x$, $z = 0$. You can rotate the surface about them into itself. There are also reflectional symmetries about the two vertical coordinate planes.

Below are two pictures. The first shows a portion near the origin which is still without self-intersections, the second features the end. One can "see" in the second picture that the Gauss map has degree 3 near infinity — the surface sits

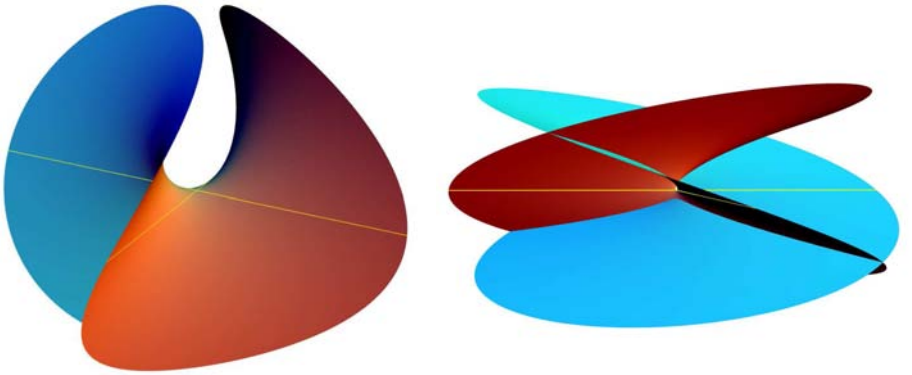


FIGURE 4. Different portions of the Enneper surface.

three times over the horizontal plane, and the Gauss map being close to infinity means that the normal vector points very much up.

MATHEMATICA NOTE 3.5. The Mathematica notebook weierstrass.nb shows how to integrate simple Weierstrass data and to plot the minimal surfaces.

3E. Local analysis at regular points. We begin our discussion of the relationship between holomorphic properties of the Weierstrass data and geometric properties of the minimal surface at regular points of the surface.

Suppose that a minimal surface is given by a conformal parametrization ϕ in a domain Ω. We have seen that this parametrization ϕ can be written in terms of data G and dh where the height differential dh is a holomorphic 1-form in Ω, and the *Gauss* map is a meromorphic function.

The Riemannian metric of a minimal surface can be computed in terms of the Weierstrass data by (3–1) and (3–2):

$$ds = \frac{1}{2}\left(|G| + \frac{1}{|G|}\right)|\,dh|\,.$$

This is precisely the conformal stretch factor of the isothermal surface parametrization. For a point $p \in \Omega$ to parametrize a regular surface point, this factor must be finite and nonzero. Hence there is a balancing condition for the orders of G and dh at regular points, namely

PROPOSITION 3.6. *At a regular point, G has a zero or pole of order n at p if and only if dh has a zero of order n.*

Otherwise, the metric would either become singular (zero or infinite) at p.

EXERCISE 3.7. *Devise simple examples of this situation and plot the corresponding surface. You can use the notebook weierstrass.nb as a starting point.*

EXAMPLE 3.8 (Surfaces defined on the entire complex plane). Let $P(z)$ and $Q(z)$ be two entire functions. Then

$$G(z) = \frac{P(z)}{Q(z)},$$
$$dh = P(z) \cdot Q(z)$$

provides Weierstrass data of a minimal surface so that all complex numbers are regular points. A very special case is given by $P(z) = z^k$ and $Q(z) = 1$. These surfaces are Enneper surfaces of higher dihedral symmetry. (Recall that the dihedral symmetry group is the symmetry group of a regular n-gon. Many minimal surfaces contain such a symmetry group).

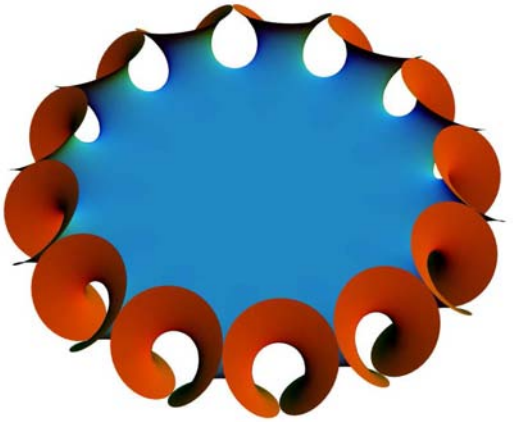


FIGURE 5. Enneper surface of order 12.

EXERCISE 3.9. *Represent the catenoid and the helicoid in this way. Hint: Use $G(z) = e^z$. Why is this possible, even though the catenoid appears to have two ends?*

MATHEMATICA NOTE 3.10. The Notebook OneEnd.nb discusses further examples of this type.

3F. The associate family. For any minimal surface Σ given by G and dh, there is a natural 1-parameter deformation Σ^t of minimal surfaces associated to it, given by G and $e^{it}dh$. The parametrization is given locally by

$$\phi^t : z \mapsto \operatorname{Re} \int^z e^{it}(\omega_1, \omega_2, \omega_3).$$

This family Σ^t of minimal surfaces is called the associate family of Σ.

All surfaces Σ^t are isometric minimal surfaces, as the square of the conformal factor of the Riemannian metric is given by

$$ds^2 = |\omega_1|^2 + |\omega_2|^2 + |\omega_2|^2$$

and thus independent of t. This has the consequence that the deformation can be carried out physically without tearing or stretching.

Also, the Σ^t have the same Gauss map by definition.

3G. The conjugate surface and symmetries. Of particular importance is the conjugate surface

$$\Sigma^* = \Sigma^{\pi/2}.$$

The simplest example of such a conjugate pair is the catenoid and the helicoid. The second simplest pair is the singly and doubly periodic Scherk surface.

Many important examples of minimal surfaces have extrinsic symmetries, that is, they are invariant under rotations about straight lines and reflections at planes. The curves which are fixed under these symmetries are landmarks of a minimal surface.

EXERCISE 3.11. *The associate family of the Enneper surface is a* 1*-parameter family of isometric surfaces in space. Show that the Enneper surface and its conjugate are isometric through a euclidean motion. This is not the case for two general members of the associate family.*

DEFINITION 3.12. Let Σ be a surface. A planar symmetry line is a curve on Σ which lies in a plane orthogonal to the surface everywhere along the curve. A straight symmetry line is a straight line lying entirely on Σ.

EXERCISE 3.13. *Discuss the symmetry lines of the catenoid and the helicoid.*

Using the Schwarz reflection principle for harmonic functions, a minimal surface which is bounded by a symmetry line of either kind can be continued by reflecting or rotating it about this line. The resulting surface will automatically be a smooth minimal surface.

This is very important, because many minimal surfaces can be decomposed by their symmetries into simply connected pieces which are simpler to deal with. Moreover, for drawing minimal surfaces, it requires less work to compute only the image of such a fundamental piece.

Furthermore, there is an intimate relationship between the two types of symmetry lines on a minimal surface and on its conjugate:

PROPOSITION 3.14. *A planar symmetry line for* Σ *becomes a straight symmetry line for* Σ^* *and vice versa. Under this correspondence, the normal vector of the symmetry plane becomes the direction vector of the straight line.*

The conjugate minimal surface plays also an important role in the construction of minimal surfaces or in proving their embeddedness. For more details, see [**Kar89**].

4. Minimal Surfaces on Punctured Spheres

We will discuss elementary properties of minimal surfaces and discuss a few examples defined on punctured spheres. It is a consequence of the López–Ros theorem (see [**LR91**]) that there are no complete embedded minimal surfaces parametrized by punctured spheres besides the plane and the catenoid. Nevertheless, these surfaces can teach us a great deal about the construction of minimal surfaces, because one of the principle difficulties, the closing of the periods, is relevant but easy to overcome.

4A. Period conditions. For a minimal surface given by holomorphic Weierstrass data $\omega = (\omega_1, \omega_2, \omega_3)$ on a domain X, the actual parametrization is obtained by the map

$$z \mapsto \operatorname{Re} \int^z \omega.$$

For this to be a single-valued map on the domain, we need that closed curves in the domain are being mapped to closed curves on the surface. This is automatically

true only for curves which are homologous to 0. In general, we want the following period condition to hold:

$$\operatorname{Re} \int_\gamma \omega = 0 \qquad \text{for all } \gamma \in H_1(X, \mathbb{Z}).$$

In many important cases the geometry of the minimal surface has preferred directions in euclidean space. For instance, if there are parallel ends, their common normal vector is such a preferred direction. Or, if there is a translational period, this is also a preferred direction. The Weierstrass representation often becomes simpler when such preferred directions are used as the coordinate vectors.

It is frequently helpful to look at the Weierstrass representation forms $G\,dh$ and $(1/G)\,dh$ instead of at the forms ω_1 and ω_2. The period conditions can then be stated as

$$\int_\gamma G\,dh = \overline{\int_\gamma \frac{1}{G}\,dh}, \qquad \operatorname{Re} \int_\gamma dh = 0,$$

for all cycles γ on X (this is easy algebra).

We will call the first condition the horizontal period condition and the second the vertical period condition.

4B. A surface with a planar end and an Enneper end. We begin our list of examples with a minimal surface parametrized by a punctured plane, with one puncture at 0 corresponding to a planar end and the puncture at ∞ corresponding to an Enneper end. This surface has no further ends. To achieve this, we make sure that the conformal stretch factors at the ends have the correct asymptotic behavior. For the Enneper end, recall the Weierstrass representation of the Enneper surface of order k was given in Example 3.8,

$$G(z) = z^k, \qquad dh = z^k\,dz$$

so that the conformal stretch factor near infinity (where we approach the end) becomes

$$ds \sim |z^{2k}| \cdot |\,dz|.$$

This makes it plausible to define an Enneper end of order k of an arbitrary minimal surface to be a region of the surface, parametrized by a neighborhood of ∞ such that the Weierstrass representation data G and dh have at ∞ the same order as in the prototype Enneper surface. Hence if we choose

$$G(z) = z^3, \qquad dh = z\,dz,$$

this surface will have an Enneper end of order 2 at infinity, and the puncture at 0 represents a planar end. To see this, we write $z = 1/w$ and obtain $G(w) = w^3$, $dh = -4w^{-3}dw$, so that near $w = \infty$ we get $ds \sim |dw|$ so that the metric becomes euclidean. Note that a planar end is distinguished from a catenoidal end (which has also euclidean growth) by the vanishing of the logarithmic growth rate, which is the residue of dh.

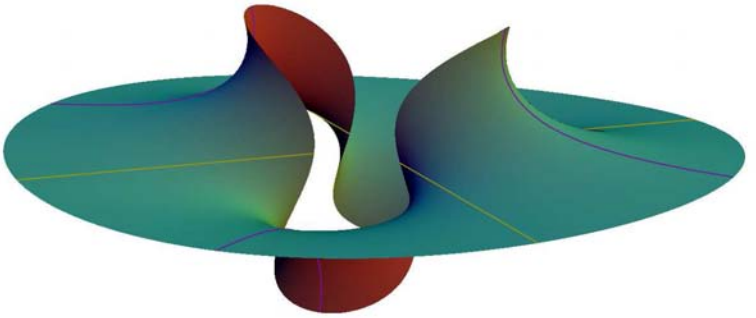


FIGURE 6. A sphere with two ends.

4C. A sphere with one planar and two catenoid ends. This example shows how the attempt to find a minimal sphere with three parallel embedded ends fails.

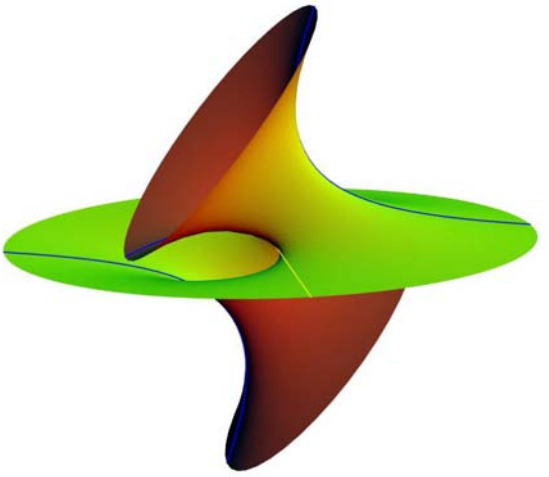


FIGURE 7. A sphere with three ends.

Let's look for a sphere with three ends, one of them planar, the other two catenoidal. Let's arrange the surface so that the Gauss map at the two catenoid ends is vertical and has two simple poles (this is an additional assumption, the other cases can be discussed in a similar way). The same will then hold for the height differential, with the residues being real.

We won't assume anything about the Gauss map at the planar end, so that we can find out whether it might be vertical.

Next let's normalize the parameter sphere so that the planar end is represented by infinity and the two catenoid ends by ± 1.

As we want a straight line on the surface, this line corresponds to the imaginary axis, and the Gauss map must be real on it.

So far we have shown that the Weierstrass representation is of the form

$$G(z) = \rho \frac{az^2 + bz + c}{z^2 - 1}, \qquad dh = \frac{az^2 + bz + c}{z^2 - 1}\, dz.$$

The condition that the Gauss map is real on the imaginary axis forces $b = 0$.

Following our philosophy that the geometric type of an end is determined by the asymptotic expansion of the Weierstrass data near the puncture representing the end, the vertical period condition at ∞ becomes

$$\operatorname*{res}_{z=\infty} dh = -b = 0$$

so that we have $b = 0$. Since the form $(1/G)\,dh = (1/\rho)\,dz$ is exact, the vertical and horizontal period conditions for cycles around ± 1 become

$$\operatorname*{res}_{z=\pm 1} G\,dh = \pm\tfrac{1}{4}(3a - c)(a + c)\rho = 0,$$
$$\operatorname*{res}_{z=\pm 1} dh = \pm\tfrac{1}{2}(a + c) = 0.$$

Hence $c = 3a$. Furthermore, a can't vanish, because then also $c = 0$, and the Weierstrass data would degenerate. Without loss we can assume that $a = 1$ and $c = 3$. This solves all period conditions and leaves us with a free parameter, the positive López–Ros parameter ρ. This is also the value of the Gauss map at ∞, the planar end. Since this parameter can't be 0 (otherwise the Gauss map would be identically 0), the Gauss map can't be vertical at the planar end. Hence the ends will never be parallel and must intersect eventually. Observe that the Weierstrass data are exact in this case.

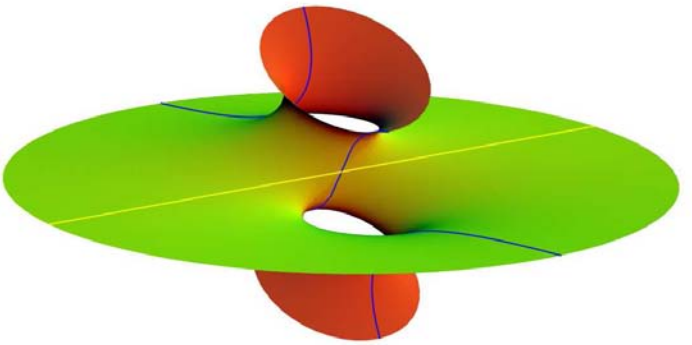


FIGURE 8. Another sphere with three ends.

4D. The k-noids of Jorge and Meeks. Let's try to find the Weierstrass data of a minimal surface Σ_k which looks like the surface below for $k = 6$:

Topologically, this is a sphere with k holes. Since each of the ends should be a catenoid end, we expect the underlying domain to be a sphere, punctured at k points p_k representing the ends. Assuming a rotational symmetry, we can choose these points to be the k^{th} roots of unity. We see and hope that the Gauss map is only vertical at the intersection of the surface with the z-axis. Again by symmetry, these two points correspond to 0 and ∞. Thus we expect $G(z) = \rho z^m$ for some integer m. Near a point where the Gauss map has degree m, the surface looks like an umbrella with $m + 1$ valleys and $m + 1$ ridges. Hence we obtain

$$G(z) = \rho z^{k-1}$$

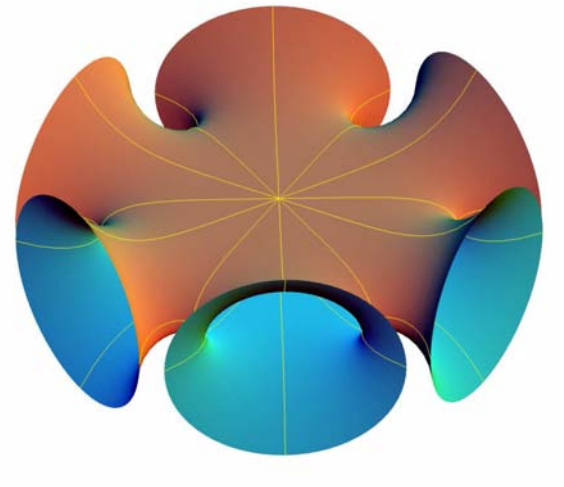


FIGURE 9. A Jorge–Meeks 6-noid.

with ρ to be determined. Now let's turn to the height differential. We claim that it has to be

$$dh = \lambda \frac{z^{k-1}}{(z^k-1)^2}\, dz.$$

This is because we need dh to have zeroes of order $k-1$ at 0 and ∞ to make these points regular and we need single order poles at the p_k to make these points represent catenoidal ends. From the residues

$$\operatorname*{res}_{z=1} G\, dh = \lambda \frac{(k-1)\,\rho}{k^2}, \qquad \operatorname*{res}_{z=1} \frac{1}{G}\, dh = \lambda \frac{1-k}{k^2\,\rho}, \qquad \operatorname*{res}_{z=1} dh = 0$$

and the period condition we conclude that

$$|\rho|^2 = \overline{\lambda}/\lambda,$$

so that we can and have to choose without loss $\lambda = 1$ and $\rho = 1$.

EXERCISE 4.1. *Verify that for $k = 2$ we get a surface which is isometric to the catenoid.*

5. The Scherk Surfaces

We discuss the classical Scherk surfaces and some of its variations.

5A. The singly periodic Scherk surface. We begin the discussion this time by starting with the Weierstrass data and explaining the geometric properties of the surface through them.

The singly periodic Scherk surface is defined on the four-punctured sphere $\widehat{\mathbb{C}} - \{\pm 1, \pm i\}$ by the meromorphic Weierstrass data

$$G(w) = w, \qquad dh = \frac{i\,w}{w^4-1}\, dw.$$

Since the Gauss map is the identity map, surface points with known normal vectors can easily be spotted in the parameter sphere. In particular, if w approaches ± 1 or $\pm i$, the Gauss map converges to these values.

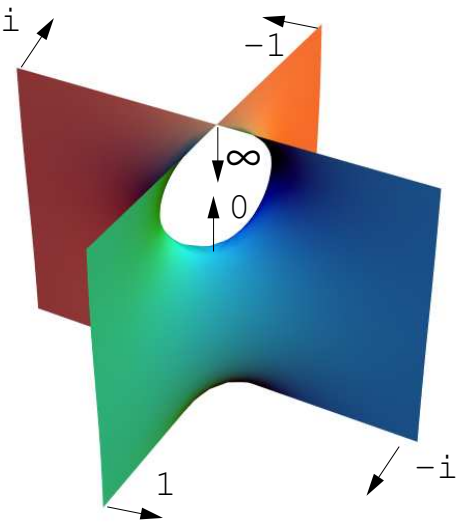


FIGURE 10. A fundamental piece of Scherk's singly periodic surface.

5B. The period conditions. We would like to understand first why this surface is invariant under a (vertical) translation in space. The reason is that closed curves γ in $\widehat{\mathbb{C}} - \{\pm 1, \pm i\}$ are mapped by the parametrization to curves connecting points on the surface which are equivalent under vertical translations.

On the other hand, the Weierstrass representation tells us that the difference vector of two such points is given by the real part of the so-called period vector

$$\oint_\gamma \left(\frac{1}{2}\Big(\frac{1}{G} - G\Big)\, dh,\ \frac{i}{2}\Big(\frac{1}{G} + G\Big),\ dh\right).$$

Since any curve on the surface is homologous to cycles around the punctures, it suffices to check that this period vector is an integral multiple of a fixed vertical vector for closed curves around the punctures $\epsilon \in \{\pm 1, \pm i\}$. These integrals can be evaluated by the residue theorem:

1. Vertical periods:

$$\oint_\epsilon dh = 2\pi i \operatorname*{res}_{w=\epsilon} \frac{iw}{w^4 - 1}\, dw = -\frac{\pi}{2\epsilon^2} \in \left\{\pm\frac{\pi}{2}\right\}.$$

2. Horizontal periods:

$$\oint_\epsilon G\, dh = 2\pi i \operatorname*{res}_{w=\epsilon} \frac{iw^2}{w^4 - 1}\, dw = -\frac{\pi}{2\epsilon},$$
$$\oint_\epsilon \frac{1}{G}\, dh = 2\pi i \operatorname*{res}_{w=\epsilon} \frac{i}{w^4 - 1}\, dw = -\frac{\pi}{2\epsilon^3}.$$

Since $\epsilon \in \{\pm 1, \pm i\}$, we have indeed

$$\operatorname{Re} \oint_\epsilon \Big(\frac{1}{G} - G\Big)\, dh = 0 = \operatorname{Re} \oint_\epsilon i\Big(\frac{1}{G} + G\Big)\, dh,$$

so that no horizontal periods occur.

As explained, the coordinate functions of a minimal surface are harmonic and hence possess locally holomorphic extensions. It turns out that the mapping behavior of the complex coordinate maps

$$z \mapsto \int^z G\,dh,$$
$$z \mapsto \int^z \frac{1}{G}\,dh$$

is very often surprisingly simple.

For the singly periodic Scherk surface, the restrictions of these maps to the first quadrant have the following image domains, consisting each of two orthogonal half-infinite parallel strips of the same width:

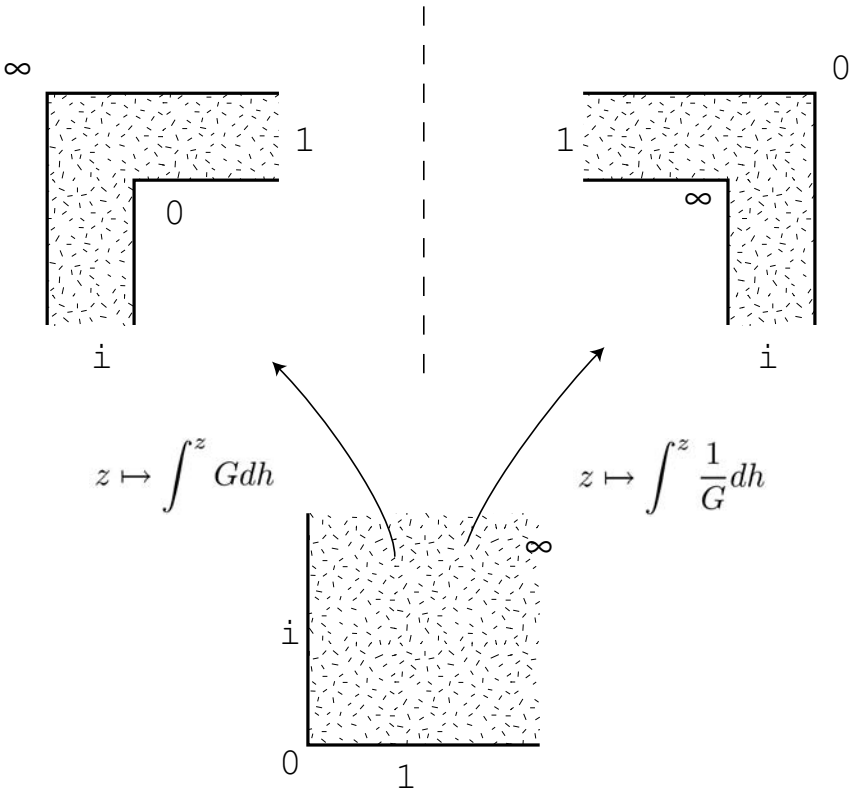


FIGURE 11. Image of the first quadrant under complex coordinate maps.

This follows from the Schwarz–Christoffel formula (see [**Ahl79**]). It is a rather surprising phenomenon that the relatively complicated minimal surface patch can be understood through pieces which have an elementary geometric description. We will explain and exploit this later again.

5C. Changing the angle between the ends. As already known to Scherk, the angle between the ends can be varied both for the singly and the doubly periodic

surface. Here is the Weierstrass representation for the singly periodic surface:

$$G(z) = z,$$
$$dh = \frac{dz}{z(z^2 + 1/z^2 - 2\cos 2\phi)} = \frac{dz}{(z^2 - e^{i2\phi})(z^2 - e^{-i2\phi})}.$$

The angle ϕ is half the angle between the Scherk ends, it varies between 0 and $\pi/2$. For $\phi = \pi/4$, we get the classical Scherk surfaces with orthogonal ends.

EXERCISE 5.1. *What do the period conditions mean in terms of Figure 12?*

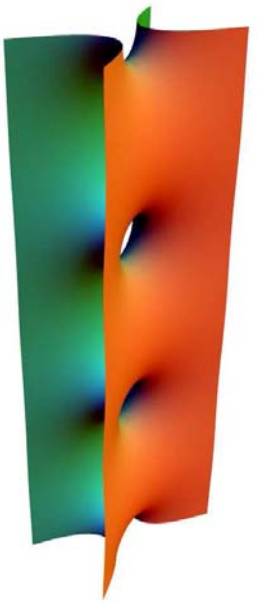


FIGURE 12. A sheared singly periodic Scherk surface.

EXERCISE 5.2. *Verify that the period conditions are satisfied.*

EXERCISE 5.3. *What happens for $\phi \to 0$?*

5D. The doubly periodic Scherk surface. Scherk wrote down his doubly periodic surface as a graph over the square $Q = (-\pi/2, \pi/2) \times (-\pi/2, \pi/2)$:

$$e^z = \frac{\cos y}{\cos x}.$$

The same formula defines a translation-invariant graph over all squares $Q + \pi k + \pi i l$ with $k + l \equiv 0 \pmod 2$ (think of these squares as the black squares of a checkerboard tessellation of the plane).

This parametrization is not very well suited for creating images of the surface. It is not conformal, and it doesn't behave nicely at the ends. Also, it doesn't show that the translational copies of the graph fit together along vertical straight lines over the corners of the squares.

How can we find a (better) conformal parametrization of the surface?

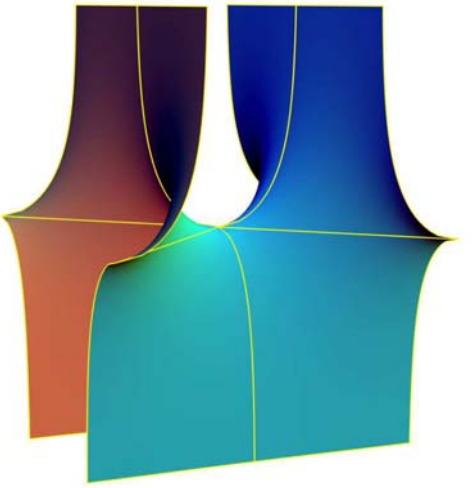


FIGURE 13. A fundamental piece of Scherk's doubly periodic surface.

MATHEMATICA NOTE 5.4. In the notebook Scherk.nb, we take the above surface parametrization and compute the Gauss map. The inverse of this map composed with the surface parametrization gives a conformal surface parametrization. This is used to derive a Weierstrass representation of the surface.

Here we will follow a more geometrical approach: Let's look at a fundamental piece of the surface, divided by its translational symmetries. It is cut into eight congruent curved pieces by straight and planar symmetry lines. In the parametrization of the surface patch as a graph over the square $(-\pi/2, \pi/2) \times (-\pi/2, \pi/2)$, these lines correspond to the diagonals and medians of the square. Note that the horizontal 'lines' at the top and bottom of the image are not truly straight lines , they just arise from cutting off the ends of the surface. Each of the three boundary curves of one eighth of the fundamental piece is mapped by the normal map to a quarter of a great circle in the image S^2, and the curves intersect under $\pi/4$ angles. Hence the symmetry curves are mapped by the Gauss map to the octahedron tessellation of the round sphere. The round S^2 is identified with the Riemann sphere (by the stereographic projection) so that the four ends of the surface become the fourth roots of unity (we might need to rotate the surface for this to hold), and the center point of the piece is mapped to 0, while the four (identified) corners are mapped to infinity.

If we use this Riemann sphere as a new domain to parametrize the surface, the Gauss map will just be the identity map.

The height differential will be a meromorphic 1-form on this Riemann sphere with two simple zeroes at 0 and ∞ where the Gauss map is vertical. Since the sum of the orders of the zeroes and poles of a meromorphic form on the Riemann sphere is -2, we need to distribute four simple poles. We are forced to put these at the ends. Hence

$$G(z) = z, \qquad dh = \lambda \frac{z}{z^4 - 1}\, dz.$$

It remains to determine λ. Since

$$\operatorname*{res}_{z=1} dz = \frac{\lambda}{4},$$

the ends cause no vertical periods if and only if λ is real. Without loss, we can put $\lambda = 1$. Observe that

$$\operatorname*{res}_{z=1} G\,dz = \operatorname*{res}_{z=1} \frac{1}{G}\,dz = \frac{1}{4},$$

so that each end produces horizontal periods, as wanted. So we see that Scherk's doubly periodic minimal surface is conjugate to his singly periodic surface. Scherk didn't know this, because the notion of conjugate surfaces was only discovered later.

By looking at the conjugate of Scherk's singly periodic surface with arbitrary angle between the ends, one obtains sheared Scherk surfaces like the one in Figure 14. Lazard-Holy and Meeks [**LHI98**] have shown that these are the only doubly periodic surfaces of genus 0 (besides the plane).

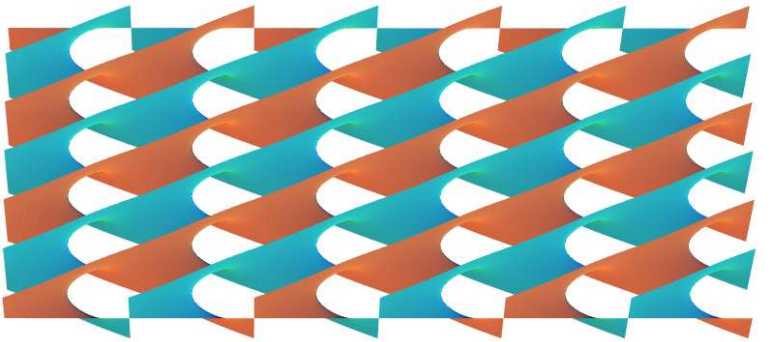


FIGURE 14. A sheared doubly periodic Scherk surface.

5E. Singly periodic Scherk surfaces with higher dihedral symmetry. We will now produce singly periodic Scherk surfaces with higher dihedral symmetry in the same way as we got the Enneper surfaces of higher symmetry. Here is the Weierstrass representation for $2k$-ended surfaces, with an additional parameter ϕ:

$$G(z) = z^{k-1},$$

$$dh = \frac{dz}{z(z^k + 1/z^k - 2\cos k\phi)} = \frac{z^{k-1}dz}{(z^k - e^{ik\phi})(z^k - e^{-ik\phi})}.$$

These surfaces are embedded as long as

$$\frac{\pi}{2} - \frac{\pi}{k} < (k-1)\phi < \frac{\pi}{2}.$$

This is, however, quite hard to see with the techniques we have developed so far.

The inequality ensures at least that the planar ends do not intersect, so the surfaces will be embedded outside of a cylinder. An additional argument like the maximum principle is required to prove embeddedness everywhere.

The most symmetric case, where the asymptotic planes of neighboring ends intersect at the same angle π/k, is obtained for $\phi = \pi/(2k)$.

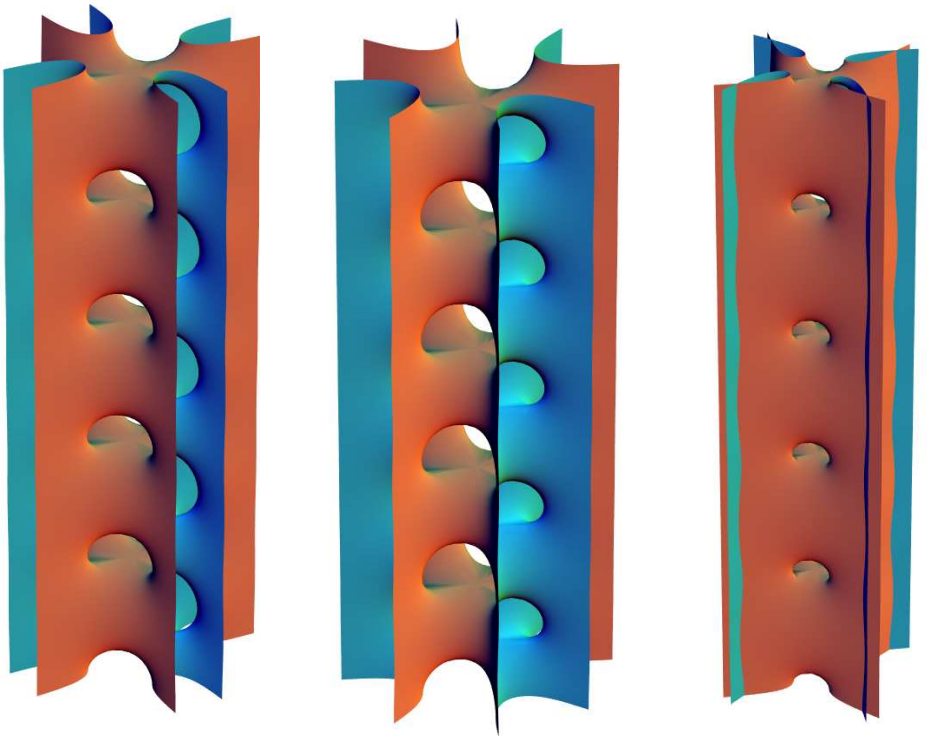


FIGURE 15. Singly periodic Scherk surface with eight ends.

EXERCISE 5.5. *Verify the last claim.*

Observe that for $\phi = 0$, we obtain the Weierstrass representation of the Jorge–Meeks surfaces. One can 'see' how the surfaces behave geometrically when $\phi \to 0$ by rescaling the surfaces correctly. The right part of Figure 15 shows a Scherk surface close to several very large 4-noids. If we fix the size of the 'holes', the holes move further and further apart and disappear in the limit. Of course, while plausible from the pictures, these statements require careful estimates to be proven.

One can even obtain less symmetric Scherk-like surfaces, using a conjugate Plateau construction in combination with a theorem of Jenkins and Serrin, see [**Kar88**].

5F. The twist deformation of Scherk's singly periodic surface. There is another, more complicated deformation of the singly periodic Scherk surfaces [**Kar88, Web00a**], shown in Figure 16.

These twisted Karcher–Scherk surfaces can be obtained from the classical Scherk surface by twisting them about the vertical coordinate axes. They are invariant under a vertical screw motion with twist angle $\alpha = a\pi$ and exist for $0 \leq a < \frac{1}{2}$. The quotient surface is again a four-punctured sphere with punctures at $\pm R, \pm i/R$, where R is a real parameter to be determined so that a horizontal period condition is satisfied. While the height differential is still quite simple, the

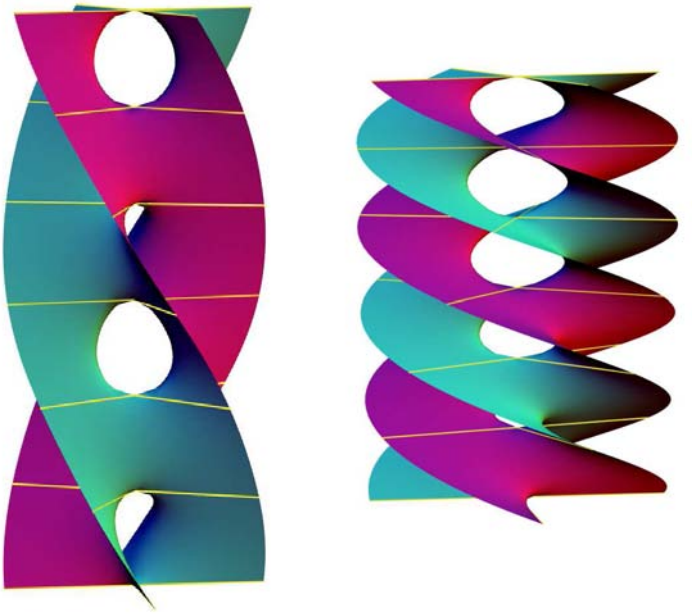


FIGURE 16. Singly periodic Scherk surface with four helicoidal ends.

Gauss map is more complicated. Here is the formula from [**Kar88**]:

$$G(w) = w \cdot \left(\frac{w^2 - R^2}{R^2 w^2 + 1} \right)^a,$$
$$dh = \frac{iw}{(w^2 - R^2)(R^2 w^2 + 1)}\, dw.$$

For $a = 0$ and $R = 1$, we obtain the classical Scherk surface.

The Gauss map has the feature that it is not anymore single valued on the quotient surface, and this makes a discussion of the horizontal period problem for this surface a bit painful. A detailed discussion using the viewpoint of Schwarz–Christoffel maps can be found in [**Web00a**].

6. Minimal Surfaces Defined on Punctured Tori

We discuss minimal surfaces defined on punctured tori. We recall basic facts from complex variables about complex tori and elliptic functions, all of which can be found in standard text books like [**Ahl79**]. For the use of geometrically defined elliptic functions in the context of minimal surfaces, see also [**HKW93**].

6A. Complex tori. A complex torus is given by a lattice Λ in $\mathbb{C}$, that is, a free abelian subgroup of rank 2 in $\mathbb{R}^2$. Such a subgroup can be written as

$$\Lambda = \{a\omega_1 + b\omega_2 : a, b \in \mathbb{Z}\}.$$

Here ω_1 and ω_2 are two complex numbers which are linearly independent over $\mathbb{R}$. They form a basis of the lattice, and the parallelogram spanned by them is called a fundamental parallelogram of the torus.

A meromorphic function $f(z)$ on $\mathbb{C}$ is called elliptic if it is periodic with respect to Λ, that is if

$$f(z+\omega) = f(z) \qquad \text{for all } \omega \in \Lambda.$$

Associated to each elliptic function is its divisor

$$(f) = \sum n_j P_j,$$

which is a formal linear combination of the zeroes and poles P_j of f in $T = \mathbb{C}/\Lambda$ with their multiplicities n_j.

The degree of a divisor is

$$deg(f) = \sum n_j.$$

The existence of meromorphic functions with given divisors is settled by an important result:

THEOREM 6.1 (Abel). *There is an elliptic function $f(z)$ on a torus $T = \mathbb{C}/\Lambda$ with divisor $\sum n_j P_j$ if and only if*

(1) $\sum n_j = 0$,
(2) $\sum n_j P_j \in \Lambda$.

EXAMPLE 6.2. On any torus, there is a meromorphic function with a double order pole at the corners of the fundamental parallelogram and a double order zero at the center.

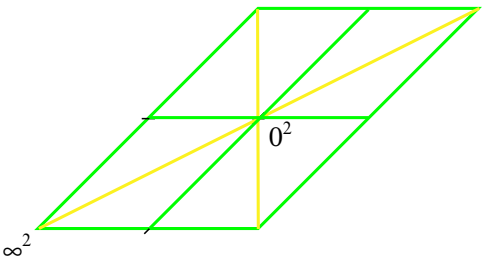


FIGURE 17. A divisor.

6B. Algebraic equations. Take an elliptic function $p(z)$ with double order pole and double order zero on a torus T. Then $p'(z)$ has a triple order pole and three zeroes, at which p takes the values $0, a, b$. Divisor considerations show that $p(z)$ must satisfy a differential equation

$$p'(z)^2 = c \cdot p(z) \cdot (p(z) - a) \cdot (p(z) - b).$$

We can scale $p(z)$ and the lattice such that

$$p'(z)^2 = p(z)(p(z) - 1)(p(z) - \lambda).$$

The constant λ is called the modular invariant of the torus T.

In principle, every elliptic function on T can be expressed algebraically in terms of $p(z)$ and $p'(z)$. Instead of working with z as a coordinate of T and computing elliptic functions by computing first $p(z)$ and $p'(z)$, we can as well use $v = p(z)$ as a coordinate and compute $w = p'(z)$ using

$$w^2 = v(v-1)(v-\lambda). \tag{6–1}$$

This equation can indeed be suitably understood as defining a set of points which is complex isomorphic to T.

Hence: Working with elliptic functions on T is equivalent to working with rational functions in v and w such that 6–1 holds.

This approach has the advantage that it generalizes easily to higher genus Riemann surfaces. But it is limited insofar as it is not always easy to express a given elliptic function in terms of p and p'.

6C. Theta functions. On the Riemann sphere, the meromorphic functions are just the rational functions, which are products and fractions of linear functions. It is not quite as simple for meromorphic functions on tori, but we can come close to it if we use theta functions instead of linear functions.

To any lattice, spanned by 1 and τ, we associate an entire function $\theta(z;\tau)$:

$$\theta(z;\tau) = \sum_{n=-\infty}^{\infty} e^{\pi i(n+1/2)^2\tau + 2\pi i(n+1/2)(z+1/2)}.$$

This complicated-looking expression has the advantage that it converges rapidly. Here are the important properties of θ:

$$\begin{aligned} \theta(0;\tau) &= 0 \qquad \text{(simple zero)}, \\ \theta(z+1;\tau) &= -\theta(z;\tau), \\ \theta(z+\tau;\tau) &= e^{2\pi i(z+\frac{\tau+1}{2})}\theta(z;\tau). \end{aligned}$$

In the following, we will omit the second argument τ. In a fundamental parallelogram, $\theta(z)$ has no further zeroes, and this allows the construction of meromorphic functions with given transformation laws with respect to the lattice periods:

LEMMA 6.3. *Let $a_i, b_i \in \mathbb{C}, i = 1 \dots n$. Then*

$$f(z) = \prod_{i=1}^{n} \frac{\theta(z-a_i)^{\alpha_i}}{\theta(z-b_i)^{\beta_i}}$$

has simple zeroes at a_i, simple poles at b_i and satisfies

$$\begin{aligned} f(z+1) &= (-1)^{\sum \alpha_i - \beta_i} f(z), \\ f(z+\tau) &= e^{2\pi i \sum \alpha_i a_i - \beta_i b_i} f(z). \end{aligned}$$

EXERCISE 6.4. *Use this to prove the existence part of Abel's theorem!*

6D. Elliptic functions via Schwarz–Christoffel maps. Any conformal torus can be specified by taking a parallelogram in the euclidean plane and identifying opposite edges. This torus comes with a flat metric which also defines the conformal structure. It will be helpful to think of this flat metric as a special *flat geometric structure* in the sense that it can be given by local coordinate charts into the euclidean plane $\mathbb{E}$ (which are the identity maps on open subsets of the interior of the parallelogram) so that the change of coordinate maps are just translations.

On the other hand, given such a geometric structure, the exterior derivatives of the coordinate charts fit together to define a globally well-defined holomorphic (with respect to the inherited conformal structure) 1-form on the torus. In our example, this is just the 1-form dz inherited from the euclidean plane. It is worth noting that the periods of this 1-form are *visible* in the geometry of the flat structure, namely as the edge vectors of the parallelogram.

We will now repeat this construction in a slightly more complicated situation. Take again a parallelogram in the euclidean plane, but now remove it from the plane and identify again opposite edges from the remainder (figure 18).

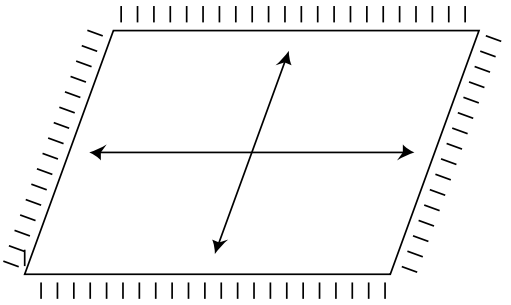


FIGURE 18. Cone metric construction of meromorphic 1-forms.

Topologically, we get a punctured torus, with the puncture corresponding to the point ∞ of the extended plane. Again this torus carries a flat geometric structure of the same type as before, but this time with two singular points: One corresponds to the point ∞ in the euclidean plane, the other to the vertices of the parallelogram which are identified to a single conical point with cone angle 6π. Away from these singularities, this flat structure defines again a holomorphic 1-form on the torus by taking the exterior derivative of local charts.

To analyze the behavior at the singular points, we first look at a neighborhood of ∞. Here the geometry is precisely the same as for dz at ∞ in the euclidean plane, which has a double order pole there. Hence our holomorphic 1-form extends meromorphically to ∞ with a double order pole. The singularity at the other singular point (the parallelogram vertex) is removable, as the exterior derivative of the coordinate function nearby is bounded. Hence the 1-form also extends to this point, and by the residue theorem, it has a double-order zero there. In fact, in the next section we will see that a cone angle α leads to a zero of order $\frac{\alpha}{2\pi} - 1$.

The periods of this 1-form are again easy to read off as the parallelogram edges. To avoid confusion: The parallelogram torus is in general *not* the same conformal torus as the 'removed parallelogram torus'. A notable exception is the square torus.

At first glance, surprisingly, this family of 'removed parallelogram tori' does not exhaust the set of tori with meromorphic 1-forms with double order pole and zero. Firstly, the parallelogram can degenerate without the torus degenerating!

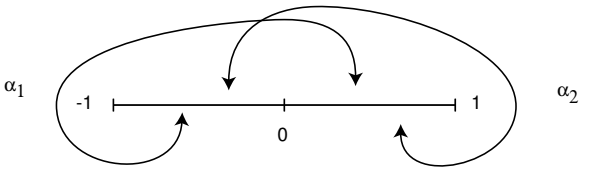


FIGURE 19. A rhombic torus with real periods.

To see this, take once again the euclidean plane and slit it along the interval $[-1, 1]$. Choose a number $a \in [0, 1)$ and glue the $[-1, -a]$ portion of the upper

(resp. lower) part of the slit to the $[a,1]$ portion of the lower (resp. upper) part of the slit.

In Figure 19, we have indicated this for $a = 0$. This identification gives again a flat torus with two singular points with the same cone angles as before. The periods of the corresponding 1-form are now both real, namely $-a$ and $1-a$. Within this 1-parameter family there is precisely one rhombic torus, corresponding to $a = 0$: In this case, the reflections about the real and imaginary axes are well defined on the torus, and have a connected fixed point set each, which implies rhombicity. To see that this rhombic torus is unique requires a bit more work. The key idea is that on a rhombic tori, the reflections at the two lines through the pole and zero of the 1-form conjugates the 1-form. This implies that the geometric model must be invariant under two reflections, too.

For rectangular tori, this construction can be understood more concretely in terms of Schwarz–Christoffel maps. The complement of a rectangle in the plane is the image domain of a fundamental rectangle under a map $\int^z p(z)dz$, where $p(z)$ is an elliptic function with double order pole and double order zero in the torus. We normalize the map by translation so that it has its double at the lower left corner of the fundamental rectangle.

If we use the reflectional symmetries of the rectangle to decompose the rectangle into four congruent smaller rectangles half the size of the original rectangle, each of these smaller rectangles will be mapped by $\int^z p(z)dz$ to a quarter of the complement-of-the-rectangle-domain. (Note, however, that in general we can't expect that the original rectangle and the one used in the complement construction are the same). Now both domains, the small rectangle and the complementary quarter, are euclidean polygons which are images of the same upper half plane under two Schwarz–Christoffel maps: The rectangle is obtained using

$$\int^p u^{-1/2}(u-1)^{-1/2}(u-\lambda)^{-1/2}du$$

and the complementary quarter is given by

$$\int^p u^{+\frac{1}{2}}(u-1)^{-1/2}(u-\lambda)^{-1/2}du.$$

Here λ is the modulus of the torus (or rectangle), it is a negative real number. If we use $u = p(z)$ as a torus coordinate, the identity map on the torus can be rewritten as

$$z = \int^z 1\,dz = \int^p \frac{1}{p'(z)}\,du = \int^p \frac{1}{\sqrt{u(u-1)(u-\lambda)}}\,du,$$

and similarly, the map to the exterior of the rectangle can be rewritten as

$$\int^z p(z)\,dz = \int^p \frac{p(z)}{p'(z)}\,du = \int^p \frac{\sqrt{u}}{\sqrt{(u-1)(u-\lambda)}}\,du.$$

This suggests that the function p, which we have just abused as a coordinate function, maps the quarter-rectangle conformally to the upper half plane. One can prove this easily by studying its boundary behavior or using the above computations.

It makes it in fact possible to define $p(z)$ by this property, see [**HKW93**].

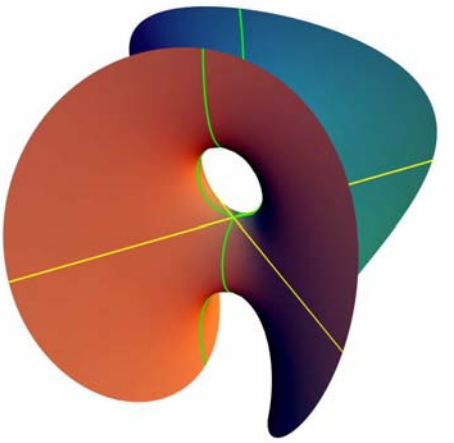


FIGURE 20. The Chen–Gackstatter surface.

6E. The Chen–Gackstatter surface. The Chen–Gackstatter surface was the first complete minimal surface of finite total curvature defined on a torus. It has the same symmetries as the Enneper surface, which implies that the only possible torus it can live on is the square torus. It has just one end of Enneper type at which the Gauss map can be assumed to have a simple zero and the height differential a triple order pole. We can put the puncture representing the end at the lower left corner of the fundamental square. The straight lines on the surface become the diagonals of the square and the planar symmetry lines become the edge bisectors and boundary edges. They intersect on the surface in the end at infinity and in the three points with vertical normal, and on the square at the four half-period points. This determines the divisor of G and dh as follows:

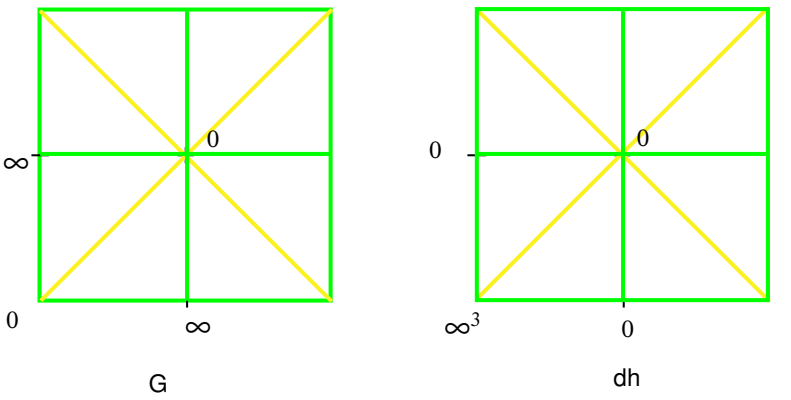


FIGURE 21. Divisors of the Weierstrass data.

This suggests that we use the algebraic description, since both dh and G are easily expressible in terms of w and v. The underlying torus T is the square

torus, given by the algebraic $w^2 = v(v-1)(v+1)$. The Gauss map G and height differential dh are given by

$$G = \rho v/w, \qquad dh = dv,$$

where ρ can be explicitly determined as

$$\rho = \frac{\Gamma(\frac{1}{4})}{\sqrt{6}\Gamma(\frac{3}{4})}$$

to satisfy the horizontal period condition.

MATHEMATICA NOTE 6.5. The notebook ChenGackstatter.nb integrates the Weierstrass data for the Chen–Gackstatter surface in terms of hypergeometric functions, computes the López–Ros parameter and plots the surface.

If we restrict our attention to the quarter of the surface which is given by the upper v-plane (which is the piece of the surface one gets if one cuts it apart by the vertical coordinate planes), the complex coordinate maps can be expressed as Schwarz–Christoffel maps:

$$\int^p G\,dh = \int^p v^{1/2}(v-1)^{-1/2}(v+1)^{-1/2}\,dv,$$
$$\int^p \frac{1}{G}\,dh = \int^p v^{-1/2}(v-1)^{1/2}(v+1)^{1/2}\,dv,$$

so that these two maps map the upper half plane to domains which are complementary regions (after a reflection). The fact that the image regions are complementary is in fact equivalent to the period conditions being satisfied. The images of fundamental cycles on the torus are mapped under $\int^p G\,dh$ and $\int^p \frac{1}{G}\,dh$ to edge vectors of the two image regions, and being complementary (after a reflection) means nothing but being conjugate, which was our reformulation of the horizontal period condition.

For the uniqueness of this surface, see [**Lop92, Web99**].

6F. Riemann's minimal surface. Riemann's minimal surface is a singly periodic surface invariant under a translation in space.

The quotient of the surface by its translational symmetry group is a torus with two ends. We arrange the surface in space so that the Gauss map is infinite at one end and zero at the other. The ends are planar, so the height differential will be regular there. Since there are no further ends, the height differential must be holomorphic, hence a constant multiple of dz. This means that the Gauss map is nowhere else vertical.

The symmetries imply that the underlying torus is rectangular, as shown in Figure 23.

The vertical period condition requires only that one of the two basic periods is imaginary, which is the case for dz. Also, the horizontal period condition needs only to be satisfied for one period. If we normalize $G(z) = \rho p(z)$ such that $G(z)$ becomes a translation of $1/G(z)$ by $(1+\tau)/2$, $G\,dh$ and $(1/G)\,dh$ have the same periods. Since for one cycle these are real, they are also complex conjugate.

A striking property of Riemann's minimal surfaces is that they are foliated horizontally by circles (or, at the level of the ends, by straight lines). See [**Rie67, MPR98**].

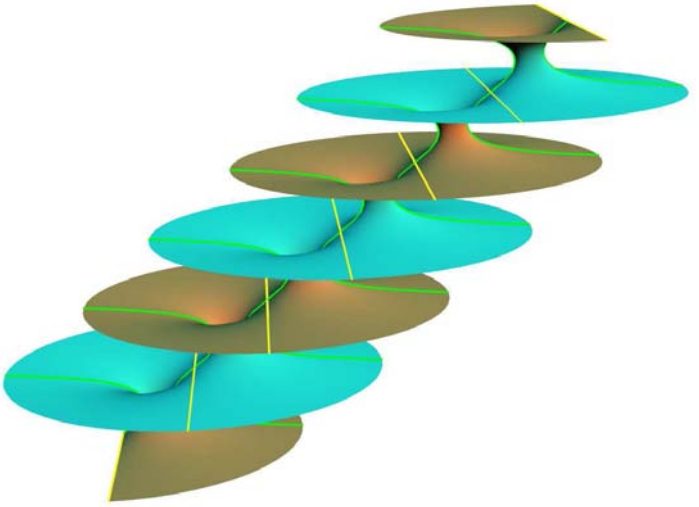


FIGURE 22. Riemann's minimal surface.

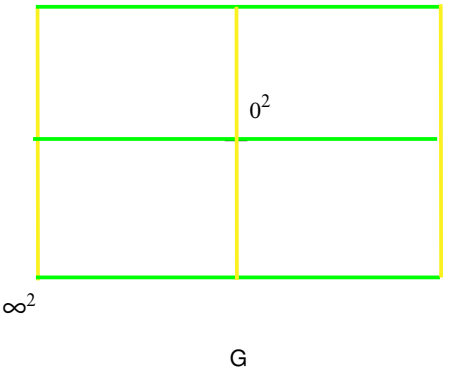


FIGURE 23. Divisor of the Gauss map.

6G. The fence of catenoids. The fence of catenoids illustrates what happens if one wants to add a handle to the catenoid. A theorem of R. Schoen [**Sch83**] implies that this is not possible. However, one is allowed to try.

Suppose we have a catenoid with an additional horizontal handle. As a finite total curvature surface, this would be represented by a twice punctured torus. Assume that the surface is symmetric with respect to reflections at the coordinate planes. This implies that the torus is rectangular, and the planar symmetry lines are represented by the horizontal and vertical symmetry lines of a fundamental rectangle.

(Recall that rectangular and rhombic tori are the only ones that allow reflective symmetries at lines. In the rectangular case, the reflective symmetry has always two components, while it is connected in the rhombic case.)

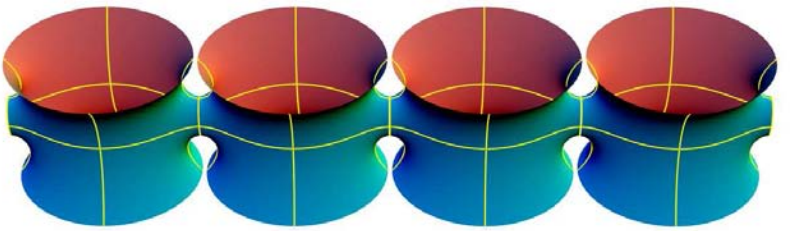


FIGURE 24. A fence of catenoids.

All special points are placed at the intersections of these points. At the catenoid ends, the Gauss map has a simple pole and zero, and the same holds for the two points on the handle where the Gauss map is vertical. On the other hand, the height differential must have simple poles at the catenoid ends with real residues, and simple zeroes at the two handle points to compensate for the Gauss map. In terms of our elliptic function p, the height differential can thus be written as dp/p (up to an irrelevant real scale factor). It remains to deal with the horizontal period condition. We see that $G(z)\,dh = \rho p(z)\,dz$ and $dh/G(z) = dz/(\rho p(z))$.

Up to a scalar factor c, these two forms are translates of each other (as their divisors are), hence for all cycles γ,

$$\int_\gamma \frac{1}{G}\,dh = c\int_\gamma G\,dh.$$

We can solve the horizontal period problem for one cycle α easily by adjusting the López–Ros parameter ρ. We can also rotate the surface in space so that both periods of $G\,dh$ and $(1/G)\,dh$ over α become real. Then the above equation implies that (in this normalization) we have $c = 1$. Hence all periods of $G\,dh$ and $(1/G)\,dh$ must be real. However, this is impossible on a rectangular torus. This follows for instance from the Legendre relation for the periods of dz and $p\,dz$ (see [**Ahl79**]):

$$\omega_1\pi_2 - \omega_2\pi_1 = 8\pi i$$

as $\omega_2 = i\omega_1$. So we are forced to leave one horizontal period unclosed. This leads to a 1-parameter family of simply periodic minimal surfaces which depends on the modulus of the rectangular torus. Depending on this parameter, the handle grows inwards as in Figure 25 or outwards as above.

6H. Costa's minimal surface. The Costa surface [**Cos82, Cos84**] was the first embedded minimal torus [**HMI85**]. It lives on the square torus with three punctures, has total absolute curvature $\kappa = 12\pi$, two catenoid ends and one planar end. Later the planar end was shown to be deformable [**HMI90**] into a catenoid end, giving rise to a 3-ended embedded minimal surface for each rectangular torus.

There are also two straight lines on the surface passing through the origin.

Assuming the existence of such a surface, we derive its Weierstrass representation as follows:

We arrange the surface so that the Gauss map is vertical at the ends.

At the two catenoid ends, dh must have simple poles with real residue. Here the Gauss map has two simple poles. At all other points, dh must be regular. Hence we expect two more zeroes. One of them will be at the origin where the two straight lines meet, and here the Gauss map must also have a pole.

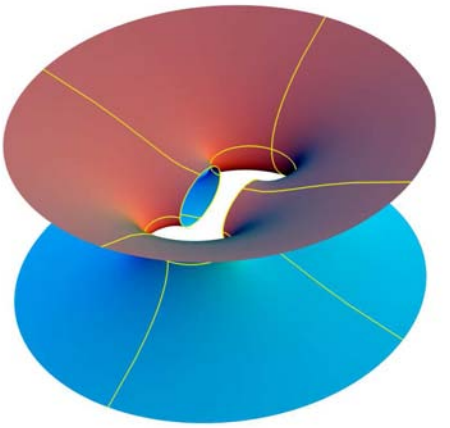


FIGURE 25. A catenoid with an attempted handle.

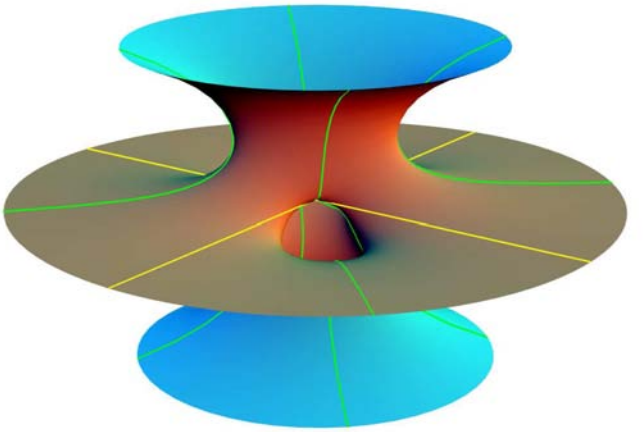


FIGURE 26. The Costa surface.

For the second zero of dh, we have two choices, forced by symmetry: Either we put it at the planar end, or we make the zero at the origin a double zero. In the latter case, the Gauss map must also have a double pole at the origin to make the metric regular there. But then the Gauss map would have a fourth order zero at the planar end and dh would be regular, contradicting the quadratic growth of the metric at planar ends. Hence we expect the second zero of dh at the planar end. This must then be compensated by a triple order zero of the Gauss map.

Again, the surface is defined on the square torus $w^2 = v(v^2 - 1)$. The point $(v, w) = (\infty, \infty)$ represents the planar end, the points $(v, w) = (\pm 1, 0)$ the catenoid ends, and $(0, 0)$ the origin. We can express the Weierstrass data in terms of v and w as follows:

$$G = \rho/w, \qquad dh = \frac{\lambda dv}{(1 - v^2)w}.$$

The free parameters are determined to kill the periods. First, the height differential must have real residues at the ends, hence λ must be real, and we can choose it to be 1 by scaling the surface. To determine ρ, we would like to evaluate two period

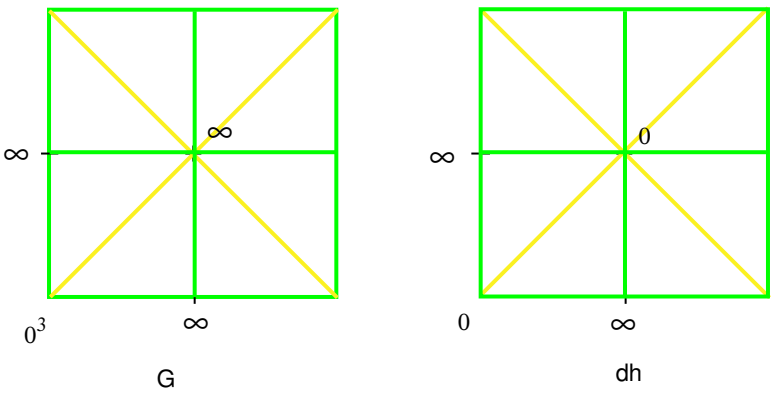


FIGURE 27. Divisors for the Costa surface.

integrals. Let α be a horizontal cycle on the torus. This becomes a closed curve γ^* encircling 0 and 1 on the v-sphere, and we can evaluate

$$\int_{\gamma^*} \frac{1}{G}\,dh = \frac{2}{\rho}\int_0^1 \sqrt{\frac{v}{1-v^2}}\,dv = \frac{4\sqrt{\pi}\,\Gamma(\frac{3}{4})}{\rho\Gamma(\frac{1}{4})}.$$

Unfortunately, it is not possible to evaluate $\int_{\gamma^*} G\,dh$ this way, because the integrand has too bad a singularity at $(v,w)=(1,0)$. To circumvent this problem, we introduce a technique which is important both for theoretical and practical purposes: We replace the 1-form $G\,dh$ by one with harmless singularities, using partial integration. First differentiating the surface equation $w^2 = v(v^2-1)$ gives

$$2wdw = (3v^2-1)dv$$

and hence

$$\begin{aligned} d\frac{v}{w} &= \frac{wdv - vdw}{w^2}, = \frac{dv}{w} - v\frac{3v^2-1}{2w^3}\,dv, = \frac{dv}{w} - \frac{3v(v^2-1)+2v}{2w^3}\,dv, \\ &= -\frac{dv}{2w} - \frac{v}{v(v^2-1)w}\,dv, = -\frac{dv}{2w} - \frac{dv}{(v^2-1)w}, \end{aligned}$$

so that

$$\int_{\gamma^*} G\,dh = -2\rho\int_0^1 \frac{dv}{2w} = \frac{2\sqrt{\pi}\,\Gamma(\frac{5}{4})}{\Gamma(\frac{3}{4})},$$

and we can determine ρ. The other period which needs to be killed is along the vertical torus cycle, or the cycle encircling 0 and -1 in the v-plane. Because of the symmetries of the construction, however, there is a symmetry of the square torus (the reflection at the diagonal) that interchanges the two cycles and leaves the Weierstrass data sufficiently invariant so that with the γ^*-period the γ-period gets killed automatically.

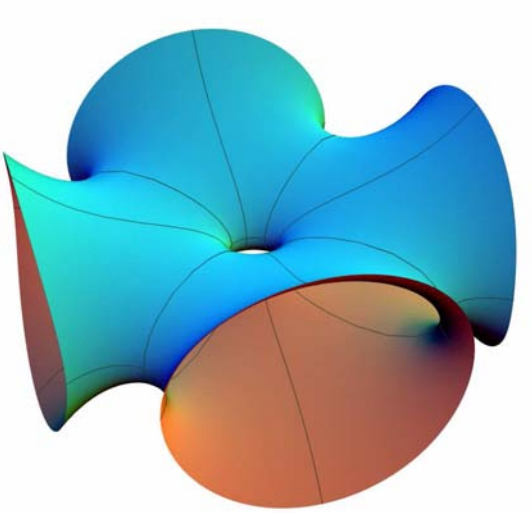


FIGURE 28. A 4-noid with handle.

6I. The Jorge–Meeks k-noids with an additional handle. The k-noids of Jorge and Meeks are complete minimal surfaces of finite total curvature defined on the sphere with catenoid ends at the punctures corresponding to the k-th roots of unity. It is possible to add a handle 'connecting' the two symmetry points. This has been observed by Karcher (see remark 5.5 on page 113 in [**Kar88**] and the diploma thesis [**Ste96**], where the existence is shown using a conjugate Plateau construction).

The underlying Riemann surface of these k-noids is a torus spanned by 1 and $k \cdot \tau$, punctured at k points. The reflectional symmetries imply that the torus is rectangular. The rotational symmetry of the surface becomes a conformal automorphism rotating the punctures, hence we can assume the punctures to be at the points $j \cdot \tau, \quad j = 0 \ldots k-1$. Denote by T' the quotient torus spanned by 1 and τ. While the height differential is invariant under the rotation and hence descends to T', the Gauss map becomes a multivalued function satisfying

$$G(z+1) = G(z),$$
$$G(z+\tau) = \zeta G(z),$$

with ζ a k-th root of unity. We can construct G using theta-functions. Take

$$G(z) = \rho \frac{\theta(z-n^-)}{\theta(z-n^+)}, \qquad dh = \frac{\theta(z-n^+)\theta(z-n^-)}{\theta(z)\theta(z-(\tau+1))}\, dz.$$

From Lemma 6.3 and the assumed periodicity we deduce from $n^\pm = \frac{1+\tau}{2} \mp \delta_0$ that $2\delta_0 = n^- - n^+ = \frac{k-1}{k}$ so that $\delta_0 = \frac{k-1}{2k}$.

This determines the divisor of dh completely. With these constructions conditions (1) and (2) are satisfied, and we are left with the period condition for dh which has never been discussed in a published paper.

MATHEMATICA NOTE 6.6. However, there is a Mathematica notebook (k-noidsg=1).nb available where the period problem is solved numerically, using theta functions.

In the same way as the ordinary k-noids are limits of singly periodic Scherk surfaces, the k-noids with a handle are limits of singly periodic Scherk surfaces with handles.

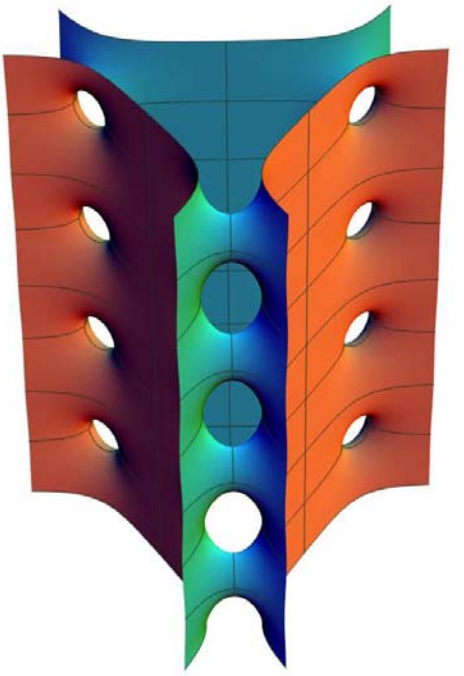


FIGURE 29. A singly periodic Scherk surface with six ends and vertical handles.

6J. Parameterizing minimal surfaces. This subsection explains how to find good parameter lines on a minimal surface. This is important, as such parameter lines often help to understand the geometry of the surface. It also yields very attractive surface models.

Given the Weierstrass representation of a surface, we have a conformal parametrization for free. So why not use the standard coordinate lines on the parameter domain? As an example, look at Costa's surface. It is clearly sufficient to draw a fundamental piece of the surface with respect to its eightfold symmetry. Using our description of the Weierstrass data, such a piece is parametrized by the first quadrant in the v-plane. Recall that the origin is mapped to the origin in $\mathbb{R}^3$, the positive imaginary axes is mapped to a straight half-line, the segment $(0, 1)$ is mapped to a planar symmetry curve which connects the origin with the top catenoid end and the segment $(1, \infty)$ is another planar symmetry curve connecting the planar end with the catenoid end.

The obvious drawback of using a rectangular coordinate grid in the first quadrant is that it is not adapted at all to the planar and the catenoid ends. In our discussion of the minimal surfaces defined on the plane or the punctured plane we have seen that the ends of a minimal surface are quite well treated by using polar coordinates centered at the punctures which represent the ends.

For our parametrization of the Costa patch, we can deal with the planar end this way by using planar coordinates in the first quadrant. This however neglects the catenoid end.

To deal with both ends simultaneously, we need a conformal map from some parameter rectangle to the first quadrant so that the left edge of the rectangle is mapped close to the point $(1, 0)$ representing the catenoid end while the right edge is mapped to a quarter circular arc close to ∞ representing the planar end. This can be achieved by taking a map from the parallel strip $\mathbb{R} \times (0, \pi)$ to the first quadrant that maps $-\infty$ to 1 and $+\infty$ to ∞, while the two sides are mapped to the edges of the first quadrant. This map, restricted to a subrectangle $(a, b) \times (0, \pi)$ of the strip, has all the desired properties.

In fact, such a map is explicitly given by

$$u \mapsto \sqrt{1 + e^u}.$$

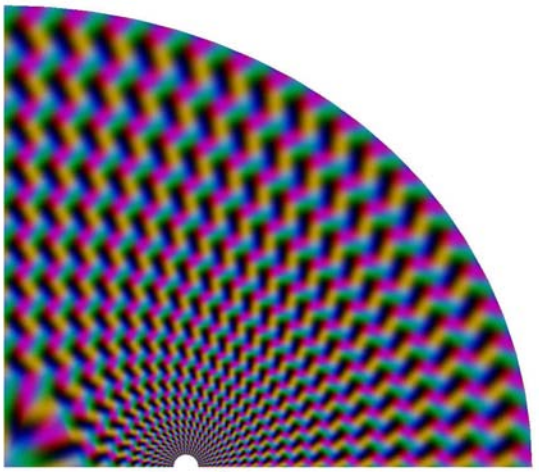


FIGURE 30. Coordinate mesh in the first quadrant.

This explicit approach works fairly well if the parameter domain of a fundamental piece of the surface is sufficiently simple so that a mapping function from a standard domain to it can be found, and so that the ends of the surface are represented by at most two boundary points of the domain. This is the case for many (though not all) of the known minimal surfaces.

7. Higher Genus Minimal Surfaces

We discuss how minimal surfaces of higher genus can be described and how one can solve the period problem numerically. For background about Riemann surfaces, see [**FK80, Gun66**].

7A. Riemann surfaces. Riemann surfaces are 2-dimensional manifolds given by an atlas of coordinate charts into $\mathbb{C}$ so that the change-of-coordinate maps are holomorphic. A convenient way to define Riemann surfaces is by specifying algebraic relations between meromorphic functions. This will also allow us to write

down meromorphic functions and 1-forms on the Riemann surface in an explicit way. Here is an example.

EXAMPLE 7.1. Consider the subset

$$X_0 = \{(x, y) \in \hat{\mathbb{C}} \times \hat{\mathbb{C}} : y^2 = x^6 - 1\}.$$

There is a natural way to turn this into a compact Riemann surface. This requires us to think carefully about points at ∞, but we will largely ignore this. The coordinate functions x and y become meromorphic functions on this surface, and all meromorphic functions can be expressed rationally in x and y.

Another way of realizing this example quite concretely is the following: Take two copies of the Riemann sphere, punctured at the sixth roots of unity (the zeroes of $x^6 - 1$). Define $f(z) = z$ on these two copies. When computing the square root $g(z) = \sqrt{f(z)^6 - 1}$ (so that we have at least locally a solution $g = y$ and $f = x$ of our equation) on one of these copies, we face a problem near the roots of unity: Analytic continuation of g along a loop around one of these points results in a sign change of g. So we can't define g globally on just one of these copies. The idea is now to glue the two copies together suitably so that whenever we continue g analytically around one of the punctures, we switch to the other copy. To do this more formally, we first restrict g to a subdomain of the punctured sphere on which we can define g globally. There are several choices for such domains: An efficient choice is to connect the six punctures in pairs by disjoint paths and then to slit the plane along these paths. Continuation of $g(z)$ around such a cut forces a double sign change of the square root so that g becomes well defined. Another choice of a suitable subdomain applies under more general circumstances: Slit the plane along cuts which emanate from the six punctures radially to infinity. Then this is a simply connected domain, there is a priori no continuation conflict. It turns out that the surface we are going to construct will be independent of the choice of cuts we make, so we will continue with the second construction.

Now take the two copies of the slit Riemann sphere and define $g(z)$ on both of them, albeit with different sign. Glue the two copies together by gluing them along the slits crosswise. This purely topological construction ensures that the functions f and g become continuous on the glued topological surface. However, it is easy to make this topological surface into a Riemann surface so that f and g become meromorphic functions: Away from the cuts, we take the identity map to the Riemann sphere as coordinates. On the cuts coordinate maps are given also by the identity maps of the two different spheres which are used for the gluing. Along the slit, these maps take the same values by construction. At the punctures, it takes a little work to construct a local coordinate which is holomorphically compatible with the other charts. Let's look at the puncture at 1. The idea is to use the function $\sqrt{z-1}$ as a local coordinate here. While this is impossible on just one copy of $\hat{\mathbb{C}}$, the two glued copies make $\sqrt{z-1}$ single-valued in a neighborhood of 1. The change of coordinate maps will be holomorphic, as $\sqrt{z-1}$ is holomorphic in neighborhoods of points away from 1. Note that we don't need to worry about holomorphicity at 1, as this point isn't covered by any other coordinate. A final issue arises with the point ∞ (to which all the slits connect). Since we have an even number of slits, the square root function does not change its sign here, so on the Riemann surface we construct there are actually two points ∞ coming from the two different Riemann spheres.

7B. Differential forms. Let's suppose we are given a Riemann surface X by an algebraic equation. For simplicity and concreteness, we assume that the equation has the form

$$y^n = P(x),$$

where P is a polynomial of degree m. Then x will be a meromorphic function of degree n on X which has zeroes precisely at the copies of the points $(0, y_j)$ with $y_j^n = P(0)$.

We will now use the functions x and y to write down meromorphic differential forms on a Riemann surface.

EXAMPLE 7.2. Let X be given by

$$y^2 = P(x) = x^6 - 1.$$

Denote the two zeroes $(0, \pm\sqrt{-1})$ by $0_{1,2}$ and the two copies of ∞ by $\infty_{1,2}$. Then x has simple zeroes at $0_{1,2}$ and simple poles at $\infty_{1,2}$. The function y has simple zeroes at the six zeroes $x_{1\ldots6}$ of $P(x)$ and triple order poles at $\infty_{1,2}$. This is consistent with the important fact that the number of zeroes of a meromorphic function on a compact Riemann surface must be equal to the number of its poles, counted with multiplicity.

We obtain meromorphic differential forms on X by taking exterior derivatives of meromorphic functions and multiplying them by other meromorphic functions. For instance, dx will have two double order poles at $\infty_{1,2}$ and simple zeroes at $x_{1\ldots6}$. Differentiating the surface equation gives

$$y\,dy = 3x^5\,dx,$$

and from this we can compute that dy has fifth order zeroes at $0_{1,2}$ and fourth order poles at $\infty_{1,2}$. We also see that the forms dx/y and $x\,dx/y$ are holomorphic on the surface.

Here are these data in a tabular form:

	$0_{1,2}$	$\infty_{1,2}$	$x_{1\ldots6}$
x	0^1	∞^1	$*$
y	$*$	∞^3	0^1
dx	$*$	∞^2	0^1
dy	0^5	∞^4	$*$
dx/y	$*$	0^1	$*$
$x\,dx/y$	0^1	$*$	$*$

EXAMPLE 7.3. Consider now the Riemann surface given by

$$y^2 = P(x) = x(x^6 - 1).$$

In this case, there is only one copy of 0 in the surface (denoted by 0) and also only one copy of ∞.

Thus x has a double zero at 0 and a double order pole at ∞. Again, the behavior of y can be read off from the equation: It has simple zeroes at the seven zeroes $x_{1\ldots7}$ of P and a seventh order pole at ∞, due to the branching at ∞.

This time, dx will have a triple order pole at ∞ and simple zeroes at $x_{1\ldots7}$. We could again compute the data for dy by differentiating the surface equation. We also see that the forms dx/y, $x\,dx/y$ and $x^2\,dx/y$ are holomorphic on the surface.

We introduce a bit of notation which we have already used for tori:

DEFINITION 7.4. Let f be a meromorphic function on a Riemann surface. Let p_i be the set of zeroes and poles of f and n_i their multiplicities. Then the formal linear combination

$$(f) = \sum n_i p_i$$

is called the divisor of f. A similar definition holds for meromorphic 1-forms. The number $\sum n_i$ is called the degree of the divisor/function/form.

REMARK 7.5. There is a form of Abel's theorem for Riemann surfaces of arbitrary genus which involves the Jacobian of the Riemann surface; see [**FK80**].

7C. The Chen–Gackstatter surface of genus 4. For each genus, there exists a minimal surface of genus g with one Enneper end and the same symmetries as the Enneper surface. These surfaces were first discussed numerically by Thayer and their existence was proven by [**Sat96**] and [**WW98**]. They are particularly simple to deal with numerically.

They are defined on Riemann surfaces of genus g, given by

$$y^2 = x \frac{\prod_{j=1}^{g/2} (x^2 - a_j^2)}{\prod_{j=1}^{(g+1)/2} (x^2 - b_j^2)}$$

with certain numbers

$$1 = a_1 < b_1 < a_2 < b_2 < \cdots.$$

The $2g + 2$ points $0, \infty, \pm a_j, \pm b_j$ are all branched points, hence have only single copies on the surface.

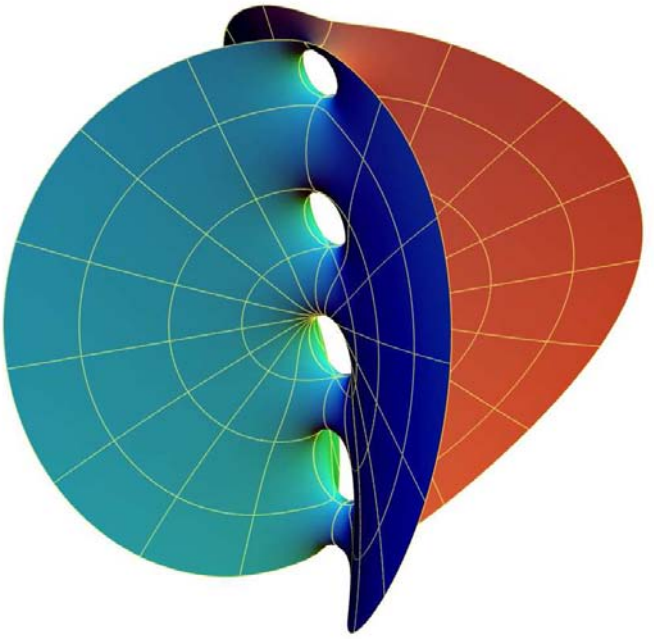


FIGURE 31. The Chen–Gackstatter surface with four handles.

The Weierstrass data are

$$G(x) = \rho y, \qquad dh = dx.$$

The height differential dh has a double order pole at ∞ and the Gauss map a single order pole or zero, depending on the parity of g. Hence ∞ represents an Enneper end, and there are no further ends.

The g free parameters a_j, b_j, ρ are needed to solve the horizontal period problem. While dh is exact, there are $2g$ cycles whose horizontal period conditions must be satisfied. But because of the imposed symmetry, only half of the conditions need to be checked. A suitable choice of these cycles is given by curves encircling the g intervals

$$(0, a_1 = 1), (a_1, b_1), (b_1, a_2), \ldots.$$

The evaluation of these integrals can be carried out on the real axes, as the singularities of the integrands at the a_i and b_i are integrable.

To draw one of these surfaces, one first has to solve the period problem numerically. This requires us to solve a g-dimensional nonlinear system of equations, for which the function evaluations are very expensive, as they involve the numerical integration of the periods.

The period problem for these surfaces can be restated in a geometrical way by looking at the following maps from the upper half plane to the complex plane given by

$$z \mapsto \int^z G\, dh,$$
$$z \mapsto \int^z \frac{1}{G}\, dh.$$

It is a surprising but easy to prove fact that the image domains of the upper half plane under these maps are staircase shaped domains as in Figure 32.

The horizontal period condition is now equivalent to the requirement that these two domains fit together.

These two domains are conformal images of the upper half plane, where the vertices correspond to 0, the $\pm a_i$, and the $\pm b_i$. This suggests another way of constructing minimal surfaces of higher genus: One defines 1-forms $G\, dh$ and

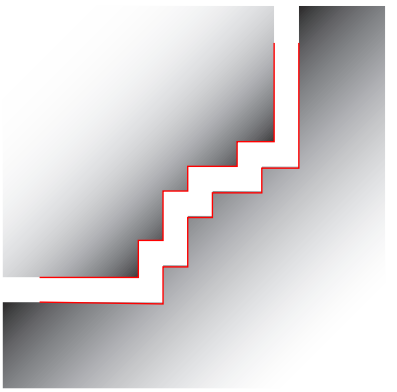


FIGURE 32. Image domains of the complex coordinate maps.

$(1/G)\,dh$ by specifying the image domains of their integrals, and using the Riemann mapping theorem. The horizontal period condition becomes a geometric condition on the shapes of the domains which needs to be solved under the assumption that the domains of $G\,dh$ and $(1/G)\,dh$ are conformal. The details are very similar to [**WW98**] and will be discussed in [**WW01**].

7D. The doubly periodic Scherk surface with handles. It is possible to add handles between each other vertical pair of ends of Scherk's doubly periodic surface. The genus one version of this was constructed by H. Karcher (see [**HKW93**]).

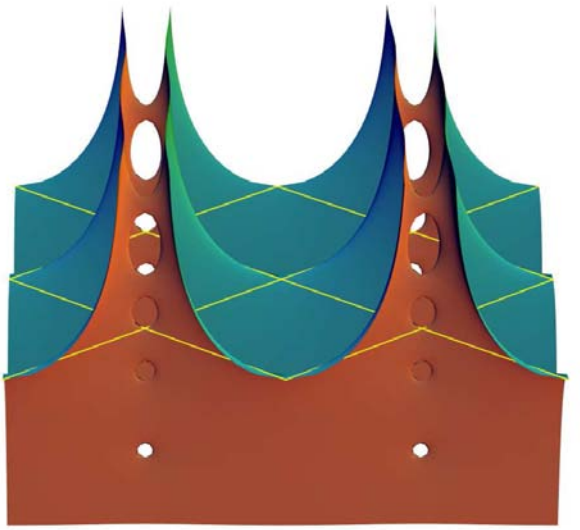


FIGURE 33. The doubly periodic Scherk surface with four handles.

For these surfaces, we use the perspective from the end of the last example. Instead of algebraically defining Riemann surfaces, we define $G\,dh$ and $(1/G)\,dh$ by the image domains of their integrals, as shown in Figure 34.

Again, the requirements on these two domains are that they fit together and that they are conformal by a vertex preserving map. Then, these domains are images of a fundamental piece of the doubly periodic Scherk surface with four handles under the maps $\int G\,dh$ and $\int (1/G)\,dh$.

Observe that the image domains can be obtained from the image domains of the classical Scherk surface by cutting corners along one boundary arc. This process of cutting corners in euclidean domains corresponds to adding handles to the minimal surface.

To obtain explicit formulas for $G\,dh$ and $(1/G)\,dh$, one expresses these forms as Schwarz–Christoffel integrands for maps to these polygonal domains. These integrands have free parameters which need to be determined numerically so that the domains fit together.

7E. Generalized Callahan–Hoffman–Meeks surfaces. The surface associated with the names of Callahan, Hoffman and Meeks [**CHMI89**] is a singly

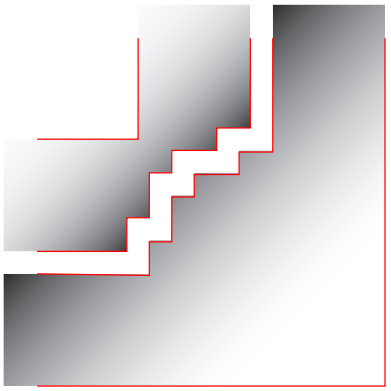


FIGURE 34. The image domains of the complex coordinates maps for the doubly periodic Scherk surface with four handles.

periodic mutant of the Costa surface: Its quotient under a vertical translation is a genus 3 surface, punctured at two points:

$$y^4 = P(x) = x^2(x^2-1)(x^2-a^2).$$

Furthermore, the Weierstrass data are given by

$$G\,dh = \frac{1}{x^2-a^2}\frac{dx}{y}, \qquad \frac{1}{G}\,dh = \frac{y\,dx}{x^2-1}.$$

7F. Period computations. It is possible to add arbitrarily many handles to each translational fundamental piece of this surface [**Web00b**]. Figure 35 shows an example with two more handles between two consecutive parallel ends; it is denoted by CHM_3. We will now explain how one can solve the period problem numerically for surfaces of this kind. Even though the method will be quite general, we will use the CHM_1-surface as the main example as here all computations can be carried out by hand.

For solving the period problem of the CHM_1-surface, we have to integrate

$$G\,dh = \frac{1}{x^2-a^2}\frac{dx}{y}$$

over closed cycles on the Riemann surface (the other Weierstrass 1-forms do not cause difficulties). These cycles are represented by closed curves in the x-plane encircling the punctures $0, \pm 1, \pm a$.

This is difficult, because the integrand $G\,dh$ is multivalued in x, and because the integration path is a curve in the complex plane. While it is tempting to replace these integrals by a real line integral between the punctures $-a, -1, 0, 1, a$, this is impossible for the integrals encircling a, as the integrand has a nonintegrable singularity there. We have encountered this kind of problem before with Costa's surface.

The idea here is to replace the 1-forms above by cohomologous ones with integrable singularities. For this, one has to find the replacements efficiently.

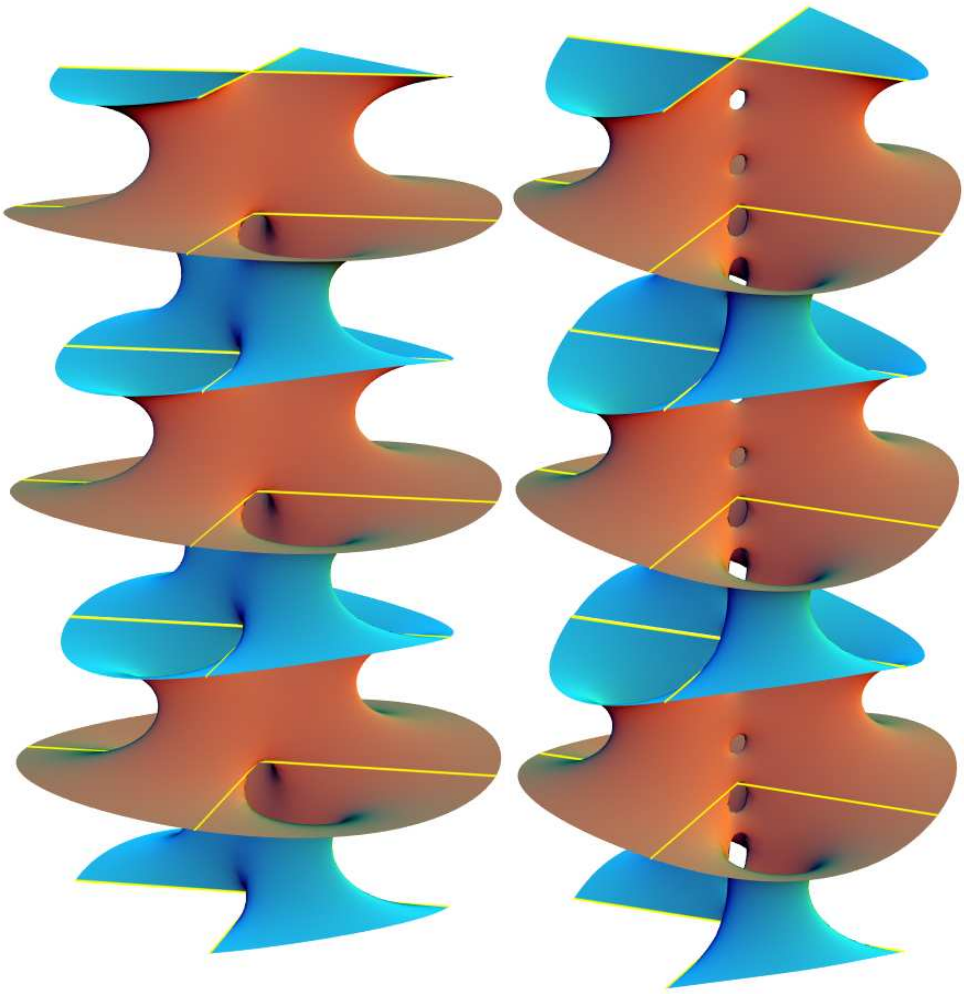


FIGURE 35. The basic Callahan–Hoffman–Meeks surface (left), and the Callahan–Hoffman–Meeks surface with two more handles (right).

We discuss this in a slightly more general way: Suppose that a Riemann surface X is given by

$$y^n = P(x) = \prod_{i=1}^{2g+1} (x - x_i)^{d_i},$$

where the x_i are real numbers and the d_i are integers. This amounts to saying that the surface is a cyclically branched cover over the Riemann sphere punctured on the real axes.

Let a meromorphic 1-form ω on X be given by

$$\omega = T(x)\frac{dx}{y}$$

with some rational function $T(x)$ such that

$$F(x) = T(x)P(x)$$

is a polynomial of some degree d. We have in mind $\omega = G\,dh$. Here the the rational function is

$$T(x) = \frac{1}{x^2 - a^2}.$$

The goal is to compute (numerically) the periods of this form ω.

From cohomological considerations it follows that we are able to write

$$\frac{F(x)}{y^{n+1}}dx - \frac{A(x)}{y}\,dx = d\frac{R(x)}{y}, \tag{7–1}$$

with new polynomials A and R. So instead of $\int_\gamma F(x)/y^{n+1}dx$ we can now compute $\int_\gamma A(x)/ydx$, which can be done along a real line segment $x_i x_{i+1}$. However, the task remains to find $A(x)$ explicitly.

By differentiating the Riemann surface equation

$$y^n = P \Rightarrow ny^{n-1}dy = P'\,dx \Rightarrow dy = \frac{P'}{ny^{n-1}}\,dx$$

we get that for an arbitrary function f on X

$$y\,d\Big(\frac{f(x)}{y}\Big) = y\frac{yf'\,dx - f\,dy}{y^2} = f'\,dx - \frac{1}{n}\frac{P'}{P}f\,dx.$$

Hence (7–1) is equivalent to

$$F - y^n A = y^n\Big(R' - \frac{1}{n}\frac{P'}{P}R\Big),$$

or yet to

$$F - P(A + R') + \frac{1}{n}P'R = 0.$$

This allows one to find the coefficients of the polynomials A and R by solving a symbolic linear system of equations.

Solving this symbolic system is still not an obvious task, if the rank of the system is large (say 10). One can get around this as follows:

Solving (7–1) for A gives

$$A = \frac{nF + P'R}{nP} - R'. \tag{7–2}$$

Since A will be a polynomial, we have to find R with $\deg R \le 2g$ such that

$$nF + P'R \equiv 0 \pmod{P}. \tag{7–3}$$

To solve this equation, we first solve

$$P'Q \equiv 1 \pmod{P}.$$

That is, we seek polynomials Q and S such that

$$QP' = SP + 1.$$

At the $2g + 1$ places x_i where P vanishes, we deduce that necessarily

$$Q(x_i) = \frac{1}{P'(x_i)},$$

so that we can write down a candidate for Q using Lagrange interpolation:

$$Q(x) = \sum_{i=1}^{2g+1} \prod_{j \neq i} \frac{x - x_j}{x_i - x_j} \frac{1}{P'(x_i)} = \sum_{i=1}^{2g+1} \frac{P(x)}{x - x_i} \lim_{x \to x_i} \frac{x - x_i}{P(x)} \frac{1}{P'(x_i)}$$

$$= P(x) \sum_{i=1}^{2g+1} \frac{1}{(x - x_i) P'(x_i)^2}.$$

Using this, we get R from (7–3) as

$$R = -nQF \pmod{P},$$

and this allows us to compute A by (7–2).

Because our F will always share many zeroes with P, it is much faster to compute R directly by Lagrange interpolation using the fact that

$$R(x_i) = -n \frac{F(x_i)}{P'(x_i)}.$$

All this is quite easy to implement. For a variation, it is conceivable that there are other kinds of replacements if one looks only at the singularities which actually cause trouble for the cycle over which one wants to integrate.

Finally, we give the results of the computation in the simplest case:

EXAMPLE 7.6 (The CHM_1 surface). Here we have for $G\,dh$,

$$P(x) = x^2(x^2 - 1)(x^2 - a^2), \qquad F(x) = x^2(x^2 - 1),$$

$$R(x) = -2\frac{x(x^2 - 1)}{a^2(a^2 - 1)}, \qquad A(x) = \frac{3x^2 - a^2}{a^2(a^2 - 1)}.$$

References

[Ahl79] L. Ahlfors. *Complex Analysis.* McGraw Hill, New York, 1979.

[CHMI89] M. Callahan, D. Hoffman, and W. H. Meeks III. Embedded minimal surfaces with an infinite number of ends. *Inventiones Math.*, 96:459–505, 1989.

[Cos82] C. Costa. *Imersões minimas en $\mathbb{R}^3$ de gênero un e curvatura total finita.* PhD thesis, IMPA, Rio de Janeiro, Brasil, 1982.

[Cos84] C. Costa. Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends. *Bull. Soc. Bras. Mat.*, 15:47–54, 1984.

[DHKW92] U. Dierkes, S. Hildebrandt, A. Küster, and O. Wohlrab. *Minimal Surfaces I.* Grundlehren der mathematischen Wissenschaften 296. Springer-Verlag, 1992.

[FK80] H. M. Farkas and I. Kra. *Riemann Surfaces.* Number 72 in Graduate texts in mathematics. Springer-Verlag, 1980.

[Gun66] R. C. Gunning. *Lectures on Riemann Surfaces.* Princeton University Press, Princeton, N.J., 1966.

[HK97] D. Hoffman and H. Karcher. Complete embedded minimal surfaces of finite total curvature. In *Encyclopedia of Mathematics*, pages 5–93, 1997. R. Osserman, editor, Springer Verlag.

[HKW93] D. Hoffman, H. Karcher, and F. Wei. The genus one helicoid and the minimal surfaces that led to its discovery. In *Global Analysis and Modern Mathematics.* Publish or Perish Press, 1993. K. Uhlenbeck, editor, p. 119–170.

[HMI85] D. Hoffman and W. H. Meeks III. A complete embedded minimal surface in $\mathbb{R}^3$ with genus one and three ends. *Journal of Differential Geometry*, 21:109–127, 1985.

[HMI90] D. Hoffman and W. H. Meeks III. Embedded minimal surfaces of finite topology. *Annals of Mathematics*, 131:1–34, 1990.

[Kar88] H. Karcher. Embedded minimal surfaces derived from Scherk's examples. *Manuscripta Math.*, 62:83–114, 1988.

[Kar89] H. Karcher. Construction of minimal surfaces. *Surveys in Geometry*, pages 1–96, 1989. University of Tokyo, 1989, and Lecture Notes No. 12, SFB256, Bonn, 1989.

[LHI98] H. Lazard-Holly and W.H. Meeks III. Classification of doubly-periodic minimal surfaces of genus zero. preprint, 1998.

[Lop92] F. J. Lopez. The classification of complete minimal surfaces with total curvature greater than -12π. *Trans. Amer. Math. Soc.*, 334(1):49–74, 1992.

[LR91] F. J. Lopez and A. Ros. On embedded complete minimal surfaces of genus zero. *Journal of Differential Geometry*, 33(1):293–300, 1991.

[MPR98] W.H. Meeks, J. Pérez, and A. Ros. Uniqueness of the Riemann minimal examples. *Invent. math.*, 131:107–132, 1998.

[Oss86] R. Osserman. *A Survey of Minimal Surfaces*. Dover Publications, New York, 2nd edition, 1986.

[Rie67] B. Riemann. Über die Fläche vom kleinsten Inhalt bei gegebener Begrenzung. *Abh. Königl, d. Wiss. Göttingen, Mathem. Cl.*, 13:3–52, 1867.

[Sat96] K. Sato. Existence proof of one-ended minimal surfaces with finite total curvature. *Tohoku Math. J.*, 48:229–246, 1996.

[Sch83] R. Schoen. Uniqueness, symmetry, and embeddedness of minimal surfaces. *Journal of Differential Geometry*, 18:791–809, 1983.

[Ste96] M. Steffens. Doppelt periodische minimalflächen. Diplom thesis, 1996.

[Web99] M. Weber. Period quotient maps of meromorphic 1-forms and minimal surfaces on tori. preprint Bonn, 1999.

[Web00a] M. Weber. The genus one helicoid is embedded. Habilitationsschrift Bonn, 2000.

[Web00b] M. Weber. On singly periodic minimal surfaces invariant under a translation. *Manuscripta Mathematica*, 101:125–142, 2000.

[WW98] M. Weber and M. Wolf. Minimal surfaces of least total curvature and moduli spaces of plane polygonal arcs. *Geom. and Funct. Anal.*, 8:1129–1170, 1998.

[WW01] M. Weber and M. Wolf. *Minimal Surfaces and Teichmueller Theory*. In preparation.

Department of Mathematics, Indiana University, Rawles Hall, Bloomington, IN 47405

E-mail address: matweber@indiana.edu

Clay Mathematics Proceedings
Volume 2, 2004

Computational Aspects of Discrete Minimal Surfaces

Konrad Polthier

Abstract. In differential geometry the study of smooth submanifolds with distinguished curvature properties has a long history and belongs to the central themes of this field. Modern work on smooth submanifolds, and on surfaces in particular, relies heavily on geometric and analytic machinery which has evolved over hundreds of years. However, nonsmooth surfaces are also natural mathematical objects, even though there is less machinery available for studying them. Consider, for example, the pioneering work on polyhedral surfaces by the Russian school around Alexandrov [**1**], or Gromov's approach of doing geometry using only a set with a measure and a measurable distance function [**10**]. Also in other fields, for example in computer graphics and numerics, we nowadays encounter a strong need for a discrete differential geometry of arbitrary meshes.

These tutorial notes introduce the theory and computation of discrete minimal surfaces which are characterized by variational properties, and are based on a part of the authors Habilitationsschrift [**27**]. In Section 1 we introduce simplicial surfaces and their function spaces. Laplace–Beltrami harmonic maps and the solution of the discrete Cauchy–Riemann equations are introduced on simplicial surfaces in Section 2. These maps are the basis for an iterative algorithm to compute discrete minimal and constant mean curvature surfaces which is discussed in Section 3. There we define the discrete mean curvature operator, derive the associate family of discrete minimal surfaces in terms of conforming and nonconforming triangle meshes, and present some recently discovered complete discrete surfaces, the family of discrete catenoids and helicoids.

1. Introduction to Polyhedral Meshes

Polyhedral meshes belong to the most basic structures for the representation of geometric shapes not only in numerics and computer graphics. Especially the finiteness of the set of vertices and of their combinatorial relation makes them an ideal tool to reduce infinite dimensional problems to finite problems. In this section we will review the basic combinatorial and topological definitions and state some of their differential geometric properties.

In practice, a variety of different triangle and other polyhedral meshes are used. In this introduction we restrict ourselves to simplicial complexes, or conforming meshes, where two polygons must either be disjoint or have a common vertex or a common edge. Or for short, a polygon is not allowed to contain a vertex of another polygon in the interior of one of its edges. This restriction avoids discontinuity

©2004 Clay Mathematics Institute

problems in the shape, so-called hanging nodes. Further, we restrict our discussion to piecewise linear meshes although many concepts extend to meshes with piecewise higher order polynomial order. Often it is too restrictive to work solely in the space of conforming triangulations, and, in later sections, we will enlarge the function space to include discontinuous, nonconforming meshes as well.

In many situations a property of a polyhedral surface can be associated to depend either on the geometric shape or on the combinatorial or topological properties of the mesh. Therefore, it is important to distinguish between the topology of a mesh and its geometric shape which is determined by the geometric position of the vertices. For example, assume that all points of a compact surface are collapsed to a single geometric position, then we would still like to derive the topological genus from the combinatorial properties of the surface. This forces us to introduce slightly more abstract definitions of polyhedral surfaces.

Introductions to polyhedral manifolds are given in most books on algebraic topology, for example by Munkres [**22**], in the book by Ziegler [**38**] on combinatorial aspects of polytopes, or by Bloch [**2**] on topological and differential geometric problems. But note, there are slight differences depending on the purpose. The standard approach in topology introduces simplices and simplicial complexes as embeddings into Euclidean space while we allow immersions with self-intersections. Good sources of applications of polyhedral manifolds to problems in differential geometry are also the books by A.D. Alexandrov and Zalgaller [**1**] and Reshetnyak [**34**].

1.1. Simplicial complexes. We begin the introduction of polyhedral surfaces with a combinatorial point of view, which means that for the moment we do not care about the specific nature of points but consider them as abstract entities. In the combinatorial setup the most basic entities of polyhedral shapes are points, line segments, triangles, tetrahedrons, and their higher dimensional analogues, called simplices:

DEFINITION 1.1. Let $\mathfrak{V} = \{\mathfrak{v}_0, \ldots, \mathfrak{v}_m\}$ be a finite set of $m+1$ abstract points. The (unordered) set $[\mathfrak{v}_0, \ldots, \mathfrak{v}_m]$ is called a *combinatorial m-simplex*, or *combinatorial simplex* in short. The number m is called the *dimension of the simplex.*

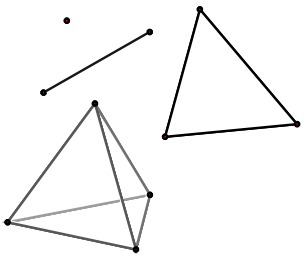


FIGURE 1. Low dimensional simplices.

DEFINITION 1.2. A *face* f of a simplex $\sigma = [\mathfrak{v}_0, \ldots, \mathfrak{v}_m]$ is a simplex determined by a nonempty subset of $\{\mathfrak{v}_0, \ldots, \mathfrak{v}_m\}$. A k-face has $k+1$ points. A *proper* face is any face different from σ.

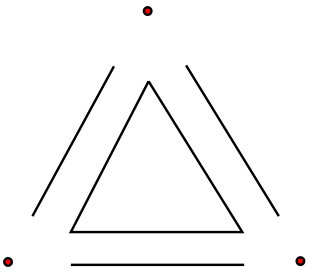


FIGURE 2. Faces of a triangle.

For example, a 0-simplex is a combinatorial point, a 1-simplex is a line segment, a 2-simplex is a triangle, and a 3-simplex is a tetrahedron. There exist seven faces of a triangle $[\mathfrak{v}_0, \mathfrak{v}_1, \mathfrak{v}_2]$: the triangle itself $[\mathfrak{v}_0, \mathfrak{v}_1, \mathfrak{v}_2]$, its three edges $[\mathfrak{v}_0, \mathfrak{v}_1]$, $[\mathfrak{v}_1, \mathfrak{v}_2]$, $[\mathfrak{v}_2, \mathfrak{v}_0]$ and its three points $[\mathfrak{v}_0]$, $[\mathfrak{v}_1]$, $[\mathfrak{v}_2]$, where the last six faces are proper. A 0-simplex has no proper face.

To perform the transition from combinatorics to geometry, we use the so-called standard simplex which serves as a geometric representative associated to each combinatorial simplex:

DEFINITION 1.3. The *standard simplex* $\Delta^m \subset \mathbb{R}^{m+1}$ is the convex hull of the endpoints $\{e_0, \ldots, e_m\}$ of the unit basis vectors in $\mathbb{R}^m$ which are given by $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$. Formally,

$$\Delta^m = \left\{ \sum_{i=0}^{m} \lambda_i e_i \;\middle|\; 0 \le \lambda_i \le 1,\ \sum_{i=0}^{m} \lambda_i = 1 \right\}.$$

The standard simplex not only is a set of points but includes the "interior" points. For example, the standard triangle Δ^2 in $\mathbb{R}^3$ is the planar triangle spanned by the three points $(1,0,0)$, $(0,1,0)$, $(0,0,1)$. Nevertheless, the standard simplex is simply a technical term. It provides the ground to formulate the next notion, defined by any set of $m+1$ points in Euclidean space $\mathbb{R}^n$ (where n might be different from m):

DEFINITION 1.4. A *geometric simplex* $\sigma = [p_o, \ldots, p_m]$ is a set $V = \{p_o, \ldots, p_m\}$ of $m+1$ points in $\mathbb{R}^n$, together with an affine map

$$\varphi : \Delta^m \to \text{convex hull}(p_0, \ldots, p_m), \qquad \varphi(e_i) = p_i.$$

The number m is called the *dimension of the simplex.*

The difference between an abstract and a geometric simplex is the existence of the geometric realization provided by the map φ, that is, the embedding of the simplex in a vector space.

DEFINITION 1.5. Let $\mathfrak{V} = \{\mathfrak{v}_1, \mathfrak{v}_2, \ldots\}$ be a set of abstract points. Then an *abstract simplicial complex* K is a set of simplices S formed by finite subsets of $\mathfrak{V}$ such that if $\sigma \in S$ is a simplex, then every subset $\tau \subset \sigma$ is also a simplex of K.

If two, or more, simplices of K share a common face, they are called *adjacent* or *neighbors.* The *boundary* of K is formed by any proper face that belongs to only one simplex, and its faces.

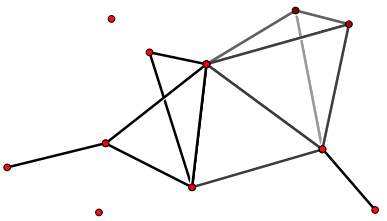


FIGURE 3. Simplicial complex.

The simplicial complex K formally represents the connectivity of a mesh, and its simplices represent the points, edges, triangles, and higher dimensional simplices. The number of points in a complex may be infinite. By associating the set of abstract points with geometric points in some $\mathbb{R}^n$ we obtain a geometric shape consisting of piecewise flat simplices. Note, the following definition does not require an embedding but allows that the geometric realization may have self-intersections. By allowing immersions this definition is nonstandard in the sense of algebraic topology which usually requires embeddings.

DEFINITION 1.6. A *simplicial complex* (K,V) of an abstract simplicial complex K is a geometric realization uniquely given by

(1) a set of geometric points $V = \{p_1, p_2, \dots\} \subset \mathbb{R}^n$ with a bijection

$$\Phi : \mathfrak{V} \to V, \qquad \mathfrak{v}_i \to p_i;$$

(2) for each k-simplex $\sigma = [p_{i_0}, \dots, p_{i_k}]$ an affine map from the standard simplex

$$\varphi : \Delta^k \to \text{convex hull}(p_{i_0}, \dots, p_{i_k}), \qquad \varphi(e_j) = p_{i_j}.$$

These definitions ensure a strict separation between the combinatorial properties of a mesh specified by K and its geometric shape determined by V, which is also expressed by adding V to the notation of the simplicial complex (K,V). The identification of abstract and geometric vertices is uniquely performed by the bijection Φ which relates the abstract points $\mathfrak{V}$ of K and the set of geometric points V. Any embedding of the abstract complex K into a Euclidean space induces a topology on the simplicial complex.

DEFINITION 1.7. The *underlying (topological) space* $|K|$ of a simplicial complex K immersed into $\mathbb{R}^n$ is the topological space consisting of the subset of $\mathbb{R}^n$ that is the union of all geometric realizations of simplices in K with the topology induced from any embedding of K.

Important examples of simplicial complexes are simplicial disks and balls.

DEFINITION 1.8. A *simplicial n-ball* B^n is a simply connected simplicial complex such that $|B^n|$ is homeomorphic to the solid unit ball in $\mathbb{R}^n$, and a *simplicial n-sphere* S^n is homeomorphic to the boundary sphere of the solid unit ball in $\mathbb{R}^{n+1}$. For $n = 2$, B^2 is also called a *simplicial disk*, and S^2 is a *simplicial sphere*. For $n = 1$, S^1 is a *simplicial circle*.

For example, an icosahedron is a simplicial sphere, and any simply closed polygon is a simplicial circle.

In some cases it makes sense to identify a simplicial complex (K, V) with its underlying set $|K|$ in a Euclidean space $\mathbb{R}^n$, for example, a polytope can always be recovered from its set of vertices. In the general case one should keep in mind that (K, V) has more the character of an immersion. For example, if the immersion of a polygonal circle intersects geometrically at a point shaping a figure-eight, it may still be a combinatorial or topological circle. The topology of such a shape cannot be recovered solely from its shape.

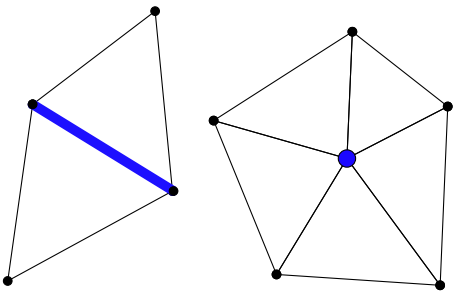


FIGURE 4. Star of an edge and a vertex.

DEFINITION 1.9. Let (K, V) be a simplicial complex. Then a subset $L \subset K$ is a *subcomplex* of K if L is a simplicial complex itself. For example, let $\sigma \in K$ be a simplex, then

$$\operatorname{star} \sigma := \{\eta \in K \text{ that contains } \sigma, \text{ plus all faces of } \eta\}$$

and

$$\operatorname{link} \sigma := \{\eta \in \operatorname{star} \sigma \mid \eta \cap \sigma = \varnothing\}$$

are subcomplexes of K.

Simplicial surfaces extend the notion of a topological 2-manifold to the simplicial world.

DEFINITION 1.10. A *simplicial surface* S is a simplicial complex consisting of a finite set $\mathfrak{T}$ of triangles such that

(1) any point $p \in S$ lies in at least one triangle $T \in \mathfrak{T}$;
(2) the star of each point $p \in S$ is a simplicial disk.

Note, in the definition one may allow a denumerable set of triangles under the additional assumption that the simplicial complex is locally finite, that is, the star of each vertex consists of a finite number of triangles.

A *polyhedral surface* is more general than a simplicial surface and may include flat faces with more than three vertices. Figure 5 illustrates several pitfalls and degenerate situations which arise in practical implementations. The first row shows two nonmanifold situations. The second row is a hanging node where adjacent faces do not join a common edge. The third row shows a valid simplicial surface consisting of four triangles where the pairwise adjacency of triangle pairs is indicated by two small lines (the right-hand drawing makes it clear how the middle edge belongs to all four triangles). Care must be taken to avoid the first two situations in practical implementations. The third situation can be resolved with an additional neighborhood information.

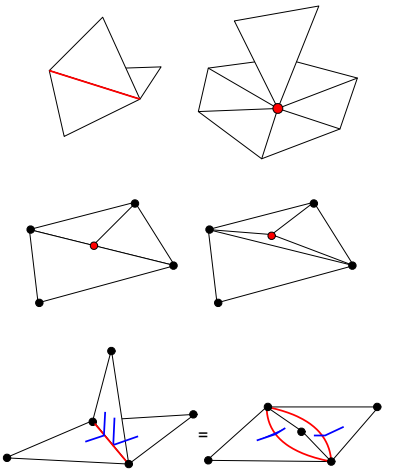


FIGURE 5. Degenerate situations and nonmanifold surfaces.

DEFINITION 1.11. Let $M \subset \mathbb{R}^n$ be a topological surface. Then a simplicial surface S *triangulates* M if there exists a homeomorphism

$$t : |S| \to M.$$

The simplicial complex S together with the homeomorphism t is called a *triangulation* of M.

Smooth surfaces and simplicial surfaces are related through the following theorem, compare [**20**]:

THEOREM 1.12. *The following facts hold for two-dimensional surfaces:*

(1) Any compact topological surface M in $\mathbb{R}^n$ can be triangulated, i.e., there exists a simplicial surface which triangulates M.

(2) If a topological surface is triangulated by two simplicial surfaces K_1 and K_2, then K_1 and K_2 have simplicially isomorphic subdivisions.

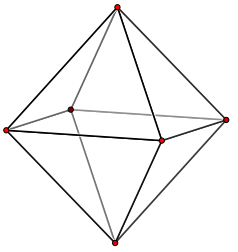


FIGURE 6. Octahedron.

1.2. Distance and metric. For metric measurements the interior of simplicial faces must be uniquely defined. Therefore, we prefer simplicial instead of polyhedral surfaces, or assure that we work with piecewise flat polygons. The metric of a surface may, for example, be induced from an immersion into a Euclidean space, or the metric may be defined in a more abstract way, say, by assigning a

length to each edge which fulfills the triangle identity on each triangle. In a locally Euclidean metric the distance between two points is measured along curves whose length is measured segment-wise on the open edges and triangles:

DEFINITION 1.13. A curve γ on a simplicial complex M is called *rectifiable*, if for every simplex $\sigma \in M$ the part $\gamma_{|\sigma}$ is rectifiable with respect to the smooth metric of σ. Then the *length* of γ is given by

$$L(\gamma) := \sum_{\sigma \in M} L(\gamma_{|\mathring{\sigma}}) \tag{1–1}$$

as the sum of the lengths on each open simplex.

The area of a simplicial surface is defined in a similar way:

DEFINITION 1.14. Let M be a simplicial surface. Then we define

$$\operatorname{area} M := \sum_{T \in M} \operatorname{area} M_{|T}. \tag{1–2}$$

Most of our considerations apply to a more general class of length spaces. Each face may have an arbitrary metric as long as the metrics of two adjacent faces are compatible, i.e., if the common edge has the same metric in both faces, and the triangle inequality holds.

In many practical applications simplicial complexes have a metric induced from an immersion into a Euclidean $\mathbb{R}^n$. For example, take a polyhedral surface in $\mathbb{R}^3$ and consider the two adjacent faces of an edge. Each face has the metric induced from $\mathbb{R}^3$, i.e., the length of any curve on a face is equal to the length of the same curve measured in $\mathbb{R}^3$. In this case, any neighborhood of a point on the edge is isometric to a planar domain, since both faces can be unfolded to $\mathbb{R}^2$.

When considering the approximation of a smooth surface M with a sequence of polyhedral surfaces $\{M_{h,i}\}$, one should be aware that higher order terms such as area may not converge as expected. The *Schwarz lantern* [**35**] is an example of a sequence of polyhedral surfaces which converges uniformly to a cylinder while the corresponding area grows to infinity.

1.3. Grids in numerics and graphics. In recent years an enormous effort went into the design of efficient grids in numerics and computer graphics. Adaptive grids and hierarchical representations became very important in numerical applications, and are nowadays complemented with subdivision surfaces in computer graphics modeling packages. Among the current issues is the construction of specialized encodings for efficient data compression.

This section recalls some important types of meshes used in numerical computations and computer graphics. The choice of a suitable grid depends on a number of criteria such as the shape of the domain, the type of the numerical method, or even the hardware, for example, to support parallelization of algorithms.

Structured grids tessellate a rectangle $[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}] \subset \mathbb{R}^2$ into regular quadrilaterals of the same size $h = (h_x, h_y)$. The grid Ω_h

$$\Omega_h = \left\{ (x_i, y_j) \;\middle|\; \begin{array}{ll} x_i = x_{\min} + ih_x & i \in [0, m-1] \\ y_j = y_{\min} + jh_y & j \in [0, n-1] \end{array} \right\}$$

is implicitly determined by the two extremal vertices $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$ and the number of subdivisions (m, n). *Multiblock grids* use several structured grids

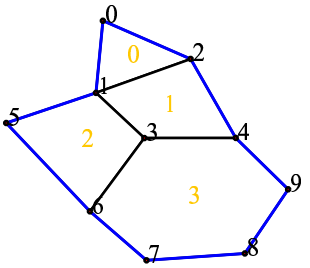
at possibly different resolutions to cover the different regions of the domain. *Multi-grids* and *sparse grids* are hierarchical representations which allow a considerable reduction of the number of grid points.

Parametric grids are obtained as images of other grid types under a continuous map Φ and thus are suitable for the discretization of more general domains. Important examples of parametric maps are Möbius maps and the Schwarz–Christoffel map, both are angle-preserving, i.e., conformal maps. Circle packings, remarkably applied by Thurston and others to problems with three-manifolds, are nowadays a promising concept in practical implementations, for example, for the flattening of rather general surfaces [**14**].

Unstructured or irregular grids may consist of rather general nonoverlapping polygons. Such grids are determined by a set of points, i.e., the vertices of the polygons,

$$P = \{P_0, P_2, \ldots, P_9\}$$

and connectivity information where each polygon is given as an ordered list of its vertices, or more efficiently, of its vertex indices. Additional information of a



$$E_0 = \{0,1,2\} \qquad N_0 = \{1,-1,-1\}$$
$$E_1 = \{2,1,3,4\} \qquad N_1 = \{2,3,-1,0\}$$
$$E_2 = \{5,6,3,1\} \qquad N_2 = \{3,1,-1,-1\}$$
$$E_3 = \{3,6,7,8,9,4\} \qquad N_3 = \{-1,-1,-1,-1,1,2\}.$$

FIGURE 7. Unstructured Grid.

structured grid may be stored in order to achieve faster access of information, or to clarify ambiguous situations. For example, a list of neighbor faces which have a common edge with the current face. The following neighbor array has for each element E_i a list of indices of adjacent elements N_i where $N_i[j]$ denotes the element adjacent to the edge $E_i[j+1]E_i[j+2]$ of E_i (indices are modulo number of vertices of E_i).

The naming convention has its origin in triangle meshes where the edge $E_i[j+1]E_i[j+2]$ is opposite to the vertex E_i. The rule just stated allows us to use the same programming code for both, simplicial as well as polyhedral surfaces.

An alternative to vertex based formats are *facet-edges formats*. Here a set of edges is given as above by specifying pairs of vertex indices. Then higher dimensional cells are defined through their boundary, which means a two-dimensional element is determined by a set of edge indices. Such formats are useful if all cells of a cell-complex play an active role and have associated information.

A wealth of meshes is used in computer graphics and numerics for different purposes. *Progressive meshes* introduced by Hoppe [**12**] are based on vertex-split and edge-collapse operations for adaptive refinement and coarsening. In recent years these data types have been very popular in computer graphics, especially since they allow topology changes. They are a special class of *multi-resolution*

grids or *hierarchical grids* which store different levels of resolution of a shape. Often a smooth transition between different hierarchical resolutions is incorporated in the data structure. *Normal meshes* [**11**] were designed to describe shapes locally as graphs over a coarser resolution of the same mesh. This technique is especially suitable for *subdivision surfaces* or multi-resolution surfaces obtained from a wavelet decomposition where the finer resolutions are obtained algorithmically.

The fast and incremental transmission of shapes over low-bandwidth connections plays an increasing role nowadays. Here specialized representations of meshes allow a compressed encoding. For example, the algorithm by Taubin and Rossignac [**36**], which is incorporated into the MPEG-4 standard, encodes the connectivity of a triangle mesh with about $2-3$ bits per vertex compared to 96 bits used in the index based representation mentioned above.

1.4. Finite element spaces. Piecewise polynomial functions on simplicial surfaces conceptually fall into the category of finite element spaces. Here we briefly recall the most basic function spaces relevant for our later work. See the books [**5**][**4**] for an introduction.

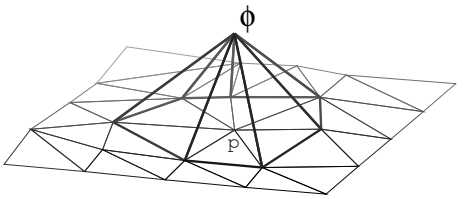


FIGURE 8. Basis function on a manifold.

DEFINITION 1.15. On a simplicial surface M_h we define the function space S_h of *conforming finite elements*:

$$S_h := \big\{ v : M_h \to \mathbb{R}^d \;\mid\; v \in C^0(M_h) \text{ and } v \text{ is linear on each triangle} \big\}.$$

S_h is a finite dimensional vector space spanned by the Lagrange basis functions $\{\varphi_1, \ldots, \varphi_n\}$ corresponding to the set of vertices $\{p_1, \ldots, p_n\}$ of M_h, that is for each vertex p_i we have a function

$$\begin{gathered} \varphi_i : M_h \to \mathbb{R},\ \varphi_i \in S_h, \\ \varphi_i(p_j) = \delta_{ij} \quad \forall i,j \in \{1,\ldots,n\}, \\ \varphi_i \text{ is linear on each triangle.} \end{gathered} \tag{1–3}$$

Then each function $u_h \in S_h$ has a unique representation

$$u_h(p) = \sum_{j=1}^{n} u_j \varphi_j(p) \quad \forall\, p \in M_h$$

where $u_j = u_h(p_j) \in \mathbb{R}^d$. The function u_h is uniquely determined by its nodal vector $(u_1, \ldots, u_n) \in \mathbb{R}^{dn}$.

Sometimes we will also use piecewise higher-order polynomial representations described in a similar way with different basis functions. Note that any component function of a function $v \in S_h$ has bounded Sobolev H^1 norm.

1.4.1. *Nonconforming finite elements.* In our later investigations the following space of nonconforming finite elements (see [**5**][**4**] for a detailed discussion) plays an important role. Since these spaces include discontinuous functions their use was often called a *variational crime* in the finite element literature. In our settings, nonconforming functions naturally appear as the correct spaces for our later considerations on constant mean curvature surfaces.

DEFINITION 1.16. For a simplicial surface M_h, we define the space of *nonconforming finite elements* by

$$S_h^* := \left\{ v : M_h \to \mathbb{R}^d \;\middle|\; \begin{array}{l} v_{|T} \text{ is linear for each } T \in M_h, \text{ and} \\ v \text{ is continuous at all edge midpoints} \end{array} \right\}.$$

The space S_h^* is no longer a finite dimensional subspace of $H^1(M_h)$ as in the case of conforming elements, but S_h^* is a superset of S_h. Let $\{m_i\}$ denote the set of edge midpoints of M_h, then for each edge midpoint m_i we have a basis function

$$\begin{array}{c} \psi_i : M_h \to \mathbb{R} \quad \psi_i \in S_h^*, \\ \psi_i(m_j) = \delta_{ij} \quad \forall\, i,j \in \{1,2,\dots\}, \\ \psi_i \text{ is linear on each triangle.} \end{array} \tag{1–4}$$

The support of a function ψ_i consists of the (at most two) triangles adjacent to the edge e_i, and ψ_i is usually not continuous on M_h. Each function $v \in S_h^*$ has a representation

$$v_h(p) = \sum_{\text{edges } e_i} v_i \psi_i(p) \quad \forall\, p \in M_h$$

where $v_i = v_h(m_i)$ is the value of v_h at the edge midpoint m_i of e_i.

Let $M_h \subset \mathbb{R}^m$ be a conforming triangulation with vertices $V = \{p_1, p_2, \dots\}$ and edge midpoints $E = \{m_1, m_2, \dots\}$. For a given triangle $t \in M_h$ with vertices $\{p_{t_1}, p_{t_2}, p_{t_3}\}$ and edge midpoints $\{m_{t_1}, m_{t_2}, m_{t_3}\}$ we have the elementary correspondence

$$\frac{1}{2} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} p_{t_1} \\ p_{t_2} \\ p_{t_3} \end{pmatrix} = \begin{pmatrix} m_{t_1} \\ m_{t_2} \\ m_{t_3} \end{pmatrix}, \tag{1–5}$$

and

$$\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} m_{t_1} \\ m_{t_2} \\ m_{t_3} \end{pmatrix} = \begin{pmatrix} p_{t_1} \\ p_{t_2} \\ p_{t_3} \end{pmatrix}. \tag{1–6}$$

We will also use the term *nonconforming surface* to denote a simplicial surface where adjacent triangles are connected at the midpoint of their common edge but may be twisted. Later we also require that the corresponding edge of two adjacent triangles must have the same length. Nonconforming surfaces may be considered as images of a nonconforming map from a conforming surface, therefore, we often do not distinguish between a nonconforming surface and a nonconforming map.

2. Conjugation of Discrete Harmonic Maps

Discrete harmonic maps appear as a basic model problem in finite element theory and differential geometry for the discretization of smooth concepts. Beyond

that, discrete harmonic maps have a wide range of nontrivial applications in computer graphics, for example to smoothen noisy meshes, or in differential geometry to compute constant mean curvature surfaces.

Several discrete operators on simplicial surfaces are related to discrete harmonic maps. For example, the area gradient, the mean curvature, or the divergence operator on vector fields. The main topic of this section is the construction of pairs of conjugate discrete Laplace–Beltrami harmonic maps on polyhedral surfaces. We start to derive the definitions and properties of discrete harmonic maps in a geometric setting which will then allow us to develop other discrete geometric operators and to solve problems related to minimal and constant mean curvature surfaces in Section 3.

Harmonic maps on surfaces also have practical importance, for example, we derive in Section 3 efficient numerical algorithms for solving free boundary value problems for unstable minimal surfaces and constant mean curvature surfaces. In the algorithms [**25**] and [**23**], the conjugate of a minimal surface is obtained via the conjugation of a discrete harmonic map. Conjugate harmonic maps are originally defined on the dual graph of the edge graph of the original surface, but one should consider them as nonconforming functions. The results of the present section provide a thorough understanding of the geometric constructions used in Pinkall and Polthier [**25**] and in Oberknapp and Polthier [**23**] by relating the discrete conjugation of surfaces to nonconforming finite element spaces.

Convergence of conforming harmonic maps has been shown by Tsuchiya [**37**]. As a more general result for surfaces, Dziuk and Hutchinson [**8**] obtained optimal convergence results in the H^1 norm for the finite element procedure of the Dirichlet problem of surfaces with prescribed mean curvature. Compare Müller et al. [**21**] for harmonic maps on planar lattices using the five-point Laplacian.

In a subsequent section we will apply the duality between discrete harmonic maps and their conjugates to define discrete conformal maps. We will extend a conformal energy proposed by Hutchinson [**15**] to the discrete spaces $S_h \times S_h^*$ and show that the discrete holomorphic maps have zero conformal energy, a property generically not available for conforming piecewise linear maps.

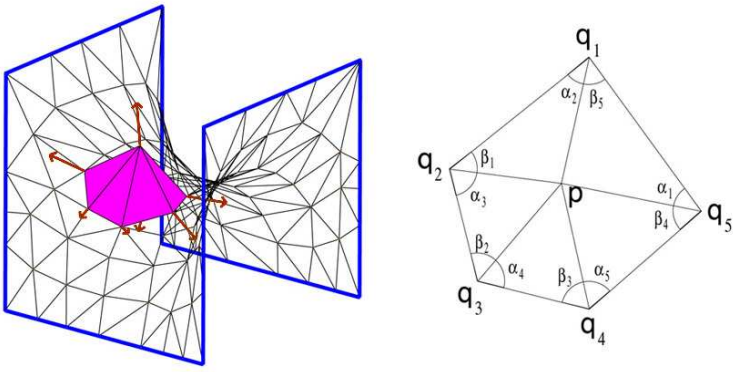


FIGURE 9. Discrete mean curvature vector on a polyhedral surface given as a Laplace–Beltrami operator of the identity map from the surface to itself.

We start with a review of the Dirichlet problem of harmonic maps in Section 2.1 followed by the discretization using conforming Lagrange elements in Section 2.2. In Section 2.3 we discretize the same Dirichlet problem using the nonconforming Crouzeix–Raviart elements, and derive a pointwise expression of the discrete minimality condition. Section 2.4 contains the main results of this section, namely, identifying solutions in both finite element spaces as pairs of discrete conjugate harmonic maps. Applications of the results are given in Section 3 to the conjugation of discrete minimal and constant mean curvature surfaces.

2.1. Review of smooth harmonic maps. On a Euclidean domain, the Laplace operator is given by the second partial derivatives

$$\Delta = \frac{\partial}{\partial x_1^2} + \cdots + \frac{\partial}{\partial x_n^2}.$$

Harmonic maps $u : \Omega \to \mathbb{R}$ on an open set Ω in $\mathbb{R}^n$ are solutions of the Laplace equation

$$\Delta u = 0 \text{ in } \Omega, \tag{2–1}$$

which often appears with prescribed boundary conditions. Dirichlet conditions prescribe fixed boundary values in the form of a function g,

$$u_{|\partial\Omega} = g \text{ on } \partial\Omega,$$

and Neumann conditions prescribe the derivative of u in the direction of the normal ν of the boundary

$$\partial_\nu u_{|\partial\Omega} = \mu \text{ on } \partial\Omega.$$

Dirichlet and Neumann boundary conditions may appear simultaneously on disjoint segments of the boundary.

The Laplace operator of vector-valued maps, and thereby the harmonicity of vector-valued maps, is defined component-wise on each coordinate function. For functions $u : M \to \mathbb{R}$ on a manifold M with a Riemannian metric g the Laplace–Beltrami operator Δ_g is a generalization of the Laplace operator. Assume normal coordinates around a point p on M and let $\{e_1, \ldots, e_n\}$ be the induced orthonormal frame in the tangent space of M, then

$$\Delta_g = \nabla_{e_1}\nabla_{e_1} + \cdots + \nabla_{e_n}\nabla_{e_n}.$$

Harmonic maps also appear as minimizers of the *Dirichlet energy*

$$E_D(u) = \frac{1}{2}\int_M |\nabla u|^2 \, dx \tag{2–2}$$

with Dirichlet conditions (or Neumann) at the boundary, since the Laplace equation (2–1) is the Euler–Lagrange equation of the Dirichlet energy. To see this, let $u(t) := u_0 + t\phi : M \to \mathbb{R}$ be any C^1-variation of a function u_0 whose variation function has compact support $\phi_{|\partial M} = 0$. Then by differentiation and integration by parts we obtain

$$\frac{d}{dt}_{|t=0} E_D(u(t)) = \int_M \langle \nabla u, \nabla \phi \rangle = -\int_M \Delta u \cdot \phi + \int_{\partial M} \partial_v u \cdot \phi,$$

where v is the exterior normal along ∂M. Since ϕ has compact support, the last integrand vanishes identically. Since the above equation holds for any C^1-variation we derive

$$\nabla E_D(u) = 0 \Longleftrightarrow \Delta u = 0$$

from the fundamental lemma of the calculus of variations. The minimizer $u_{\min}$ is unique since

$$\begin{aligned} E_D(u_{\min} + \phi) &= E_D(u_{\min}) + E_D(\phi), \\ &> E_D(u_{\min}) \quad \forall\, \phi_{|\partial M} = 0, \phi \neq 0, \end{aligned}$$

where the cross term vanishes because of the minimality condition for $u_{\min}$.

2.2. Discrete Dirichlet energy. There are different equivalent ways to introduce discrete harmonic maps. Here we use the characterization of harmonic maps as minimizers of the Dirichlet energy, since this approach also provides an efficient numerical algorithm to solve the boundary value problems for discrete harmonic maps.

DEFINITION 2.1. Let M_h be a simplicial surface in $\mathbb{R}^m$ and S_h the set of polyhedral maps on M_h. Then the *Dirichlet energy* of a function $u_h \in S_h$ with $u_h : M_h \to \mathbb{R}^d$ is given by

$$E_D(u_h) := \frac{1}{2} \sum_{T \in \mathfrak{T}_h} \int_T |\nabla u_h|^2 \, dx. \tag{2–3}$$

That is, the Dirichlet energy of u_h is the sum of the Dirichlet energies of the smooth atomic maps $u_{h|T}$ on each triangle T.

Now we consider critical points of the Dirichlet energy. For simplicity, we restrict to interior variations which keep the boundary values fixed.

DEFINITION 2.2. A *variation* $\phi(t) \in S_h$, $t \in [0, \varepsilon)$, is a family of functions differentiable in t such that each map $u_h \in S_h$ gives rise to a family of maps $u_h(t) \in S_h$ with

$$u_h(t) = u_h + \phi(t).$$

Basically, a variation of a function $u_h \in S_h$ is a modification of its values at each vertex p_i of the triangulation M_h given by $u_h(t)(p_i) = u_h(p_i) + \phi(t)(p_i)$. For simplicity, we restrict to Dirichlet boundary conditions, that is, the variations $\phi(t)$ are zero along the boundary of M_h.

DEFINITION 2.3. A critical point u_h in S_h of the Dirichlet energy (2–3) in S_h with respect to Dirichlet boundary conditions is called a *discrete harmonic map*.

In the following we derive an explicit representation of the Dirichlet energy of polyhedral maps and a system of equations for the discrete minimizers which characterize discrete harmonic maps.

Let $T = \{p_1, p_2, p_3\}$ be a triangle of a simplicial surface and oriented edges $\{c_1, c_2, c_3\}$ with $c_i = p_{i-1} - p_{i+1}$, and $\varphi_i : T \to \mathbb{R}$ be the Lagrange basis function at vertex p_i with $\varphi_i(p_j) = \delta_{ij}$. Then its gradient is

$$\nabla \varphi_{i|T} = \frac{1}{2 \operatorname{area} T} J c_i, \tag{2–4}$$

where J denotes rotation by $\pi/2$ oriented such that Jc_i points into the triangle. Note, that $\varphi_1 + \varphi_2 + \varphi_3 = 0$ implies $\nabla \varphi_i = -\nabla \varphi_{i-1} - \nabla \varphi_{i+1}$. The basis functions have mutual scalar products given by

$$\langle \nabla \varphi_{i-1}, \nabla \varphi_{i+1} \rangle = -\frac{\cot \alpha_i}{2 \operatorname{area} T}, \qquad \langle J \nabla \varphi_i, \nabla \varphi_{i+1} \rangle = \frac{1}{2 \operatorname{area} T},$$

$$|\nabla \varphi_i|^2 = \frac{\cot \alpha_{i-1} + \cot \alpha_{i+1}}{2 \operatorname{area} T}.$$

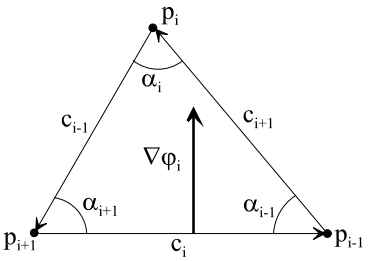


FIGURE 10. Gradient of basis function.

Since each function $u_h \in S_h$ has a representation

$$u_h(p) = \sum_{j=1}^{n} u_j \varphi_j(p) \quad p \in M_h,$$

where $u_j = u_h(p_j)$ denotes the function value of u_h at the vertex p_j of M_h, on a single triangle T the gradient of $u_{h|T} : T \longrightarrow \mathbb{R}^d$ is given by

$$\nabla u_{h|T} = \frac{1}{2 \operatorname{area} T} \sum_{j=1}^{3} u_j J c_j. \tag{2–5}$$

THEOREM 2.4. *Let M_h be a simplicial surface and S_h the set of continuous and piecewise linear functions on M_h. Then the discrete Dirichlet energy of any function $u_h \in S_h$ is given by*

$$E_D(u_h) = \frac{1}{4} \sum_{edges\ (x_i, x_j)} (\cot \alpha_{ij} + \cot \beta_{ij}) \left| u_h(p_i) - u_h(p_j) \right|^2. \tag{2–6}$$

Further, the minimizer of the Dirichlet functional (2–3) is unique and solves

$$\frac{d}{du_i} E_D(u_h) = \frac{1}{2} \sum_{x_j \in n(x_i)} (\cot \alpha_{ij} + \cot \beta_{ij})(u_h(p_i) - u_h(p_j)) = 0 \tag{2–7}$$

at each interior vertex p_i of M_h. The first summation runs over all edges of the triangulation, and the second summation over all edges emanating from p_i. The angles α_{ij} and β_{ij} are vertex angles lying opposite to the edge (p_i, p_j) in the two triangles adjacent to (p_i, p_j).

PROOF. Using the explicit representation (2–4) of the basis functions and the identity $\nabla \varphi_i = -\nabla \varphi_{i-1} - \nabla \varphi_{i+1}$, we obtain the Dirichlet energy of $u_{h|T}$:

$$\begin{aligned} E_D(u_{h|T}) &= \frac{1}{2} \int_T - \sum_{j=1}^{3} \left| u_{j+1} - u_{j-1} \right|^2 \langle \nabla \varphi_{j-1}, \nabla \varphi_{j+1} \rangle \\ &= \frac{1}{4} \sum_{j=1}^{3} \cot \alpha_j \left| u_{j+1} - u_{j-1} \right|^2. \end{aligned}$$

Summation over all triangles of M_h and combining the two terms corresponding to the same edge leads to equation (2–6).

At each interior vertex p_i of M_h, the gradient of E_D with respect to variations of $u_i = u_h(p_i)$ in the image of u_h is obtained by partial differentiation and easily derived from

$$\frac{d}{du_i} E_D(u_h) = \int_\Omega \langle \nabla u_h, \nabla \varphi_i \rangle \, .$$

Since S_h is a finite dimensional space, the quadratic minimization problem for the Dirichlet energy has a unique solution u_h in S_h. □

The definition of the Dirichlet energy of vector-valued maps $F_h : M_h \to N_h \subset \mathbb{R}^d$ is in full coherence with the definition of Dirichlet energy of scalar-valued maps. Namely, if the map $F_h = (f_1, \ldots, f_d)$ has component functions $f_i : M_h \to \mathbb{R}$, then we have

$$E_D(F_h) = \sum_{i=1}^{d} E_D(f_i)$$

since $|\nabla F_h|^2 = |\nabla f_1|^2 + \cdots + |\nabla f_d|^2$. Vector-valued harmonic maps are defined as critical values of the Dirichlet functional in the same way as in the scalar-valued case. Therefore, the balancing condition for scalar-valued harmonic maps directly gives a balancing formula for vector-valued discrete harmonic maps too.

The following definition includes more general boundary conditions. Neumann boundary conditions constrain the derivative of a function in direction of the exterior normal of the domain. Later we will make use of other boundary conditions which are useful for maps from a simplicial surface M_h to another surface N_h.

DEFINITION 2.5. A solution $u_h \in S_h$ of the Dirichlet problem (2–7) in S_h is called a *discrete harmonic map*. To include symmetry properties into this definition we allow in some cases also variation of boundary points:

- if a domain boundary arc and its corresponding image boundary arc are straight lines, then the interior boundary points may vary along the straight line in image space;
- if both corresponding arcs are planar symmetry curves restricted to planes we allow variation of interior boundary points in the image plane. This models also free boundary value problems;
- in all other cases the image boundary points remain fixed.

REMARK 2.6. At each vertex x_i equation (2–7) can be geometrically interpreted as a balancing condition for the weighted edges emanating from the vertex x_i. The weight of each edge solely depends on the angles in the base surface M_h, i.e., the weights depend only on the conformal structure of M_h.

Examples of Discrete Harmonic Maps. Simple examples of discrete harmonic maps are derived from the observation that on the integer grid $\mathbb{Z} \times \mathbb{Z}$ in $\mathbb{R}^2$ the interpolants of some smooth harmonic functions are discrete harmonic:

EXAMPLE 2.7. On a rectangular $\mathbb{Z} \times \mathbb{Z}$ grid in $\mathbb{R}^2$, which is triangulated by subdividing along either diagonal of each rectangle, the interpolating functions of $\operatorname{Re} z$, $\operatorname{Re} z^2$, $\operatorname{Re} z^3$ and $\operatorname{Im} z^4$ are discrete harmonic maps, and so are the interpolants of some other polynomials.

EXAMPLE 2.8. On a rectangular $\mathbb{Z} \times \mathbb{Z}$ grid in $\mathbb{R}^2$, the weight of each diagonal is $\cot \frac{\pi}{2}$, and it vanishes independent of the chosen diagonal in each square. Therefore,

at each grid point (i,j) only the discrete values of the *five-point stencil*

$$\{(i,j),(i,j-1),(i-1,j),(i+1,j),(i,j+1)\} \quad i,j \in \mathbb{Z}$$

of the finite difference Laplacian contribute to the Dirichlet gradient.

The next example leads to discrete harmonic maps on a simplicial surface using linear maps:

DEFINITION 2.9. Let M_h be a polyhedral surface in $\mathbb{R}^m$. A map $u_h \in S_h(M_h)$ from M_h to $\mathbb{R}^d$ is called a *linear map* if u_h is the restriction $u_{|M}$ of a linear map $u : \mathbb{R}^m \to \mathbb{R}^d$, i.e.,

$$u_h = u_{|M} : M_h \to \mathbb{R}^d.$$

For example, any coordinate function $x_i : M_h \to \mathbb{R}$ on a polyhedral surface M_h is a linear map, and, more generally, let $a \in \mathbb{R}^m$ be a constant vector, then

$$u_h(p) := \langle a, p \rangle \quad \forall\, p \in M_h$$

is linear.

On an arbitrary simplicial surface $M_h \subset \mathbb{R}^m$ the following geometric assumption on the underlying domain surface M_h leads to discrete harmonic functions:

EXAMPLE 2.10. All linear maps $u_h : M_h \to \mathbb{R}^d$ on a polyhedral surface M_h are discrete harmonic if and only if M_h is a discrete minimal surface.

PROOF. Using the Lagrange basis functions $\varphi_i : M_h \to \mathbb{R}$ associated to each vertex p_i of M_h we have the representation

$$u_h(x) = \sum_{p_i \in M_h} u_h(p_i)\varphi_i(p), \quad p \in M_h.$$

The gradient of the Dirichlet energy can be transformed using the linearity of u_h:

$$\begin{aligned}\frac{d}{du_i}E_D(u_h) &= \frac{1}{2}\sum_{j\in n(i)} (\cot\alpha_{ij} + \cot\beta_{ij})\,(u_h(p_i) - u_h(p_j)) \\ &= u_h\Big(\frac{1}{2}\sum_{j\in n(i)} (\cot\alpha_{ij} + \cot\beta_{ij})\,(p_i - p_j)\Big) = u_h\Big(\frac{d}{dp_i}E_D(\mathrm{id}_h\, M_h)\Big).\end{aligned}$$

Therefore, u_h is a critical value of the Dirichlet energy if and only if the identity map of M_h is discrete harmonic. The harmonicity of the identity map of a discrete minimal surface is shown in Corollary 3.7. □

Mean Value Property and Maximum Principle. Among the two most important properties of smooth harmonic maps are the mean value property and the maximum principle.

Mean Value Property: Let $p \in M$ and $U_\varepsilon(p)$ be a disk with radius ε around p. Then the value of a smooth harmonic function u at the center p is the average of the values along the boundary of the disk

$$u(p) = \frac{1}{2\pi\varepsilon}\int_{|q-p|=\varepsilon} u(q).$$

We obtain a discrete version for polyhedral maps if we replace the disk with a regular polygon.

LEMMA 2.11. *Let u_h be a discrete harmonic map defined on a simplicial surface M_h. If the* star *of a vertex p consists of congruent isosceles triangles centered at p, then*

$$u_h(p) = \frac{1}{\#\operatorname{link} p} \sum_{q_j \in \operatorname{link} p} u_h(q_j)$$

is the center of mass of the function values $\{u_h(q_j)\}$ on the link of p.

PROOF. All vertex angles appearing in equation (2–7) are the same. □

Maximum Principle: Since smooth harmonic maps solve an elliptic differential equation they satisfy a maximum principle. This means, in any open domain $U \subset M$ the maximum and minimum of u is attained at the boundary ∂U. In the discrete case, a similar statement for the star of a vertex does not hold in general, for example, it may fail if the spatial domain contains angles larger than 90 degrees.

LEMMA 2.12. *Let u_h be a discrete harmonic map defined on a spatial domain of a simplicial surface M_h formed by the points $\{q_j\}$ around a vertex p. If the triangles around p are all acute, then $u_h(p)$ is contained in the convex hull of the points $\{u_h(q_j)\}$.*

PROOF. From the local harmonicity condition (2–7) we see that $u_h(p)$ can be represented as a linear combination of the points $\{u_h(q_j\}$. Since all relevant angles are acute the weights of the $u_h(q_j)$ are in the interval $(0, 1)$, and $u_h(p)$ is a convex combination. □

The two previous lemmas do not hold if we allow more general domains. For example, if the domain contains obtuse triangles as in Figure 15 then neither the mean value nor the convex hull property may be valid.

The nonconvexity of discrete harmonic maps will lead to interesting counterexamples of the maximum principle of minimal surfaces in Section 3. In practical applications, for example, when smoothing meshes with a Laplace filtering or mapping surfaces onto a planar domain, then one would often like to ensure convexity. In these case the mesh parametrization by Floater [**9**] might be a useful strategy since it ensures convexity.

2.3. Nonconforming harmonic maps. Nonconforming maps on simplicial surfaces were introduced in Section 1.4.1 as another natural set of discrete maps. Let M_h be a simplicial surface; then we state the Dirichlet energy in the space S_h^* as in the previous section.

DEFINITION 2.13. Let M_h be a simplicial surface in $\mathbb{R}^m$. Then the *Dirichlet energy* of a function $v_h \in S_h^*$ with $v_h : M_h \to \mathbb{R}^d$ is given by

$$E_D(v_h) := \frac{1}{2} \sum_{T \in M_h} \int_T |\nabla v_h|^2 \, dx.$$

That is, the Dirichlet energy of v_h is the sum of the Dirichlet energies of the smooth atomic maps $v_{h|T}$ on each triangle T.

Now we consider critical points of the Dirichlet energy, and again, for simplicity, we restrict to interior variations which keep the boundary values fixed.

DEFINITION 2.14. A *variation* $\Psi(t) \in S_h^*$, $t \in [0, \varepsilon)$, is a family of functions differentiable in t such that each map $v_h \in S_h^*$ gives rise to a family of maps $v_h(t) \in S_h^*$ with

$$v_h(t) = v_h + \Psi(t).$$

Basically, a variation of a function $v_h \in S_h^*$ is a modification of its values at each edge midpoint m_i of the simplicial surface M_h given by $v_h(t)(m_i) = v_h(m_i) + \Psi(t)(m_i)$. For simplicity, we restrict to Dirichlet boundary conditions, that is, the variations $\Psi(t)$ are zero at midpoints of boundary edges of M_h.

DEFINITION 2.15. A critical point v_h in S_h^* of the Dirichlet energy (2–3) in S_h^* with respect to Dirichlet boundary conditions is called a *(nonconforming) discrete harmonic map.*

Using the identities in a Euclidean triangle T with vertices $\{p_1, p_2, p_3\}$ and oriented edges $\{c_1, c_2, c_3\}$ with $c_i = p_{i-1} - p_{i+1}$, we obtain on T the following representation of the basis functions $\psi_i \in S_h^*$ corresponding to edge c_i :

$$\nabla \psi_i = -2 \nabla \varphi_i = \frac{-1}{\operatorname{area} T} J c_i, \tag{2–8}$$

where $\varphi_i \in S_h$ is the conforming basis function corresponding to the triangle vertex p_i opposite to the edge c_i, and J is the rotation of an edge by $\pi/2$ such that Jc points in the opposite direction of the outer normal of the triangle.

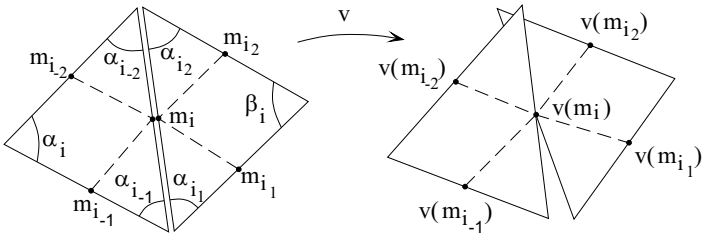


FIGURE 11. A nonconforming map is given by its values on edge midpoints.

THEOREM 2.16. *Let $v \in S_h^*$ be a nonconforming function on a simplicial surface M_h. Then the Dirichlet energy of v_h has the explicit representation*

$$E_D(v) = \sum_{\text{all edges } c_i} \cot \alpha_i \left| v_{i_{-2}} - v_{i_{-1}} \right|^2 + \cot \beta_i \left| v_{i_1} - v_{i_2} \right|^2, \tag{2–9}$$

where $\{i_{-2}, i_{-1}, i_1, i_2, \}$ denote indices of adjacent edge midpoints as shown in Figure 11, and v_{i_j} denotes the value $v(m_{i_j})$. The angles are measured on M_h.

The unique minimizer of the Dirichlet functional on M_h solves a system of equations such that at each edge midpoint m_i we have

$$\begin{aligned} \frac{d}{dv_i} E_D(v) &= 2(\cot \alpha_{i_{-2}}(v_i - v_{i_{-1}}) + \cot \alpha_{i_{-1}}(v_i - v_{i_{-2}}) \\ &\qquad + \cot \alpha_{i_1}(v_i - v_{i_2}) + \cot \alpha_{i_2}(v_i - v_{i_1})) \\ &= 0. \end{aligned} \tag{2–10}$$

PROOF. Since $\nabla\psi_i = -2\nabla\varphi_i$, the representation of the Dirichlet energy is a consequence of the explicit representation for conforming elements (2–6). On a single triangle T,

$$\begin{aligned} E_D(v_{|T}) &= \frac{1}{2}\int_T -\sum_{j=1}^{3} |v_{j+1}-v_{j-1}|^2 \langle \nabla\psi_{j-1}, \nabla\psi_{j+1}\rangle \\ &= \sum_{j=1}^{3} \cot\alpha_j \, |v_{j+1}-v_{j-1}|^2 . \end{aligned}$$

The support of a component of the gradient of the Dirichlet energy consists of those two triangles adjacent to the edge corresponding to this variable. Equation (2–10) follows directly from the representation on a single triangle T with edges $\{c_1, c_2, c_3\}$ and $c_1 + c_2 + c_3 = 0$,

$$\begin{aligned} \frac{d}{dv_i} E_D(v_{|T}) &= \int_T \langle \nabla v_{|T}, \nabla\psi_i\rangle = \frac{1}{\operatorname{area} T} \sum_{j=1}^{3} v_j \langle c_j, c_i\rangle \\ &= 2\cot\alpha_{i-1}(v_i - v_{i+1}) + 2\cot\alpha_{i+1}(v_i - v_{i-1}) \end{aligned} ,$$

by combining the expression for the two triangles in the support of ψ_i. □

2.4. Conjugate harmonic maps. Discrete harmonic maps have been well studied as a basic model problem in finite element theory, while the definition of the conjugate of a discrete harmonic map was not completely settled. In this section we are interested in pairs of discrete harmonic maps on a Riemann surface M which are both minimizers of the Dirichlet energy

$$E(u) = \frac{1}{2}\int_M |\nabla u|^2 \, dx,$$

and are conjugate, i.e., solutions of the Cauchy–Riemann equations

$$dv = *du.$$

We note that generically such pairs do not exist in the space of piecewise linear conforming Lagrange finite elements S_h, but the problem naturally leads to the space of piecewise linear nonconforming Crouzeix–Raviart elements S_h^*. S_h alone is too rigid to contain the conjugate of a generic discrete harmonic function.

We define the conjugate harmonic maps of discrete harmonic maps in S_h and in S_h^*. A smooth harmonic map $u : M \to \mathbb{R}$ on an oriented Riemannian surface M and its conjugate harmonic map $u^* : M \to \mathbb{R}$ solve the Cauchy–Riemann equations

$$du^* = *du$$

where $*$ is the Hodge star operator with respect to the metric in M. In the discrete version, we denote by J the rotation through $\pi/2$ in the oriented tangent space of M, and start with a locally equivalent definition as Ansatz:

DEFINITION 2.17. Let $u \in S_h$, respectively S_h^*, be a discrete harmonic map on a simplicial surface M_h with respect to the Dirichlet energies in S_h, respectively S_h^*. Then its conjugate harmonic map u^* is defined by the requirement that it locally fulfills

$$\nabla u^*_{|T} = J\nabla u_{|T} \quad \forall \text{ triangles } T \in M_h. \tag{2–11}$$

The remainder of the section is devoted to prove that the discrete conjugate map is well-defined by showing the closedness of the differential $*du$, and to prove the harmonicity properties of its integral u^*.

To avoid case distinctions we represent each function with respect to the basis functions ψ_i of S_h^* such that on each triangle

$$u_{|T} = \sum_{i=1}^{3} u_i \psi_i,$$

where u_i is the function value of u at the midpoint of edge c_i. We use the same notation for $u^*_{|T}$, and obtain by (2–11)

$$\sum_{i=1}^{3} u_i^* \nabla \psi_i = \sum_{i=1}^{3} u_i J \nabla \psi_i. \tag{2–12}$$

LEMMA 2.18. *Let T be a triangle with oriented edges $\{c_1, c_2, c_3\}$, $c_1+c_2+c_3 = 0$. A pair of linear functions u and u^* related by equation (2–12), has values at edge midpoints related by*

$$\begin{pmatrix} u_3^* - u_1^* \\ u_3^* - u_2^* \end{pmatrix} = \begin{pmatrix} \cot\alpha_3(u_2 - u_1) + \cot\alpha_1(u_2 - u_3) \\ \cot\alpha_3(u_2 - u_1) + \cot\alpha_2(u_3 - u_1) \end{pmatrix}. \tag{2–13}$$

PROOF. The representation (2–8) of $\nabla\psi_i$ converts equation (2–12) to

$$\sum_{i=1}^{3} u_i^* J c_i = \sum_{i=1}^{3} u_i c_i.$$

Using $-c_3 = c_1 + c_2$, we express the left side of the above equation as a vector in the span of $\{Jc_1, Jc_2\}$,

$$(u_3^* - u_1^*)\, Jc_1 + (u_3^* - u_2^*)\, Jc_2 = \sum_{i=1}^{3} u_i c_i.$$

If the triangle T is nondegenerate, then the matrix (Jc_1, Jc_2) has rank 2, and scalar multiplication with c_1 and c_2 yields

$$\begin{pmatrix} u_3^* - u_1^* \\ u_3^* - u_2^* \end{pmatrix} = \frac{2}{\operatorname{area} T} \sum_{i=1}^{3} u_i \begin{pmatrix} \langle c_2, c_i \rangle \\ -\langle c_1, c_i \rangle \end{pmatrix},$$

which easily transforms to equation (2–13). □

Now we consider a discrete harmonic map $u \in S_h$ and prove local exactness of its discrete conjugate differential.

PROPOSITION 2.19. *Let M_h be a simply connected simplicial surface and $u \in S_h$ with $u : M_h \longrightarrow \mathbb{R}^d$ an edge continuous discrete harmonic function. Then the discrete Cauchy–Riemann equations (2–11) have a globally defined solution $u^* : M_h \longrightarrow \mathbb{R}^d$ with $u^* \in S_h^*$. Two solutions u_1^* and u_2^* differ by an additive integration constant.*

PROOF. We define the discrete differential du^* of u^* by the property that, on each triangle T,

$$du^*_{|T} := *du_{|T}.$$

Since $u_{|T}$ is a linear map, the conjugate differential $du^*_{|T}$ is well defined and there exists a unique smooth solution $u^*_{|T}$ of the smooth Cauchy–Riemann equations on T, up to an additive constant. By Lemma 2.18, $u^*_{|T}$ is explicitly given in terms of $u_{|T}$ and T.

If $u \in S_h$ is a discrete harmonic map, it turns out that du^* is closed along closed paths on M_h that cross edges only at their midpoints. Since du^* is closed inside each triangle, it is sufficient to prove closedness for a path γ in the vertex star of a vertex $p \in M_h$ such that $\gamma_{|T}$ linearly connects the midpoints of the two edges of T having p in common; see Figure 12. Let $\{m_1, \ldots, m_k\}$ be the sequence

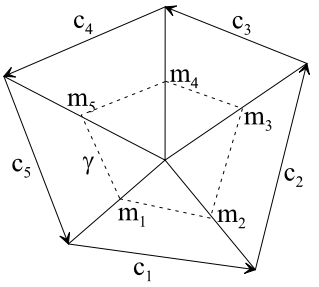


FIGURE 12. Dual edge graph γ around a vertex.

of edge midpoints determining γ. The edges $d_j := m_{j+1} - m_j$ of γ are parallel to c_j with $c_j = 2d_j$. We use equation (2–13) in each triangle to derive

$$\int_\gamma du^* = \sum_{j=1}^{k} \int_{\gamma_{|T_j}} *du_{|T_j} = \sum_{j=1}^{k} \langle J\nabla u_{|T_j}, d_j \rangle = -\frac{1}{2} \sum_{j=1}^{k} \langle \nabla u_{|T_j}, Jc_j \rangle = 0,$$

since u is harmonic in S_h; see equation (2–7). Therefore, du^* is closed along the dual edge graph through the edge midpoints of M_h, and $u^* \in S_h^*$ is globally defined on simply connected regions of M_h. □

For a harmonic map $u \in S_h$, the following proposition proves harmonicity of the conjugate map $u^* \in S_h^*$.

PROPOSITION 2.20. *Let $u \in S_h$ be a discrete harmonic map on a simplicial surface M_h and let $u^* \in S_h^*$ be a solution of the discrete Cauchy–Riemann equations (2–11) given by Proposition 2.19. Then u^* has the same Dirichlet energy as u, and u^* is discrete harmonic in S_h^*.*

PROOF. Let u^* be the solution of the discrete Cauchy–Riemann equations (2–11) for a discrete harmonic map $u \in S_h$. Then we show that u^* is a critical point of the nonconforming Dirichlet energy in S_h^* by rewriting the Dirichlet gradient (2–10) of u^* in terms of values of u.

On a single triangle T with midpoint m_i on edge c_i, we note that

$$\langle J\nabla u_{|T}, \nabla \psi_i \rangle = \frac{2}{\operatorname{area} T}(u(m_{i-1}) - u(m_{i+1})) \ \forall i \in \{1,2,3\}, \tag{2–14}$$

which follows directly from $\nabla u = \sum_{j=1}^{3} u(m_j)\nabla\psi_j$ and

$$\langle J\nabla\psi_j, \nabla\psi_i\rangle = \begin{cases} 0 & \text{for } j = i, \\ \dfrac{2}{\text{area } T} & \text{for } j = i-1, \\ \dfrac{-2}{\text{area } T} & \text{for } j = i+1. \end{cases}$$

Let $T_1 \cup T_2$ denote the two triangles forming the support of ψ_i as shown in Figure 13. Using equation (2–14) we obtain

$$\begin{aligned} \frac{d}{du_i^*} E_D(u^*) &= \int_{T_1 \cup T_2} \langle \nabla u^*, \nabla\psi_i \rangle \\ &= 2(u(m_{i_{-2}}) - u(m_{i_{-1}})) + 2(u(m_{i_1}) - u(m_{i_2})). \end{aligned}$$

Since u is linear we can rewrite the differences at edge midpoints as differences of u at vertices on the common edge of T_1 and T_2, and obtain

$$\frac{d}{du_i^*} E_D(u^*) = u(V_{j_{-1}}) - u(V_{j_{-2}}) + u(V_{j_2}) - u(V_{j_1}). \tag{2–15}$$

This equation relates the energy gradient of u^* to the function values at vertices of u. We emphasize the fact that the derivation of the equation does not use edge continuity of u, which will allow us to use (2–15) in the proof of Theorem 2.21. The

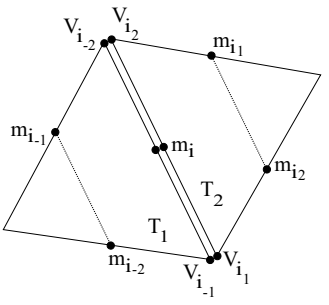


FIGURE 13. Notation of edge midpoints in a pair of triangles.

right-hand side of (2–15) vanishes if and only if

$$u_{|e_i \text{ in } T_1} = u_{|e_i \text{ in } T_2} + \textit{constant}.$$

Therefore, the harmonicity of u^* follows from, and is equal to, the edge continuity of $u \in S_h$. □

The following main theorem states the complete relationship between harmonic maps in S_h and S_h^*, and includes the previous propositions as special cases.

THEOREM 2.21. *Let M_h be a simplicial surface in $\mathbb{R}^m$ and S_h respectively S_h^* the space of conforming, respectively nonconforming, maps from M_h into $\mathbb{R}^d$. Then we have the following duality of Laplace–Beltrami harmonic maps on M_h:*

(1) *Let $u \in S_h$ be a minimizer of the Dirichlet energy in S_h. Then its conjugate map u^* is in S_h^* and is discrete harmonic.*

(2) *Let $v \in S_h^*$ be a minimizer of the Dirichlet energy in S_h^*. Then its conjugate map u is in S_h and discrete harmonic.*
(3) *Let $u \in S_h$ and S_h^* be discrete harmonic in S_h and S_h^*, respectively. Then $u^{**} = -u$.*

PROOF. (1) The first statement was proved in Propositions 2.19 and 2.20.

(2) Let $v \in S_h^*$ given by $v = \sum v_i \psi_i$ be discrete harmonic. Along the lines of the proof for the corresponding Proposition 2.19 concerning S_h, we define $v^*_{|T}$ (up to an additive integration constant) as the well-defined integral of

$$dv^*_{|T} := *dv_{|T} \quad \forall\, T \in M_h,$$

which uniquely exists since $v_{|T}$ is linear. Using the same arguments as in the proof of Proposition 2.20 and $\nabla v^* = J \nabla v$, we derive an equation for v that is identical to equation (2–15) for u:

$$\frac{d}{dv_i} E_D(v) = v^*(V_{j_{-1}}) - v^*(V_{j_{-2}}) + v^*(V_{j_2}) - v^*(V_{j_1}),$$

where V_{j_k} are vertices as denoted in Figure 13. Since v is harmonic, we can choose the integration constants of v^* such that v^* becomes edge continuous and lies in S_h.

The harmonicity property of v^* follows from the closedness of v. Let $v^* = \sum v_i^* \varphi_i \in S_h$, and then splitting $\nabla \psi_i = -\nabla \psi_{i_j} - \nabla \psi_{i_{j+1}}$ in each triangle, we obtain

$$\begin{aligned}
\frac{d}{dv_i^*} E_D(v^*) &= \int_{M_h} \left\langle \nabla v^*, \frac{d}{dv_i^*} \nabla v^* \right\rangle = \int_{star(p_i)} \langle J \nabla v, \nabla \varphi_i \rangle \\
&= \sum_j \int_{T_{i_j}} \left\langle J \nabla v, -\frac{1}{2} (\nabla \psi_{i_j} + \nabla \psi_{i_{j+1}}) \right\rangle \\
&= \sum_j \int_{T_{i_j}} \frac{1}{\operatorname{area} T_{i_j}} ((v_{i_{j+1}} - v_{i_{j-1}}) + (v_{i_{j-1}} - v_{i_j})) = \sum_j v_{i_{j+1}} - v_{i_j} = 0,
\end{aligned}$$

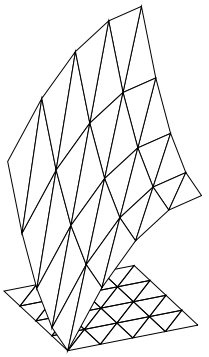
since $v \in S_h^*$ is closed on the path around each vertex p_i. Therefore v^* is critical for the Dirichlet energy in S_h.

(3) The third statement is a direct consequence of applying the $*$ operator twice, which rotates the gradient in each triangle by π in the plane of the gradient. □

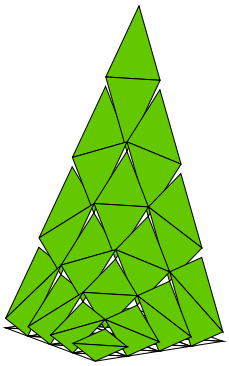
COROLLARY 2.22. *Conjugation is a bijection between discrete Laplace–Beltrami harmonic maps in S_h and S_h^*, where each pair (u, v) fulfills the discrete Cauchy–Riemann equations. Further, corresponding maps have the same Dirichlet energy.*

PROOF. The proof of Theorem 2.21 and the previous propositions show that, for a pair (u, v) of harmonic conjugate functions $u \in S_h$ and $v \in S_h^*$, the harmonicity condition of u is equal to the closedness condition of v, and the closedness condition of u is equal to the harmonicity condition of v.

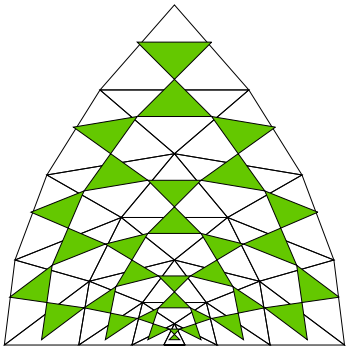
The equality of the Dirichlet energies follows directly from the Cauchy–Riemann equations. □



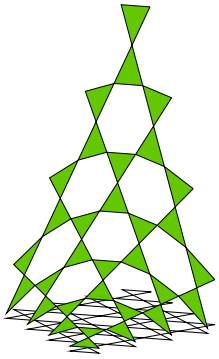
Discrete harmonic map
$u \in S_h$ interpolating $\operatorname{Re} z^2$



Conjugate harmonic $u^* \in S_h^*$
is a nonconforming map



Holomorphic pair (u, u^*) and
exact solution as full grid



u^* applied to center
quarter of each triangle.

2.5. Minimizing with conjugate gradients. For completeness we will mention some of the numerical methods to practically solve the variational problems which we discussed so far. These methods apply to both the conforming and nonconforming meshes.

Let u_h be a map from a simplicial surface M_h satisfying a Dirichlet boundary value problem

$$u_h : M_h \to \mathbb{R}^3, \qquad u_{h|\partial M} = \Gamma.$$

With respect to the Lagrange basis functions, u_h is given by

$$u_h(x) = \sum_{i=1}^{n} u_i \varphi_i(x).$$

Assume that we ordered the set vertices of M_h by interior and boundary vertices $\{p_1, \ldots, p_I, p_{I+1}., p_{I+B}\}$. Then the harmonicity condition at each interior vertex p_j

is

$$\begin{aligned}\frac{d}{du_j}E_D(u_h) &= \sum_{i=1}^{n} u_i \int_{M_h} \langle \nabla\varphi_i, \nabla\varphi_j\rangle \\ &= \sum_{i=1}^{I} u_i \int_{M_h} \langle \nabla\varphi_i, \nabla\varphi_j\rangle + \sum_{i=I+1}^{I+B} u_i \int_{M_h} \langle \nabla\varphi_i, \nabla\varphi_j\rangle \\ &= 0 \quad \forall\, j \in \{1,\ldots,I\}\,.\end{aligned}$$

This system of equations is equivalent to a single matrix equation

$$Au = -B,$$

where $A = (a_{ji})$ is an $I \times I$ matrix, the so-called *stiffness matrix*, and $u = (u_i)$ and $B = (b_j)$ are I-dimensional vectors with

$$a_{ji} = \int_{M_h} \langle \nabla\varphi_i, \nabla\varphi_j\rangle\,, \qquad b_j = \sum_{i=I+1}^{I+B} u_i \int_{M_h} \langle \nabla\varphi_i, \nabla\varphi_j\rangle\,.$$

In practice it is usually more efficient not to solve the matrix system but to employ a conjugate gradient method, which is an iterative method with a fast convergence, especially during the first iteration steps. See the comments of Brakke [**3**], who compared our method with other minimization algorithms built into the Surface Evolver.

The method of *steepest descent* is an iterative algorithm which incrementally reduces the energy by modifying the function u_h a small distance ε in the direction of the negative of the energy gradient:

$$u_0 := u_h, \qquad u_{i+1} := u_i - \varepsilon \nabla E_p(u_i).$$

The *conjugate gradient method* is a more efficient method where the direction vector is modified such that previous optimizations are not spoiled. It uses a sequence of *line minimizations*: given $p \in \mathbb{R}^n$, direction $n \in \mathbb{R}^n$ and an energy functional $E : \mathbb{R}^n \to \mathbb{R}$. Find a scalar λ that minimizes

$$E(p + \lambda n) \to \min,$$

and then replace p by $p + \lambda n$. If the energy functional is differentiable, the gradient of E is an obvious choice for a direction. Such a gradient method can be made more efficient by incorporating second-order information, which avoids spoiling of previous results.

The Taylor expansion around p gives

$$E(x) = E(p) + \nabla E_p(x) + \tfrac{1}{2}\nabla^2 E_p(x,x) + \cdots \approx c - bx + \tfrac{1}{2}x^t Ax.$$

For a quadratic function E the gradient can be written as

$$\nabla E(x) = Ax - b.$$

How does the gradient change along some direction ν?

$$\partial_v \nabla E = A \cdot \partial_v x = Av.$$

The idea of the conjugate gradient method can be summarized as follows: assume we have moved along some direction u to a minimum and now want to move along a new direction v. Then v shall not spoil our previous minimization, i.e., the change of the gradient shall be perpendicular to u:

$$0 = \langle u, \partial_v (\nabla E)\rangle = uAv.$$

The vectors u and v are called *conjugate directions* which can be constructed using the following Gram–Schmidt biorthogonalization procedure employed in the methods of Fletcher–Reeves and Polak–Ribiere [**26**][**33**].

Let A be a positive-definite, symmetric $n \times n$ matrix. Let g_0 be an arbitrary vector, and $h_0 = g_0$. For $i = 0, 1, 2, \ldots$ define the two sequences of vectors

$$g_{i+1} = g_i - \lambda_i A h_i, \qquad h_{i+1} = g_{i+1} + \gamma_i h_i, \tag{2–16}$$

where λ_i, respectively γ_i, are chosen to obtain mutually orthogonal vectors $g_{i+1} \cdot g_i = 0$, respectively mutually conjugate directions $h_{i+1} A h_i = 0$ that is,

$$\lambda_i = \frac{g_i \cdot g_i}{g_i A h_i}, \qquad \gamma_i = -\frac{g_{i+1} A h_i}{h_i A h_i}.$$

If denominators are zero take $\lambda_i = 0$ or $\gamma_i = 0$, respectively. Then

$$g_i \cdot g_{i+1} = 0 \qquad \text{and} \qquad h_i A h_j = 0 \qquad \forall i \neq j$$

and the biorthogonalization procedure has produced a sequence g_i where each g_i is orthogonal and each h_i is conjugate to its set of predecessors.

Generally, the Hessian matrix A is not known. In this case the following observation provides the essential hints. Assume E is a quadratic functional and we take

$$g_i := -\nabla E|_{p_i} \quad \text{for some point } p_i.$$

Then we proceed from p_i along the direction h_i to the local minimum of E which is located at some point p_{i+1}. If we set again $g_{i+1} := -\nabla E|_{p_{i+1}}$, then this vector g_{i+1} is exactly the vector which would have been obtained by the above equations (2–16) but without the knowledge of the Hessian A. More precisely, the matrix A never needs to be computed.

Summarizing, the conjugate gradient method computes a set of directions h_i using only line minimizations, the evaluations of the energy gradient, and an auxiliary vector to store the recent vectors g_i. In practice, further optimizations are obtained through preconditioning.

2.6. Discrete Laplace operators. The discretization of the second order Laplace operator for smooth functions to simplicial meshes may be pursued in different ways. Depending on the structure of and information about the underlying mesh, the Laplace operator may include more combinatorial or more geometric information. Here we review some basic combinatorial Laplacians and then relate them with the Laplace–Beltrami operator in the context of the function spaces used in this section.

Combinatorial Laplacian. The purely combinatorial point of view ignores metric information like edge length or vertex angles of a mesh. All information about a combinatorial mesh is contained in its connectivity. For theoretical purposes it is convenient to express the connectivity in the form of the adjacency matrix.

DEFINITION 2.23. Let $\{p_1, \ldots, p_n\}$ be the vertices of a mesh. Then the *adjacency matrix* A of the mesh connectivity is an $n \times n$ matrix given by

$$A_{ij} = \begin{cases} 1 & \text{if } p_i p_j \text{ is an edge,} \\ 0 & \text{else.} \end{cases}$$

The matrix A is sparse, and the sum of the i-th row or i-th column is equal to the valence d_i of the vertex i. Note, in practical applications one would never explicitly store the full matrix.

DEFINITION 2.24. Let D be an $n \times n$ diagonal matrix with entries $d_{ii} := \frac{1}{d_i}$ where d_i is the valence of the vertex p_i, then the matrix

$$L := \mathrm{id} - DA, \qquad L_{ij} = \begin{cases} 1 & \text{if } i = j, \\ -\dfrac{1}{d_i} & \text{if } p_i p_j \text{ is an edge}, \\ 0 & \text{else} \end{cases}$$

is the *combinatorial Laplacian* of the mesh, or in short, the *mesh Laplacian.*

Let e_i be the vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 at the i-th position which is associated to p_i. Then

$$Le_i = e_i - \frac{1}{d_i} \sum_{j \in n(i)} e_j$$

where $n(i)$ denotes the set of vertices adjacent to p_i excluding p_i.

Karni and Gotsmann [**19**] extend the mesh Laplacian in the framework of mesh compression to include distance information

$$GL(p_i) = p_i - \frac{\displaystyle\sum_{j \in n(i)} \frac{1}{|p_i - p_j|} p_j}{\displaystyle\sum_{j \in n(i)} \frac{1}{|p_i - p_j|}}.$$

Five-Point Laplacian. The five-point Laplacian is the $2d$-extension of the finite difference Laplacian on the real axis. Consider a real-valued function $f : \mathbb{R} \to \mathbb{R}$ on an interval of the real axis. Then the smooth Laplacian Δf is defined as second derivative of f. In the discrete case, let $\{u_i\}$ be a uniform knot vector on the axis, for example, $u_i := i$, then $f_h''(x_i)$ can be approximated using finite differences:

$$\begin{aligned} f_h''(x_i) &= \tfrac{1}{2}\left(f_h'(x_{i+1}) - f_h'(x_i)\right) = \tfrac{1}{2}\left(\left(f(x_{i+1}) - f(x_i)\right) - \left(f(x_i) - f(x_{i-1})\right)\right) \\ &= \tfrac{1}{2}\left(f(x_{i+1}) - 2f(x_i) + f(x_{i-1})\right). \end{aligned}$$

This can be written in matrix representation

$$\Delta f_h = -A f_h$$

with $f = (f(x_0), f(x_1), \ldots)$ and the matrix

$$A = \frac{1}{2} \begin{pmatrix} 0 & & & & & & & & \\ -1 & 2 & -1 & & & & & & \\ & -1 & 2 & -1 & & & & & \\ & & & \cdots & & & & & \\ & & & & \cdots & -1 & 2 & 1 & \\ & & & & & & & 0 & \end{pmatrix}.$$

A smoothing operation of f can be performed by a so-called Gaussian filtering method

$$f_h^{j+1} := f_h^j + \lambda \Delta f_h^j$$

with a scalar factor $0 < \lambda < 1$. Other values of λ will enhance the variation of f. In matrix form we have

$$f_h^{j+1} = f_h^j - \lambda A f_h^j.$$

Discrete Laplace–Beltrami Operator on Surfaces. Let M_h be a simplicial surface in $\mathbb{R}^m$. We now define the Laplace operator for piecewise linear functions in S_h, respectively S_h^*, similar to the derivation of the discrete Dirichlet energy. Since second derivatives are involved the discrete Laplace–Beltrami operator will be a function on the vertices and edge midpoints, and it will not extend as a piecewise linear function over the whole triangulation.

DEFINITION 2.25. Let $u \in S_h$ be a map $u : M_h \to \mathbb{R}^d$ on a simplicial surface M_h with set of vertices V_h. Then the (total) *discrete Laplacian* $\Delta_h u(p) \in \mathbb{R}^d$ at each vertex $p \in V_h$ is defined as

$$\Delta_h u(p) := -\int_{\text{star } p} \langle \nabla u, \nabla \varphi_p \rangle \,. \tag{2–17}$$

Similarly, let $u \in S_h^*$ be a nonconforming map, then $\Delta_h^* : S_h^* \to V_h^*$ at an edge midpoint m is given by

$$\Delta_h^* u(m) := -\int_{\text{star } m} \langle \nabla u, \nabla \psi_m \rangle \tag{2–18}$$

with basis functions $\varphi_p \in S_h$ and $\psi_m \in S_h^*$.

Explicitly, we have at an interior vertex p and an interior edge midpoint m,

$$\begin{aligned}
\Delta_h u(p) &= -\frac{1}{2} \sum_{q_i \in n(p)} (\cot \alpha_i + \cot \beta_i)(u(p) - u(q_i)), \\
\Delta_h^* u(m) &= -2(\cot \alpha_{-2}(u(m) - u(m_{-1})) + \cot \alpha_{-1}(u(m) - u(m_{-2})) \\
&\quad + \cot \alpha_1 (u(m) - u(m_2)) + \cot \alpha_2 (u(m) - u(m_1))),
\end{aligned}$$

where $\{q_i\}$ is the set of vertices on the link of p, and $\{m_i\}$ the set of vertices on the link of m in counter-clockwise order and vertex angles α_i opposite to m_i in each triangle.

3. Discrete Minimal Surfaces

Minimal surfaces are characterized by having least area compared to nearby surfaces with the same boundary. This variational property, which was the original interest in minimal surfaces, was soon relaxed to include unstable critical points as well. Equivalently, these surfaces can be geometrically characterized by having vanishing mean curvature.

Examples have played a central part in the development of the minimal surface theory and fruitfully complemented the theoretical research. In recent years many new examples were studied experimentally using elaborate calculations for the analytic continuation of complex functions and the integration of the Weierstraß representation formulas. Although these methods allow one to compute any surface given by its Weierstraß representation, this analytic approach has the drawback that the Weierstraß formulas must be known in advance. Since the existence of many unstable minimal surfaces was mathematically proved indirectly via the

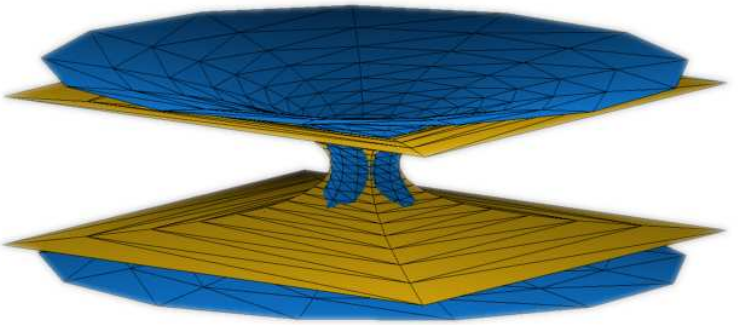


FIGURE 14. Asymptotic growth of two complete discrete catenoids depends on the dihedral symmetry.

so-called *conjugate surface construction* there was a strong need to develop a numerical scheme and actually compute the conjugate surface of a minimal surface [**17**][**18**].

The numerical method developed in [**25**] jointly with Pinkall was the first scheme to compute the conjugate of a numerically computed minimal surface. The key insight came from a new understanding of the geometric and variational properties of triangle nets. The method was later extended in [**23**] jointly with Oberknapp to the computation of constant mean curvature surfaces via a conjugation of minimal surfaces in the unit sphere S^3 in $\mathbb{R}^4$.

The main theoretical result in this section is a new precise understanding of the variational properties of pairs of discrete conjugate minimal surfaces, which was not known in the original works, by working in the function space of nonconforming triangle meshes. This section also introduces discrete minimal surfaces and derives their variational properties; we define the mean curvature normal as an operator on the discrete mesh.

Another important result is an explicit description of some complete discrete minimal surfaces which were jointly discovered with Rossman [**32**]. For example, these descriptions allow one to construct unstable discrete surfaces whose shape is given by exact coordinates, a fact which is particularly useful for the study of higher order properties like the index of minimal surfaces.

3.1. Review of the smooth variation of area. Let $F : \Omega \to M \subset \mathbb{R}^3$ be a parameterized surface of a domain $\Omega \subset \mathbb{R}^2$. A variation of M is a family of surfaces given by a differentiable map

$$G : \Omega \times (-\varepsilon, \varepsilon) \to \mathbb{R}^3, \qquad G(x, 0) = F(x) \quad x \in \Omega.$$

The induced vector field on M,

$$Y : \Omega \to TM, \qquad Y(x) = \frac{d}{d\varepsilon} G(x, \varepsilon)_{|\varepsilon=0},$$

is called the *first variation* of G.

LEMMA 3.1. *For a given surface M and a variation vector field Y the first variation of the area functional at M in the direction of Y is defined by*

$$\delta \operatorname{area}(M, Y) := \frac{d}{d\varepsilon} \operatorname{area}(M_\varepsilon)_{|\varepsilon=0} \in \mathbb{R}$$

and given by

$$-\delta \operatorname{area}(M, Y) = \int_{\partial M} \langle Y, \nu \rangle \, ds + 2 \int_M \langle Y, N \rangle \, H dA \ ,$$

where ν is the outer normal along ∂M.

PROOF. See [**6**] or [**16**]. □

If $Y = \lambda N$ is a normal variation then the boundary component vanishes and we have

$$\delta \operatorname{area}(M, Y) = -2 \int_M \lambda H dA.$$

Further, if $\lambda \equiv 1$ and H is constant we obtain

$$H = -\frac{\delta \operatorname{area}(M, N)}{\operatorname{area}(M)}.$$

3.2. First variation of the discrete area and volume. A variation of a polyhedral surface is determined by a variation of its vertices with the same mesh connectivity. For simplicity we require a C^2 variation but often a differentiability of lower order is sufficient.

DEFINITION 3.2. Let $\mathcal{P} = \{p_1, \ldots, p_m\}$ be the set of vertices of a discrete surface M_h. A *variation* $M_h(t)$ of M_h is defined as a C^2 variation of the vertices p_i

$$p_i(t) : [0, \epsilon) \to \mathbb{R}^d \quad \text{so that} \quad p_i(0) = p_i \ \forall i = 1, \ldots, m.$$

The straightness of the edges and the flatness of the triangles are preserved as the vertices move.

Formally, we have for each t that $p_i(t) \in S_h$ and $p_i(0) = \mathrm{id}(M_h)$ is the identity map of M_h.

Up to first order a variation is given by a set of vectors $\mathcal{V} = \{v_1, \ldots, v_m\}$, $v_i \in \mathbb{R}^d$ defined on the vertices $\mathcal{P} = \{p_1, \ldots, p_m\}$ of M_h. Often we restrict a variation to interior vertices by assuming $v_i = 0 \in \mathbb{R}^d$ along the boundary, or add special constraints on the boundary of M_h. The vectors v_{p_j} are the *variation vector field* such that the variation has the form

$$p_j(t) = p_j + t \cdot v_{p_j} + \mathcal{O}(t^2), \tag{3–1}$$

that is, $p_j'(0) = v_{p_j}$. We define the vector $\vec{v} \in \mathbb{R}^{dm}$ by

$$\vec{v}^t = (v_1^t, \ldots, v_m^t). \tag{3–2}$$

In the following we will restrict to $d = 3$ which allows the use of a well-defined normal vector although many results hold in higher codimension too.

In the smooth situation, the variation at interior points is typically restricted to normal variation since the tangential part of the variation only performs a reparametrization of the surface. However, on discrete surfaces there is an ambiguity in the choice of normal vectors at the vertices, so we allow arbitrary variations.

In the following we derive the evolution equations for some basic discrete operators under variation $M_h(t)$ of a discrete surface M_h.

Recall that the area of a discrete surface is

$$\operatorname{area} M_h := \sum_{T \in \mathcal{T}} \operatorname{area} T,$$

where $\operatorname{area} M_h$ denotes the Euclidean area of the triangle T as a subset of $\mathbb{R}^3$.

At each vertex p of M_h, the gradient of area is

$$\nabla_p \operatorname{area} M_h = \frac{1}{2} \sum_{T=(p,q,r)\in \operatorname{star} p} J(r-q), \tag{3–3}$$

where J is rotation of angle $\pi/2$ in the plane of each oriented triangle T. The first derivative of the surface area is then given by the chain rule

$$\frac{d}{dt} \operatorname{area} M_h = \sum_{p\in\mathcal{V}} \langle p', \nabla_p \operatorname{area} M_h \rangle. \tag{3–4}$$

The volume of an oriented surface M_h is the oriented volume enclosed by the cone of the surface over the origin in $\mathbb{R}^3$:

$$\operatorname{vol} M_h := \frac{1}{6} \sum_{T=(p,q,r)\in M_h} \langle p, q \times r \rangle = \frac{1}{3} \sum_{T=(p,q,r)\in M_h} \langle \vec{N}, p \rangle \cdot \operatorname{area} T,$$

where p is any of the three vertices of the triangle T and

$$\vec{N} = (q-p) \times (r-p)/|(q-p) \times (r-p)|$$

is the oriented normal of T. It follows that

$$\nabla_p \operatorname{vol} M_h = \sum_{T=(p,q,r)\in \operatorname{star} p} q \times r/6 \tag{3–5}$$

and

$$\frac{d}{dt} \operatorname{vol} M_h = \sum_{p\in\mathcal{P}} \langle p', \nabla_p \operatorname{vol} M_h \rangle. \tag{3–6}$$

REMARK 3.3. Note also that $\nabla_p \operatorname{vol} M_h = \sum_{T=(p,q,r)\in \operatorname{star} p} (2 \cdot \operatorname{area} T \cdot \vec{N} + p \times (r-q))/6$. Furthermore, if p is an interior vertex, then the boundary of $\operatorname{star} p$ is closed and $\sum_{T\in \operatorname{star} p} p \times (r-q) = 0$. Hence the $q \times r$ in equation (3–5) can be replaced with $2 \cdot \operatorname{area} T \cdot \vec{N}$ whenever p is an interior vertex.

3.3. Discrete mean curvature. The mean curvature vector on smooth surfaces provides a measure of how much the surface area changes compared to nearby surfaces, that is, if a surface is moved at constant speed along the surface normal. In the polyhedral case we will use a similar approach to obtain a discrete version of the mean curvature vector. Similar to the definition of a discrete Gauss curvature, the polyhedral mean curvature will measure the curvature of a small region. Later it will turn out that the mean curvature vector can be interpreted as the discrete Laplace–Beltrami operator on surfaces which was introduced in Section 2.

The area of a polyhedral surface is defined as the sum of the area of all elements. Let T be a triangle spanned by two edges v and w emanating from a vertex, then its area is given by the relation $4 \operatorname{area}^2 T = |v|^2 |w|^2 - \langle v, w \rangle^2$. In the following we

prefer an expression of the area in terms of vertices and vertex angles of the surface. Let T be a triangle with vertices q_i and vertex angles α_i. Then

$$\text{area}\, T = \frac{1}{4} \sum_{j=1}^{3} \cot \alpha_j \, |q_{j-1} - q_{j+1}|^2 . \tag{3–7}$$

For practical applications we derive a simple formula of the area gradient in intrinsic terms of the polyhedral mesh, see [**25**].

LEMMA 3.4. *Let p be an interior vertex of a simplicial surface M_h. Then the gradient of the area with respect to variation of vertices can be expressed in the cotangent formula*

$$\nabla_p \, \text{area}\, M_h = \frac{1}{2} \sum_j (\cot \alpha_j + \cot \beta_j)(p - q_j). \tag{3–8}$$

PROOF. The area gradient is the sum of the individual area gradients of all triangles containing p. In each triangle the area gradient of p is parallel to the height vector point toward p with length $|c|$. If c is the oriented edge opposite to p and J the rotation in the oriented plane of the triangle by $\pi/2$, then the gradient can be expressed by $\frac{1}{2} Jc$. Summing over all triangles containing p we obtain

$$\nabla_p \, \text{area}\, M_h = \frac{1}{2} \sum_j Jc_i .$$

Using the explicit representation of Jc on a single triangle with edges $c = a - b$ and vertex angles α and β at the end points of c,

$$Jc = a \cot \alpha + b \cot \beta ,$$

one obtains the proposed equation. □

This formula easily generalizes to nonmanifold surfaces where, for example, three triangles join at a common edge.

If $M_h(t)$ is a variation of simplicial surfaces such that each vertex $p(t)$ is a differentiable function for $t \in (-\varepsilon, \varepsilon)$, then

$$\frac{d}{dt} \, \text{area}\, M_h(t) = \sum_{p \in P} \langle p', \nabla_p \, \text{area}\, M_h \rangle .$$

The mean curvature of a smooth surface measures the variation of area when changing to parallel surfaces in normal direction. In the discrete case there exists no unique normal vector, but, as first derived in [**25**], if we choose as normal vector the direction of the area gradient, then the following definition leads to a discrete mean curvature vector which has similar properties as the smooth mean curvature vector.

DEFINITION 3.5. The discrete mean curvature at the vertex p of a simplicial surface M_h is a vector-valued quantity

$$\vec{H}(p) := \nabla_p \, \text{area}\, M_h . \tag{3–9}$$

Note that this mean curvature operator is an integrated operator and measures the total mean curvature in the vicinity of a vertex. Therefore, when computing the total mean curvature of a surface one simply needs to sum up the mean curvature of all vertices instead of integrating over the surface. In this sense, the mean curvature

is a measure at vertices similar to the (total) Gauss curvature discussed in [**1**]. This contrasts to the use of nontotal discrete mean curvatures in [**13**] in the experimental study of minimizers of the Willmore integral.

3.4. Properties of discrete minimal surfaces. In the previous section we introduced the notion of mean curvature vector as the gradient of the discrete area functional. Here we will study the critical values of the area functional in more detail, that is, surfaces with $H \equiv 0$.

DEFINITION 3.6. A simplicial surface M_h is a *discrete minimal surface* if and only if the discrete area functional of M_h is critical with respect to variations of any set of interior vertices. To include symmetry properties into this definition we sometimes allow a constrained variation of boundary points:

- if a boundary arc is a straight line, then its interior points may vary along the straight line;
- if a boundary arc is a planar curve, then its interior points may vary within the plane;
- in all other cases the boundary points always remain fixed.

Note that the above definition is equivalent to saying that the area of M_h is critical with respect to variations of any interior vertex. The relaxed boundary constraints allow us to simulate free boundary value problems, and to extend minimal surfaces by reflection.

COROLLARY 3.7. *A simplicial surface M_h is minimal if and only if at each interior vertex p,*

$$\nabla_p \operatorname{area} M_h = \frac{1}{2} \sum_j (\cot \alpha_j + \cot \beta_j)(p - q_j) = 0, \tag{3–10}$$

where $\{q_j\}$ denotes the set of vertices of $\operatorname{link} p$ *and α_j,β_j denote the two angles opposite to the edge pq_j. At boundary vertices on symmetry arcs the area gradient is constrained to be tangential to the straight line or to the plane.*

PROOF. This equation follows directly from the representation of the area gradient as the discrete mean curvature vector. □

The following properties of discrete minimal surfaces derived in [**25**] are similar to equivalent properties of harmonic maps.

LEMMA 3.8. *Let M_h be a discrete minimal surface. If the star of an interior vertex p consists of congruent isosceles triangles, then p lies in the center of mass of the vertices of its link.*

PROOF. The weights in equation (3–10) are all equal, therefore, p is the mean of its adjacent vertices $\{q_i\}$. □

The convex hull property for discrete minimal surfaces holds as long as the surface consists only of acute triangles.

LEMMA 3.9. *Let M_h be a discrete minimal surface. If the star of an interior vertex p consists of acute triangles then p lies in the convex hull of its star.*

PROOF. The weights in equation (3–10) are all positive; therefore, p is a convex combination of its adjacent vertices $\{q_i\}$:

$$p = \frac{\sum_j (\cot \alpha_j + \cot \beta_j) q_j}{\sum_j (\cot \alpha_j + \cot \beta_j)}$$

and lies within the convex hull of its link spanned by $\{q_i\}$. □

The previous lemma does not hold in a more general case. The following configuration is a counterexample to the maximum principle and the convex hull property of discrete minimal surfaces. Its construction in [**32**] jointly with Rossman is based on the existence of obtuse triangles. See also the model at [**30**] which contains an interactive applet to analyze the dependence on the boundary configuration.

The counterexample is a special configuration of the 1-parameter family of discrete minimal surfaces. The vertices are $q_0 = (-u, 0, -u)$, $q_1 = (u, 0, -u)$, $q_2 = (-1, 1, 0)$, $q_3 = (1, 1, 0)$, $q_4 = (-1, -1, 0)$, $q_5 = (1, -1, 0)$, $q_6 = (0, 0, h(u))$, while the faces are 062, 632, 613, 046, 456, 516. The parameter u varies in $(0, \infty)$ and the function $h(u)$ determines the vertical height of the center vertex. For $u \in [0, 2]$ the central vertex lies within the convex hull of the boundary after minimization. The remarkable fact is that this property does not hold for $u > 2$, when the minimum position of the central vertex is outside the convex hull of the boundary.

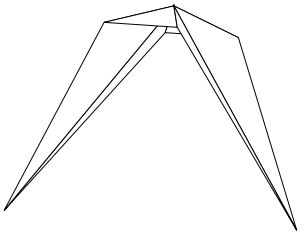


FIGURE 15. Counterexample to the maximum principle of discrete minimal surfaces. The drawing corresponds to the case $u = 5$. The center vertex lies outside the convex hull of its link.

Since the identity map of a discrete minimal surface is a discrete harmonic map, this example also demonstrates that the mean value property and convex hull property of discrete harmonic maps do not hold. Further note that both properties hold in special situations where all triangles have all vertex angles in $[-\pi/2, \pi/2]$. In this example, the center vertex lies on the convex hull exactly at $u = 2$ which is the situation when the first vertex angle becomes $\pi/2$. Increasing u further leads to an increasing angle.

Note that the discrete maximum principle does hold for the five-vertex Laplacian defined over the special rectangular $\mathbb{Z} \times \mathbb{Z}$ grid [**4**].

3.5. Computing discrete minimal surfaces. A direct minimization of the area functional is a nonlinear problem because of the angle terms in equation (3–10). Another effect, which may spoil numerical convergence, is the invariance of the area functional with respect to reparametrizations of the image surface. This may lead to tangential motions in an area minimization procedure.

The following observation leads to an effective method for area minimization which in fact minimizes the Dirichlet energy in an iteration process. This method

was first employed by Dziuk [**7**] for the mean curvature flow and later used in the context of discrete minimal surfaces by Pinkall and Polthier [**25**]. For a smooth map $F : M \to \mathbb{R}^3$ from a Riemann surface M we have the estimate

$$\operatorname{area} F(M) \leq \frac{1}{2} \int_M |\nabla F|^2 \, dx =: E_D(F)$$

with equality if and only if F is a conformal map. Following a proposal of Hutchinson [**15**] we call the difference

$$E_C(F) := E_D(F) - \operatorname{area} F(M)$$

the *conformal energy* of the map F, since for a Euclidean (x, y)-domain Ω one has

$$E_C(F) = \frac{1}{2} \int_\Omega |JF_x - F_y|^2 , \tag{3–11}$$

where J is the rotation by $\pi/2$ in the oriented tangent plane. E_C is a natural measure of failure for a map to be conformal. In the following we will introduce a discrete analogue of these relationships.

LEMMA 3.10. *The gradient of the Dirichlet energy of the identity map* id *of a simplicial surface M_h is equal to the area gradient, that is, at any interior vertex $p \in M$ we have*

$$\nabla_p \operatorname{area} M_h = \nabla_p E_D(\mathrm{id}).$$

PROOF. The statement follows directly by applying Theorem 2.4 to the id map and comparing its Dirichlet gradient with the area gradient of M_h. □

COROLLARY 3.11. *A simplicial surface M_h is minimal if and only if the identity map* $\mathrm{id}_h : M_h \to M_h$ *is discrete harmonic.*

As a consequence, we have a simplicial equivalent for the conformal energy of smooth maps given in equation (3–11).

DEFINITION 3.12. Let $F_h : M_h \to N_h$ be a map between two simplicial surfaces, then its discrete conformal energy is given by

$$E_C(F_h) := E_D(F_h) - \operatorname{area} F_h(M_h). \tag{3–12}$$

COROLLARY 3.13. *Let $F_h : M_h \to N_h$ be a map between two simplicial surfaces, then the discrete conformal energy and its gradient are*

$$E_C(F_h) = \frac{1}{4} \sum_{p_i p_j \text{ edges}} (\Delta\alpha_{ij} + \Delta\beta_{ij}) \, |F_h(p_i) - F_h(p_j)|^2 ,$$

$$\nabla_{F_h(p_i)} E_C(F_h) = \frac{1}{2} \sum_{p_j \in \operatorname{link} p_i} (\Delta\alpha_{ij} + \Delta\beta_{ij})(F_h(p_i) - F_h(p_j)) \tag{3–13}$$

with the shortcuts

$$\Delta\alpha_{ij} := \cot\alpha_{ij} - \cot\bar{\alpha}_{ij}, \qquad \Delta\beta_{ij} := \cot\beta_{ij} - \cot\bar{\beta}_{ij}$$

where α, β denote vertex angles on M_h and $\bar{\alpha}, \bar{\beta}$ denote vertex angles on N_h in triangles opposite to the edge $p_i p_j$.

PROOF. The relations follow immediately from the expressions of the discrete area in equation (3–7),

$$\nabla_{F_h(p_i)} \operatorname{area} F_h(M_h) = \frac{1}{2} \sum_{p_i p_j \text{ edges}} \left(\cot \bar{\alpha}_{ij} + \cot \bar{\beta}_{ij}\right) \left(F_h(p_i) - F_h(p_j)\right)$$

and the Dirichlet energy in Theorem 2.4,

$$\nabla_{F_h(p_i)} E_D F_h(M_h) = \frac{1}{2} \sum_{p_i p_j \text{ edges}} \left(\cot \alpha_{ij} + \cot \beta_{ij}\right) \left(F_h(p_i) - F_h(p_j)\right). \qquad \square$$

Note, a map has vanishing conformal energy if and only if angles of domain and image triangles are equal. But critical values of the conformal energy are much less constrained. For example, Hutchinson [**15**] noticed that minimizing the conformal energy leads to nice triangulations since it avoids decreasing the surface area which occurs when minimizing the Dirichlet energy.

The following algorithm uses a sequence of discrete harmonic maps. In short, let M_0 be an initial simplicial surface and let a sequence of simplicial surfaces $\{M_i\}$ be defined as images of a sequence of maps

$$F_i : M_i \to M_{i+1}, \qquad \Delta_h F_i = 0, \qquad \partial F_i(M_i) = \Gamma,$$

which are discrete harmonic on M_i. If the limit surface $M := \lim M_i$ exists, then the limit function $F : M \to M$ is harmonic and conformal, therefore, $F(M)$ is minimal.

The algorithm makes essential use of the fact that minimizing the Dirichlet energy also minimizes the surface area in first order. The major advantages of minimizing the Dirichlet energy compared to minimizing surface are, first, that the minimization process has a unique solution, and, second, that tangential motions can be ignored during the first iterations. Compare the comments of Brakke on this issue [**3**].

ALGORITHM 3.14. Solve the boundary value problem for discrete minimal surfaces (either Dirichlet or Neumann conditions):

(1) Choose an arbitrary initial surface M_0 with boundary $\partial M_0 = \Gamma$ as the first approximation of M, set i to 0.

(2) Let M_i be a surface with boundary Γ, then compute the surface M_{i+1} as minimizer of the Dirichlet energy

$$\int_{M_i} \left|\nabla(F_i : M_i \to M_{i+1})\right|^2 = \min_{M, \partial M = \Gamma} \int_{M_i} \left|\nabla(F : M_i \to M)\right|^2.$$

This uniquely defines a Laplace–Beltrami harmonic function F_i whose image $F_i(M_i) = M_{i+1}$ will be taken as the domain surface in the next iteration.

(3) Set i to $i+1$ and continue with step 2, for example, until

$$|\operatorname{area} M_i - \operatorname{area} M_{i+1}| < \epsilon.$$

In practice, this algorithm converges very quickly during the first iteration steps. It slows down if the surface is close to a critical point of the area functional, probably because then the area gradient no longer approximates a "good" surface normal. In any case, if the algorithm converges to a nondegenerated surface, then the limit is discrete minimal. The next convergence statement shown in [**25**] is

merely a theoretical observation, rather than having use in practical applications since the degeneracy assumption can hardly be ensured in advance.

PROPOSITION 3.15. *The algorithm converges to a solution of the problem, if no triangles degenerate.*

PROOF. The condition "no triangles degenerate" means that we assume all triangle angles for all surfaces of the sequence to be uniformly bounded away from 0 and π. From the construction the sequences $\{\text{area}\, M_i\}$ and$\{E_D(F_i : M_i \to M_{i+1})\}$ are monotone decreasing:

$$\begin{aligned} \text{area}\, M_i = E_D(\text{id}_{|M_i}) &\geq E_D(F_i : M_i \to M_{i+1}) \\ &= \text{area}\, M_{i+1} + E_C(F_i) \\ &\geq E_D(\text{id}_{|M_{i+1}}) = \text{area}\, M_{i+1}. \end{aligned}$$

If no triangles degenerate we minimize in a compact set of surfaces. Therefore, a subsequence of $\{M_i\}$ converges uniformly to a limit surface M with respect to the norm assumed in the space of surfaces.

Since the identity map of the limit surface M is discrete harmonic, the area gradient of M vanishes everywhere, and that means M is discrete minimal. □

Other Methods for Solving the Plateau Problem. The Plateau problem looks for a minimal surface M spanned by a given boundary curve $\Gamma \subset \mathbb{R}^3$. As an overview we mentioned three popular methods to compute a numerical solution.

Minimal graph: If the surface is known to be a graph over a plane, then there exists a scalar-valued function z over a planar domain $\Omega \subset \mathbb{R}^2$ with boundary $\partial\Omega$,

$$z : \Omega \to \mathbb{R}, \qquad z_{|\partial\Omega} = g_1 \text{ or } \partial_v z_{|\partial\Omega} = g_2,$$

where g_1 are prescribed Dirichlet boundary values, or g_2 are Neumann boundary conditions which prescribe the directional derivative of z in direction of the outer normal along $\partial\Omega$. Such a graph is area minimizing with respect to variations with compact support if it fulfills a nonlinear elliptic partial differential equation, the minimal surface equation [**6**]

$$(1 + z_y^2)z_{xx} - 2z_x z_y z_{xy} + (1 + z_x^2)z_{yy} = 0.$$

Mean curvature flow allows us to gradually decrease surface area. Let $M(t)$, with $\partial M(t) = \Gamma$ be a 1-parameter family of C^2 surfaces which is differentiable in t. Then $M(t)$ flows by mean curvature if it fulfills the following parabolic partial differential equation

$$\frac{\partial}{\partial t} M(t) = H(t) \cdot N(t) = \Delta_g M$$

where $H(t)$ is the mean curvature and $N(t)$ the surface normal of $M(t)$. If the flow does not run into a singularity and if it stops, then this limit surface is minimal.

3.6. Conjugate pairs of discrete minimal surfaces. Here we combine the results on nonconforming meshes of Section 2 and on simplicial minimal surfaces to derive the variational properties of pairs of conjugate discrete minimal surfaces.

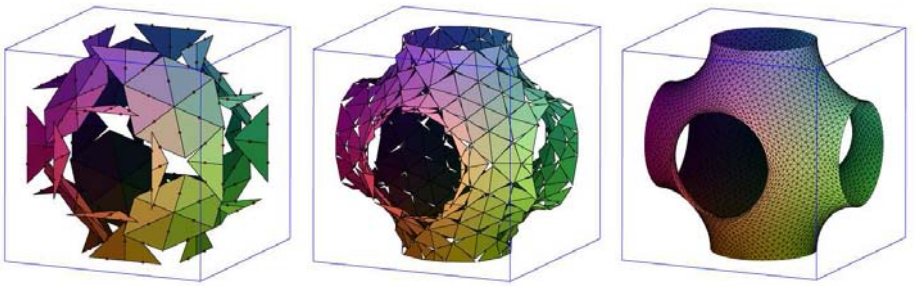


FIGURE 16. Free-boundary value problem of Schwarz P-surface in a cube solved via discrete conjugate surface construction. Even the very coarse resolution of the nonconforming mesh gives qualitatively good results.

Review of Smooth Minimal Surfaces. Among the fundamental observations in the theory of smooth minimal surfaces was the fact that each minimal surface comes in a family of minimal surfaces, the so-called *associate family* or *Bonnet family.* The simplest and most known example is the associate family which transforms the catenoid C into the helicoid H, both are given by

$$C(u,v) = \begin{pmatrix} \cos v \cosh u \\ \sin v \cosh u \\ u \end{pmatrix}, \qquad H(u,v) = \begin{pmatrix} \sin v \sinh u \\ -\cos v \sinh u \\ v \end{pmatrix}.$$

Their associate family $F^\varphi(u,v)$ consists of all minimal surfaces given by

$$F^\varphi(u,v) = \cos\varphi \cdot C(u,v) + \sin\varphi \cdot H(u,v).$$

The surface $F^{\frac{\pi}{2}}$ is called the *conjugate surface* of F^0, and more generally, all pairs F^φ and $F^{\varphi+\frac{\pi}{2}}$ are conjugate to each other. Applying the conjugate twice leads to $F^\pi = -F$ which is obtained from F^0 by reflection in the origin.

A more appropriate notation of the associate family follows from the representation of minimal surfaces as complex curves in $\mathbb{C}^3$. Recall the basic fact in minimal surface theory that the three coordinate functions $F = (f_1, f_2, f_3)$ of a minimal surface $F : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^3$ are harmonic maps if F is a conformal parameterization. Therefore, there exist three conjugate harmonic maps f_i^* which describe another minimal immersion $F^* = (f_1^*, f_2^*, f_3^*) : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^3$. If we introduce complex coordinates $z = u + iv$ in Ω, then combination of both maps to a holomorphic curve $F + iF^* : \Omega \to \mathbb{C}^3$ with holomorphic coordinate functions gives a family of immersions $F^\varphi = \mathrm{Re}(e^{-i\varphi} \cdot (F + iF^*)$ called the associate family of F or of F^*. In the above example the introduction of complex coordinates leads to the representation of the associate family of catenoid and helicoid given by

$$F^\varphi(z) = \mathrm{Re}(e^{-i\varphi} \cdot (C(z) + i \cdot H(z))) = \mathrm{Re}\left(e^{-i\varphi} \cdot \begin{pmatrix} \cosh z \\ -i \sinh z \\ z \end{pmatrix}\right).$$

The following theorem summarizes the most important properties of the associate family of smooth minimal surfaces without proof.

THEOREM 3.16. *Let $C, H : \Omega \to \mathbb{R}^3$ be a pair of conformally parametrized conjugate minimal surfaces. Then the associate family $F^\varphi : \Omega \to \mathbb{R}^3$ has the following properties:*

(1) *All surfaces F^φ of the associate family are minimal and isometric.*
(2) *The surface normal at each point $F^\varphi(u, v)$ is independent of φ.*
(3) *The partial derivatives fulfill the following correspondence:*

$$F^\varphi_u(u, v) = \cos\varphi \cdot C_u(u, v) - \sin\varphi \cdot C_v(u, v),$$
$$F^\varphi_v(u, v) = \sin\varphi \cdot C_u(u, v) + \cos\varphi \cdot C_v(u, v);$$

in particular, the partials of a conjugate pair C and H satisfy the Cauchy–Riemann equations:

$$C_u(u, v) = H_v(u, v), \qquad C_v(u, v) = -H_u(u, v).$$

This relation can be written in a compact form $dH = *dC$ using the Hodge $*$ operator.

(4) *If a minimal patch is bounded by a straight line, then its conjugate patch is bounded by a planar symmetry line and vice versa. This can be seen in the catenoid-helicoid examples, where planar meridians of the catenoid correspond to the straight lines of the helicoid.*
(5) *Since at every point the length and the angle between the partial derivatives are identical for the surface and its conjugate (i.e., both surfaces are isometric) we have as a result, that the angles at corresponding boundary vertices of surface and conjugate surface are identical.*

The last two properties are most important for the later conjugate surface method.

Review of the Conjugate Surface Construction. Over the last decade the conjugate surface method has been established as one of the most powerful techniques to construct new minimal surfaces with a proposed shape in mind. One of the major drawbacks of the method is the so-called period problem which often prevents a rigorous existence proof of the examples. In those situations where theoretical techniques fail up to now, a numerical approach is required to allow experiments.

The major obstacle for a numerical simulation of the conjugate surface method is the fact that the minimal surfaces are usually unstable. Currently, the conjugation method based on discrete minimal surfaces is the only numerical method to compute the conjugate of a polyhedral minimal surface with satisfactory results.

3.6.1. *Discrete conjugate minimal surface.* In this section we develop the notion of the conjugate and the associate family of a discrete minimal surface. In [**25**] the discrete conjugation algorithm is based on the concept of discrete harmonic maps, but the method did not unveil the variational properties of the conjugate surface. In the following we first show the area minimality of the conjugate discrete minimal surface, and second, describe a practical algorithm by reformulating the conjugation method of [**25**] in terms of the conjugation of harmonic maps using conforming and nonconforming functions derived in Section 2.

Currently, the method [**25**] seems to be the only method to allow the conjugation of a numerically computed discrete minimal surface with reasonable results. The main difficulties are to provide accurate C^1 information, which is required for the conjugation, from numerically obtained minimal surfaces.

The remaining part of this section shows that the conjugate minimal surface is well-defined, and derives some important properties. Most results follow from properties of the conjugate harmonic coordinate functions.

Let us review some properties of the differential of a polyhedral map $F : M_h \to \mathbb{R}^d$ where either $F \in S_h$ or $F \in S_h^*$. At each point $p \in M_h$ the differential $\nabla_p F : T_p M_h \to T_{F(p)} F(M_h)$ is given by

$$\nabla_p F(v) = \begin{pmatrix} \langle \nabla_p f_1, v \rangle \\ \dots \\ \langle \nabla_p f_d, v \rangle \end{pmatrix} \quad \forall\, v \in T_p M_h$$

if $F = (f_1, \dots, f_d)$ are the coordinate functions. A map F is said to be harmonic if all coordinate functions are harmonic with respect to the metric of M_h. Recalling the definition of the Hodge $*$ operator directly leads to the following definition by applying the operator on the component functions. We say that a simplicial surface M_h is in S_h, respectively S_h^*, if the triangulation is edge continuous, respectively edge-midpoint continuous.

Definition 3.17. Let $F = (f_1, \dots, f_d) : M_h \to \mathbb{R}^d$ be a simplicial map in S_h or S_h^*. The Hodge star operator is defined by

$$*dF_{|p}(v) := \begin{pmatrix} *df_{1|p}(v) \\ \dots \\ *df_{d|p}(v) \end{pmatrix} = \begin{pmatrix} \langle J \nabla_p f_1, v \rangle \\ \dots \\ \langle J \nabla_p f_d, v \rangle \end{pmatrix} \quad \forall\, v \in T_p M_h$$

where J is the rotation by $\pi/2$ in the oriented tangent space of each triangle of M_h with respect to the metric in M_h.

For example, if $F = \mathrm{id} : M_h \to M_h$ is the identity map of a simplicial surface, then we obtain on each triangle

$$*d\,\mathrm{id}_{|p}(v) := -Jv \;\forall\, v \in T_p M_h. \tag{3–14}$$

Now we are ready to extend the results on discrete harmonic maps of the previous section to the conjugation of simplicial minimal surfaces. In the following theorem we show that the differential $*d\,\mathrm{id}$ is closed on simplicial minimal surfaces, and that its integral gives the conjugate minimal surface:

Definition 3.18. Let M_h be a simplicial minimal surface in S_h (or in S_h^*). Then a *discrete conjugate minimal surface* M_h^* is a solution of equation (3–14).

The following theorem justifies this definition and states the general relation between conjugate pairs of discrete minimal surfaces.

Theorem 3.19.
(1) *Let $M_h \subset \mathbb{R}^d$ be a discrete minimal surface in S_h. Then there exists a conjugate surface $M_h^* \subset \mathbb{R}^d$ in S_h^* which is critical for the area functional in S_h^*.*
(2) *Let $M_h \subset \mathbb{R}^d$ be a discrete minimal surface in S_h^*. Then there exists a conjugate surface $M_h^* \subset \mathbb{R}^d$ in S_h which is critical for the area functional in S_h.*
(3) *M_h^* is uniquely determined by M_h up to translation.*
(4) *M_h and M_h^* are isometric and have the same Gauss map in the sense that corresponding triangles are congruent and parallel.*

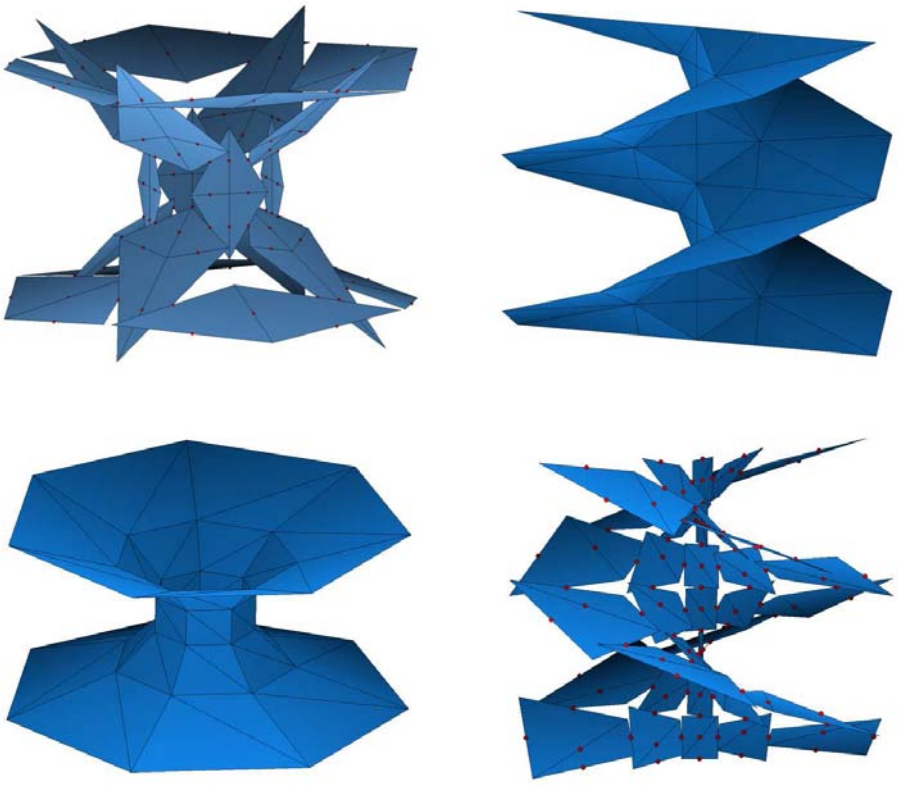


FIGURE 17. Conforming and nonconforming catenoids and helicoids (horizontal pairs are discrete conjugate).

(5) *Applying conjugation twice leads to*

$$M_h^{**} = -M_h$$

for a suitably chosen origin.

PROOF. Since M_h is critical for the area functional, the identity map

$$\mathrm{id} : M_h \to M_h$$

is a discrete harmonic map by Corollary 3.11. Therefore, Theorem 2.21 in Section 2 proves that there exist conjugate harmonic component functions that give rise to a map on M_h,

$$\mathrm{id}^* : M_h \to \mathbb{R}^d$$

with $M_h^* := \mathrm{id}^* M_h$.

It remains to show that M_h^* is a discrete minimal surface. Here we assume that M_h is in S_h—the case M_h in S_h^* would work with the same words.

We show that M_h^* fulfills the balancing condition. Let $p^* \in M_h^*$ be an interior vertex, then by harmonicity of $\mathrm{id}^* \in S_h^*$ we have

$$\begin{aligned} \frac{d}{dm^*} E_D(\mathrm{id}^*) &= 2(\cot\alpha_{-2}(m^* - m_{-1}^*) + \cot\alpha_{-1}(m^* - m_{-2}^*) \\ &\quad + \cot\alpha_1(m^* - m_2^*) + \cot\alpha_2(m^* - m_1^*)) \\ &= 0 \end{aligned} \tag{3–15}$$

where m^* and m_i^* are the images of id^* of edge midpoints in M_h.

Since on each triangle id^* is a rotation by $\pi/2$, corresponding triangles of M_h and M_h^* are isometric and have the same angles. Therefore, equation (3–15) also is the criticality condition of the Dirichlet energy of the identity map of M_h^* which lies in S_h^*. Thus M_h^* is a discrete minimal surface in S_h^*.

The uniqueness follows from the uniqueness of the conjugate harmonic map and its integration constants. □

Summarizing, the theorem shows that a conjugate pair of discrete minimal surfaces does not exist in the space of piecewise linear conforming elements S_h but naturally leads to the space of piecewise linear nonconforming Crouzeix–Raviart elements S_h^*. S_h alone is too rigid to contain the conjugate of a minimal surface too.

In other words, if M_h is a simplicial minimal surface in S_h or in S_h^*, its discrete conjugate minimal surface M_h^* is the image of the conjugate harmonic $\mathrm{id}^* : M_h \to \mathbb{R}^d$ map of the identity map of $\mathrm{id} : M_h \to M_h$, that is, id and id^* satisfy

$$d\,\mathrm{id}^* = *d\,\mathrm{id}\,.$$

The usage of the same domain M_h for both identity maps seems to distinguish M_h from M_h^* but only the conformal structure of the domain surface is relevant for the minimality condition. Therefore, we may instead use M_h^* or, more appropriate, use $\mathrm{id} : M_h \to M_h$ and $\mathrm{id}^* : M_h^* \to M_h^*$.

3.6.2. *Numerical conjugation.* In practical applications the conjugation of a simplicial minimal surface by rotating each triangle and reassembling the rotated copies requires that the simplicial minimal surface has been computed very exactly. Often, minimal surfaces are computed by solving a variational problem where the numerical method stops before reaching the absolute zero of the gradient. A much more stable procedure has been suggested in [**25**] to circumvent this difficulty: in a minimization procedure based on the Dirichlet energy there exists an accurately computed harmonic map F_i between the last two computed surfaces M_i and M_{i-1}. Instead of applying the conjugation to the approximation M_i of the limit minimal surface, it is more stable to compute the harmonic conjugate map

$$F_i^* : M_{i-1} \to M_i^*.$$

The following algorithm summarizes the procedure:

ALGORITHM 3.20. To compute the conjugate M_h^* of the Plateau problem M_h with Dirichlet boundary condition Γ:

(1) Follow the minimization algorithm above to compute a sequence of discrete harmonic maps $F_i : M_i \to M_{i+1}$.
(2) Compute the harmonic conjugate F_i^* of $F_i : M_i \to M_{i+1}$.
(3) Set $M_h := M_{i+1}$ as numerical approximation of the Plateau solution, and set $M_h^* := F_i^*(M_i)$ as approximation of the conjugate minimal surface.

This algorithm generates a sequence of discrete surfaces $\{M_i\}$ and vector-valued harmonic maps $\{F_i : M_i \to M_{i+1}\}$ which converges to a minimal surface if no degeneration occurs. In order to extend the conjugation technique of the previous sections to the computation of the conjugate of a minimal surface, we allow the surfaces M_i to be either all conforming or all nonconforming triangulations. In this case the coordinate functions of each F_i are discrete harmonic functions either in S_h or S_h^*, and the image $F_i^*(M_i)$ of the conjugate harmonic of F_i is a good approximation of the conjugate minimal surface. The two approximations M_h and M_h^* are either a conforming and a nonconforming triangulation, or vice-versa.

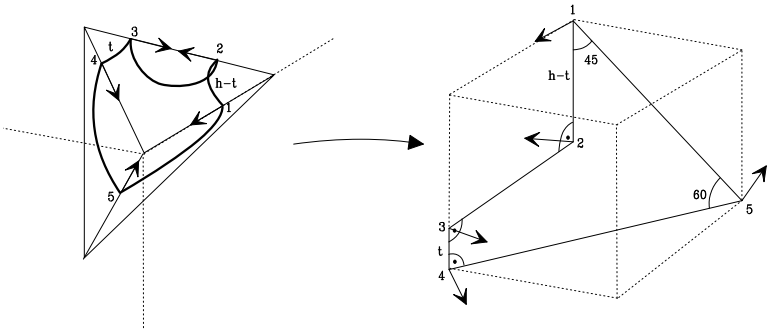


FIGURE 18. Transformation of a free boundary value problem into a family of Dirichlet boundary value problems with a fixed contour.

REMARK 3.21. The output of the numerical conjugation Algorithm 3.20 is a conforming mesh if one uses a nonconforming triangulation to solve the Dirichlet boundary conditions. The resulting stability problems when minimizing the nonconforming mesh are solved in [**28**].

3.7. Discrete minimal catenoid. Examples are important building blocks in the development of a mathematical theory. The first smooth minimal surfaces were found already in the 18th century when Lagrange formulated the variational characterization of minimal surfaces. The French geometer and engineer Jean Baptiste Meusnier (1754–1793) recognized the first nontrivial examples of minimal surfaces: the catenoid found by Euler in 1744, also called the chain surface, because it is the surface swept out when one rotates the catenary curve that corresponds to a freely hanging chain about a suitable horizontal line, and the helicoid, or screw surface. Already the discovery of the next examples in 1835 was regarded as so sensational that its discoverer Heinrich Ferdinand Scherk (1798–1885), Professor at Kiel and Bremen, won a prize at the Jablonowski Society at Leipzig in 1831.

The discovery of this discrete minimal catenoid by Polthier and Rossman [**32**] was driven by a very practical need, namely the provision of an unstable discrete minimal surface for investigations on the index of minimal surfaces. The numerical eigenvalue computations require a very accurate unstable surface as input, which is hardly produced by means of minimization methods. Here the explicit formulae allows us to create unstable catenoids of arbitrary resolution. The model [**29**] at the EG-Models journal includes an interactive applet to study the whole family of discrete catenoids.

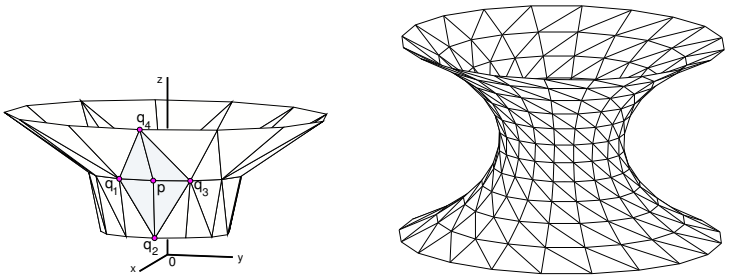


FIGURE 19. Discrete catenoid with essential stencil.

The strategy for the construction of an explicit formula for embedded complete discrete minimal catenoids is to assume that the vertices lie on congruent planar polygonal meridians and that the meridians are placed so that the traces of the surfaces will have dihedral symmetry. Under these assumptions we find that the vertices of a discrete meridian lie equally spaced on a smooth hyperbolic cosine curve. Furthermore, these discrete catenoids will converge uniformly in compact regions to the smooth catenoid as the mesh is made finer.

We now state an explicit formula for discrete minimal catenoids, by specifying the vertices along a planar polygonal meridian. Then the traces of the surfaces will have dihedral symmetry of order $k \geq 3$. The surfaces are tessellated by planar isosceles trapezoids like a $\mathbb{Z}^2$ grid, and each trapezoid can be triangulated into two triangles by choosing a diagonal of the trapezoid as the interior edge. Either diagonal can be chosen, as this does not affect the minimality of the catenoid.

The discrete catenoid has two surprising features. First, the vertices of a meridian lie on a scaled smooth cosh curve (just as the profile curve of smooth catenoids lies on the cosh curve), and there is no apriori reason to have expected this. Secondly, the vertical spacing of the vertices along the meridians is constant.

THEOREM 3.22 (joint with W. Rossman). *There exists a four-parameter family of embedded and complete discrete minimal catenoids $C = C(\theta, \delta, r, z_0)$ with dihedral rotational symmetry and planar meridians. If we assume that the dihedral symmetry axis is the z-axis and that a meridian lies in the xz-plane, then, up to vertical translation, the catenoid is completely described by the following properties:*

(1) *The dihedral angle is $\theta = \frac{2\pi}{k}$, $k \in \mathbb{N}$, $k \geq 3$.*
(2) *The vertices of the meridian in the xz-plane interpolate the smooth* cosh *curve*

$$x(z) = r \cosh\left(\frac{1}{r} a z\right),$$

with

$$a = \frac{r}{\delta} \operatorname{arccosh}\left(1 + \frac{1}{r^2} \frac{\delta^2}{1 + \cos\theta}\right),$$

where the parameter $r > 0$ is the waist radius of the interpolated cosh *curve, and $\delta > 0$ is the constant vertical distance between adjacent vertices of the meridian.*
(3) *For any given arbitrary initial value $z_0 \in \mathbb{R}$, the profile curve has vertices of the form $(x_j, 0, z_j)$ with*

$$z_j = z_0 + j\delta, \qquad x_j = x(z_j),$$

where $x(z)$ is the meridian in item 2 above.

(4) *The planar trapezoids of the catenoid may be triangulated independently of each other.*

COROLLARY 3.23. *There exists a two-parameter family of discrete catenoids $C_1(\theta, z_0)$ whose vertices interpolate the smooth minimal catenoid with meridian $x = \cosh z$.*

PROOF. The waist radius of the scaled cosh curve must be $r = 1$. Further, we must choose the parameter $a = 1$ which is fulfilled if θ and δ are related by $1 + \cos\theta + \delta^2 = (1 + \cos\theta)\cosh\delta$. The offset parameter z_0 may be chosen arbitrarily leading to a vertical shift of the vertices along the smooth catenoid. Note that if $z_0 = 0$, we obtain a discrete catenoid that is symmetric with respect to a horizontal reflection. □

COROLLARY 3.24. *For each fixed r and z_0, the profile curves of the discrete catenoids $C(\theta, \delta, r, z_0)$ approach the profile curve $x = r\cosh\frac{z}{r}$ of a smooth catenoid uniformly in compact sets of $\mathbb{R}^3$ as $\delta, \theta \to 0$.*

PROOF. This is a direct consequence of the explicit representation of the meridian. Since

$$\lim_{\delta\to 0} \frac{1}{\delta} \operatorname{arccosh}\left(1 + \frac{1}{r^2}\frac{\delta^2}{1+\cos\theta}\right) = \frac{\sqrt{2}}{r\sqrt{1+\cos\theta}},$$

it follows that the profile curve of the discrete catenoid converges uniformly to the curve

$$x = r\cosh\frac{\sqrt{2}z}{r\sqrt{1+\cos\theta}}$$

as $\delta \to 0$. Then, as $\theta \to 0$ we approach the profile curve $x = r\cosh\frac{z}{r}$. □

3.8. Discrete minimal helicoid. We continue with the derivation of explicit discrete helicoids which are a natural second example of a complete, embedded discrete minimal surface. The full construction of the surface is given in [**32**]. An interactive data set of the model is available at the EG-Models site at [**31**].

In the smooth setting, there exists an isometric deformation through conjugate surfaces from the catenoid to the helicoid (see, for example, [**24**]). So, one might first try to make a similar deformation from the discrete catenoids in Theorem 3.22 to discrete minimal helicoids. But such a deformation is impossible in the space of edge-continuous triangulations. In fact, in order to make an associate family of discrete minimal surfaces, one must allow noncontinuous triangle nets having greater flexibility.

Therefore, we adopt a different approach for finding discrete minimal helicoids. The helicoids will be comprised of planar quadrilaterals, each triangulated by four coplanar triangles; see Figures 19 and 17. Each quadrilateral is the star of a unique vertex, and none of its four boundary edges are vertical or horizontal, and one pair of opposite vertices in its boundary have the same z-coordinate, and the four boundary edges consist of two pairs of adjacent edges so that within each pair the adjacent edges are of equal length.

THEOREM 3.25 (joint with W. Rossman). *There exists a family of complete embedded discrete minimal helicoids, with the connectivity as shown in Figure 19.*

The vertices, indexed by $i, j \in \mathbb{Z}$, are the points

$$\frac{r \sinh(x_0 + j\delta)}{\sin\theta}(\cos(i\theta), \sin(i\theta), 0) + (0, 0, ir) \ ,$$

for any given real numbers $\theta \in (0, \frac{\pi}{2})$ and $r, \delta \in \mathbb{R}$.

Note that these surfaces are invariant under the screw motion that combines vertical upward translation of distance $2r$ with rotation about the x_3-axis by an angle of 2θ. The term x_0 determines the offset of the vertices from the z-axis (if $x_0 = 0$, then the z-axis is included in the edge set), and δ determines the horizontal spacing of the vertices. The homothety factor is r, which equals the vertical distance between consecutive horizontal lines of edges.

References

[1] A. D. Aleksandrov and V. A. Zalgaller. *Intrinsic Geometry of Surfaces*, volume 15 of *Translations of Mathematical Monographs*. AMS, 1967.

[2] E. D. Bloch. *A First Course in Geometric Topology and Differential Geometry*. Birkhäuser, Boston, 1997.

[3] K. Brakke. *Surface Evolver Manual v2.14*, 1999. Available at http://www.susqu.edu/facstaff/b/brakke/evolver.

[4] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer Verlag, 1994.

[5] P. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, 1978.

[6] U. Dierkes, S. Hildebrandt, A. Küster, and O. Wohlrab. *Minimal Surfaces*, volume 1 of *Grundlehren der Mathematik*. Springer Verlag, 1992.

[7] G. Dziuk. An algorithm for evolutionary surfaces. *Numer. Math.*, 58:603–611, 1991.

[8] G. Dziuk and J. E. Hutchinson. Finite element approximations and the Dirichlet problem for surfaces of prescribed mean curvature. In H.-C. Hege and K. Polthier, editors, *Mathematical Visualization*, pages 73–87. Springer, Heidelberg, 1998.

[9] M. S. Floater. Parametrization and smooth approximation of surface triangulations. *Computer Aided Geometric Design*, 14:231–250, 1997.

[10] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser, 1999.

[11] I. Guskov, K. Vidimce, W. Sweldens, and P. Schröder. Normal meshes. In *Computer Graphics Proceedings (Siggraph '00)*, pages 95–102, 2000.

[12] H. Hoppe. Progressive meshes. *Computer Graphics (SIGGRAPH '96 Proceedings)*, pages 99–108, August 1996.

[13] L. Hsu, R. Kusner, and J. Sullivan. Minimizing the squared mean curvature integral for surfaces in space forms. *Experim. Math.*, 1(3):191–207, 1992.

[14] K. Hurdal, P. L. Bowers, K. Stephenson, D. W. L. Sumners, K. Rehm, K. Schaper, and D. A. Rottenberg. Quasi-conformally flat mapping the human cerebellum. In C. Taylor and A. Colchester, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 279–286. Springer Verlag, 1999.

[15] J. E. Hutchinson. Computing conformal maps and minimal surfaces. *Proc. Centr. Math. Anal., Canberra*, 26:140–161, 1991.

[16] H. Lawson Jr. *Lectures on Minimal Submanifolds*. Publish or Perish Press, 1971.

[17] H. Karcher. Construction of minimal surfaces. *Surveys in Geometry, University of Tokyo*, pages 1–96, 1989.

[18] H. Karcher and K. Polthier. Construction of triply periodic minimal surfaces. In J. Klinowski and A. L. Mackay, editors, *Curved Surfaces in Chemical Structures*, volume 354 (1715) of *A*, pages 2077–2104. The Royal Society, London, Great Britain, phil. trans. r. soc. lond. edition, September 1996.

[19] Z. Karni and C. Gotsman. Spectral compression of mesh geometry. In *Computer Graphics Proceedings (Siggraph '00)*, pages 279–286. ACM SIGGRAPH, 2000.

[20] E. E. Moise. *Geometric Topology in Dimensions 2 and 3*. Springer Verlag, 1977.

[21] S. Müller, M. Struwe, and V. Sverak. Harmonic maps on planar lattices. Technical Report 17, MPI Leipzig, 1997.
[22] J. R. Munkres. *Elements of Algebraic Topology*. Addison Wesley, 1984.
[23] B. Oberknapp and K. Polthier. An algorithm for discrete constant mean curvature surfaces. In H.-C. Hege and K. Polthier, editors, *Visualization and Mathematics*, pages 141–161. Springer Verlag, Heidelberg, 1997.
[24] R. Osserman. *A Survey of Minimal Surfaces*. Dover, 1986.
[25] U. Pinkall and K. Polthier. Computing discrete minimal surfaces and their conjugates. *Experim. Math.*, 2(1):15–36, 1993.
[26] E. Polak. *Computational Methods in Optimization*. Academic Press, 1971.
[27] K. Polthier. Polyhedral surfaces of constant mean curvature. Habilitationsschrift, Technische Universität Berlin, 2002.
[28] K. Polthier. Unstable periodic discrete minimal surfaces. In S. Hildebrandt and H. Karcher, editors, *Nonlinear Partial Differential Equations*, pages 127–143. Springer Verlag, 2002.
[29] K. Polthier and W. Rossman. Discrete catenoid. *Electronic Geometry Models*, 2000. Available at http://www.eg-models.de/2000.05.002/.
[30] K. Polthier and W. Rossman. Counterexample to the maximum principle for discrete minimal surfaces. *Electronic Geometry Models*, 2001. http://www.eg-models.de/2000.11.040/.
[31] K. Polthier and W. Rossman. Discrete minimal helicoid. *Electronic Geometry Models*, 2001. http://www.eg-models.de/2001.01.046/.
[32] K. Polthier and W. Rossman. Discrete constant mean curvature surfaces and their index. *J. reine angew. Math*, 549(47–77), 2002.
[33] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipies in C: The Art of Scientific Computing*. Cambridge University Press, 1993. `http://www.nr.com/`.
[34] Y. G. Reshetnyak. *Geometry IV*, volume 70 of *Encyclopaedia of Mathematical Sciences*, chapter 1. Two-Dimensional Manifolds of Bounded Curvature, pages 3–164. Springer Verlag, 1993.
[35] , Hermann A. Schwarz. Sur une définition erronée de l'aire d'une surface courbe, 1890. Reprinted in *Gesammelte Mathematische Abhandlungen*, vol. 2, 309–311, Springer, Berlin.
[36] G. Taubin and J. Rossignac. Geometric compression through topological surgery. *ACM Transactions on Graphics*, 17(2):84–115, 1998.
[37] T. Tsuchiya. Discrete solutions of the Plateau problem and its convergence. *Math. of Comp.*, 49:157–165, 1987.
[38] G. M. Ziegler. *Lectures on Polytopes*. Springer Verlag, 1998. 2nd ed.

Technische Universität Berlin, FB 3 Mathematik, MA 8-3, Str. des 17. Juni 136, 10623 Berlin
E-mail address: `polthier@math.tu-berlin.de`

Clay Mathematics Proceedings
Volume 2, 2004

Mean Curvature in Riemannian Geometry and General Relativity

Richard Schoen

1. Introduction

In this series of three lectures we describe some connections between the theory of minimal submanifolds and mean curvature with Riemannian Geometry and General Relativity. Our goal will be to put the known results in context, and to outline problems which remain open. We begin with a discussion of the first and second variation formulae for volume. We place particular emphasis on the notion of stability for minimal submanifolds because it is primarily through stability and the Jacobi operator that the ambient curvature influences the behavior of minimal submanifolds. By definition a minimal submanifold is stable if the second variation of volume is nonnegative for all compactly supported deformations. We summarize the known analytic estimates and compactness theorems for stable hypersurfaces, and then proceed to discuss connections with scalar curvature and general relativity.

We describe the constraint equations and asymptotic flatness for initial data sets for Einstein's equations. We then discuss the ADM mass and state the positive mass theorem. We discuss the Schwarzschild spacetime, and formulate the Penrose inequality. We then describe the Huisken–Ilmanen and Bray work on the Riemannian Penrose inequality. We emphasize throughout the close connection of these results with the mean curvature theory; indeed they are really sharp quantitative statements about the area of minimal surfaces in three-dimensional manifolds with nonnegative scalar curvature.

In the final part of this paper we discuss issues connected with area minimizing and stable submanifolds in higher codimensions. We describe the theory of calibrations, and place particular emphasis on the holomorphic and special lagrangian calibrations. We introduce the lagrangian Plateau problem and outline its connection with constructions of special lagrangian and minimal lagrangian submanifolds of Kähler–Einstein manifolds.

We thank Weiyang Qiu who took notes during the lectures and prepared a draft of this paper.

This work was partially supported by NSF grant DMS-0104163.

©2004 Clay Mathematics Institute

2. Background and Notation

Let (M^{n+1}, g) be a Riemannian manifold, and let D be an affine connection. D is called the Levi-Civita connection of g if

(1) $D_X g = 0$, for any vector field X;
(2) D satisfies the torsion free condition: $D_X Y - D_Y X = [X, Y]$, for any vector fields X and Y.

It is a standard result that given a metric g, there exists a unique Levi-Civita connection.

Using the Riemannian metric g, we can define the length of a curve and the Riemannian distance function for a pair of points x and y,

$$d(x, y) := \inf\{L(\gamma) : \gamma \text{ is a rectifiable curve connecting } x \text{ and } y\}.$$

Geodesics are defined to be the critical curves for length functional. Locally they are length minimizing, but in the large they tend to be unstable critical points in many cases.

The Riemann curvature tensor may be considered a $(0,4)$ tensor $R_{ijkl} = \langle R(e_i, e_j)e_k, e_l\rangle$ where R is the curvature tensor of the Levi-Civita connection D and $\{e_1, \ldots, e_{n+1}\}$ is a basis for $T_x M$. We define the Ricci tensor to be the trace of the curvature tensor, $R_{ij} = \sum_{k,l} g^{kl} R_{ikjl}$, and the scalar curvature R to be the trace of Ricci tensor $R = \sum_{i,j} g^{ij} R_{ij}$.

Now we study hypersurfaces in M^{n+1}. Let Σ^n be a hypersurface in M. We define the induced Riemannian connection ∇ on Σ by

$$\nabla_X Y := (D_X Y)^T,$$

where $(D_X Y)^T$ is the projection of $D_X Y$ onto the tangent space of Σ. It is easy to check that ∇ is an affine connection and is indeed the Levi-Civita connection for the induced metric on Σ.

Let ν be a chosen unit normal vector field for Σ. We define the scalar-valued *second fundamental form* h as

$$h(X, Y) := \langle D_X Y, \nu\rangle,$$

for tangent vector fields X and Y on Σ. We can show that h is in fact a symmetric $(0,2)$ tensor.

The *mean curvature* H is defined to be the trace of second fundamental form, that is,

$$H := \sum_{i,j} g^{ij} h_{ij},$$

where $h_{ij} = h(e_i, e_j)$.

DEFINITION 2.1. *A hypersurface is called minimal if the mean curvature is identically zero, i.e., $H = 0$.*

The following fundamental equations relate the intrinsic geometry (defined by g) and the extrinsic geometry (defined by h) for a hypersurface. Let $\Sigma^n \subset M^{n+1}$ be a hypersurface and $e_1, \ldots, e_n$ be a orthonormal basis on Σ.

PROPOSITION 2.2 (Gauss Equation). *The Riemann curvature tensor of the induced metric on Σ is given by*

$$R^{\Sigma}_{ijkl} = R^{M}_{ijkl} + h_{ik} h_{jl} - h_{il} h_{jk},$$

for $1 \leq i,j,k,l \leq n$.

PROPOSITION 2.3 (Codazzi Equation). *The compatibility conditions*

$$\nabla_{e_i} h_{jk} - \nabla_{e_j} h_{ik} = R^M_{ijk(n+1)},$$

hold for $1 \leq i,j,k,l \leq n$.

The basic philosophy of this paper is to study the geometric properties of the ambient manifold M using minimal hypersurfaces. In particular, it will be important to understand how curvature properties of M influence the behavior of the minimal submanifolds in M.

Notice that the theory of minimal surfaces in R^{n+1} may be viewed as the "local theory". This is because the complete minimal submanifolds in R^{n+1} arise by rescaling a potential "blow-up" sequence in a manifold M^{n+1}, and then the global behavior of this submanifold reflects local properties of the original submanifolds.

A final remark here is that in applications, the minimal submanifolds we study will typically arise from a minimizing procedure, and hence will have a minimizing or stability property. We proceed to discuss the variational theory in more detail.

3. First and Second Variation

In this section we record the first and second variation formulae. Let $\Sigma^k \subset M^n$ be a submanifold. Let X be a vector field in M, and F_t be the flow generated by X. A standard computation (see [**L**]) gives the first and second variation formulae:

LEMMA 3.1 (First Variation).

$$\delta\Sigma(X) := \frac{d}{dt} Vol(F_t(\Sigma))_{t=0} = \int_\Sigma \mathrm{div}_\Sigma(X)\, d\mu_\Sigma$$

where $\mathrm{div}_\Sigma(X) := \sum_{i,j} g^{ij} \langle D_{e_i} X, e_j \rangle = -\langle X, H \rangle + \mathrm{div}(X^T)$. *Moreover if* $X = 0$ *on* $\partial\Sigma$, *then*

$$\delta\Sigma(X) = -\int_\Sigma \langle X, H \rangle\, d\mu_\Sigma.$$

A submanifold with vanishing first variation for all compactly supported vector fields X is said to be *stationary*, and we see from the first variation formula that a smooth submanifold Σ is stationary if and only if it is minimal.

LEMMA 3.2 (Second Variation).

$$\frac{d^2}{dt^2} Vol(F_t(\Sigma))|_{t=0} = \int_\Sigma \Big(\sum_{i=1}^k |D^\perp_{e_i} X|^2 + \mathrm{div}_\Sigma(D_X X) + \sum_{i=1}^k R^M(e_i, X, e_i, X) + \sum_{i=1}^n \langle D_{e_i} X, e_i \rangle^2 - \sum_{i,j=1}^k \langle D_{e_i} X, e_j \rangle \langle D_{e_j} X, e_i \rangle \Big)\, d\mu_\Sigma.$$

If Σ is a stationary hypersurface, then the second variation formula above has a simpler form. Assume the the normal bundle of Σ is trivial and let ν be a unit normal vector field. Define a vector field $X = \varphi(x)\nu$ where $\varphi(x)$ is a smooth function on Σ such that $\varphi = 0$ on $\partial\Sigma$. If we assume that Σ is minimal, then the second variation formula becomes:

LEMMA 3.3 (Second variation for hypersurface). *If $X = \varphi\nu$, then*

$$\delta^2\Sigma(X) = -\int_\Sigma \varphi L\varphi \, d\mu_\Sigma,$$

where $L\varphi := \Delta\varphi + (\|h\|^2 + Ric(\nu,\nu))\varphi$.

A minimal submanifold Σ is called *stable* if the second variation is nonnegative, i.e., $\delta^2\Sigma(X) \geq 0$ for any vector field X with compact support on Σ. In the hypersurface case, stability is equivalent to the condition that $\lambda_1(-L,\Omega) \geq 0$ for any compact domain Ω in Σ, where λ_1 is the first Dirichlet eigenvalue.

4. Curvature Estimates and Compactness Theorems

In this section we state (without proof) some results concerning curvature estimate for minimal stable hypersurfaces. Let Σ^n be a minimal stable hypersurface in M^{n+1} and let h be the second fundamental form. The following estimate is proven in [**S1**].

THEOREM 4.1 (Schoen). *If $n = 2$, then $\|h(x)\| \leq c(M, d(x,\partial\Sigma))$. If M is $\mathbb{R}^3$, then $\|h(x)\| \leq c \cdot d(x,\partial\Sigma)^{-2}$ for an absolute constant c.*

In the cases $n = 3,4,5$ the curvature estimate holds with a constant depending also on the volume of Σ (see [**SSY**]).

THEOREM 4.2 (Schoen–Simon–Yau). *If $n = 3,4,5$ and Σ is stable immersion, then*

$$\|h(x)\| \leq c(M, |\Sigma|, d(x,\partial\Sigma)),$$

where $|\Sigma|$ is the volume of Σ. If M is $\mathbb{R}^{n+1}$, then $\|h(x)\| \leq c \cdot d(x,\partial\Sigma)^{-2}$ for an absolute constant c. If $n = 6$ and Σ is a proper embedding, then the same estimate holds.

Curvature estimates of this type imply strong compactness theorems on large classes of stable hypersurfaces. The first theorem implies that the limit of immersed stable two-dimensional minimal surfaces in a fixed three-manifold is a stable minimal lamination (provided we stay a fixed distance from the boundaries). Because singularities are generally present even in area minimizing hypersurfaces for $n \geq 7$, we do not expect the same estimate to hold in general dimensions. The following compactness theorem holds in all dimensions (see [**SS**]).

THEOREM 4.3 (Schoen–Simon). *For arbitrary n, let*

$$\mathcal{G} := \{\Sigma : \Sigma \text{ is a proper embedded stable minimal hypersurface in } M \text{ with } |\Sigma| \leq C\}.$$

Then any element in the closure of $\mathcal{G}$ has singular set of Hausdorff dimension no greater than $n-7$.

It is an unsolved question whether Theorem 4.1 holds for $n = 3,4,5,6$. It would be very interesting to have such an estimate. It is also unknown whether there is a version of Theorem 4.3 without the volume bound. Very recently Neshan Wickramasekera [**Wic**] in his Stanford Ph.D. thesis has given a partial extension of Theorem 4.3 in the case that the hypersurfaces are immersed rather than embedded.

5. Geometry and Second Variation

The simplest example of the philosophy which is involved in the use of minimal submanifold theory to study curvature is the following theorem.

THEOREM 5.1. *If $Ric_M > 0$, then there is no compact stable minimal hypersurface in M.*

PROOF. Assume Σ is a compact stable minimal hypersurface in M. Let $\varphi \equiv 1$ in Lemma 3.3. The stability of Σ gives

$$\int_\Sigma \left(\|h\|^2 + Ric(\nu,\nu)\right) d\mu_\Sigma \leq 0,$$

which contradicts the fact that $Ric_M > 0$. □

If we combine this with the existence results (see [**F**]) which represent integral homology classes with area minimizing submanifolds, we get as a special case the vanishing of the n-th Betti number, $b_n(M^{n+1}) = 0$. This result is dual to Bochner's theorem on the vanishing of harmonic 1-forms under the positive Ricci curvature assumption.

Now we will study the case where only the scalar curvature is positive. In this setting the minimal hypersurface theory becomes much more powerful. Assume $R_M \geq 0$, and apply the Gauss Equation to express the intrinsic scalar curvature of Σ:

$$R^\Sigma_{ijij} = R^M_{ijij} + h_{ii}h_{jj} - h^2_{ij}.$$

Summing for $1 \leq i,j \leq n$ we get

$$R^\Sigma = \sum_{i,j=1}^{n} R^M_{ijij} - |h|^2 = R^M - 2Ric(\nu,\nu) - |h|^2.$$

Therefore,

$$|h|^2 + Ric(\nu,\nu) = \tfrac{1}{2}|h|^2 + \tfrac{1}{2}R^M - \tfrac{1}{2}R^\Sigma.$$

To illustrate the usefulness of this expression, consider the case $n = 2$ and let $\varphi \equiv 1$. The stability assumption then implies

$$\int_\Sigma R^\Sigma \, d\mu > 0,$$

which by the Gauss–Bonnet formula gives

$$\chi(\Sigma) > 0.$$

This gives us the following result (see [**SY1**]).

THEOREM 5.2. *Let (M^3, g) be a Riemannian manifold with positive scalar curvature $R^M > 0$. Then any compact stable minimal surface in M is topologically a sphere.*

For arbitrary dimension, the corresponding result is that the induced metric on Σ is conformally equivalent to a metric of positive scalar curvature (see [**SY2**]). This makes it possible to inductively study the topological structure of manifolds of positive scalar curvature. There has been extensive work in this direction (see [**GL1**], [**GL2**], [**SY2**], [**SY3**], [**St**]) which uses both the stable hypersurface approach and the dual approach using harmonic spinors. One of the basic unresolved questions in this theory is whether a compact $K(\pi,1)$ manifold of dimension four

or more can carry a metric of positive scalar curvature. There is a more refined theory which has to do with the Yamabe invariants of compact manifolds (see [**S2**], [**LB**], [**P**] for a general discussion and some results).

There is a rather direct relationship between the scalar curvature theory and problems about the Einstein equations of General Relativity. We now describe this connection.

6. General Relativity

A spacetime in General Relativity is a Lorentz four-dimensional manifold $(N^4, \bar{g})$, where $\bar{g}$ has signature type $(-,+,+,+)$. The evolution of the gravitational field $\bar{g}$ is then determined by the Einstein Equations

$$\bar{R}_{\mu\nu} - \tfrac{1}{2}\bar{R}\bar{g}_{\mu\nu} = 8\pi T_{\mu\nu},$$

where $\bar{R}_{\mu\nu}$ is the Ricci tensor, $\bar{R}$ is the scalar curvature, and $T_{\mu\nu}$ is the stress-energy tensor of any matter fields which are present.

6.1. Dominant energy condition. An observer moves tangent to a timelike vector e_0, and if e_0, e_1, e_2, e_3 is a Lorentz frame, then the observed energy-momentum density is represented by the vector $\sum T_{0\mu} e_\mu$. The condition that this vector is forward pointing and timelike for every observer is the *dominant energy condition*:

$$T_{00} \geq \sqrt{\sum_{i=1}^{3} T_{0i}^2}.$$

Since the Einstein equations form an evolution equation of hyperbolic type, a spacetime is determined by initial data given on a three-dimensional spacelike hypersurface in N. It is readily observed from the Gauss equations that a totally geodesic hypersurface in a Ricci flat manifold has vanishing scalar curvature; thus we expect such initial data to satisfy scalar curvature conditions. Using the Gauss and Codazzi equations together with the Einstein equations we first rewrite the dominant energy condition. Let M^3 be a space-like hypersurface, g be the induced metric and p be the second fundamental form. We then see that $\mu = T_{00}$ is given by

$$\mu := \frac{1}{16\pi}[R_M - \|p\|^2 + (\operatorname{Tr} p)^2],$$

and $J = \sum_{i=1}^{3} T_{0i} e_i$ is

$$J := \frac{1}{8\pi}(\operatorname{div} p - \nabla \operatorname{Tr} p)$$

where the covariant derivatives are taken in the induced geometry on M. Thus the dominant energy condition becomes:

$$\mu \geq \|J\|. \tag{6–1}$$

An important special case is $p = 0$, and in this case we see that (6–1) is equivalent to the condition that M has nonnegative scalar curvature $R(g) \geq 0$. In this paper we usually restrict our attention to this case, which we refer to as the Riemannian case.

6.2. Asymptotic flatness. The condition that a spacetime be an isolated gravitating system with all other matter/gravity at infinite distance is called asymptotic flatness. An initial data set (M^3, g, p) is said to be *asymptotically flat* if there is a compact set $K \subset M$ such that $M^3 \setminus K$ is diffeomorphic to $\mathbb{R}^3 \setminus B_1$(where B_1 is the unit ball in $\mathbb{R}^3$) and such that under this diffeomorphism, the metric and second fundamental form of $M^3 \setminus K$ can be written as

$$g_{ij} = \delta_{ij} + O(|x|^{-1}), \qquad p_{ij} = O(|x|^{-2}).$$

One can define the total ADM mass (and linear momentum) for an asymptotically flat initial data set (see for example [**HE**]). We first consider an important example of an asymptotically flat manifold.

6.3. Schwarzschild initial data. The Schwarzschild manifold is

$$(\mathbb{R}^3 \setminus \{0\}, s),$$

where s is a metric given by $s_{ij} = (1+m/2r)^4 \delta_{ij}$ and m is a positive constant and is by definition the total mass of the manifold. This manifold has zero scalar curvature everywhere and hence defines initial data for the vacuum Einstein equations (see [**HE**] for a discussion of this spacetime). Physically the Schwarzschild corresponds to the gravitational field exterior to a rotationally symmetric black hole of mass m. For simplicity, in our general discussion of asymptotic flatness, we can think of asymptotically flat manifolds as being asymptotic to a Schwarzschild manifold, i.e., we can write the asymptotic condition on g as

$$g_{ij} = \left(1 + \frac{m}{2r}\right)^4 \delta_{ij} + O(r^{-2}).$$

That this is no loss of generality in discussing the ADM mass is shown in [**SY7**]. We can then simply define the ADM mass of the manifold to be m. (More generally, the ADM mass is defined as a certain boundary integral over large coordinate spheres.)

7. Positive Mass Theorem

In this section we give a brief discussion of the Positive Mass Theorem both in the Riemannian case and the general case.

7.1. Positive Mass Theorem: Riemannian case. With the terminology we have set up, we may state the Positive Mass Theorem (Riemannian case) which was first proven by Schoen and Yau [**SY4**] using minimal surface theory, and later by Witten [**Wit**], [**PT**] using the theory of the Dirac operator:

THEOREM 7.1 (Positive Mass Theorem: Riemannian Case). *Let (M^3, g) be a complete, smooth asymptotically flat manifold with nonnegative scalar curvature and total mass m. Then*

$$m \geq 0,$$

with equality if and only if (M^3, g) is isometric to $\mathbb{R}^3$ with the standard Euclidean metric.

Now we outline the idea of the minimal surface proof of this theorem:

Step 1. We may normalize the asymptotic data so that g is conformal to the Euclidean metric near infinity. More precisely $g_{ij} = u^4 \delta_{ij}$ near ∞ where u is a function such that

$$\triangle u \leq 0, \qquad u = 1 + \frac{m}{2|x|} + O(|x|^{-2}).$$

Step 2. Assuming $m < 0$, then $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : |x_3| \leq C\}$ is a mean convex region for large C. This is a direct calculation using the asymptotic assumption on g. It follows that there exists a stable minimal surface Σ asymptotic to a plane $x_3 = a$ for some $a \in (-C, C)$. This follows from solving the Plateau problem with boundary a large circle in the x_1x_2 plane and using the mean convexity obtained in the previous step to show that we may let the radius of this circle go to infinity and find a limit of these area minimizing surfaces which is a stable minimal surface asymptotic to a horizontal plane.

Step 3. We show that this is inconsistent with the stability of Σ and the non-negativity of the scalar curvature; in fact, stability of Σ gives (we have to justify choosing $\varphi = 1$)

$$\int_\Sigma (2^{-1}\|h\|^2 + 2^{-1}R_M)\, da \leq \int_\Sigma K_\Sigma\, da, \tag{7–1}$$

where K_Σ is the Gauss curvature of the surface Σ. Then because of the fact that Σ is asymptotic to a plane, the Gauss–Bonnet formula gives

$$2^{-1}\int_\Sigma K_\Sigma\, da = 2\pi\chi(\Sigma) - 2\pi \leq 0.$$

On the other hand by a conformal argument we may assume the scalar curvature of M is positive, i.e., $R_M > 0$. Then (7–1) implies that

$$2^{-1}\int_\Sigma K_\Sigma\, da > 0,$$

a contradiction.

7.2. Positive Mass Theorem: general case. We discuss briefly an unpublished proof due to the author of the general case of the Positive Mass Theorem. This proof follows along the same lines as the Riemannian case. The original proof [**SY5**] is more complicated, but the development there has other useful applications particularly to the existence of apparent horizons assuming largeness conditions on the initial data ([**SY6**],)

Trapped Surfaces and the horizon equation: Let S be a 2-dimensional surface in an initial data set M^3 with induced metric g and second fundamental form p. Let e_0, e_1, e_2 and e_3 be a Lorentz frame. Let ν be the outward unit normal vector field to S in M. Then $\nu \pm e_0$ are orthogonal null vectors. We say that S is *trapped for* $\nu \pm e_0$ if

$$H_S \pm \mathrm{Tr}_S(p) < 0.$$

Then the apparent horizon equation for S is the marginally trapped condition

$$H_S \pm \mathrm{Tr}_S(p) = 0. \tag{7–2}$$

The general positive mass theorem says that the total energy-momentum vector (E, P_1, P_2, P_3) is a forward pointing time-like vector, i.e.,

$$E \geq \sqrt{\sum_{i=1}^{3} P_i^2},$$

where E is the total energy and $P = (P_1, P_2, P_3)$ is the total linear momentum (defined precisely below). We outline an analogous proof as for the Riemannian case.

Step 1. Without loss of generality we may assume for g and p the asymptotic condition

$$g_{ij} = (1 + \frac{E}{2|x|})^4 \delta_{ij} + O(|x|^{-2}),$$

$$p_{ij} = \mathcal{L}\Big(2\|P\| \cdot |x|^{-1} \frac{\partial}{\partial x_3}\Big) + O(|x|^{-3}),$$

where P is the total linear momentum vector given by

$$P_i = \frac{1}{8\pi} \oint_{S_\infty} \left((P_{ij}\nu^j - (\operatorname{Tr} P)\nu^i\right) dA,$$

and E is the total energy given by

$$E = \frac{1}{16\pi} \oint_{S_\infty} (g_{ij,j} - g_{jj,i})\nu_i dA,$$

and $\mathcal{L}$ is an operator defined by

$$\mathcal{L}X := L_X g - 2^{-1} \operatorname{div} X.$$

Step 2. Assume for the sake of contradiction that $E < |P|$. Then, by a calculation based on the asymptotics above, we have $H + \operatorname{Tr} p < 0$ on the boundary of the region $|x_3| < C$ for C large. Therefore this region is trapped, and there exists an apparent horizon Σ which is asymptotic to a plane $x_3 = a$ and with $H_\Sigma = -\operatorname{Tr} p$. (The existence here is complicated by the fact that the horizon equation is not a variational equation.)

Step 3. The solution which is constructed in the previous step satisfies the condition that its linearized operator has nonnegative first eigenvalue on compact sets, and an analogous stability-type argument produces a contradiction.

8. The Penrose Inequality (Riemannian Case)

We will call an outermost minimal sphere in an asymptotically flat manifold M^3 with nonnegative scalar curvature a *horizon.* We see that the Schwarzschild manifold has one horizon and an explicit calculation shows that the surface area A of the horizon satisfies:

$$m = \sqrt{\frac{A}{16\pi}}.$$

In general, the existence theory for minimal surfaces implies that there exists a finite collection of outermost minimal spheres so that if we remove the regions interior to them, then the manifold becomes diffeomorphic to $\mathbb{R}^3$ with a finite number of disjoint balls removed. Thus we may assume that our initial data set M is topologically the exterior of balls, and that the boundary spheres are minimal. We may furthermore assume that there are no compact minimal surfaces in the interior of M (the existence of one would produce another outermost minimal sphere). Now the Penrose inequality asserts that the total area of these outermost minimal spheres is bounded above by the horizon area for a Schwarzschild metric with the same total mass.

THEOREM 8.1 (Penrose Inequality). *Let (M^3, g) be a complete, smooth asymptotically flat manifold with nonnegative scalar curvature and total mass m whose outermost minimal spheres have total surface area A. Then*

$$m \geq \sqrt{\frac{A}{16\pi}},$$

with equality if and only if (M^3, g) is isometric to $(\mathbb{R}^3 \setminus \{0\}, s)$, the Schwarzschild metric of mass m, outside their respective horizons.

The Penrose Inequality was first conjectured by Penrose [**Pen**] in 1973 and was proven in a slightly weaker form by Huisken and Ilmanen [**HI**] in 1997 using inverse mean curvature flow and in full by Bray [**B**] in 1998 using a totally different method. (Gibbons, Tod, Bartnik, Herzlich and Bray had obtained earlier partial results.) We note that the general case of the Penrose Inequality (for arbitrary initial data sets) is still open.

8.1. The Geroch and Jang/Wald approach. We first describe the Hawking mass which for certain surfaces in M gives a reasonable definition of the gravitational mass enclosed by that surface.

DEFINITION 8.2 (Hawking Mass). *Let $\Sigma^2 \subset M^3$. The* Hawking mass of Σ , $m_H(\Sigma)$ *is defined to be*

$$m_H(\Sigma) := \sqrt{\frac{\text{Area}\,\Sigma}{16\pi}}(1 - \frac{1}{16\pi}\int_\Sigma H^2\, da).$$

In the Schwarzschild manifold, it is easily seen that the Hawking mass of any sphere S_r is equal to the mass m, and for general asymptotically flat M,

$$\lim_{r\to\infty} m_H(S_r) = m$$

where S_r denotes a coordinate sphere of large radius. A calculation (see [**BS**]) shows that if g is rotationally symmetric and the scalar curvature $R_M \geq 0$, then $m_H(S_r)$ is an increasing function of r.

Geroch [**G**] showed that if M^3 has nonnegative scalar curvature, then the Hawking mass of Σ is nondecreasing when the surface Σ flows outward at a speed equal to the inverse of the mean curvature. More precisely, assume Σ_t is a family of *connected* surfaces evolving by the equation

$$\frac{\partial x}{\partial t} = \frac{1}{H}\nu(x), \tag{8–1}$$

where H is the mean curvature of Σ_t and ν is the unit vector which is opposite to the mean curvature direction. Geroch then derived the important monotonicity property

$$\frac{d}{dt} m_H(\Sigma_t) \geq 0.$$

Using this monotonicity result, Jang and Wald [**JW**] gave a formal proof of the Penrose inequality in case there is a single outermost minimal sphere. Their formal argument supposes that Σ_0 is an outermost minimal sphere. Assume that the inverse mean curvature flow equation (8–1) with initial data Σ_0 has a family of smooth solution Σ_t for $0 \leq t < \infty$ such that for large t, the surface Σ_t is asymptotic to a coordinate sphere with large radius. It then follows that

$$\lim_{t\to\infty} m_H(\Sigma_t) = m.$$

On the other hand, Geroch monotonicity implies that

$$m_H(\Sigma_0) = \sqrt{\frac{\text{Area}\,\Sigma_0}{16\pi}} \leq m_H(\Sigma_t).$$

Hence we get

$$m \geq \sqrt{\frac{A}{16\pi}},$$

which gives the Penrose inequality in the case when M has only one outermost minimal sphere.

The main difficulty in making this formal argument rigorous is to prove the existence and regularity of the inverse mean curvature flow. In general singularities occur in this flow, and these must be understood in order to have a hope of giving a rigorous proof. Huisken and Ilmanen succeeded in rigorizing this argument by constructing an appropriate weak solution of the inverse mean curvature flow. We now give an outline of their argument.

9. The Huisken–Ilmanen Approach

To understand the difficulties which may arise, we first analyze some special cases for the inverse mean curvature flow.

Example 1. The initial surface is a coordinate sphere in $\mathbb{R}^3$, i.e., $\Sigma_0 = S_{r_0}$ for some r_0. Then we have

$$\frac{d}{dt}(\operatorname{Area}\Sigma_t) = \operatorname{Area}\Sigma_t.$$

So we have

$$\Sigma_t = S_{e^{t/2}r_0}.$$

Thus in this example the flow exists and has the desired behavior. We next consider two examples for which the flow becomes singular.

Example 2. The initial surface Σ_0 is a "thin" torus of revolution. In this case, the mean curvature vector points into the solid torus and Σ_t exists for a short time, but at some finite time we will have $\min_{\Sigma_t} H \to 0$ and $1/H$ goes to infinity so the flow does not make sense after that. Indeed, there must be a topology change if the large time flow is to resemble large spheres.

Example 3. The initial surface is the disjoint union of two spheres, i.e., $\Sigma_0 = S_{r_1}(P) \cup S_{r_2}(Q)$. In this case, the classical flow Σ_t must develop self intersections in a finite time.

From the examples above, we see that in order for a flow to exist for all t, we must allow jumps and changes in topology.

9.1. The level set approach. To construct a solution of the inverse mean curvature flow, Huisken and Ilmanen rewrite the flow as an equation for the level sets of a function. Let $u(x)$ be a function such that $\Sigma_t := \{x : u(x) = t\}$ is smooth for all t. By direct calculation we see that if Σ_t is a solution to the inverse mean curvature flow, then

$$\operatorname{div}\Big(\frac{\nabla u}{|\nabla u|}\Big) = |\nabla u|. \tag{9–1}$$

REMARK 9.1. *The level set formulation allows jumps in a natural way, since if u is constant on an open set Ω, then the level sets "jump" across Ω.*

DEFINITION 9.2. *$\Sigma^2 \subset M^3$ is called outer minimizing if* $\operatorname{Area}\Sigma \leq \operatorname{Area}\Sigma_1$ *for any surface Σ_1 enclosing Σ. Σ is called strictly outer minimizing if equality holds if and only if $\Sigma = \Sigma_1$.*

The following illustrates the connection between the inverse mean curvature flow and the outer minimizing property.

LEMMA 9.3. *If $u(x)$ is a solution to the inverse mean curvature flow equation (9–1), then for any $t > 0$, the level set Σ_t is outer minimizing.*

PROOF. Let Σ be any surface enclosing Σ_t and Ω be the region between Σ and Σ_t. Integration by parts gives

$$\int_\Omega \operatorname{div}\Big(\frac{\nabla u}{|\nabla u|}\Big)\, dv = \int_\Sigma \nu \cdot \frac{\nabla u}{|\nabla u|}\, da - \operatorname{Area}\Sigma_t.$$

On the other hand (9–1) implies that

$$\operatorname{div}\Big(\frac{\nabla u}{|\nabla u|}\Big) = |\nabla u| \geq 0.$$

Thus we conclude that

$$\operatorname{Area}\Sigma_t \leq \operatorname{Area}\Sigma.$$

(Equality implies that $u \equiv t$ on Ω.) □

The existence of outer minimizing sets follows from the Plateau theory.

LEMMA 9.4. *For any $\Sigma \subset M$, there exists a unique smallest strictly outer minimizing surface $\hat{\Sigma}$. We will call $\hat{\Sigma}$ the strict minimizing hull of Σ.*

REMARK 9.5. *Although Σ might be enclosed by more than one outer minimizing surface, the strictly minimizing hull is the maximal such surface.*

Assume $u(x)$ is a global solution of the inverse mean curvature flow equation (9–1). Let

$$\Omega_t := \{x : u(x) < t\}, \qquad \Omega_t^+ := \{x : u(x) \leq t\}.$$

Then $\Sigma_t = \partial\Omega_t$ is outer minimizing and $\Sigma_t^+ = \partial\Omega_t^+$ is strictly outer minimizing. In fact Σ_t^+ is the strict minimizing hull of Σ_t. The surfaces Σ_t and Σ_t^+ differ precisely at jumps, i.e., when $u(x) = t$ on a set of positive measure.

9.2. Heuristic description of the Huisken–Ilmanen flow. A valid way to think of the Huisken–Ilmanen modified inverse mean curvature flow is as follows: If $\hat{\Sigma}_t = \Sigma_t$, then continue with the classical inverse mean curvature flow, but at any instant for which $\hat{\Sigma}_t \neq \Sigma_t$, jump to $\hat{\Sigma}_t$ and continue with the classical inverse mean curvature flow. The rigorous construction is different from this, but it captures the main idea, and explains the examples given above.

Example 2. The "thin" torus: After some time t_0 before the mean curvature goes to zero the surface Σ_{t_0} will cease to be outer minimizing. At this time the surface Σ_{t_0} will jump to its strict minimizing hull which will be a sphere, and from then on the flow will be smooth and asymptotic to large coordinate spheres.

Example 3. The disjoint union of spheres: A similar phenomenon will occur, and after time t_0, Σ_{t_0} becomes a single sphere enclosing the two.

LEMMA 9.6 (Monotonicity of Hawking Mass). *The Hawking mass $m_H(\Sigma_t)$ is increasing for the Huisken–Ilmanen flow.*

PROOF. This is a rough sketch. It suffices to show that the Hawking mass increases at the jumps. Assume $\hat{\Sigma}_t \neq \Sigma_t$, then we see $\operatorname{Area}\Sigma_t = \operatorname{Area}\hat{\Sigma}$, $\hat{H} = H$ on $\Sigma_t \cap \hat{\Sigma}_t$ and $\hat{H} = 0$ on $\hat{\Sigma}_t \setminus \Sigma_t$. Therefore

$$\int_{\hat{\Sigma}} \hat{H}^2 \, da \leq \int_{\Sigma} H^2 \, da.$$

This implies $m_H(\hat{\Sigma}_t) \geq m_H(\Sigma_t)$. □

A major problem with the Huisken–Ilmanen flow is how to rigorize the construction, (e.g., the jump times are not known to be discrete). We give a hint as to how that is done.

9.3. Elliptic regularization. Consider the following perturbed version of the inverse mean curvature flow equation (9–1)

$$\operatorname{div}\left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon^2}}\right) = \sqrt{|\nabla u|^2 + \varepsilon^2}. \tag{9–2}$$

We notice that the perturbed equation (9–2) is a nondegenerate elliptic equation and it has the following geometric meaning. If $u(x)$ is a solution to the perturbed equation (9–2), then

$$G_t(x) := \operatorname{graph}\Big(\frac{u(x)}{\varepsilon} - \frac{t}{\varepsilon}\Big) \subset M \times \mathbb{R}$$

is a solution to the unperturbed equation (9–1).

We now sketch Huisken–Ilmanen's proof of a version of the Penrose Inequality.

Step 1. There is a smooth solution u_ε of the perturbed equation (9–2) satisfying $u_\varepsilon = 0$ on the initial outermost minimal sphere Σ_0 and u_ε has suitable behavior near infinity.

Step 2. The $\{u_\varepsilon\}$ above satisfy local uniform Lipschitz bounds and hence a subsequence converges to some function u uniformly.

Step 3. The limit function u is a Lipschitz weak solution of the unperturbed equation (9–1) in the sense that

$$\int_\Omega (|\nabla u| + u|\nabla u|)\, d\mu \leq \int_\Omega (|\nabla v| + v|\nabla u|)\, d\mu,$$

for any v such that $v - u$ has compact support in Ω for any compact set Ω in M. Moreover there can be no other Lipschitz weak solution whose zero set is Σ_0.

Step 4. The surface Σ_t is connected and is $C^{1,\alpha}$ for any $t > 0$, and Σ_t becomes asymptotic to a coordinate sphere for t large. Therefore by Lemma 9.6, $m_H(\Sigma_t)$ can be defined and is monotone increasing in t. Using a similar argument as for the formal proof of Jang/Wald, we can show that the Riemannian Penrose Inequality holds assuming there is one outermost minimal sphere. More generally the statement is as follows.

THEOREM 9.7 (Huisken–Ilmanen). *If M^3 is a complete, asymptotically flat Riemannian manifold with nonnegative scalar curvature, with total mass m and with outermost minimal spheres $\Sigma_1, \Sigma_2, \ldots, \Sigma_k$. Then*

$$m \geq \sqrt{\frac{A_{max}}{16\pi}},$$

where

$$A_{max} := \max\{\text{Area}\,\Sigma_1, \ldots, \text{Area}\,\Sigma_k\}.$$

Moreover, equality holds if and only if the region of M outside Σ_1, Σ_2, ..., Σ_k is isometric to the exterior Schwarzschild metric (in particular, $k = 1$).

10. Bray's Approach

After the Huisken/Ilmanen was written, H Bray [**B**] found a clever argument to use the Positive Mass Theorem [**SY4**] to prove the full version of the Penrose Inequality (Riemannian case). In this section we will discuss his approach.

We first may simplify the assumptions. By Schoen–Yau's argument in the proof of the Positive Mass Theorem [**SY7**] we can assume $R_g \equiv 0$ and

$$g_{ij} = u^4 \delta_{ij}$$

outside a compact set, where u is a function satisfying

$$\begin{cases} \triangle_g u = 0, \\ u(x) = 1 + m/(2|x|) + O(|x|^{-2}). \end{cases}$$

Bray defines a continuous family of conformal metrics $\{g_t\}$ on M^3, where

$$g_t = u_t^4 g_0,$$

for some suitable function $u_t(x)$ (described later) and $u_0(x) \equiv 1$. Given the metric g_t, define

$$\Sigma(t) := \text{the outer minimal area enclosure of } \Sigma_0 \text{ in } (M^3, g),$$

where Σ_0 is the union of the original outer-minimizing spheres in (M^3, g_0). The time rate of change of u_t is given by v_t where v_t satisfies

$$\text{(10–1)} \qquad \begin{cases} \triangle_{g_0} v_t = 0, & \text{outside } \Sigma_t, \\ v_t = 0, & \text{on } \Sigma(t), \\ \lim_{x\to\infty} v_t(x) = -e^{-t}, \end{cases}$$

and set $v_t \equiv 0$ inside $\Sigma(t)$. Then we set $d/dt(u_t) = v_t$, or,

$$u_t(x) = 1 + \int_0^t v_s(x)\, ds.$$

THEOREM 10.1. *With the notation above, there exists a solution $u_t(x)$ and Σ_t for any $t \geq 0$ such that $u_t(x)$ is Lipschitz in t, C^1 in x globally and C^∞ in x outside $\Sigma(t)$. Moreover $\Sigma(t_1) \cap \Sigma(t_2) = \varnothing$ for $t_1 \neq t_2$.*

The monotonicity associated with this flow is given as follows.

PROPOSITION 10.2. *The mass $m(t)$ of (M^3, g_t) is nonincreasing.*

PROOF. Since the flow has an autonomous character, it suffices to show $m'(0) \leq 0$. Assume

$$\text{(10–2)} \qquad u_t(x) = a(t) + \frac{b(t)}{2|x|} + O(|x|^{-2}).$$

Note that $g_t = u_t^4 g_0$ is asymptotically flat with the expansion

$$g_t = \left(\left(a(t) + \frac{b(t)}{2|x|}\right)^4 \left(1 + \frac{m(0)}{2|x|}\right)^4\right)\delta_{ij} + O(|x|^{-2}).$$

Computing, we get

$$m(t) = a(t)(b(t) + m(0)b(t)).$$

From the definition of v_t in (10–1) we find

$$v_0 = -1 + \frac{C_0}{2|x|} + O(|x|^{-2}), \tag{10–3}$$

where C_0 is the Newtonian capacity defined by

$$C_0 = \inf\{\frac{1}{2\pi}\int_{M\setminus\Sigma(0)} |\nabla\varphi|^2 : \varphi = 0 \text{ on } \Sigma(0) \text{ and } \varphi = 1 \text{ at } \infty\}.$$

Now since $u_0 \equiv 1$ and $d/dt(u_t) = v_t$, comparing (10–2) and (10–3), we see

$$a(0) = 1, \qquad \dot{a}(0) = -1, \qquad b(0) = 0, \qquad \dot{b}(0) = C_0.$$

Thus

$$m'(0) = C_0 - 2m(0).$$

We now prove the following inequality which will complete the proof of monotonicity. □

PROPOSITION 10.3. $m(0) \geq 2^{-1}C_0$.

PROOF. This will come by applying the positive mass theorem. We double M by reflection along Σ_0 and extend v_0 to the doubled manifold by odd reflection. Define

$$\tilde{g}_0 = (2^{-1}(1 - v_0))^4 g_0.$$

We can then show that the mass $\tilde{m}$ of $\tilde{g}$ is in fact $m(0) - 2^{-1}C_0$. The positive mass theorem therefore gives

$$\tilde{m} = m(0) - 2^{-1}C_0 \geq 0$$

which is the desired conclusion. □

We may now complete a sketch of Bray's proof of the Penrose Inequality: By Theorem 10.1, we see that Σ_{t_2} strictly encloses Σ_{t_1} for $t_2 > t_1 \geq 0$. Also we can prove that $A(t) :=$ total area of Σ_t is constant, i.e., $A(t) \equiv A_0$ for any $t \geq 0$, and we have shown that the mass $m(t)$ is nonincreasing for $t \geq 0$. It is shown by Bray that $(M \setminus \Sigma(t), g_t)$ converges to a Schwarzschild metric. This, together with Proposition 10.2, implies that

$$m = m(0) \geq \lim_{t\to\infty} m(t) = m_{sch} = \sqrt{\frac{A}{16\pi}},$$

which gives the Penrose Inequality.

11. Higher Codimensions

The minimal submanifold theory used up to now has been all in codimension one. We now consider some applications of minimal submanifolds with higher codimension. One of the main motivations for many of the phenomena in the subject is the Bernstein Theorem:

THEOREM 11.1 (Bernstein Theorem). *Any entire minimal graph in* $\mathbb{R}^3$ *is a plane.*

While the original proof involved a PDE type argument the theorem has been generalized in more geometric ways. Osserman [**O**] gave a generalization using the fact that the Gauss map is conformal, and replaced the graphical assumption with the assumption that the Gauss image omit a sufficiently large set on the two-sphere. This theory was improved dramatically by Fujimoto [**Fu**]. A great achievement for the codimension 1 theory was to prove the Bernstein theorem for entire minimal graphs of dimension seven or less, and to find counterexamples in higher dimensions (see [**F**] for an account of this). Schoen, Simon and Yau [**SSY**] proved this result in dimensions up to 6 for complete stable minimal hypersurfaces with a volume bound using curvature estimate.

It is natural to ask if there is an analogue of the Bernstein theorem for higher codimension minimal submanifolds. The theory of Osserman concerning the size of the omitted set for complete minimal surfaces was generalized to higher codimension (see [**CO**]). As a first guess, one might expect that entire graphs could be holomorphic with respect to an orthogonal complex structure on $\mathbb{R}^n$. That this is not the case was shown by Osserman [**O**]. One may then ask what is the suitable global hypothesis for a minimal surface in higher codimensions to be holomorphic. Only partial results of this type are known.

It turns out that in various higher dimensions and codimensions there are classes of submanifolds which satisfy first order reductions of the minimal submanifold equation and are automatically volume minimizing like the holomorphic submanifolds. We now give a general discussion of these calibrated submanifolds.

12. Calibrations

To construct volume minimizing submanifolds, Harvey and Lawson in their paper [**HL**] defined the concept of a calibration.

DEFINITION 12.1. *A k-form α on a Riemannian manifold (M^n, g) is called a calibration if*

1. $d\alpha = 0$;
2. $|\alpha(\pi_x)| \leq 1$, *for any k-dimensional subspace π_x in T_xM and any $x \in M$.*

DEFINITION 12.2. *Let α be a calibrating k-form on M. A k-dimensional submanifold Σ^k is said to be calibrated by α if $\alpha(T_x\Sigma) = 1$ for any $x \in \Sigma$. In other words, Σ is calibrated by α if the restriction of α to Σ is the volume form on Σ.*

The most important result about calibrated submanifolds is the following minimizing property for calibrated submanifolds.

PROPOSITION 12.3. *Let α be a calibrating k-form on M and let Σ^k be a submanifold calibrated by α. Then Σ is volume minimizing in its (relative) integral homology class, i.e.,*

$$|\Sigma| = \inf\{|\Sigma_1| : \Sigma_1 \text{ is homologous to } \Sigma,\ \partial\Sigma_1 = \partial\Sigma\}$$

where $|\cdot|$ denotes the volume.

PROOF. Since Σ and Σ_1 are homologous (with $\partial\Sigma = \partial\Sigma_1$), we assume $\Sigma - \Sigma_1 = \partial C$, where C is a $(k+1)$-dimensional chain. Then Stokes' Theorem gives

$$\int_{\Sigma-\Sigma_1} \alpha = \int_{\partial C} \alpha = \int_C d\alpha = 0.$$

On the other hand since Σ is calibrated by α, we see that

$$|\Sigma| = \int_\Sigma \alpha = \int_{\Sigma_1} \alpha \leq |\Sigma_1|,$$

where, in the last inequality, we used property (2) of the definition of calibrating form. □

Complex submanifolds in $\mathbb{R}^{2m}$ can be viewed as calibrated submanifolds. A complex structure on $\mathbb{R}^{2m}$ is a linear isomorphism $J : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$ such that $J^2 = -I$, where I is the identity map. A complex structure J is said to be compatible with the Euclidean metric if J is also an isometry, i.e., $J(v) \cdot J(w) = v \cdot w$, where $\cdot$ is the Euclidean inner product of $\mathbb{R}^{2m}$. For a compatible complex structure J, we define the *Kähler form (or symplectic form)* ω by $\omega(v, w) := J(v) \cdot w$. A $2k$-dimensional subspace V^{2k} in $\mathbb{R}^{2m}$ is called *complex with respect to* J if it is invariant under J, i.e., $J(V) = V$. A $2k$-dimensional submanifold Σ^{2k} is called *complex (or holomorphic) with respect to* J if $T_x\Sigma$ is complex for every $x \in \Sigma$.

The following Wirtinger Inequality (for proof, see [**L**]) implies that $\omega^k/k!$ is a calibrating form.

LEMMA 12.4 (Wirtinger Inequality). *Let ω be the Kähler form for a compatible complex structure J in $\mathbb{R}^{2m}$. Define $\alpha := \omega^k/k!$, where $\omega^k = \omega \wedge \cdots \wedge \omega$. Then $|\alpha(V)| \leq 1$ and $\alpha(V) = 1$ if and only if V is complex.*

An immediate consequence of the Wirtinger Inequality is

COROLLARY 12.5. *The form α defined in the above lemma is a calibrating form which calibrates complex submanifolds. Therefore any complex submanifold is volume minimizing in its relative homology class (i.e., for its boundary).*

We now give the standard complex structure on $\mathbb{R}^{2m}$. Let $x_1, \ldots x_m, y_1, \ldots y_m$ be the standard coordinates in $\mathbb{R}^{2m}$. Define J by

$$J\Big(\frac{\partial}{\partial x_j}\Big) := \frac{\partial}{\partial y_j}, \quad J\Big(\frac{\partial}{\partial y_j}\Big) := -\frac{\partial}{\partial x_j}.$$

Therefore the Kähler form $\omega = \sum_{j=1}^m dx_j \wedge dy_j$. Let $z_j = x_j + \sqrt{-1}y_j$ be the standard complex coordinates.

Now consider a graph over the $z_1, \ldots z_k$ plane given by

$$z_\alpha = f_\alpha(z_1, \ldots z_k), \quad \alpha = k+1, \ldots, m.$$

It is easy to check that the graph is holomorphic if $\partial f_\alpha/\partial \bar{z}_j = 0$ for $j = 1, \ldots k$. From the above discussion we see that the graph of f is a volume minimizing submanifold.

Corollary 12.5 says that every complex submanifold is volume minimizing; however, there has been very little success in showing that volume minimizing submanifolds are complex even in situations where one may expect them to be. There are a few results in dimension 2.

THEOREM 12.6 (Siu–Yau [**SiY**]). *Let M^{2m} be a Kähler manifold with positive bisectional curvature. Any stable minimal 2-sphere is either holomorphic or anti-holomorphic.*

THEOREM 12.7 (Micallef [**M**]). *In $\mathbb{R}^4$, any 2-dimensional complete stable minimal surfaces with quadratic area growth is holomorphic for some compatible complex structure J. Also any 2-dimensional entire stable minimal graph in $\mathbb{R}^4$ is holomorphic for some compatible complex structure.*

THEOREM 12.8 (Micallef [**M**]). *In R^{2m}, any 2-dimensional stable minimal surface with genus zero and with finite total curvature is holomorphic.*

REMARK 12.9. *Recently, Arezzo and Micallef* [**AM**] *gave examples for n large (near 20) of stable minimal Σ with genus 1 and finite total curvature which are not holomorphic.*

There is also the following existence result for holomorphic disks which is important in symplectic geometry. The proof uses the $\bar{\partial}$ operator directly, and up to this time there is no proof which uses area minimization.

DEFINITION 12.10. *A submanifold Σ^k in R^{2k} is called lagrangian if $\omega|_\Sigma = 0$, where ω is the Kähler form.*

THEOREM 12.11 (Gromov [**Gr1**]). *Let Σ^k be a compact embedded lagrangian submanifold in R^{2k}. Then there exists a holomorphic 2-disk D such that $\partial D \subset \Sigma$.*

12.1. The special lagrangian calibration. Another important calibrating form is the special lagrangian form, which was originally defined in Harvey and Lawson's paper [**HL**]. This form and the associated class of special lagrangian submanifolds can be defined generally on a Calabi–Yau manifold. For simplicity we discuss here the flat case. In $\mathbb{R}^{2m}$ let $z_1, \ldots, z_m$ be the standard complex coordinates. Define a real m-form $\alpha := \mathrm{Re}(dz_1 \wedge \cdots \wedge dz_m)$. Standard linear algebra shows that $|\alpha(V)| \leq 1$ for any m-dimensional subspace V (see [**HL**] for a detailed proof). Thus we have the following result.

LEMMA 12.12. *$\alpha := \mathrm{Re}(dz_1 \wedge \cdots \wedge dz_m)$ is a calibrating m-form in $\mathbb{R}^{2m}$.*

DEFINITION 12.13. *An m-subspace V in $\mathbb{R}^{2m}$ is called special lagrangian if it is calibrated by α.*

We have the following characterization of special lagrangian subspace. We refer the reader to [**HL**] for a proof.

THEOREM 12.14. *Let $\alpha = \mathrm{Re}(dz^1 \wedge dz^2 \wedge \cdots \wedge dz^m)$ and $\mu = \mathrm{Im}(dz^1 \wedge dz^2 \wedge \cdots \wedge dz^m)$ in $\mathbb{C}^m$. Then α is a calibrating m-form. Moreover, if P is an oriented m-plane in $T_x C^m$, the following statements are equivalent:*

(a) *P is special lagrangian;*
(b) *$\mu(P) = 0$ and P is lagrangian;*
(c) *There is a linear map $A \in SU(m)$ such that A maps the x-plane (the m-plane in $\mathbb{C}^m$ spanned by $\partial/\partial x^1, \ldots, \partial/\partial x^m$) onto P.*

REMARK 12.15. *If P is any lagrangian subspace, then we see that $|dz(P)| = 1$. Thus we can write $dz(P) = e^{\sqrt{-1}\beta}$ for some angle β (called lagrangian angle) which is well-defined mod 2π. Therefore we see that P is special lagrangian if and only if P is lagrangian and $\beta = 0$.*

This leads us to the definition of special lagrangian submanifolds. These bear certain formal analogies to the class of holomorphic submanifolds.

DEFINITION 12.16. *A submanifold $\Sigma^m \subset \mathbb{R}^{2m}$ is called special lagrangian if each tangent plane of Σ is special lagrangian.*

Note that there is an S^1 family of calibrating forms in $\mathbb{C}^n$ associated with the form dz. Let

$$\alpha_\theta = \mathrm{Re}(e^{\sqrt{-1}\theta} dz^1 \wedge dz^2 \wedge \cdots \wedge dz^m), \tag{12–1}$$

and

$$\mu_\theta = \mathrm{Im}(e^{\sqrt{-1}\theta} dz^1 \wedge dz^2 \wedge \cdots \wedge dz^m). \tag{12–2}$$

As above, the form α_θ is a calibrating form and there is an associated class of submanifolds calibrated by α_θ. We will refer to these as *special lagrangian with respect to α_θ*. One of the reasons we are interested in this S^1 family of special lagrangian geometry is the following theorem (see [**HL**] or the discussion in the next section).

THEOREM 12.17. *Let Σ be a smooth submanifold of real dimension m in $\mathbb{R}^{2m}$. Then Σ is both minimal and lagrangian if and only if Σ is special lagrangian with respect to α_θ for some choice of θ.*

13. Existence Theory for Special Lagrangian Submanifolds

The existence of special lagrangian submanifolds is of great interest in both geometry and string theory. Theorem 12.17 reduces this problem to the existence of minimal lagrangian submanifolds. To construct such submanifolds, one idea is to minimize volume among lagrangian competitors. The author's joint work with Jon Wolfson [**SW2**] explores this variational approach in detail and gives an approach to the construction of a smooth minimal lagrangian submanifold in each lagrangian homology class in a 4-dimensional Kähler–Einstein manifold. Here we provide some background and elementary properties related to this approach.

13.1. Hamiltonian stationary submanifolds. To solve this lagrangian variational problem, we need to find suitable deformations which preserve the lagrangian condition. Ambient symplectic deformations (i.e., deformations of $\mathbb{R}^{2m}$ which preserve the symplectic form $\omega = \sum_{j=1}^m dx_j \wedge dy_j$) are good candidates. One family of such deformations consists of the hamiltonian deformations which arise from a smooth ambient function.

DEFINITION 13.1. *Let $h(x_1, \ldots x_m, y_1, \ldots y_m)$ be a smooth function on R^{2m} with compact support. The hamiltonian vector field associated to h is defined to be*

$$X_h := J\nabla h = \sum_{j=1}^m \Big(\frac{\partial h}{\partial x_j} \frac{\partial}{\partial y_j} - \frac{\partial h}{\partial y_j} \frac{\partial}{\partial x_j} \Big).$$

It is not difficult to check that the flow generated by any hamiltonian vector field preserves the symplectic form. Therefore the image of any lagrangian submanifold under a hamiltonian deformation is lagrangian.

DEFINITION 13.2. *A lagrangian submanifold Σ is called hamiltonian stationary if the first variation of Σ under any compactly supported hamiltonian deformation is zero, i.e.,*

$$\delta|\Sigma|(X_h) = 0$$

for any smooth function h on R^{2m} whose restriction to Σ has compact support.

The following proposition says that the mean curvature vector of any lagrangian submanifold is a (multi-valued) hamiltonian vector field (see [**HL**] or [**SW1**] for a proof).

PROPOSITION 13.3. *Let $\Sigma^m \subset R^{2m}$ be a lagrangian submanifold. We then have $dz(T_x\Sigma) = e^{\sqrt{-1}\beta}$ for some angle β (the lagrangian angle) defined mod 2π. Moreover, $H = J\nabla^\Sigma \beta$ where H is the mean curvature and ∇^Σ is the induced connection on Σ.*

The above proposition implies that if we minimize volume among a class for which the mean curvature is an allowable variation, then the solution will be minimal lagrangian (hence special lagrangian with respect to some α_θ).

13.2. The Euler–Lagrange equations. In this section, we give two versions of the Euler–Lagrange equations for hamiltonian stationary submanifolds. The first is the following geometric version.

PROPOSITION 13.4. *Let $\Sigma^m \subset \mathbb{R}^{2m}$ be a lagrangian submanifold which is hamiltonian stationary. Let H be the mean curvature vector. Define a one-form σ_H on Σ by $\sigma = H\rfloor\omega$, where $\rfloor$ denotes the interior product and ω is the standard symplectic form on $\mathbb{R}^{2m}$. Then we have the Euler–Lagrange equations*

$$d\sigma_H = 0, \qquad \delta\sigma_H = 0.$$

PROOF. From above we have that locally $\sigma_H = d\beta$ (this is equivalent to saying $H = J\nabla\beta$). Thus we have $d\sigma_H = 0$ for any lagrangian submanifold. The first variation formula and the hamiltonian stationarity then imply that

$$0 = \delta\Sigma(X_h) = -\int_\Sigma X_h \cdot H \, d\mu = -\int_\Sigma J\nabla h \cdot J\nabla\beta \, d\mu = -\int_\Sigma \nabla h \cdot \nabla\beta \, d\mu.$$

Hence $\triangle\beta = 0$, that is, β is a harmonic function on Σ. This is equivalent to the condition $\delta\sigma_H = 0$. □

Now we give the analytical version of the Euler–Lagrange equations. Standard lagrangian geometry implies that any lagrangian submanifold which is graphical over a lagrangian plane, can be written as the graph of the gradient of a potential function defined on the plane. More precisely, if Σ is a lagrangian submanifold which is graphical over the x-plane, then there is a function $u(x)$ such that the graph may be defined by $y = \nabla u$, i.e.,

$$y_i = \frac{\partial u}{\partial x_i} \qquad \text{for } i = 1, \ldots, m.$$

If we in addition assume that Σ is also hamiltonian stationary, a standard first variation computation gives the Euler–Lagrange equation

$$\sum_{j=1}^m \frac{\partial}{\partial x_j}\Big(\triangle_g \frac{\partial u}{\partial x_j}\Big) = 0. \tag{13–1}$$

Here the induced metric is given by $g_{ij} = \delta_{ij} + \sum_k u_{ik}u_{jk}$ with subscripts denoting partial derivatives of u. This is a fourth order quasilinear scalar equation for u which is of bi-harmonic type.

13.3. Examples. In this section we give some examples of hamiltonian stationary submanifolds.

Example 1. If $m = 1$, hamiltonian stationarity implies that the lagrangian angle β is a linear function of s (the arclength parameter). Thus the hamiltonian stationary curves in R^2 are the lines and circles.

Example 2 (Clifford Tori). Consider $\mathbb{R}^{2m}$ as $\mathbb{C}^m$, and define $\Sigma = S^1(r_1) \times \cdots \times S^m(r_m)$, where $S^i(r_i)$ is the circle with radius r_i in the i-th copy of $\mathbb{C}$. These are called the *Clifford tori*, and they are hamiltonian stationary. It is unknown whether for $m > 1$ these minimize volume in their hamiltonian isotopy class. For $m = 1$ this is true and is equivalent to the isoperimetric inequality for plane regions.

Example 3 (Helein/Romon). If $m = 2$, then Helein and Romon [**HR**] showed that there are infinitely many distinct hamiltonian stationary tori in R^4. Their proof uses explicit representation formulae for hamiltonian stationary surfaces in $\mathbb{R}^4$ which arise from the theory of integrable systems.

Example 4 (Hamiltonian stationary cones). Let

$$S^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}$$

be the unit 3-sphere in $\mathbb{C}^2$. Let $\pi : S^3 \to S^2$ be the standard Hopf map defined by

$$\pi((z_1, z_2)) = \frac{z_1}{z_2} \in \mathbb{C} \cup \{\infty\} = S^2.$$

The fiber of this projection is a great circle, i.e., for $p \in S^2$,

$$\pi^{-1}(p) = \{e^{\sqrt{-1}\theta}(z_1, z_2) : \theta \in [0, 2\pi)\}$$

for any point (z_1, z_2) in $\pi^{-1}(p)$. For a point $q \in S^3$ we will call the great circle $\{e^{\sqrt{-1}\theta}q : \theta \in [0, 2\pi]\}$ the *Hopf fiber through* q and simply denote it as $e^{\sqrt{-1}\theta}q$. Now let $\gamma(t) \subset S^3$ be a curve satisfying:

1. γ is horizontal in the sense that $\gamma'(t) \perp T_{\gamma(t)}(e^{\sqrt{-1}\theta}\gamma(t))$ (i.e., γ is perpendicular to the Hopf fibers);
2. $\pi(\gamma)$ is a circle in S^2.

It is shown in [**SW1**] that there are infinitely many closed curves in S^3 satisfying these two properties. These are precisely the curves γ in S^3 with the property that the corresponding cones in $\mathbb{R}^4$ over γ are hamiltonian stationary.

14. The Lagrangian Plateau Problem in $\mathbb{R}^4$

Let Γ be a smooth Jordan curve in R^4. We would like to find a lagrangian disk bounded by γ which has the least area among all such disks.

First we need to ask whether γ bounds any lagrangian disk. One can easily derive a necessary condition using Stokes' Theorem. Let $\eta = \sum_{i=1}^{2}(x_i\, dy_i - y_i\, dx_i)$. Clearly η is a primitive of the standard symplectic form ω. If Γ bounds an oriented lagrangian surface Σ, then Stokes' Theorem implies that

$$0 = \int_\Sigma \omega = \int_\Gamma \eta.$$

Therefore, a necessary condition for a closed curve Γ to bound a lagrangian disk is $\int_\Gamma \eta = 0$.

It turns out that this is also a sufficient condition. Quantitative results obtained by Gromov[**Gr2**] and Allcock[**A**] show that if Γ is a closed curve in $\mathbb{R}^4$ such that $\int_\Gamma \eta = 0$, then Γ bounds a (singular) lagrangian disk D with area bounded in terms

of the length of Γ, i.e., $\operatorname{Area} D \leq c(\operatorname{Length}\Gamma)^2$, where c is an absolute constant. (The situation is completely different if we allow nonorientable surfaces. Qiu [**Q1**] proved that any closed curve in R^4 bounds a lagrangian Möbius band with similarly bounded area).

We state the following result (whose proof is contained in [**SW2**]) concerning the existence and regularity of the Lagrangian Plateau Problem:

THEOREM 14.1. *Let Γ be a Jordan curve in R^4 such that $\int_\Gamma \eta = 0$. Then there exists a map $F : D \to R^4$, where D is the unit disk in R^2 such that $F|_{\partial D}$ is a $1-1$ parametrization of Γ. Also F satisfies:*

1. *F is lagrangian, i.e., $F^*\omega = 0$ and F has the least area among all lagrangian disks bounded by Γ;*
2. *F is weakly conformal;*
3. *F is Lipschitz in D and continuous in $\bar{D}$;*
4. *F is a smooth immersion except at a discrete set of points. Those singularities are either branch points or points at which $F(D)$ has a non-flat tangent cone described in Example 4 above.*

REMARK 14.2. *For the Plateau boundary condition, $\Sigma = F(D)$ is hamiltonian stationary, but will not be minimal since the mean curvature H does not vanish along the boundary Γ and hence is not an allowable variation for the problem. The condition that a curve in $\mathbb{R}^4$ bound a minimal lagrangian surface is much more restrictive than the condition that it bound a lagrangian disk.*

15. The Free Boundary Problem

As remarked above, a solution of the Lagrangian Plateau Problem need not be minimal. In order to produce minimal lagrangian surfaces, we need to consider more flexible boundary conditions. We describe one of these here.

Let S be a complex (real two-dimensional) surface in R^4 and Γ be a nontrivial (in the relative homotopy sense) curve on S. We call a surface Σ with boundary $\partial\Sigma \subset S$ *a solution to the free boundary problem with respect to* (S, Γ) if

$$\operatorname{Area}\Sigma = \inf\{\operatorname{Area}\Sigma_1 : \Sigma_1 \text{ lagrangian such that } \partial\Sigma_1 \subset S \text{ and is homotopy to } \Gamma\}.$$

LEMMA 15.1. *Let S be a complex surface in R^4 and Γ be a nontrivial (in the relative homotopy sense) curve on S. A smooth solution Σ for the free boundary problem with respect to (S, Γ) is minimal lagrangian.*

PROOF. Clearly since Σ is a solution to the free boundary problem with respect to (S, Γ), the first variation gives that

$$\int_\Sigma \langle X, H\rangle \, d\mu = 0,$$

for any vector field X along Σ such that X is tangent to S along $\partial\Sigma$. Let X be a hamiltonian vector field X_h. Since $H = J\nabla\beta$, where β is the lagrangian angle, we see that

$$\int_\Sigma \langle \nabla h, \nabla\beta\rangle \, d\mu = 0.$$

Integrating by parts and using $\triangle\beta = 0$, we have

$$\int_{\partial\Sigma} h\frac{\partial\beta}{\partial\nu}\, ds = 0,$$

where ν is the conormal vector along $\partial\Sigma$. Since S is complex, h can be arbitrary on $\partial\Sigma$ (i.e., any h defined on $\partial\Sigma$ can be extended to an ambient function on R^4 such that $J\nabla h$ is tangent to S along $\partial\Sigma$). We see that

$$\frac{\partial\beta}{\partial\nu} = 0.$$

Hence $\nabla\beta$ is tangent to $\partial\Sigma$, and in particular tangent to S. Since S is complex, $H = J\nabla\beta$ is tangent to S so H is an allowable variation. Therefore the first variation gives

$$\int_\Sigma \langle H, H\rangle \, d\mu = 0,$$

and Σ is minimal. □

REMARK 15.2. *The preceding argument shows more generally that if Σ is a compact lagrangian stationary submanifold, then we can conclude $H = 0$ using the mean curvature as the variational vector field. In other words, it does not use the minimizing property of Σ.*

REMARK 15.3. *The boundary regularity for this free boundary problem has been studied by Weiyang Qiu* [**Q2**] *in his Stanford Ph.D. dissertation.*

References

[A] Allcock, D., *An isoperimetric inequality for the Heisenberg groups*, GAFA, 8 (1998) 219–233.

[AM] Arezzo, C., Micallef, M., *Minimal surfaces in flat tori*, Geom. Funct. Anal. 10 (2000), 679–701.

[B] Bray, H., *Proof of the Riemannian Penrose conjecture using the positive mass theorem*, preprint.

[BS] Bray, H., Schoen, R., *Recent proofs of the Riemannian Penrose conjecture*, Current Developments in Math, Harvard Univ., 1999.

[CO] Chern, S. S., Osserman, R., *Complete minimal surfaces in euclidean n-space*, J. Analyse Math. 19 (1967), 15–34.

[F] Federer, H., *Geometric measure theory*, Springer, New York, 1969.

[Fu] Fujimoto, H., *Nevanlinna theory and minimal surfaces*, Encyclopedia of Math. Sci., 90, Springer, Berlin, 1997.

[G] Geroch, R., *Energy Extraction*, Ann. New York Acad. Sci. 224 (1973) 108–17.

[Gr1] Gromov, M., *Pseudoholomorphic curves in symplectic manifolds*, Invent. Math. 82 (1985), 307–347.

[Gr2] Gromov, M., *Carnot–Carathéodory spaces seen from within*, Progress in Mathematics 144 (1996), 79–323.

[GL1] Gromov, M., Lawson, H. B., *The classification of simply connected manifolds of positive scalar curvature*, Ann. of Math. 111 (1980), 423–434.

[GL2] Gromov, M., Lawson, H. B., *Positive scalar curvature and the Dirac operator on complete Riemannian manifolds*, IHES Publ. Math. 58 (1983), 83–196.

[HL] Harvey, R., Lawson, H. B. Jr, *Calibrated Geometries*, Acta. Math, 148 (1982), 47–157.

[HE] Hawking, S. W., Ellis, G. F. R., *The Large-Scale Structure of Spacetime*, Cambridge Univ. Press, Cambridge, 1973.

[HR] Hélein, F., Romon, P., *Weierstrass representation of Lagrangian surfaces in four-dimensional space using spinors and quaternion*, Comment. Math. Helv. 75 (2000), 668–680.

[HI] Huisken, G., Ilmanen, T., *The inverse mean curvature flow and the Riemannian Penrose Inequality*, J. Differential Geom. 59:3 (2001), 353–437.

[JW] Jang, P. S., Wald, R. M., *The positive energy conjecture and the cosmic censor hypothesis*, J. Math. Phys. 18 (1977) 41–44.

[L] Lawson, H. B., *Lectures on Minimal Submanifolds. Vol. I, Second edition*, Mathematics Lecture Series 9, Publish or Perish, Inc., Wilmington, Del., 1980.

[LB] LeBrun, C., *Yamabe constants and the perturbed Seiberg-Witten equations*, Comm. Anal. Geom. 5 (1997), 535–553.

[O] Osserman, R., *A survey of minimal surfaces*, Van Nostrand Reinhold Math. Study, 1969.
[M] Micallef, M., *Stable minimal surfaces in Euclidean space*, J. Diff. Geom. 19 (1984), 57–84.
[PT] Parker, T., Taubes, C., *On Witten's proof of the positive energy theorem*, Comm. Math. Phys. 84 (1982), 223–238.
[Pen] Penrose, R., *Naked singularities*, Ann. NY Acad. Sci. 224 (1973), 125–134.
[P] Petean, J., *The Yamabe invariant of simply connected manifolds*, J. Reine Angew. Math. 523 (2000), 225–231.
[Q1] Qiu, W. Y., *Non-orientable Lagrangian surfaces with controlled area*, Math. Res. Lett. 8 (2001), 693–701.
[Q2] Qiu, W., Stanford Ph.D. dissertation directed by R. Schoen, 2002.
[S1] Schoen, R., *Estimates for stable minimal surfaces in three-dimensional manifolds*, Seminar on Minimal Submanifolds, Ann. of Math. Stud., Princeton Univ. Press, Princeton, NJ, 1983, 111–126.
[S2] Schoen, R., *Variational theory for the total scalar curvature functional for Riemannian metrics and related topics*, Topics in calculus of variations (Montecatini Terme, 1987), Lecture Notes in Math., 1365, Springer, Berlin, 1989, 120–154.
[SS] Schoen, R., Simon, L., *Regularity of stable minimal hypersurfaces*, Comm Pure Appl. Math. 34 (1981), 741–797.
[SSY] Schoen, R., Simon, L., Yau, S. T., *Curvature estimates for minimal hypersurfaces*, Acta Math. 134 (1975), 275–288.
[SW1] Schoen, R., Wolfson, J., *Minimizing volume among Lagrangian submanifolds*, Differential equations: La Pietra 1996 (Florence), Proc. Sympos. Pure Math., 65, Amer. Math. Soc., Providence, RI, 1999, 181–199.
[SW2] Schoen, R., and Wolfson, J., *Minimizing area among Lagrangian surfaces: the mapping problem*, preprint.
[SY1] Schoen, R., Yau, S. T., *Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds with nonnegative scalar curvature*, Ann. of Math. 110 (1979), 127–142.
[SY2] Schoen, R., Yau, S. T., *On the structure of manifolds with positive scalar curvature*, Manuscripta Math. 28 (1979), 159–183.
[SY3] Schoen, R., Yau, S. T., *Complete three-dimensional manifolds with positive Ricci curvature and scalar curvature*, Seminar on Differential Geometry, Ann. of Math. Stud. 102, Princeton Univ. Press, Princeton, NJ, 1982, 209–228.
[SY4] Schoen, R., Yau, S. T., *On the proof of the positive mass conjecture in General Relativity*, Comm. Math. Phys. 65 (1979) 45–76.
[SY5] Schoen, R., Yau, S. T., *Proof of the positive mass theorem. II*, Comm. Math. Phys. 79 (1081), 231–260.
[SY6] Schoen, R., Yau, S. T., *The existence of a black hole due to condensation of matter*, Comm. Math. Phys. 90 (1983), 575–579.
[SY7] Schoen, R., Yau, S. T., *Positivity of the total mass of a general space-time*, Phys. Rev. Lett. 43 (1979), 1457–1459.
[SiY] Siu, Y. T., Yau, S. T., *Compact Kähler manifolds of positive bisectional curvature*, Invent. Math. 59 (1980), 189–204.
[St] Stolz, S., *Simply connected manifolds of positive scalar curvature*, Ann. of Math. 136 (1992), 511–540.
[Wic] Wickramasekera, N., Stanford Ph.D. dissertation directed by L. Simon, 2002.
[Wit] Witten, E., *A new proof of the positive energy theorem*, Comm. Math. Phys. 80 (1981), 381–402.

Department of Mathematics, Stanford University, Stanford, CA 94305, United States

E-mail address: schoen@math.stanford.edu

Clay Mathematics Proceedings
Volume 2, 2004

Introduction to Conjugate Plateau Constructions

Hermann Karcher

Abstract. We explain how geometric transformations of the solutions of carefully designed Plateau problems lead to complete, often embedded, minimal or constant mean curvature surfaces in space forms.

1. Introduction

The purpose of this paper is to explain to a reader who is already familiar with the theory of minimal surfaces another successful method to construct examples. This method is independent of the Weierstrass representation, which is the construction method more immediately related to the theory. This second method is called the *conjugate Plateau construction*, and we summarize it as follows: solve a Plateau problem with polygonal contour, then take its conjugate minimal surface, which turns out to be bounded by planar lines of reflectional symmetry (details in Section 2); finally use the symmetries to extend the conjugate piece to a complete (and, if possible, embedded) minimal surface. For triply periodic minimal surfaces in $\mathbb{R}^3$ this has been by far the simplest and richest method of construction. [KP] is an attempt to explain the method to a broader audience, beyond mathematicians. Large families of doubly and singly periodic minimal surfaces in $\mathbb{R}^3$ have also been obtained. By contrast, for finite total curvature minimal surfaces the Weierstrass representation has been much more successful.

Since *conjugate minimal surfaces* can also be defined in spheres and hyperbolic spaces (Section 2), the method has also been used there. However, in these applications the contour of the Plateau problem is *not* predetermined by the symmetry group with which one wants to work. Even in the simplest cases the correct contour has to be determined by a degree argument from a two-parameter family of (solved) Plateau problems. For spherical examples see [KPS]; for hyperbolic examples see [Po].

There is an even wider range of applications. In [La] two constant mean curvature one surfaces in $\mathbb{R}^3$ were constructed from minimal surfaces in $\mathbb{S}^3$. In [Ka2] it was observed that for large numbers of cases the required spherical Plateau *contours*, surprisingly, can be determined without reference to the Plateau *solutions* by using Hopf vector fields in $\mathbb{S}^3$. Then [Gb] added strings of bubbles to these examples by solving Plateau problems not in $\mathbb{S}^3$ but in mean convex domains such as the universal cover of the *solid* Clifford torus. He also obtained nonperiodic limits.

©2004 Clay Mathematics Institute

More generally, one can start with minimal surfaces in a spaceform $M^3(k)$ (having constant curvature k). From these it is possible to obtain in a similar way surfaces of constant mean curvature c in the space $M^3(k-c^2)$. However, one has the same problem explained for conjugate minimal surfaces: the required Plateau contours have to be chosen from at least two-parameter families and the choice depends on properties of the solutions. In those cases where conjugate minimal pieces have been obtained by solving degree arguments, for example in [KPS], [Po], a trivial perturbation argument shows that constant mean curvature deformations of the constructed minimal surfaces also exist.

Constant mean curvature one surfaces in $\mathbb{H}^3$, also called Bryant surfaces, have to be mentioned separately. They are obtained from minimal surfaces in $\mathbb{R}^3$. Again, the contours in $\mathbb{R}^3$ are not determined by the symmetries with which one wants the Bryant surface to be compatible. Nevertheless the situation still is a bit simpler than, for instance, the spherical or hyperbolic conjugate minimal surfaces, because one of the parameters of the $\mathbb{R}^3$-contours is just scaling. For example, to prove existence of Bryant surfaces having the symmetries of a Platonic tessellation of $\mathbb{H}^3$, the scaling parameter allowed success with a once iterated intermediate value argument instead of a general winding number argument [Ka3].

Frequently we will argue: "Take the Plateau solution and...". This may suggest difficulties which are not encountered. For our applications it is necessary to know more about the Plateau solution than just their existence. In particular, all our Plateau contours will be polygons in $\mathbb{R}^3$ or piecewise geodesic in $\mathbb{S}^3$ or $\mathbb{H}^3$. In many cases these solutions can be obtained as graphs over convex polygons with piecewise linear Dirichlet boundary data. It will be convenient to include also projections for which certain edges are projected to points. A reference for such cases is [Ni], where Dirichlet problems for graphs over convex domains are solved even if the boundary values have jump discontinuities. For example over two opposite edges of a square one can prescribe the value $-n$ and over the other pair the value $+n$. The Nitsche graph has then vertical segments over the vertices of the square as boundary, i.e., it is the Plateau solution of the polygon which has two horizontal edges at height $-n$, two horizontal edges at height $+n$ and four vertical edges of length $2n$. The fact that such Dirichlet problems have graph solutions implies that the tangent planes along the vertical edges have to rotate in a monotone way (otherwise the surface could not be a graph over the interior). This implies that the corresponding boundary arcs of the conjugate piece are convex arcs, a very helpful qualitative control of the conjugate piece. This work was extended in [JS], and has yielded sufficient conditions for including *infinite* Dirichlet boundary values. For example on a convex $2n$-gon with *equal* edge lengths (i.e., allowing angles $\leq \pi$) one can alternately prescribe the boundary values $+\infty, -\infty$. The conjugate pieces of these Jenkins–Serrin graphs generate a rich family of generalizations of Scherk's singly periodic saddle towers; see Section 2 and [Ka1, pp. 90–93].

This paper has three sections. In the first we do not yet use the conjugate minimal surfaces. We discuss the construction of complete embedded minimal surfaces by extending polygonally bounded minimal surface pieces, not necessarily of disk type. I concentrate on a new family of hyperbolic minimal surfaces in an attempt to show how quantitative facts about hyperbolic geometry are essential for existence and embeddedness. The second section repeats, for the convenience of the reader, those parts of minimal surface theory which are relevant for the conjugation

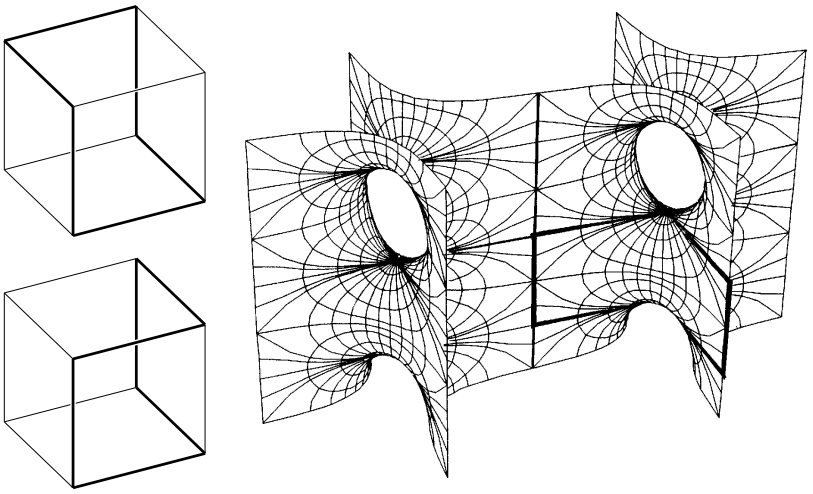


FIGURE 1. Left: the polygonal contours $P1$, $P2$ on the boundary of a cube (a special brick). Right: the contour $P2$ with its Plateau solution and an extension by 180° rotations around boundary edges. It is named Schwarz's CLP-surface.

constructions in space forms. Also, I try to indicate, how varied the applications to minimal surfaces in $\mathbb{R}^3$ are. In the third section I concentrate on constant mean curvature one surfaces in $\mathbb{R}^3$ since I feel one has to understand these before one can work on the less explicit problems mentioned in Section 2.

Acknowledgment. Karsten Große-Brauckmann read an earlier version of this paper and supplied a detailed list of the passages he found unnecessarily obscure. I have improved all criticized portions and I thank Karsten very much for his help.

2. Extension of Minimal Surfaces Across Boundary Segments

Minimal surfaces in $\mathbb{R}^3$. More than 50 years before the general Plateau problem was solved, solutions for certain polygonal contours were found by B. Riemann and by H. A. Schwarz, in terms of integrals of multivalued functions. In today's terminology they used the Weierstrass representation on nontrivial Riemann surfaces. The following particularly simple examples are due to H. A. Schwarz. Consider the following hexagonal polygons made of edges of a brick with edge lengths a, b, c (Figure 1, left):

$$P_1 : (0,0,0) \to (a,0,0) \to (a,b,0) \to (a,b,c) \to (0,b,c) \to (0,0,c) \to (0,0,0),$$

$$P_2 : (0,0,0) \to (a,0,0) \to (a,b,0) \to (a,b,c) \to (a,0,c) \to (0,0,c) \to (0,0,0).$$

In both cases the contour has a one-to-one convex projection (in fact many). The Plateau problem has therefore a unique solution, which is a graph over the interior of the chosen convex projection. Moreover, by the maximum principle, every compact minimal surface lies in the convex hull of its boundary. In particular, our Plateau solutions are inside the brick with edge lengths a, b, c. Imagine a black and white ("checkerboard") tessellation of $\mathbb{R}^3$ by these bricks, and imagine our Plateau solution to be in a black brick. The Schwarz reflection theorem says

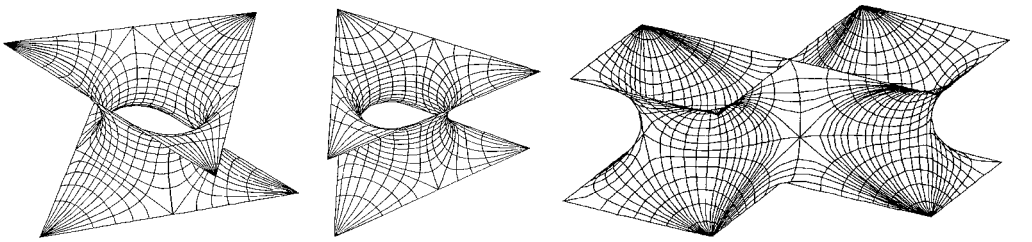


FIGURE 2. Three polygonally bounded minimal annuli in $\mathbb{R}^3$, one with noninjective boundary. The middle one is called Schwarz's H-surface.

that 180° rotation around any of the boundary edges of the Plateau piece gives an analytic continuation of the initial minimal surface piece. Because of the 90° angles of the polygons we can repeatedly extend across the edges which leave from one polygon vertex and thus obtain a smooth extension containing that vertex as an interior point. Notice that all the extensions are inside black bricks, and if the extensions lead us back into the first brick, the whole brick comes back in its original position. This shows that the extensions lead to embedded triply periodic minimal surfaces. The first one is called the *Schwarz D-surface* and the second the *Schwarz CLP-surface*; see [DHKW, vol. I, plate V(a)–(c)]. The names, given by A. Schoen, are well known in the crystallographic literature.

The crystallographers Fischer and Koch listed the crystallographic groups that contain enough 180° rotations so that from segments of the rotation axes polygons can be formed with the property that the Plateau solutions of these polygons extend to embedded triply periodic minimal surfaces. Their work includes examples of fairly complicated such polygons.

Minimal surfaces in spheres and hyperbolic spaces. Since the Schwarz reflection theorem extends to minimal surfaces in spheres and hyperbolic spaces, one can extend the above idea to space forms. Lawson's minimal surfaces in spheres [La] are constructed in this way from disk type Plateau solutions bounded by great circle quadrilaterals. I will now construct new embedded minimal surfaces in hyperbolic space which have compact *annular* fundamental domains. The purpose of these examples is to illustrate the use of barriers and of basic hyperbolic geometry. Euclidean analogues of such annuli, where the annulus is bounded by a pair of equilateral triangles in parallel planes, or a pair of squares in parallel planes, were already constructed by H. A. Schwarz; see [DHKW, vol. I fig. 22(a),(b),(d) and fig. 23(a),(b)]. For more complicated minimal annuli see [Ka2, pp. 342–343]. All these extend to embedded triply periodic minimal surfaces.

The existence construction. We have to deal with the following problem: if two circles in parallel planes are too far apart, then there is no catenoid annulus which joins them. The problem is dealt with by barriers. In $\mathbb{R}^3$ for example, assume that two convex polygons in parallel planes are so close together that a catenoid exists which meets only the interior of the two polygons. Then the two polygons are the boundary of a minimal annulus which surrounds the catenoid. In principle this would work also in hyperbolic space $\mathbb{H}^3$, but the meridians of hyperbolic catenoids

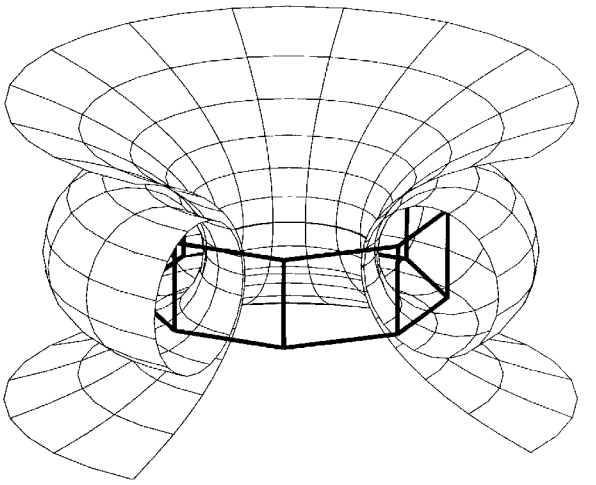


FIGURE 3. Catenoid, mean-convex torus and polygonally bounded annulus inside.

are given by differential equations, not by formulas in terms of well-known functions. Therefore one cannot readily decide whether some prism can surround a catenoid or whether it is too high and cannot. Instead we will rely on tori as (less optimal but more explicit) barriers. A torus of revolution has its mean curvature vector pointing into the solid torus if the radius of the meridian circle is less than half the soul radius. Such a torus is called *mean convex*.

Minimization techniques from calculus of variation can be applied to the set of surfaces which lie in a domain with mean convex boundary and which have the desired boundary. We can therefore choose any two different latitude circles of the torus and find an *annular minimal surface* (with those two latitudes as boundary) inside the solid torus. This minimal annulus together with the exterior planar domains of the two latitudes bounds another mean convex domain. In this mean convex domain we find a rich collection of minimal annuli: any two simply closed curves which lie in the exterior domains of the two latitude circles, and are homotopic to them, bound a minimal annulus which can be found by minimization in our mean convex domain. We apply this to the top and bottom polygon of the following regular and tessellating prisms: start with a regular geodesic n-gon, $n \geq 8$, with 90° angles, lying in a totally geodesic hyperbolic plane in $\mathbb{H}^3$. (There also exist hyperbolic five-, six- and sevengons with 90° angles but for those the torus barriers do not work.) Consider next the infinite prism orthogonal to the n-gon in $\mathbb{H}^3$. Cut this infinite prism above and below its symmetry plane so that the dihedral angles along the top and bottom rim are 60°. The following hyperbolic computation shows that the two rims of this prism lie in a mean convex torus (see Figure 3 above) and can therefore be joined by a minimal annulus (of course inside the convex prism). And analytic extension of this annulus by repeated 180° rotations around boundary edges produces a complete minimal surface which turns out to be embedded.

My reference for *hyperbolic trigonometry* is [Bu, pp. 31–42]. It is helpful to know that general trigonometric formulae simplify more than one expects from the

general expressions if one specializes to simple figures. To emphasize this I only quote and use two such special cases. First, for right triangles with edge lengths a, b, c and angles $\alpha, \beta, \gamma = \pi/2$ we have

$$\cosh c = \cosh a \cosh b = \cot\alpha \cot\beta,$$
$$\sinh a = \sin\alpha \sinh c = \cot\beta \tanh b,$$
$$\cos\alpha = \cosh a \sin\beta = \tanh b \coth c.$$

The formulas $\cos\alpha = \cosh a \sin\beta$, $\cosh c = \cot\alpha\cot\beta$ imply that a regular n-gon with $90°$ angles has inradius $a = r_i$ and outer radius $c = r_o$ given by

$$\cosh r_i = \frac{\cos(\pi/4)}{\sin(\pi/n)}, \qquad \cosh r_o = \cot(\pi/4)\cot(\pi/n).$$

Our second simple figure is the trirectangle, a quadrilateral with three angles $\pi/2$, edge lengths a, b, α, β and fourth angle ϕ between edges α, β. Here the general formulas simplify to

$$\cos\phi = \sinh a \sinh b = \tanh\alpha \tanh\beta,$$
$$\cosh a = \cosh\alpha \sin\phi = \tanh\beta \coth b,$$
$$\sinh\alpha = \sinh a \cosh\beta = \coth b \cot\phi.$$

Next consider the prism over the above $90°$ n-gon, which has at its top rim a dihedral angle of $\phi = \pi/3$. We determine its inner height h_i (the distance between the top and bottom plane) and its outer height h_o (the distance between the bottom n-gon and the top n-gon, i.e., the distance between the midpoints of the top and bottom edge of a vertical face). A plane through the (vertical) symmetry axis and the midpoint of a (horizontal) edge intersects the prism in a trirectangle with $a = h_i$, $b = r_i$, $\alpha = h_o$, $\phi = \pi/3$. With the formulas $\cos\phi = \sinh a \sinh b$ and $\sinh\alpha = \coth b \cot\phi$ we get

$$\sinh h_i = \frac{\cos(\pi/3)}{\sinh r_i}, \qquad \sinh h_o = \frac{\cot(\pi/3)}{\tanh r_i}.$$

Finally we have to check that the two $60°$ rims of the double prism lie in a mean convex domain as above. For this we choose the midpoint M of the meridian circle of the torus on the extension of the edge r_i of the trirectangle and at a distance $r_s = 1.1 \cdot r_i$ (the soul radius) from the symmetry axis. The meridian radius r_m is determined by $\cosh r_m = \cosh h_o \cosh(0.1 r_i)$ The condition for mean convexity of the torus was $r_s \geq 2r_m$ and this is satisfied for $n \geq 8$. Therefore we have the *existence of a minimal annulus* which is bounded by the top and bottom rim of our convex and *tessellating* hyperbolic prism. For fivegons to sevengons either this barrier computation is not good enough or the top and bottom rims are indeed too far apart for the minimal annulus to exist. Note another quantitative aspect of this computation: we do not have other families of such prisms, because the sum of the three dihedral angles at a top vertex of the prism has to be greater than π; if the sum of the dihedral angles is $= \pi$, the vertices of the prism are on the sphere at infinity, and if the sum of the dihedral angles is $< \pi$, the vertical edges do not meet the top face.

Analytic extension of the annular piece by $180°$ rotations. By repeated $180°$ rotation around boundary edges, the annular fundamental piece is analytically extended to a complete minimal surface. The embeddedness proof has two parts:

(i) the constructed annulus is embedded and (ii) the continuation does not create selfintersections. We omit the first part because the arguments are disjoint from the topic of this paper. For part (ii) we have to understand the tessellation of hyperbolic space by our prisms; such geometric discussions are always part of a construction of a complete surface from Plateau pieces. The idea is to color the prisms of the tessellation in red, green and blue so that the color changes across a face and 180° rotation around an edge of a prism maps that prism to one of the same color. If that can be achieved, then we have the complete surface contained only in the prisms of one color. Since along each edge only two prisms of the same color meet we have avoided self-intersections. We start by making the first prism red, the two neighbors above and below we make green and the n neighbors across vertical faces we make blue. Next we describe how the 24 prisms meet which have one vertex in common. Consider how a small sphere around that vertex meets the adjacent prisms: each prism intersects the sphere in a geodesic triangle whose angles are the dihedral angles at the three edges of the prism at that vertex, i.e., the angles are $\pi/2, \pi/3, \pi/3$. Four such triangles around the $\pi/2$-corner fit together to a spherical square with angles $2\pi/3$. Six such squares tessellate the sphere; it is the same tessellation obtained by central projection of a cube to its circumsphere. The colors of the prisms which meet at one vertex are by definition the same as the colors of the 24 triangles of the spherical tessellation and vice versa. Initially we colored four prisms at each vertex of the first prism. Now consider the spherical triangulation which describes the neighborhood of one vertex of the prism; one can view it on a cube after subdividing each face by its diagonals into four triangles. Our coloring prescription says that opposite triangles on one face have the same color, since they correspond to two prisms whose position differs by a 180° rotation about a 90° edge. Also, each triangle has neighbors of both other colors. These two observations imply that the different colors of two neighboring triangles determine the colors of all the other triangles on the cube uniquely. Therefore we can extend the coloration to all the prisms which meet a vertex of the first prism. We continue the coloration and because of the unique extension of a partial coloration to the full sphere we can assign a unique color to each prism and thus complete the proof.

Further remarks. Minimal annuli in $\mathbb{S}^3$ bounded by two geodesic quadrilaterals and extending to complete embedded minimal surfaces by repeated 180° rotations around boundary edges are described in [Ka2, p. 344]. The complete surface is invariant under a rotation of $\mathbb{S}^3$ which has two orthogonal great circles as axes. Near each axis this $\mathbb{S}^3$-rotation looks like a screw motion in $\mathbb{R}^3$ and the constructed surface can therefore be thought of as a generalization to $\mathbb{S}^3$ of the twisted Scherk saddle towers (see [Ka1, pp. 94–99] and Figure 2 below).

The unbounded Dirichlet problems over convex domains from [JS] yield doubly periodic embedded minimal surfaces which are graphs over the full plane minus parallel lines, see below. (The projection of Scherk's doubly periodic surfaces covers only half the plane.)

A list of the known *disk type* Plateau solutions for polygonal boundaries, which extend to *embedded* minimal surfaces, is rather short. However, we see in the next section that the properties of conjugate minimal surfaces offer very flexible possibilities for the construction of embedded minimal surfaces. We emphasize that the conjugate surface method of the next section requires disk type minimal surfaces; the construction of embedded minimal surfaces from annular (or still higher genus)

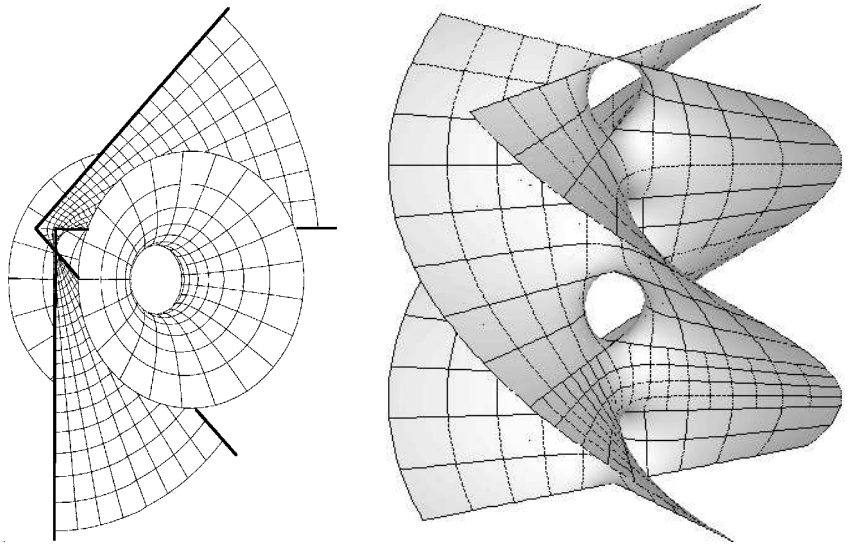


FIGURE 4. Existence proof of the twisted Scherk tower. The surface to the right is generated by a fundamental domain which is a strip (left) between two broken lines (here, each broken line consists of two half-lines which meet under 90°). One can make a mean convex domain from the two quarter planes spanned by the broken lines and from two congruent pieces of a helicoid whose axis is the segment between the corners of the quarter planes. The strip between the broken lines is to be constructed as a limit of Plateau solutions which look like the left picture: each pair of half-lines is connected by an arc in the mean convex domain, preferably by the helices on the two barrier helicoids. Moving these connecting arcs outward gives a monotone family of minimal surfaces (the earlier ones are barriers for the later ones). Finally, the shown catenoid is the barrier which prevents this family from converging against the two quarter-planes.

fundamental domains, which we have seen in this first section, is not compatible with the definition of the conjugate surface.

3. Conjugate Minimal Surfaces

Some basic theory. When the theory of minimal surfaces developed in the nineteenth century, it was realized early that the three coordinate functions of a *conformal* parametrization of a minimal surface are the *real parts of holomorphic functions*! Since the existence of conformal parametrizations of surfaces in $\mathbb{R}^3$ is a hard theorem which is rarely completely explained in differential geometry courses, we observe that the above discovery can also be stated without a conformal parametrization. If one imagines an atlas of conformal coordinates for the surface, then it makes sense to multiply a tangent vector by the complex number i. But this multiplication by i is, on each tangent space, the positive 90° rotation, and this is

a geometric description of the multiplication by i which does not use a conformal parametrization. The endomorphism field of tangent space wise 90° rotations is therefore also called the *complex structure* and is denoted by J. A differentiable map from the surface to the complex numbers, $f : M^2 \to \mathbb{C}$ is then *holomorphic* if its differential Tf satisfies $Tf(J \cdot X) = i \cdot Tf(X)$ for each tangent vector X. We write f in terms of its real and imaginary parts, $f = u + i \cdot v$. Holomorphicity of f then can be expressed as $Tu(J \cdot X) = -Tv(X)$, or $Tv(J \cdot X) = Tu(X)$. These *Cauchy–Riemann equations* say that the differential of the imaginary part can be computed from the differential of the real part and J, namely $Tv = -Tu \circ J$. Moreover, just as in $\mathbb{C}$, a differentiable function u is (locally) the real part of a holomorphic function if and only if the differential form $\omega := -Tu \circ J$ is closed. Therefore we can now formulate, without reference to a conformal parametrization, the mentioned discovery of the 19th century, namely that the coordinate functions of minimal surfaces are, locally, real parts of holomorphic functions. To see why this is true requires the use of the *surface equations*, which we are about to state.

Let $F : M^2 \to \mathbb{R}^3$ be a local immersion, $N : M^2 \to \mathbb{S}^2$ the normal Gauss map, S the shape operator, $g(X, Y) = \langle TF(X), TF(Y)\rangle_{\mathbb{R}^3}$ the Riemannian metric, Γ the Christoffel map (or symbol) defined by $T^2F(X, Y)^{\text{tang}} = TF \circ \Gamma(X, Y)$, ∇ the covariant derivative of the metric g and K its curvature. Then the following equations hold:

Surface equations for TF, N with data $\{g, S\}$.

Weingarten equation $\quad TN = TF \circ S,$

Gauss equation $\quad T^2F(X, Y) = TF \circ \Gamma(X, Y) - g(S \cdot X, Y)N,$

Covariant version $\quad \nabla^2 F(X, Y) := T^2F(X, Y) - TF \circ \Gamma(X, Y) = -g(S \cdot X, Y)N.$

Integrability conditions.

Symmetry $\quad g(S \cdot X, Y) = g(S \cdot Y, X),$

Codazzi equation $\quad \nabla_X S \cdot Y = \nabla_Y S \cdot X,$

Gauss equation $\quad \det S = K,$

Minimality condition $\quad \operatorname{trace} S = 0.$

The close connection between minimal surfaces and holomorphic maps. The minimal and self-adjoint shape operator S has, with respect to an orthonormal basis, a trace free and symmetric matrix $\left(\begin{smallmatrix} a & b \\ b & -a \end{smallmatrix}\right)$ This implies $-K = -\det(S) = a^2 + b^2$, hence $S^2 = -K \cdot \mathrm{id}$. This proves *the conformality of the normal Gauss map* since scalar products change only by the scaling factor $-K$:

$$\langle TN(X), TN(Y)\rangle = \langle TF(S \cdot X), TF(S \cdot Y)\rangle = g(S \cdot X, S \cdot Y) = -K \cdot g(X, Y).$$

In particular, we obtain the meromorphic Gauss map G if we compose the normal Gauss map N (which is orientation reversing but angle preserving) with an orientation reversing (and angle preserving) stereographic projection $\mathrm{St} : \mathbb{S}^2 \to \mathbb{C}$ as follows $G := \mathrm{St} \circ N$.

The complex structure J has with respect to any orthonormal basis the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. This and the matrix of S (which was implied by minimality and symmetry) give $J \circ S = \left(\begin{smallmatrix} -b & a \\ a & b \end{smallmatrix}\right)$. This implies first that $S^* := J \circ S$ is again symmetric and has trace zero. Secondly, $\det J = 1$ implies that the Gauss equations for S and S^* are

equivalent. And finally $\nabla J = 0$, that is, $\nabla S^* = J \cdot \nabla S$, implies that the Codazzi equations for S and S^* are also equivalent. *Therefore g, S^* satisfy the integrability conditions and define another minimal immersion F^** (of simply connected pieces or coverings). These two pairs of surface data, $\{g, S\}$ and $\{g, S^*\}$, belong to minimal immersions F, F^* which fit together as real and imaginary parts of a holomorphic map $F + i \cdot F^*$ for the following reason. We rewrote the second surface equation in its covariant form. This shows immediately that the covariant derivative of the ($\mathbb{R}^3$-valued) 1-form TF, which is equal to $-g(S \cdot X, Y)N$, has trace 0 so that *the three coordinate functions F^j are Riemannian harmonic.* Together with $\nabla J = 0$ the covariant surface equation shows further that the derivative of the ($\mathbb{R}^3$-valued) 1-form $TF \cdot J$, which is equal to $-g(S \cdot J \cdot X, Y)N$, is again symmetric (equal to $-g(X, S \cdot J \cdot Y)N$), i.e., the exterior derivative of $TF \circ J$ is 0. Therefore $TF \circ J$ is (on simply connected domains again) the derivative of some other map. If we define F^* by $TF^* := -TF \circ J$, then we have proved that the surface equations for $\{TF, N\}$ with data $\{g, S\}$ imply immediately that $\{TF^*,\ N^* := N\}$ are solutions for the surface equations with data $\{g, S^*\}$. F^* is called the *conjugate minimal immersion.* Moreover, $F + iF^* : M^2 \to \mathbb{C}^3$ *is not only differentiable but even holomorphic* because the Cauchy–Riemann equations $T(F+iF^*)\cdot J = i \cdot T(F+iF^*)$ are satisfied. For numerical computations it has been extremely convenient that F^* can be obtained by one integration from *first* derivative data, namely from $TF^* := -TF \circ J$. (The *second* order surface data $\{g, J \circ S\}$, which are numerically more difficult to obtain, determine the surface via an ODE.)

Extension of conjugate minimal surfaces to $\mathbb{S}^3$ and $\mathbb{H}^3$. Some of these observations carry over to minimal surfaces in spheres $\mathbb{S}^3(c^2)$ or hyperbolic spaces $\mathbb{H}^3(-c^2)$ of curvature c^2 and $-c^2$, respectively. One cannot speak of harmonic coordinate functions of minimal surfaces in these spaces. But the surface equations and the integrability conditions are almost the same. One only has to interpret N as a unit normal field along the immersion F, then the surface equations hold. The first two integrability conditions stay the same and the Gauss equation needs only a small adjustment:

Gauss equation in $M^3(k)$ $\qquad\qquad k + \det(S) = K.$

Therefore: if $\{g, S\}$ are minimal surface data in a space $M^3(k)$ of constant curvature k, then $\{g, (\mathrm{id}\cos\alpha + J\sin\alpha)\cdot S\}$ is a one-parameter family of isometric (and in general noncongruent) minimal surface data in $M^3(k)$, the so-called *associate family.*

Extension to constant mean curvature surfaces. In fact, minimal surface data $\{g, S\}$ in one space form $M(k)$ provide constant mean curvature $\pm c$ surface data $\{g, S \pm c \cdot \mathrm{id}\}$ in another space form $M(k - c^2)$. I learnt this from [La]; I am told Lawson heard this from Calabi and in any case, it is immediate from the surface equations and their integrability conditions since $\det(S \pm c \cdot \mathrm{id}) = \det(S) + c^2$. We will use this in Section 3.

Symmetry lines of minimal surfaces. Before we can exploit this geometric transformation of a simply connected minimal surface to its conjugate minimal surface, we need one more piece of geometric information. The usual Frenet theory of curves in 3-dimensional space forms fails at points where the curve has curvature $= 0$; in particular it cannot handle geodesics. This problem goes away for curves

on surfaces. Given a unit speed curve γ in the domain of a parametrized surface $F : D^2 \to M^3$ with normal field $N : D^2 \to TM^3, N(p) \perp \text{image}(TF_p)$. We then choose as frame along the image curve $c := F \circ \gamma$ the tangent field $e_1 := TF(\dot{\gamma})$, the conormal field $e_2 := TF(J \cdot \dot{\gamma})$, and the surface normal field $e_3 := N \circ \gamma$. We then have, with geodesic curvature κ_g, normal curvature $k_n = \langle \dot{e}_3, e_1 \rangle = g(S\dot{\gamma}, \dot{\gamma})$ and normal torsion $\tau_n = \langle \dot{e}_3, e_2 \rangle = g(S\dot{\gamma}, J \cdot \dot{\gamma})$:

Frenet equations for curves on surfaces.

$$\dot{e}_1 = \kappa_g \cdot e_2 - k_n \cdot e_3, \qquad \dot{e}_2 = -\kappa_g \cdot e_1 - \tau_n \cdot e_3, \qquad \dot{e}_3 = k_n \cdot e_1 + \tau_n \cdot e_2;$$

Data on the conjugate surface $\qquad \kappa_g^* = \kappa_g,\ k_n^* = -\tau_n,\ \tau_n^* = k_n.$

We discuss these equations for geodesics, i.e., $\kappa_g = 0$. Curves are principal curvature lines if and only if $\tau_n = 0$ and asymptote lines (vanishing normal curvature) if and only if $k_n = 0$. Observe that a principal curvature line on a minimal immersion is an asymptote line on the conjugate immersion and vice versa. But a geodesic asymptote line has vanishing tangential and normal curvature, hence is even a geodesic in the space form $M(k)$. Consider next a geodesic curvature line; the Frenet equations show (i) that the surface normal $N \circ \gamma$ is the principal curvature normal of the curve and (ii) that its torsion in $M(k)$ is 0, i.e., such a curve is planar (or lies in a 2-dimensional totally geodesic subspace). Finally, for minimal surfaces in any space form we have:

Reflection principle for minimal surfaces in $M^3(k)$.

(a) 180° rotation around a geodesic asymptote line (in fact a geodesic in $M^3(k)$) is a congruence of the minimal surface.
(b) Reflection in the plane of a geodesic principle curvature line is a congruence of the minimal surface.
(c) Plateau solutions in polygonal contours are sufficiently regular at the boundary so that the symmetry from (a) can be used to analytically extend the Plateau piece across each boundary segment. If the conjugate surface is considered, then this extension is transformed into the symmetry (b).

Embeddedness criterion. The integration of the surface equations usually does not say whether the obtained immersion is in fact an embedding. The following result is an easily applied criterion which covers many interesting cases.

R. Krust's conjugate graph theorem in $\mathbb{R}^3$. *If a minimal surface is a graph over a convex domain, then all surfaces of the associate family are graphs (usually not over convex domains) and hence embedded.* See [DHKW, pp. 118–119].

Some singly periodic examples in $\mathbb{R}^3$. (I give more details for the triply periodic case.) The conjugates of Jenkins–Serrin graphs over *equilateral* convex $2n$-gons were already quoted in the introduction. Being equilateral is a trivial sufficient condition under which the results of [JS] can be applied. And the fact that the strips between the vertical lines of the graph all have the same width implies that, on the conjugate graph, all the pairs of horizontal symmetry lines have the same vertical distance. These conjugate patches are fundamental domains for a rich family of

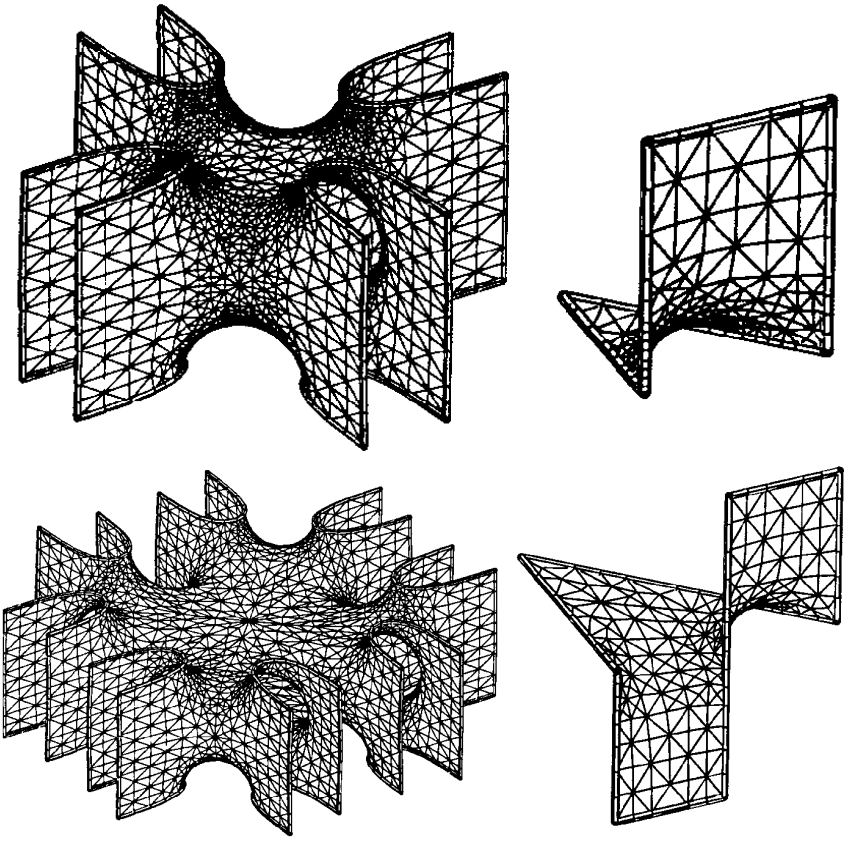


FIGURE 5. Examples of saddle towers with many wings. To obtain the singly periodic examples to the left one has to imagine the polygonal contours to be infinitely high and get minimal graphs with [JS]. The conjugate of such a patch is a fundamental piece for the saddle tower to its left. In the first case one can extend the Jenkins–Serrin piece by repeated rotation around its two horizontal edges to a graph over a square, then reduce the height of two adjacent strips from $\pm\infty$ to $\pm n$. This is still a Jenkins–Serrin contour. The conjugate of its minimal graph generates a saddle tower where one pair of wings has finite length. Two such surfaces fit together along the pair of symmetry lines between them after one adjusts one parameter, see text below.

deformations of the singly periodic Scherk saddle towers. Embeddedness follows from Krust's theorem.

Each pair of adjacent saddles of the most symmetric of these saddle towers were connected by a "catenoid like" handle in [Ka1, p. 107–110], giving toroidal saddle towers.

The Jenkins–Serrin criterion allows certain angles of the convex polygon to be equal to π. Corresponding neighboring wings of the saddle tower are then parallel. This suggests to "cut one pair of such parallel wings off, modify the cuts to be

planar symmetry lines and reflect to get other toroidal saddle towers". Technically this is done by replacing the boundary values $\pm\infty$ of the pair of edges with angle π between them by finite boundary values $\pm n$. The boundary symmetry lines corresponding to the two horizontal edges at height $\pm n$ are no longer straight. Therefore they connect two horizontal symmetry lines the distance of which is *less* than the distance between all other neighboring horizontal symmetry lines (which are all equal to the edge length of the equilateral polygon). This has to be remedied by making the edges with finite Dirichlet data a bit longer. If we make the convex polygon *symmetric* (with respect to the symmetry line of the angle π) then this is a one-parameter problem which is solved with the intermediate value argument, effortless if we permit ourselves to take n very large. The following figures are taken from [KP, p. 2098]; these surfaces are computed by Polthier as discrete minimal surfaces in the sense of Pinkall and Polthier. They suggest further modifications: instead of double connections one can take triple connections or more. These higher parameter problems have not been treated.

Conjugate construction of some doubly periodic examples in $\mathbb{R}^3$. The top part of Figure 5, taken from [KP, p. 2098], also suggests one-parameter problems to obtain doubly periodic surfaces. We only have to take two such pairs of edges with angle π in symmetric position, for instance by subdividing a rectangle with edge lengths $2+\epsilon$ and k. All the neighboring infinite horizontal symmetry lines then have vertical distance 1 and the distance of the finite horizontal symmetry lines from their neighbors is controlled by ϵ. Higher parameter versions of this idea, suggested by the bottom figure, have not been treated. Another possibility is to increase for instance the parameter a in the contours P_1, P_2 of Section 1; see [Ka1, pp. 102–107]. Also, the idea of gluing catenoid like handles into already known doubly periodic surfaces has succeeded, for one handle in Scherk's doubly periodic surface; see [HKW, pp. 138–141]. Many handles have been constructed by Wei, Thayer, Wohlgemuth or Weber (mostly unpublished work), generally using the Weierstrass representation.

Description of many triply periodic examples in $\mathbb{R}^3$. Here already the one-parameter possibilities are richer than I can cover. I begin the detailed explanations with the contour P_1 of Section 1 which needs no parameter adjustment. The disk type Plateau solution has a hexagonal polygon as boundary, i.e., a boundary made up of six geodesic asymptote lines. The conjugate simply connected piece is therefore bounded by six geodesic principal curvature lines and one can analytically extend the piece by reflection in the planes of these boundary arcs. At each vertex the two boundary arcs meet with a $\pi/2$-angle; repeated reflections in the arcs which leave from one vertex therefore extend the piece to a larger one with the vertex as a smooth interior point. Now we consider the six symmetry planes of the boundary arcs and we claim that they are the boundary planes of a brick. For this it is crucial that a minimal surface and its conjugate have the same Gauss map. If we orient the contour P_1 so that the edges are parallel to the coordinate axes in $\mathbb{R}^3$, then it follows that the symmetry planes of the conjugate piece are orthogonal to the coordinate axes. Since this nonplanar conjugate piece also sits in the convex hull of its boundary, we have that the symmetry planes are indeed the boundary planes of a brick which contains the six symmetry arcs on its six faces and the interior of the conjugate piece in its interior.

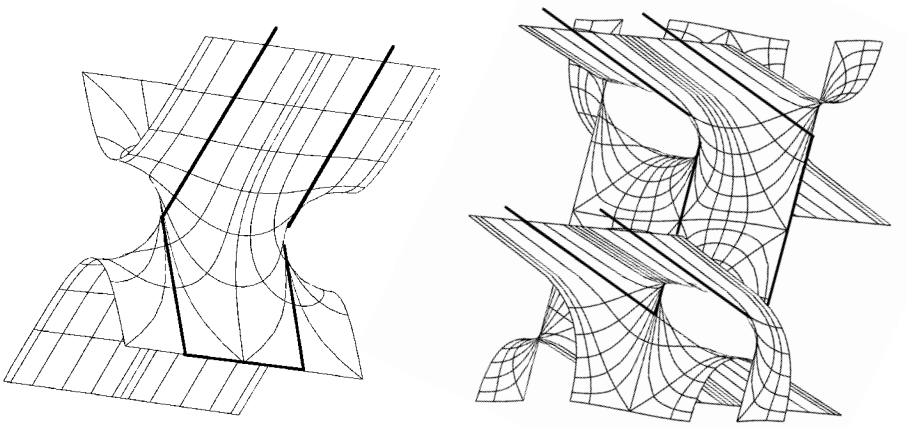


FIGURE 6. Two minimal surfaces that are analytically continued from Jenkins–Serrin graphs. Left: The Dirichlet data on the edges of a rectangle are $0, 0, 0, \infty$. Right: The Dirichlet data on the edges of a rectangle are $0, \infty, 0, \infty$. The conjugate surfaces of both of these are also doubly periodic embedded minimal surfaces. If ∞ is replaced by a finite height, one obtains triply periodic surfaces.

The next claim is that each of the six boundary arcs has a monotonically rotating normal and the total rotation is $\pi/2$, i.e., each is one quarter of a closed convex curve with two orthogonal symmetry axes. We look again at the polygonal Plateau contour. It follows from [Ni] that the interior is a graph over the interior of the convex projection of the boundary *also in the limiting case where certain edges project to a point.* This implies that the tangent planes of the Plateau piece along such a vertical edge must rotate monotonically since otherwise the interior could not be a graph. And since the Plateau piece is also in the convex hull of its boundary this monotone rotation can only be through $\pi/2$, which is the angle of the projected contour (at the vertex to which the vertical edge projects). This proves the claim since the normal rotates through the *same angle* along the boundary arc of the conjugate piece.

Now we have the complete picture: eight of the conjugate pieces fit together to a translational fundamental domain of the complete surface. This larger building block sits in a brick with twice the edge lengths of the previous brick (around the conjugate piece); and each face of this brick is met by a closed convex symmetry line of the minimal surface. A. Schoen named the member of this family with cubical symmetry *Schwarz P-surface* (where the P is referring to the primitive cubical lattice). See [DHKW, fig. 22(a)–(c), plates II(b), V(d)] and Figure 7 left.

Summary of similar examples in $\mathbb{R}^3$. In the same way one can have minimal surfaces which meet all the faces of prisms over a regular hexagon or over a regular triangle in convex curves; see [DHKW, plate III(a)–(b)]. But these two surfaces are in fact only different views of the same surface; see the illustration in [Ka2, p. 298]. The conjugate contour of a fundamental piece consists of five edges

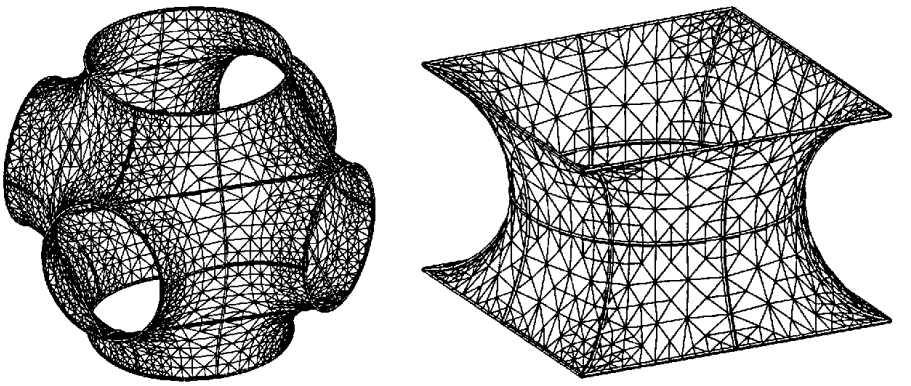


FIGURE 7. Left: Translational fundamental domain of a deformation of Schwarz P-surface (the top hole is bigger!). Right: The minimal annulus between parallel squares is part of this P-surface. Again, these are discrete minimal surfaces by Polthier.

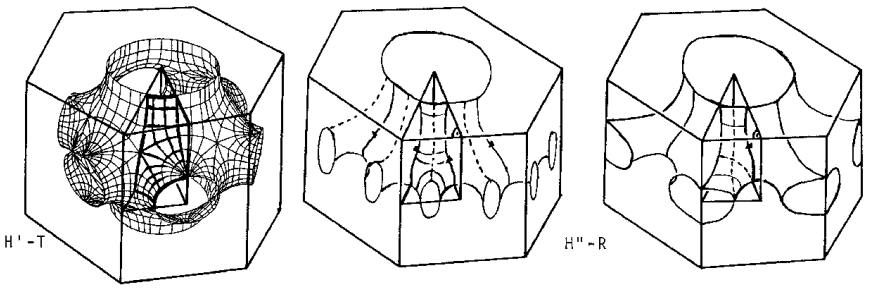


FIGURE 8. At the left and right are translational fundamental domains of two minimal surfaces of A Schoen, named H'-T- and H''-R-surface. The middle one is a one-parameter modification. (Conjugate contours for these surfaces are given below.)

of the prism over a $(30, 60, 90)$-triangle; see [Ka2, p. 330] and below. A. Schoen named it the *H'-T-surface.*

The rhombic dodecahedron is another tessellating polyhedron; the conjugate construction with the other contour on [Ka2, p. 330] gives a minimal surface which meets all the faces of the rhombic dodecahedron in convex curves. A. Schoen named it *F-Rd-surface.* In all these cases we call the catenoid-like connections to the neighboring crystallographic cells *Schwarz-handles.* In what follows the emphasis is on the construction of minimal surfaces which *combine* features of better known simpler surfaces. The term Schwarz-handle is meant to direct the attention to one such feature.

E. R. Neovius, a student of H. A. Schwarz, has constructed the following minimal surface with a translational fundamental domain in a cube. It is connected

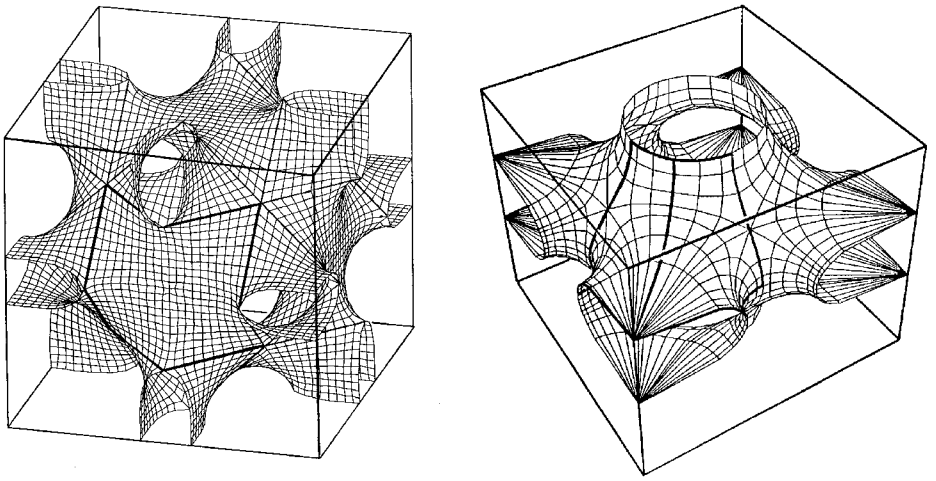


FIGURE 9. Neovius minimal surface with cubical symmetry carries the same lines as Schwarz P-surface. To the right is A. Schoen's mixture of Schwarz handles and Neovius handles, his S'-S''-surface.

with the neighboring crystallographic cells by twelve "arms" crossing the midpoints of the edges; one can also say, a handle like a thickened cross connects the surface pieces in the four cubes around one edge. We call this connection a *Neovius-handle.* Conjugate contours are discussed in [Ka2]; for more illustrations see [DHKW, plate VII(b)] and [Ka2, p. 300]. A. Schoen's name for the *Neovius-surface* is $C(P)$-*surface.*

Also without parameter adjustments of the Plateau contours we can obtain minimal surfaces which continue to meet the top and bottom face of a prism over a square, over a regular hexagon or over a regular triangle in closed convex curves (Schwarz-handles), but the vertical edges are met by (4-fold, 3-fold, 6-fold, respectively) Neovius handles. The conjugate contours consist of five edges of the prism over an $(\alpha, \pi/2 - \alpha, \pi/2)$-triangle with $\alpha = \pi/4, \pi/3, \pi/6$; see [Ka2, p. 308]. For illustrations see [DHKW, plates IV(e)] and [Ka2, p. 299]. Schoen's names are S'-S''-, H''-R-, T'-R'-*surfaces.*

We mentioned the annular minimal surface, *Schwarz's H-surface*, bounded by two parallel regular triangles. A translational fundamental domain of this surface meets the top and bottom face of a hexagonal prism in closed convex curves and every *second* vertical edge is met by a 3-fold Neovius-handle. A contour for a conjugate construction is in [Ka2, p. 328]. For more illustrations see [DHKW, plate VI(a)–(d)] and [Ka2, p. 297].

Another way to fit a minimal surface into a crystallographic cell does not occur in Schwarz's school, the first example is A. Schoen's *I-WP-surface.* The translational fundamental domain in a cube sends arms towards all the eight vertices so that each arm is cut by the cube in three convex arcs on the faces around one vertex of the cube. The conjugate contour is in [Ka2, p. 331], illustrations are in [DHKW, plate VII(a)] and [Ka2, p. 297]. We call the handle *I-WP-handle.* This

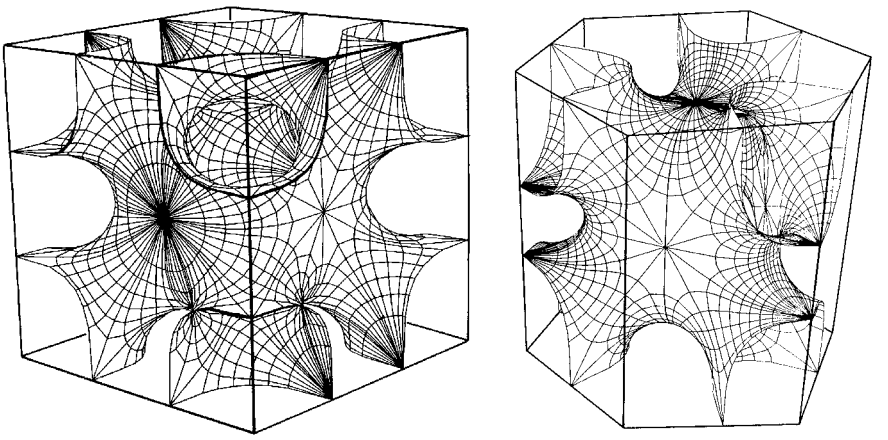


FIGURE 10. Two triply periodic minimal surfaces in crystallographic cells of $\mathbb{R}^3$. Left: Schoen's I-Wp-surface. Right: A similar hexagonal one. The conjugate contours are quadrilateral and hexagonal polygons.

conjugate contour depends on an angle α, with $\alpha = \pi/4$ for Schoen's surface and with $\alpha = \pi/3$ for a surface whose translational fundamental domain sits in a hexagonal prism with Schwarz-handles to the top and bottom faces and 3-fold Neovius handles to every *second* vertex. Illustrations: [DHKW, plate VI(e)], [Ka2, p. 297].

Many more examples in $\mathbb{R}^3$ via one-parameter adjustments. From now on we consider the previous examples as construction material and we want to use them for more complicated examples, but we only want to solve one-dimensional intermediate value problems. I cannot exhaust the possibilities, and it will also be clear that with one-parameter problems one should expect to find a huge collection of further examples. In [Ka2, pp. 350–356] I have given fifteen one-parameter contours designed to mix the handles we have seen so far. The paper [KP] was written to explain the conjugate Plateau construction to a wider audience and with Polthier's discrete minimal surface programs we also illustrated this mixing of handles: [KP, p. 2099] shows how I-WP-handles grow out of a Schwarz-P-surface until they are long enough to reach the faces of the cube around the P-surface; the result is a surface which combines the handles of the Schwarz-P- and the Schoen-I-WP-surfaces. In [KP, p. 2100] we combine the handles of the I-WP- and the Neovius surface and on p. 2101 we combine the handles of the Schwarz-P- and the Neovius surface. The contours given in [Ka2] also add to the Schwarz-handles of Schoen's F-Rd-surface either I-WP-handles to all the rhombic dodecahedron's 4-valent vertices or to all its 3-valent vertices. Other contours give (i) both types of I-WP-handles to all vertices of the rhombic dodecahedron (the Schwarz-handles are omitted since they require a further parameter) or (ii) Neovius-handles to all edges of the rhombic dodecahedron (further handles need more parameters to be adjusted).

Another modification is illustrated in [KP, pp. 2096–2097]. We insert vertical catenoid-like handles into all the horizontal Schwarz-handles of the P-surface (2096)

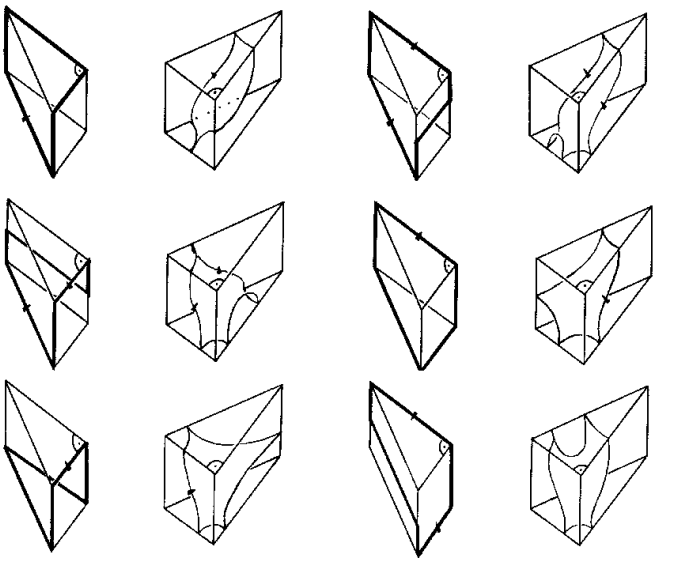


FIGURE 11. Illustration of intermediate contours. On the boundary of a 30°-60°-90°-prism we see six polygonal contours and to the right of each we see a sketch of the conjugate minimal patch in the same 30°-60°-90°-prism. These conjugate patches can be extended to triply periodic complete embedded minimal surfaces by repeated reflection in the faces of the prism. The three pentagonal contours lead to surfaces of A. Schoen without period killing. The three hexagons have one horizontal edge on a face of the prism and one has to use the intermediate value theorem to find the correct height of this edge. The contour to the right in the first row gives the surface between Schoen's H'-T- and H''-R-surfaces above. The other two intermediate contours add vertical handles towards the horizontal faces of the hexagonal prism.

or into all the horizontal Schwarz-handles of the H'-T-surface (see the contours in Figure 11). This works because the conjugate contours allow us to define one-parameter families of contours which give already known minimal surfaces (without parameter adjustments) at the endpoints of the parameter range. In the family we have, in general, one undesired period, a symmetry arc along which the normal rotates through 180°, but the parallel normals at the endpoints are not on the same *line*. On both parameter boundary values the normal of that arc rotates only through 90°, i.e., this symmetry arc converges (at the parameter boundary) to a *rising*, resp. *falling*, convex arc. The intermediate value theorem applies, giving one parameter value where also the symmetry line with the 180° rotating normal closes to a convex curve. The idea of adding Schwarz-handles at suitable places can often be accomplished by adjusting just one parameter for the length of the handle.

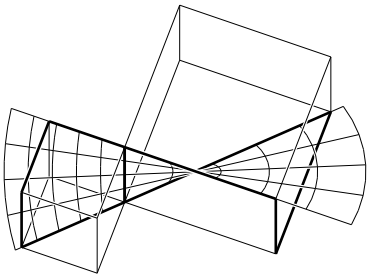


FIGURE 12. Combining two contours. The Plateau solution for the pentagon in the smaller (left) quadratic prism is conjugate to a piece of Schwarz's P-surface, and the Plateau solution for the pentagon in the larger prism is conjugate to a piece of Schoen's S-S'-surface. If one removes the common edge of the two, then one obtains a contour whose Plateau solution is a candidate for the surface which looks like the one obtained by stacking the two prisms (one with a P- the other with an S-S'-surface) on top of each other. Note that the new contour still has one convex projection (in the direction of one edge of the squares) so that its Plateau solution is a Nitsche graph [Ni], in particular unique and hence depending continuously on the edge lengths. The picture also shows a helicoid (with the removed edge as axis) that is a barrier for this contour. Such barriers can be used if more precise control of the Plateau solution is needed.

One can also insert a 4-fold Neovius-handle into the vertical necks of the Schwarz-P-surface. This splits each vertical catenoid like neck into four thinner parallel ones, [KP, p. 2104]. Analogous modifications apply to other surfaces of the "construction material".

In all the highly symmetric crystallographic prisms we have seen surfaces with vertical Schwarz handles but with *different* types of horizontal handles (Schwarz- and Neovius-handles). In the quadratic prism we have Schwarz's P-surface and Schoen's S-S'-surface; in the hexagonal prism we have Schoen's H'-T- and H''-R- and Schwarz's H-surfaces and similar examples exist in the prisms over equilateral triangles.

One observes that the convex symmetry lines on the top and bottom faces of the prisms are almost circles, see also [KP, pp. 2096–2097]. This suggests to glue one of these on top of the other, minimally of course. Figure 12 explains why this is an easy task for the conjugate contour: just put the two contours together along the edge which corresponds to the horizontal symmetry line; then omit that edge. One problem remains to be solved by a one-parameter adjustment: the two types of horizontal handles must have the same length, i.e., their vertical symmetry planes (which are by construction parallel) must coincide. If the common height of the two prisms converges to zero, the Plateau solutions converge to the union of two squares, and so do their conjugate patches. This says that those handles, which come from the part of the contour over the smaller square, are the shorter ones, at

least for very small heights of the prisms. Continuity and the intermediate value theorem finish the proof.

I hope this is enough to convince the reader that the conjugate Plateau method for minimal surfaces in $\mathbb{R}^3$ is so flexible that already with contours, where only one parameter has to be adjusted, we can obtain a very large collection of triply periodic minimal surfaces. In most cases embeddedness comes for free because the Plateau contour is a graph, by Krust's theorem the conjugate piece is embedded and the different congruent pieces sit in different fundamental cells for the symmetry group in question. One would also expect that the possibilities for two-parameter adjustments are more than one would like to describe.

4. Conjugate Constant Mean Curvature Surfaces

Recall from Section 2 that, if we are given minimal surface data $\{g, S\}$ (Riemannian metric g and Weingarten map S) in a space $M^3(k)$ of constant curvature k, then $\{g, S \pm c \cdot \mathrm{id}\}$ are surface data for a constant mean curvature $\pm c$ surface in $M^3(k-c^2)$. Such surface data can be integrated on simply connected domains. The constant mean curvature surface has been called the *cousin* of the minimal surface. In the examples of Section 2 we have seen that the transition from a polygonally bounded Plateau piece to the conjugate surface bounded by planar lines of reflectional symmetry is the main reason for the flexibility of this method. This is true even more for the transition to constant mean curvature surfaces because only the geodesic principle curvature lines allow the use of a symmetry of the space M^3, namely a reflection, to analytically extend the original piece. Then one speaks of a constant mean curvature *conjugate cousin surface* if the data $\{g, S\}$ of the minimal surface are changed to $\{g, J \cdot S \pm c \cdot \mathrm{id}\}$, where J is the complex structure, the positive 90° rotation.

We abbreviate "constant mean curvature 1" by cmc1. Quite remarkably it turns out that the derivative TF^* of the cmc1 conjugate cousin in $\mathbb{R}^3$ can be obtained explicitly from the derivative TF of the corresponding minimal surface in $\mathbb{S}^3$ and from the complex structure J; one does not have to relate these through the second order surface data. In conjugate cousin constructions, one always assumes that the minimal piece has a geodesic polygon as boundary so that the conjugate cousin is bounded by planar symmetry lines. Of course, the angles between the geodesic edges are the same as the angles between the planar symmetry arcs at corresponding vertices, and these angles have to be of the form $2\pi/k$, $k \in \mathbb{N}$ so that the extended surface (by repeated reflections) has the vertices as *smooth* interior points.

We first need to understand why the general case is so much more difficult than the case of conjugate minimal surfaces in $\mathbb{R}^3$ in Section 2. Then we can appreciate the extra help which we get from the group structure of $\mathbb{S}^3$ when we use the conjugate cousin method to construct constant mean curvature one surfaces in $\mathbb{R}^3$. We have the simplest case possible if we start with a disk type minimal Plateau solution in a nonplanar geodesic quadrilateral with angles of the form $2\pi/k$. The conjugate minimal piece and all the conjugate cousins are then bounded by four planar symmetry arcs. Extension by reflections in the two arcs that meet at one vertex then gives a larger surface with that vertex as smooth interior point. Moreover, if the quadrilateral was chosen with some care, the four symmetry planes of the boundary arcs are *boundary planes of a simplex in M^3 that contains the conjugate cousin piece.* This is the situation in [Sm], [KPS], parts of [Po] and also

in [Ka3]. For the construction of complete embedded minimal or cousin surfaces the difficulty begins now: *We need to guarantee that the above simplex (made from the boundary symmetry planes of the cousin piece) tessellates* $M^3(k-c^2)$.

Note that the cousin piece meets the faces of the simplex orthogonally. This means that the *dihedral angle* between any pair of symmetry planes that meet at one vertex is also the angle between the corresponding symmetry lines on the surface and therefore is one of the angles of the quadrilateral Plateau contour. In other words, four of the six dihedral angles of the simplex are known as the angles of the quadrilateral. The problem is to control the other two dihedral angles of the simplex. This is much simpler in $\mathbb{R}^3$ (where the scalar product of the normals of the planes gives the desired dihedral angle) than in a curved space. In the applications of [KPS], [Po] and [Ka3] the simplex is a fundamental domain for the symmetry group of a platonic polyhedron. This means that it has three dihedral angles equal to $\pi/2$; these and a fourth one are the angles given by the angles of the conjugate Plateau quadrilateral, as we said before. The simplification caused by the $\pi/2$-angles is that the remaining two dihedral angles of the simplex are also face angles of this simplex. But each such face angle is *the angle between the normals at the endpoints of a boundary arc* because these normals are edges of the simplex.

In $\mathbb{R}^3$ such an angle is the same as the total curvature of the symmetry arc and therefore also the same as the total rotation of the tangent plane of the Plateau piece along the edge under consideration. This means in particular that *in* $\mathbb{R}^3$ *this angle is determined by the Plateau contour alone, without reference to the Plateau solution.*

In spheres and hyperbolic spaces the Gauss–Bonnet theorem says that the total curvature and the desired angle between the end point normals *differ by the area of the curved triangle that is bounded by the symmetry arc and its two endpoint normals.* Therefore the remaining dihedral angles are not computable from the Plateau *contour* alone, but they can be estimated if one has good bounds for the Plateau *solution.*

A quadrilateral with prescribed angles has two free parameters in spheres, Euclidean and hyperbolic spaces and we have to choose those so that the two remaining dihedral angles of the simplex around the conjugate cousin have the correct values. One therefore has to find a closed curve in the 2-dimensional parameter space so that the image curve of the pairs of dihedral angles has winding number $\neq 0$ with respect to the correct pair of dihedral angles. This argument gets simpler for Bryant surfaces since one of the domain parameters can be specified as a function of the other, such that along the corresponding curve one of the two dihedral angles is always correct while the other is too small at one end and too large at the other. The completion of these arguments requires so much work that only the mentioned simplest cases are treated in [KPS], [Po] and [Ka3]. One is far from the flexibility which we saw in the applications in Section 2 to triply periodic minimal surfaces in $\mathbb{R}^3$.

Conjugate cousins in $\mathbb{R}^3$. The preceding discussion directs some extra attention to the case of constant mean curvature one surfaces in $\mathbb{R}^3$, considered as conjugate cousins of minimal surfaces in $\mathbb{S}^3$. In these cases the remaining dihedral angles are now known to be computable from the spherical Plateau *contour* without reference to the Plateau *solution.* But how explicitly can this be done? The answer is very nice. Consider $\mathbb{S}^3$ as a group, for instance as the unit quaternions. The parallel

translations of $\mathbb{R}^3$ are replaced by the *left translations* $L_q : \mathbb{S}^3 \to \mathbb{S}^3$, $L_q(p) := q \cdot p$. With these isometries we can extend every tangent vector $X \in T_{\mathrm{id}}\mathbb{S}^3$ to a left invariant vector field X^* by $X^*(q) := TL_q|_{\mathrm{id}}(X) = q \cdot X$. Such a left invariant vector field has constant length, the angle between two such vector fields is *constant* and the integral curves of these vector fields are great circles, for instance, if $|X| = 1$, then $t \mapsto \cos(t) \cdot q + \sin(t) \cdot X^*(q)$ is the integral curve through q. And clearly, each great circle is an integral curve of exactly one unit length left invariant vector field. All this is in complete analogy to parallel vector fields on $\mathbb{R}^3$. With these notions the answer is: *Measure the total rotation of the tangent plane of the Plateau piece along a great circle edge against left invariant vector fields, then this total rotation is the same as the angle between the normals at the endpoints of the symmetry arc (which corresponds to the edge under consideration) of the conjugate cousin surface.*

This says in particular that we can explicitly determine the great circle Plateau contours if we have decided which angles between symmetry planes we want to achieve. For example all the conjugate contours for triply periodic minimal surfaces in $\mathbb{R}^3$ which did not require any parameter adjustment can now be translated into great circle polygons such that the conjugate cousins of their Plateau solutions give immersed constant mean curvature one surfaces in $\mathbb{R}^3$ with the same symmetry groups as the minimal surfaces. A slightly different picture is: Scale up the conjugate cousins made from very small spherical polygons so as to have the same periods as the minimal surfaces, we get *small constant mean curvature deformations* of the minimal surfaces. Such small deformations continue to be embedded. Moreover, there are deformations which are not interesting for the minimal surface: If we let the edge length a of the contour P_1 in Section 1 shrink to 0, we get a planar contour and a planar minimal piece. However the corresponding spherical quadrilateral does *not* lie in a great sphere, therefore we get a nontrivial Plateau piece and the corresponding doubly periodic constant mean curvature one cousin looks like one horizontal layer of the minimal surface, but with the vertical Schwarz handles between layers having shrunk to zero size.

Since we get in this way without effort many triply periodic cmc1 surfaces together with doubly periodic degenerations, it is clear that with a little effort one can get many more families of cmc1 surfaces. Among the surfaces one obtains that way are triply periodic ones which almost look like sphere packings; the large handles of the related minimal surface have shrunk to very small catenoid like connections between the spheres. In the work of N. Kapouleas, portions of his surfaces are very close to very long strings of spheres. It is therefore a natural question how close to such examples one can come with the conjugate cousin method. Here a successful idea was to solve "spherical" Plateau problems not really in the sphere but for example in the universal cover of the solid Clifford torus [Gb]. This modification indeed allows one to replace the catenoid like connectors between spheres by long strings of spheres with tiny necks between them.

We conclude this section with a proof of the explicit relation between the differentials of minimal surfaces in $\mathbb{S}^3$ and the differentials of their cmc1 conjugate cousins in $\mathbb{R}^3$. We need three steps.

Cross product and complex structure. For a 2-dimensional surface M^2 with normal field N in a 3-dimensional space M^3 we have a simple relation between the complex structure J of M^2 and the normal N via the cross product in the tangent spaces of M^3. Usually the orientations are chosen so that, for each tangent

vector X of M^2,

$$X \times JX = N \qquad \text{or} \qquad JX = N \times X.$$

An immediate application is that J is a covariantly parallel endomorphism field. Recall that a tangent vector field $t \mapsto X(t)$ to M^2 along a curve $t \mapsto c(t)$ is called *covariantly parallel along c*, if and only if the covariant derivative $\frac{\nabla}{dt}X(t)$ in M^3 is orthogonal to M^2. As in $\mathbb{R}^n$ extend this definition and call an *endomorphism field covariantly parallel* if and only if it maps parallel vector fields to parallel vector fields. The endomorphism field J has this property because of

$$\frac{\nabla}{dt}(JX) = \frac{\nabla}{dt}(N \times X) = \frac{\nabla}{dt}N \times X + N \times \frac{\nabla}{dt}X \sim N + 0 \perp M^2.$$

This parallelism is important because it says that for the covariant differentiation D/dt of M^2 (which equals the M^2-tangential component of ∇/dt) the composition with J behaves as multiplication of complex functions by the constant i behaves:

$$\frac{D}{dt}(JX) = J\frac{D}{dt}X \qquad \text{or} \qquad \frac{D}{dt}(JSX) = J\Big(\frac{D}{dt}S\Big)X + JS\frac{D}{dt}X.$$

Quaternions and left translation on $\mathbb{S}^3$. We consider $\mathbb{S}^3$ as the group of unit quaternions. The tangent space at $1 \in \mathbb{S}^3$ is $T_1\mathbb{S}^3 = \mathrm{Im}\mathbb{H} \perp 1$. Left invariant vector fields are given, for each $X \in T_1\mathbb{S}^3$, by using quaternion multiplication $\bullet$ as follows:

$$X^*(q) := q \bullet X.$$

The fact that the *imaginary part* of the quaternionic product of two imaginary quaternions is their *cross product* in $\mathrm{Im}\mathbb{H}$ is easily checked on the basis $\{i, j, k\}$ with $i \bullet j = k = i \times j$, $\mathrm{Im}(i \bullet i) = 0$, etc. The covariant derivative $\frac{\nabla}{dt}$ (in $\mathbb{S}^3$) of a left invariant vector field along a curve $t \mapsto q(t)$, by definition the tangential part of the ordinary derivative in $\mathbb{H} = \mathbb{R}^4$, can therefore be expressed by the cross product with the tangent vector $q'(t)$ of the curve as follows:

$$\frac{\nabla}{dt}X^*(q(t)) = ((q(t) \bullet X)')^{\mathrm{tang}} = (q'(t) \bullet X)^{\mathrm{tang}},$$

and because the normal part of $(q'(t) \bullet X)$ is proportional to q and $q^{-1} \bullet q \in \mathbb{R}$ this expression is equal to

$$q \bullet \mathrm{Im}\Big(\frac{q'}{q} \bullet X\Big) = q \bullet \Big(\frac{q'}{q} \times X\Big).$$

This formula says that left invariant vector fields "turn towards the right" of covariantly constant vector fields in the following sense: If we look in the direction q' of the curve, we see the vector $X^*(q)$ and the covariant derivative of the vector field X^*, namely the vector $q \bullet ((q'/q) \times X)$, points 90° to the right of $X^*(q)$.

The conjugate cousin relation. Now assume that $F : M^2 \to \mathbb{S}^3$ is a minimal immersion with unit normal field $N : M^2 \to T\mathbb{S}^3$, $N(p) \perp \mathrm{image}(TF_p)$. Then $\{TF, N\}$ satisfy the (minimal) surface equations:

$$\nabla N(X) = TF(S \cdot X) \qquad \nabla^2 F(X, Y) = -g(SX, Y) \cdot N, \qquad \mathrm{trace}\, S = 0.$$

We left translate the vector field N and the vector valued 1-form TF to $1 \in \mathbb{S}^3$ and, also using the complex structure J, we define:

$$n(p) := -F^{-1}(p) \bullet N(p), \qquad \omega_p(\,) := F^{-1}(p) \bullet TF_p(J\,), \qquad \text{with } F^{-1} \bullet F = 1 \in \mathbb{S}^3.$$

We claim that the $\mathbb{R}^3$-valued 1-form ω has a symmetric derivative and is therefore integrable. The integral surfaces clearly have n as normal field and their Riemannian metric $g(\cdot,\cdot)$ is the same as that of the minimal surface in $\mathbb{S}^3$ since left translation is an isometry. And we also claim that the shape operator of the integral surfaces is $(JS - \mathrm{id})$, i.e., they are surfaces of constant mean curvature -1. The following computations prove these claims by showing that $\{n, \omega\}$ satisfy the surface equations in $\mathbb{R}^3$ for the given Riemannian metric g and Weingarten map $JS - \mathrm{id}$. First we use the product rule to get the derivative of the quaternionic inverse F^{-1}:

$$\frac{\nabla}{dt}\left(F^{-1}(p(t)) \bullet F(p(t))\right) = 0 \;\Rightarrow\; TF_p^{-1}(p') = -F^{-1}(p) \bullet TF_p(p') \bullet F^{-1}(p).$$

Next we differentiate n using the covariant product rule. Since $n(p)$ is in the fixed tangent space $T_{\mathrm{id}}\mathbb{S}^3 = \mathrm{Im}\,\mathbb{H}$, the ordinary derivative in that Euclidean space agrees with the covariant derivative of $\mathbb{S}^3$ (which in turn is the tangential part of the ordinary derivative in $\mathbb{H}$). Abbreviate $X := p' \in T_p\mathbb{S}^3$.

$$\nabla n_p(X) = \mathrm{Im}\left(F^{-1}(p) \bullet TF_p(X) \bullet F^{-1}(p) \bullet N_p\right) \;-\; F^{-1}(p) \bullet \nabla N_p(X)$$

(use link with the cross product from above and the surface equation for ∇N)

$$= (F^{-1}(p) \bullet TF_p(X)) \times (F^{-1}(p) \bullet N_p) \;-\; F^{-1}(p) \bullet TF_p(SX)$$

(use $X \times N = -JX$ and $J \cdot J = -\mathrm{id}$)

$$= -F^{-1}(p) \bullet TF_p(JX) \;+\; F^{-1}(p) \bullet TF_p(J \cdot JSX)$$

(finally insert the definition of ω_p)

$$= \omega_p((JS - \mathrm{id})X).$$

This gives the first surface equation for $\{n, \omega\}$, with shape operator $JS - \mathrm{id}$:

$$\nabla n_p = \omega_p \circ (JS - \mathrm{id}).$$

Similarly we use the covariant product rule to differentiate ω:

$$\nabla_X \omega(Y) = -\,\mathrm{Im}\left(F^{-1}(p) \bullet TF_p(X) \bullet F^{-1}(p) \bullet TF_p(JY)\right) + F^{-1} \bullet \nabla^2 F(X, JY)$$

(observe that $X \times JY = \det(X, JY) \cdot N$ and insert the surface equation for $\nabla^2 F$)

$$= -(F^{-1}(p) \bullet (\det(X, JY) \cdot N(p))) \;-\; F^{-1}(p) \bullet (g(X, SJY) \cdot N)$$

(use $\det(X, JY) = g(X, Y)$, the symmetry of S and the skew symmetry of J)

$$= -g(X, Y) \cdot ((F^{-1}(p) \bullet N(p))) \;+\; F^{-1}(p) \bullet (g(JSX, Y) \cdot N)$$

(finally insert the definition of $n(p)$)

$$= -g((JS - \mathrm{id})X, Y) \cdot n(p),$$

which is the second surface equation for $\{n, \omega\}$, with shape operator $JS - \mathrm{id}$:

$$\nabla_X \omega(Y) = -g((JS - \mathrm{id})X, Y) \cdot n.$$

Other signs come from other conventions, for instance between J and N or between S and ∇N.

References

[Bu] Buser, P.: Geometry and spectra of compact Riemann surfaces. Birkhäuser, Boston, 1992.

[Gb] Große-Brauckmann, K.: New surfaces of constant mean curvature. Math. Z. 214 (1992), 527–565.

[DHKW] Dierkes, U., Hildebrandt, S., Küster, A., Wohlrab, O.: Minimal surfaces I, II. Grundlehren der math. Wiss. 295, Springer, Berlin, 1992.

[HKW] Hoffman, D., Karcher, H., Fusheng, W.: The genus one helicoid and the minimal surfaces that led to its discovery. pp. 119–170 in Global Analysis in Modern Mathematics, K. Uhlenbeck (ed.), Publish or Perish, 1993.

[JS] Jenkins, H., Serrin, J.: Variational problems of minimal surface type II. Arch. Rat. Mech. Analysis 21 (1966), 321–342.

[Ka1] Karcher, H.: Embedded minimal surfaces derived from Scherk's examples. Manuscripta Math. 62 (1988), 83–114.

[Ka2] Karcher, H.: The triply periodic minimal surfaces of Alan Schoen and their constant mean curvature companions. Manuscripta Math. 64 (1989), 291–357.

[Ka3] Karcher, H.: Hyperbolic constant mean curvature one surfaces with compact fundamental domains. Preprint.

[KP] Karcher, H., Polthier, K.: Construction of triply periodic minimal surfaces. Phil. Trans. R. Soc. Lon. A 354 (1996), 2077–2104.

[KPS] Karcher, H., Pinkall, U., Sterling, I.: New minimal surfaces in $\mathbb{S}^3$, J. Diff. Geom. 28 (1988), 169–185.

[La] Lawson, B. H.: Complete minimal surfaces in S^3. Annals of Math. 92 (1970), 335–374.

[Ni] Nitsche, J.,J.: Über ein verallgemeinertes Dirichletsches Problem für die Minimalflächengleichung und hebbare Unstetigkeiten ihrer Lösungen. Math. Ann. 158 (1965), 203–214.

[Po] Polthier, K.: Geometric a priori estimates for hyperbolic minimal surfaces. Bonner Math. Schriften 263 (1994).

[Sm] Smyth, B.: Stationary minimal surfaces with boundary on a simplex. Invent. Math. 76 (1984), 411–420.

Mathematisches Institut, der Universität Bonn, Beringstrasse 1, D-53115 Bonn, Germany

E-mail address: unm416@uni-bonn.de

Clay Mathematics Proceedings
Volume 2, 2004

Parabolicity and Minimal Surfaces

Joaquín Pérez
(Joint work with Francisco J. López)

Abstract. A crucial problem in the understanding of the theory of minimal surfaces is to determine which conformal structures are allowed under appropriate geometric conditions. Parabolicity is a powerful tool in this direction, when minimal surfaces with boundary are considered. We will give an introductory look to this classical concept in the abstract setting, revise some of the known applications to minimal surface theory (due to Collin, Kusner, Meeks and Rosenberg) and explain briefly some new statements in this field.

1. Preliminaries on Parabolicity

In this section, M will denote a noncompact Riemannian surface with boundary $\partial M \neq \varnothing$, although all that follows is also valid in greater dimensions. Given a point $p \in M$, we can associate to p a measure μ_p on ∂M, called the *harmonic or hitting measure of M with respect to p*, in two equivalent ways. The first one comes from a probability point of view:

Definition 1.1. Given an interval $I \subset \partial M$, $\mu_p(I) \in [0,1]$ is the probability of a random walk that begins at p of exiting M by first time crossing at a point in I.

For instance, when $M = \{|z| \leq 1\} - A$ with A a nonempty closed subset of $\{|z| = 1\}$ and $p = 0$, then the harmonic measure μ_p assigns to each interval $I \subset \partial M$ the value

$$\mu_p(I) = \frac{\text{length}\, I}{\text{length}(\{|z| = 1\})} = \frac{1}{2\pi} \text{length}\, I.$$

In this particular case, μ_p coincides with the Lebesgue measure divided by 2π, and $\mu_p(\partial M) = 1 - (\text{length}\, A)/(2\pi)$; see Figure 1.

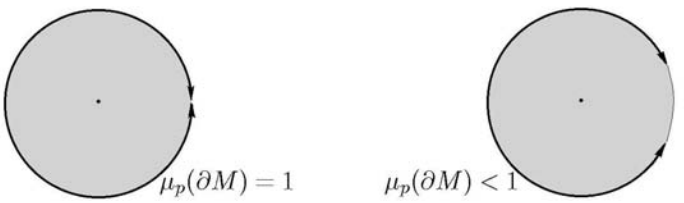


Figure 1. Left: $M = \{|z| \leq 1\} - \{1\}$. Right: $M = \{|z| \leq 1\} -$ Interval.

©2004 Clay Mathematics Institute

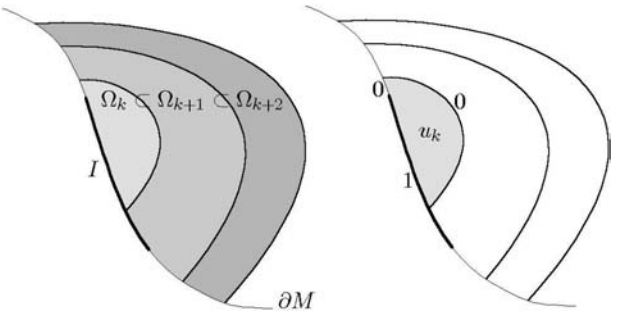


FIGURE 2. The sequences $\Omega_k \nearrow M$ and $u_k \nearrow h$.

We have avoided defining a random walk. The reason is that in what follows, we will only use a second definition for harmonic measure, this one being purely analytic.

Analytic approach to the harmonic measure. Fix an interval $I \subset \partial M$. Given a relatively compact open set $\Omega \subset M$, the noncompactness of M gives that the first eigenvalue of the Dirichlet problem for the Laplacian in Ω is positive, which implies that the following Dirichlet problem can be uniquely solved:

$$(\star)_{\Omega,I} \begin{cases} \Delta u = 0 & \text{in } \Omega, \\ u = 1 & \text{in } \partial\Omega \cap I, \\ u = 0 & \text{in } \partial\Omega - I. \end{cases}$$

Moreover, the maximum principle implies that $0 \leq u \leq 1$ in Ω.

Now consider an increasing sequence $\{\Omega_k : k \in \mathbb{N}\}$ of relatively compact open sets with $\bigcup_k \Omega_k = M$. Denote by u_k the unique solution of the Dirichlet problem $(\star)_{\Omega_k,I}$ for each k. Since $\{u_k\}_k$ is an increasing sequence (by the maximum principle) and $0 \leq u_k \leq 1$ in Ω_k, it follows that $\{u_k\}_k$ converges on compact subsets of M to a harmonic function $h : M \to \mathbb{R}$ that satisfies $h = 1$ in I, $h = 0$ in $\partial M - I$, $0 \leq h \leq 1$ in M; see Figure 2.

We will express the fact that h is a limit of functions that vanish eventually (i.e., that vanish outside a certain compact set, depending on k) by saying that h *vanishes at the ideal boundary of* M, although we will not discuss rigorously the notion of ideal boundary.

Next we check that the function h does not depend on the increasing sequence $\{\Omega_k\}_k$, but only on the chosen interval $I \subset \partial M$: Let $\tilde{h}$ be another harmonic function on M with the same boundary values as h, constructed as limit of solutions $\tilde{u}_k$ of the Dirichlet problem $(\star)_{\tilde{\Omega}_k,I}$ for another increasing sequence $\{\tilde{\Omega}_k\}_k \nearrow M$ of relatively compact open sets. Since $\tilde{h} \geq u_k$ in $\partial\Omega_k$, the maximum principle gives the same inequality in Ω_k. This is true for all k, hence taking limits we get $\tilde{h} \geq h$ in M. The inverse inequality can be deduced analogously, thus equality holds.

Thus, for each interval $I \subset \partial M$ there exists a *unique* harmonic function h_I on M with boundary values $h_I = 1$ in I, $h_I = 0$ in $\partial M - I$, vanishing at the ideal boundary of M.

DEFINITION 1.2. In the above setting, the harmonic measure μ_p with respect to the point $p \in M$ assigns to I the value $\mu_p(I) = h_I(p) \in [0,1]$. Now μ_p extends to any (Lebesgue) measurable set $A \subset \partial M$ by the standard procedure.

REMARKS 1.3. (1) The uniqueness property for h_I does not necessarily hold for bounded solutions of the Dirichlet problem

$$(\star)_I \qquad \begin{cases} \Delta u = 0 & \text{in } M, \\ u = 1 & \text{in } I, \\ u = 0 & \text{in } \partial M - I, \end{cases}$$

as is demonstrated by the counterexample of a closed disk minus a nontrivial closed interval: we can fill the ideal boundary with arbitrary data and solve uniquely the corresponding Dirichlet problem in the whole closed disk. With this in mind, h_I is the unique solution when we impose the data at the ideal boundary to be zero.

(2) When $p \in \partial M$, the definition clearly implies $\mu_p(I) = 1$ if $p \in I$ and $\mu_p(I) = 0$ if $p \in \partial M - I$, that is, $\mu_p(I) = \chi_I(p)$, where χ_I is the characteristic function of I.

(3) If $I_1 \subseteq I_2 \subseteq \partial M$, then $\mu_p(I_1) \leq \mu_p(I_2)$: this follows from the maximum principle for harmonic functions, applied to the solutions of $(\star)_{\Omega_k, I_1}$, $(\star)_{\Omega_k, I_2}$ for an increasing sequence $\Omega_k \nearrow M$ as before.

(4) Suppose there is an interior point $p \in \operatorname{Int} M$ and a measurable subset $I \subset \partial M$ such that $\mu_p(I) = 1$. Then, $\mu_q(I) = 1$ and $\mu_q(\partial M - I) = 0$ for all $q \in M$. To see this, note that $1 = \mu_p(I) = h_I(p)$; hence the harmonic function h_I attains an interior maximum at p, so it is the constant 1. Thus, $\mu_q(I) = h_I(q) = 1$ for any $q \in M$. Now the point follows immediately.

(5) Since harmonicity is independent of the Riemannian metric inside a given conformal class, we deduce that the harmonic measure μ_p does not depend on the metric on M but only on its conformal class. In other words, μ_p carries information of the conformal structure of M, not of its metric character. This observation also applies to all that follows.

DEFINITION 1.4. A surface M is said to be *parabolic* if there exists $p \in \operatorname{Int} M$ such that the harmonic measure μ_p is full, i.e., $\mu_p(\partial M) = 1$.

By Remark 1(4), if μ_p is full for some $p \in \operatorname{Int} M$, then μ_q is also full for any $q \in M$, so the above definition is independent of p. Equivalently, M is parabolic if and only if the unique bounded solution $h_{\partial M}$ of the Dirichlet problem $(\star)_{\partial M}$ that vanishes at the ideal boundary of M is the constant 1.

REMARKS 1.5. (1) If a constant is the unique bounded solution of $(\star)_{\partial M}$ (without any further conditions at the ideal boundary), then M is clearly parabolic. The converse is also true: suppose that M is parabolic and h is a bounded harmonic function on M with $h = 1$ in ∂M. We claim that h is the constant 1. To see this, take $a > 0$ such that $|h| \leq a$ on M and let $\Omega_k \nearrow M$ be an increasing sequence of relatively compact open sets. The existence and uniqueness of the solution u_k of the Dirichlet problem $(\star)_{\Omega_k, \partial M}$ gives that the problem

$$\begin{cases} \Delta w_k = 0 & \text{in } \Omega_k, \\ w_k = 1 & \text{in } \partial\Omega_k \cap \partial M, \\ w_k = a & \text{in } \partial\Omega_k - \partial M, \end{cases}$$

has $w_k = f \circ u_k$ as unique solution, where $f : \mathbb{R} \to \mathbb{R}$ is the affine function $f(t) = (1 - a)t + a$. Since M is parabolic, $\{u_k\}_k$ must converge to the constant 1, hence $\{w_k\}_k$ converges to the constant $f(1) = 1$. But the maximum principle implies $h \leq w_k$ in Ω_k, so after taking limits we have $h \leq 1$. A similar argument

shows that $h \geq 1$ (use $-a$ instead of a), so h must be the constant 1. Therefore, *M is parabolic if and only if the unique bounded solution of the Dirichlet problem $(\star)_{\partial M}$ is constant.*

This fact can be easily generalized to the uniqueness of bounded solutions of any Dirichlet problem on M, no matter what (bounded) boundary values we impose. This is commonly expressed as *M is parabolic if and only if bounded harmonic functions on M are determined by their boundary values.*

(2) On a parabolic surface M, any bounded harmonic function f satisfies the *mean value property*

$$f(p) = \int_{x \in \partial M} f(x)\, \mu_p \quad \text{for any } p \in M.$$

SKETCH OF PROOF. Both sides of the formula are bounded harmonic functions in the variable $p \in M$ (for the right-hand side, one has to compute the laplacian with respect to p under the integral sign, and use the expression of μ_p in terms of the value at p of certain harmonic functions), and they coincide along ∂M (use that when $p \in \partial M$, $\mu_p(I)$ coincides with the value at p of the characteristic function of I for all $I \subset \partial M$). Since the surface M is parabolic, bounded harmonic functions are determined by their boundary values, so both sides must coincide. The converse is obvious (apply the mean value property to the constant one); thus *M is parabolic if and only if the mean value property is true for any bounded harmonic function.* □

(3) Let $h : M \to [0,1]$ be the (unique) solution of the Dirichlet problem $(\star)_{\partial M}$ that vanishes at the ideal boundary of M. *If M is not parabolic,* $\inf_M h = 0$.

PROOF. By contradiction, suppose that $\varepsilon \leq h$ in M for some $\varepsilon > 0$. Consider the affine function $f(t) = (t-\varepsilon)/(1-\varepsilon)$, whose unique fixed point is $f(1) = 1$. Then, the function $h_1 = f \circ h$ is bounded and harmonic in M, and has boundary values $h_1 = 1$ in ∂M. Furthermore, h_1 is limit of the sequence $\{w_k = f \circ u_k\}_k$, where each u_k is the solution of the Dirichlet problem $(\star)_{\Omega_k, \partial M}$ for an increasing sequence of relatively compact open subsets $\Omega_k \nearrow M$. We claim that $h_1 = h$ in M: the boundary conditions of w_k are $w_k = 1$ in $\partial\Omega_k \cap \partial M$, $w_k = -\varepsilon/(1-\varepsilon)$ in $\partial\Omega_k - \partial M$. Thus, $w_k \leq u_k$ in $\partial\Omega_k$, so $w_k \leq u_k$ in Ω_k by the maximum principle, and taking limits, $h_1 \leq h$ in M. Conversely, $h \geq \varepsilon$ in M implies $h_1 \geq 0$ in M, from where $h_1 \geq u_k$ in $\partial\Omega_k$. Again the maximum principle gives $h_1 \geq u_k$ in Ω_k, thus taking limits $h_1 \geq h$ in M, thereby proving the claim. Finally, as $h = h_1 = f(h)$ in M and 1 is the unique fixed point of f we deduce that $h = 1$ in M, which is impossible because M is not parabolic. □

(4) *Suppose that M is parabolic, and $\Omega \subset M$ is a noncompact proper subdomain. Then Ω is parabolic.*

PROOF. It suffices to see that the unique harmonic function $h : \Omega \to [0,1]$ with boundary values $h = 1$ in $\partial\Omega$ and that vanishes at the ideal boundary of Ω is the constant 1. Take an increasing sequence of relatively compact open sets $\Omega_k \nearrow M$ and let u_k be the corresponding solutions of the problem $(\star)_{\Omega_k, \partial M}$. Comparing u_k with h we see that $u_k \leq h$ in $\partial(\Omega \cap \Omega_k)$, so again the maximum principle gives $u_k \leq h$ in $\Omega \cap \Omega_k$. Taking limits, $\lim_k u_k \leq h$ in Ω. But $\lim_k u_k$ is the constant 1 because M is parabolic, thus $h \geq 1$ which forces h to be 1. □

(5) *Parabolicity is not affected by removing or adding compact subsets.*

PROOF. If M is parabolic and $K \subset M$ is compact, then $M - K$ is parabolic by the last point. Conversely, suppose that $M - K$ is parabolic for some $K \subset M$ compact, and let us see that M is also parabolic: On the contrary, assume that M is not parabolic. Thus, the solution h of the Dirichlet problem $(\star)_{\partial M}$ that vanishes at the ideal boundary of M is not the constant 1. Since $0 < h \leq 1$ in M and K is compact, it holds that $0 < \varepsilon \leq h|_K \leq 1$ for certain $\varepsilon > 0$. Consider an increasing sequence of relatively compact subsets $\Omega_k \nearrow M$. We can suppose that $K \subset \Omega_k$ for all k. Since $M - K$ is parabolic, the sequence $\{u_k\}_k$ of solutions of the Dirichlet problem

$$\begin{cases} \Delta u_k = 0 & \text{in } \Omega_k - K, \\ u_k = 1 & \text{in } \partial(\Omega_k - K) \cap \partial(M - K), \\ u_k = 0 & \text{in } \partial(\Omega_k - K) - \partial(M - K) \end{cases}$$

converges to the constant 1 in $M - K$. But $h/\varepsilon \geq u_k$ in $\partial(\Omega_k - K)$, thus $h/\varepsilon \geq u_k$ in $\Omega_k - K$ by the maximum principle. Taking limits, $h/\varepsilon \geq 1$ in $M - K$, which contradicts that $\inf_M h = 0$; see (3) above. □

(6) *If $M = M_1 \cup M_2$ with M_1, M_2 both noncompact and parabolic and $M_1 \cap M_2$ is compact, then M is parabolic.*

PROOF. Suppose that M is not parabolic. Repeating the argument from the preceding point with $M_1 \cap M_2$ instead of K, we deduce that the solution h of the Dirichlet problem $(\star)_{\partial M}$ that vanishes at the ideal boundary of M satisfies: h is not the constant 1 and $0 < \varepsilon \leq h|_{M_1 \cap M_2} \leq 1$ for certain positive ε. On the other hand, $M_1 - (M_1 \cap M_2)$ is a noncompact proper subdomain of the parabolic surface M_1, thus $M_1 - (M_1 \cap M_2)$ is parabolic. Again from the arguments in the proof of the last point (exchange $M - K$ by $M_1 - (M_1 \cap M_2)$), we arrive at $h/\varepsilon \geq 1$ in $M_1 - (M_1 \cap M_2)$. Analogously, $h/\varepsilon \geq 1$ in $M_2 - (M_1 \cap M_2)$ thus $h \geq \varepsilon$ in M, a contradiction. □

(7) *If M carries a proper nonnegative superharmonic function $h : M \to [0, \infty)$, then M is parabolic.*

PROOF. It suffices to check that a bounded harmonic function u on M with $u = 0$ at ∂M must vanish identically. Suppose there exists $p \in M$ such that $u(p) \neq 0$. Choose $a \in \mathbb{R}$ such that $au(p) > h(p)$, and consider the superharmonic function $H = h - au : M \to \mathbb{R}$. Since h is proper and nonnegative in M and u is bounded, H must be also proper in M and positive outside a compact subset. Since $H|_{\partial M} \geq 0$, $H(p) < 0$ and H is eventually positive, H must reach an interior minimum, a contradiction. Therefore u must be identically zero. □

2. Some Known Parabolicity Results for Minimal Surfaces with Boundary

Let $X : M \to \mathbb{R}^3$ be a complete minimal immersion with compact boundary. If X has finite total curvature, then it has finite topology and each end is conformally a punctured disk. Such a surface is always parabolic, so we can see the parabolicity as a generalization of having finite total curvature.

The following theorem gives a geometric situation where parabolicity holds. Recently, it has been successfully applied in different settings to control the conformal structure of proper minimal surfaces without boundary [**6**, **7**]. Moreover,

the nature of its proof has been a source of inspiration for the statements we will explain later on.

THEOREM 2.1 (Collin, Kusner, Meeks, Rosenberg [**2**]). *Let $X : M \to \mathbb{R}^3$ be a proper minimal immersion of a surface M with nonempty boundary. If $X(M)$ is contained in a closed half-space, then M is parabolic.*

SKETCH OF PROOF. After trivial considerations, $X(M)$ can be supposed non-flat and contained in $\{x_3 \geq 0\}$. The function x_3 is harmonic on M, so it has a locally well-defined harmonic conjugate x_3^*. Outside the discrete set of points with vertical normal vector, $z = x_3 + ix_3^*$ is a local conformal coordinate for M, thus the pullback of the flat metric $|dz|^2$ through z is conformal to the induced metric ds^2 by the immersion. We label ∇, Δ as the gradient and laplacian operators with respect to $|dz|^2$ and $r = \sqrt{x_1^2 + x_2^2}$ as the horizontal distance to the x_3-axis. Using that the Weierstrass data of X with respect to z is $(g(z), \phi = dz)$, one gets

$$\Delta \log r = r^{-2} \operatorname{Re}\Big(\frac{g}{\bar{g}} e^{-2\theta i}\Big)$$

in $M - \{r = 0\}$, where $\tan\theta = x_2/x_1$. In particular, $|\Delta \log r| \leq 1/r^2$, where $r \neq 0$.

Now consider the domain $M_k = x_3^{-1}([0,k])$. Since $M_k \cap r^{-1}([0,1])$ is compact, to prove that M_k is parabolic it suffices to check that $M_k \cap r^{-1}([1,+\infty))$ is parabolic. This is proved by finding a proper, eventually positive (i.e., proper outside a compact set), superharmonic function on $M_k \cap r^{-1}([1,+\infty))$, namely $h = \log r - x_3^2$. That h is proper and eventually positive follows because $\log r$ is proper, positive and x_3 is bounded in $M_k \cap r^{-1}([1,+\infty))$, while that h is eventually superharmonic comes from the estimate

$$\Delta h = \Delta \log r - \Delta(x_3^2) \leq |\Delta \log r| - 2\|\nabla x_3\|^2 \leq \frac{1}{r^2} - 2.$$

Once parabolicity holds for M_k, the argument is finished as follows: Since x_3 is bounded and harmonic on M_k, the mean value property says that for a fixed point $p \in M$ at height 1,

$$\begin{aligned} 1 = x_3(p) = \int_{\partial M_k} x_3 \mu_p^k &= \int_{\partial M_k \cap \{0 \leq x_3 < k\}} x_3 \, \mu_p^k + \int_{\partial M_k \cap \{x_3 = k\}} k \, \mu_p^k \\ &\geq k \int_{\partial M_k \cap \{x_3 = k\}} \mu_p^k, \end{aligned}$$

where μ_p^k is the harmonic measure of M_k with respect to p. Thus $\int_{\partial M_k \cap \{x_3 = k\}} \mu_p^k \leq 1/k$. Again by the parabolicity of M_k,

$$1 = \int_{\partial M_k} \mu_p^k = \int_{\partial M_k \cap \{0 \leq x_3 < k\}} \mu_p^k + \int_{\partial M_k \cap \{x_3 = k\}} \mu_p^k \leq \int_{\partial M_k \cap \{0 \leq x_3 < k\}} \mu_p^k + \frac{1}{k}.$$

Taking limits, one gets $1 \leq \int_{\partial M} \mu_p$ thus the harmonic measure μ_p in M is full. □

Further refinements of this theorem plus arguments first used in Xavier [**8**] give rise to statements as:

THEOREM 2.2 (Meeks, Rosenberg [**6**]). *Let $X : M \to \mathbb{R}^3$ be a proper minimal immersion of a surface M with finite topology and compact boundary. If there exists a plane $\Pi \subset \mathbb{R}^3$ such that $X^{-1}(\Pi)$ has a finite number of components and a finite number of crosses, then any end of M is conformally a punctured disk, and M*

is conformally a compact Riemann surface minus a finite number of points and a finite number of open disks.

Other conditions that imply the existence of a finite plane, that is, a plane $\Pi \subset \mathbb{R}^3$ in the hypotheses of Theorem 2.2, are contained in the next result:

THEOREM 2.3 (López [**4**]). *Let $X : M \to \mathbb{R}^3$ be a proper minimal immersion of a surface M with finite topology and compact boundary. If $X(M)$ misses a subset $F \subset \mathbb{R}^3$ of one of the three types described below, then M is conformally a compact Riemann surface minus a finite number of points and a finite number of disks.*

(1) *F is the disjoint union of three closed vertical half-planes whose convex hull is $\mathbb{R}^3$.*

(2) *F is the disjoint union of three closed vertical half-planes, two of them being coplanar.*

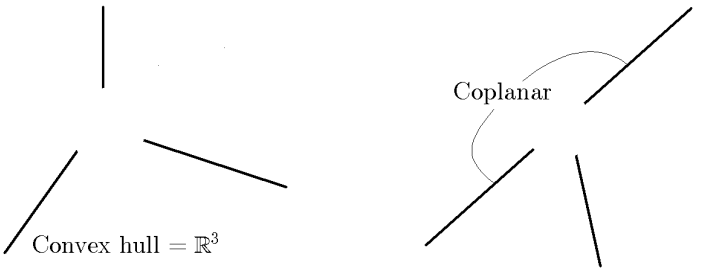


FIGURE 3. Cases 1 (left) and 2 (right) of Theorem 2.3.

(3) *Take three parallel planes in $\mathbb{R}^3$, draw a city map on one of the three planes and translate it orthogonally into the remaining planes. Then, we define F as the union of the complements of the city maps in the three planes*[1]*. Moreover, it must be imposed that the streets are narrow enough in terms of the distance between the planes* (see Figure 4).

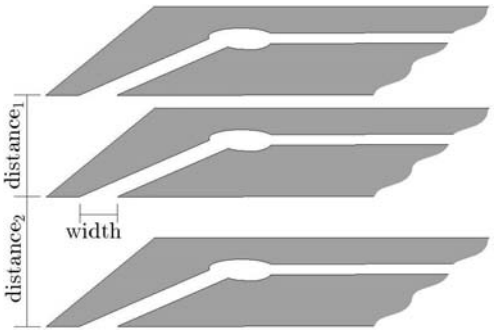


FIGURE 4. Case 3 of Theorem 2.3. The ratio width/distance$_i$ must be small.

[1]In other words, we require that the surface meet these three parallel planes along city maps.

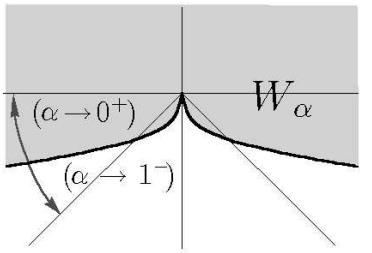


FIGURE 5. The region W_α, where $0 < \alpha < 1$.

3. Minimal Surfaces over a Negative Sublinear Graph

In this section we will explain a new result similar in nature to Theorem 2.1, and valid in a strictly larger region than a half-space, although an additional assumption on the Gauss map is needed. For details, see the paper by the author and López [**5**]. As before, call $r = \sqrt{x_1^2 + x_2^2}$ the distance from any point in space to the x_3-axis. Our aim is to deduce parabolicity for certain minimal surfaces properly immersed in the region $W_\alpha = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 > -r^\alpha\}$, where $\alpha \in (0, 1)$ (see Figure 5).

With this notation, here is our main statement:

THEOREM 3.1. *Let $X : M \to \mathbb{R}^3$ be a proper nonflat minimal immersion such that $X(M) \subset W_\alpha$ for some $\alpha \in (0, 1)$. If, up to removing a compact subset of M, the Gauss map of the surface has image set contained in a hyperbolic open subset of the sphere, then M is parabolic.*

Although the term *hyperbolic* mentioned above is classic, we give here a brief explanation for the sake of completeness. Given a Riemann surface Σ without boundary and $q \in \Sigma$, a function G_q is called a *Green's function in Σ with singularity at q* provided that:

(i) G_q is positive and harmonic in $\Sigma - \{q\}$,
(ii) If z is a local conformal coordinate centered at q, then $G_q(z) + \log|z|$ is harmonic in a neighborhood of q (this condition is independent of the conformal coordinate), and
(iii) G_q is the smallest function satisfying (i), (ii).

If G_q exists for one such point q, the Green's function G_x with singularity at x also exists for any $x \in \Sigma$, see [**3**]. Two consequences of the minimality in (iii) are that if such a function G_q exists, then it is unique, and that if Σ is an open subset of a larger Riemann surface Σ' and G_q exists for Σ, then $G_q|_{\partial\Sigma} = 0$ provided that $\partial\Sigma$ is regular enough.

The existence of Green's functions is classically related with the classification of Riemann surfaces without boundary: the Green's function exists (with singularity at any point of Σ) if and only if the surface is *hyperbolic*, in the sense that Σ carries a positive nonconstant superharmonic function. In the case $\partial\Sigma \neq \varnothing$, the classification of Riemann surfaces is more involved, see for instance [**1**]. The reader should be aware of the dichotomy between hyperbolicity and parabolicity in our setting, as parabolicity has been only defined for surfaces with *nonempty boundary*. In this nonempty boundary case, we will say that a Riemann surface is hyperbolic when it is not parabolic.

When $\Sigma = \{|z| < 1\} \subset \mathbb{C}$, the Green's function with singularity at $q = 0$ is $G_0(z) = -\log|z|$. After obvious transformations, this gives the Green's function in

any round open disk contained in the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. On the other hand, if G_q is the Green's function in a surface Σ for a point $q \in \Sigma$ and $\Omega \subset \Sigma$ is a relatively compact open set with $q \in \Omega$, then the Green's function G_q^Ω in Ω for the point q is $G_q^\Omega = G_q - u$, where u is the (unique) solution of the Dirichlet problem in Ω with boundary values $u = G_q$ in $\partial\Omega$. These elementary facts show that any open nondense subset Ω of $\overline{\mathbb{C}}$ is hyperbolic. Since $\overline{\mathbb{C}}$ minus a closed disk is biholomorphic to $\overline{\mathbb{C}}$ minus a nonconstant curve, this last surface is also hyperbolic. Nevertheless, $\overline{\mathbb{C}}$ minus any finite number of points is not hyperbolic.

Before sketching the proof of Theorem 3.1, we state here a direct consequence of it.

COROLLARY 3.2. *Let $M \subset \mathbb{R}^3$ be a proper minimal graph defined on a domain $D \subset \mathbb{R}^2$. If $M \subset W_\alpha$ for some $\alpha \in (0,1)$, then M is parabolic. In particular, any proper minimal graph lying above a half-catenoid $\{x_3 = -a\cosh\sqrt{x_1^2+x_2^2}\}$, $a > 0$, is parabolic.*

OUTLINE OF PROOF OF THEOREM 3.1. Let $R = \sqrt{x_1^2+x_2^2+x_3^2}$ be the norm of X, a positive proper C^∞-function in $M - X^{-1}(\{0\})$. As before, $z = x_3 + ix_3^*$ is a local conformal coordinate for M around points where the gradient of x_3 does not vanish, where x_3^* stands for a (locally well-defined) harmonic conjugate function x_3^*. Let $|dz|^2, \nabla, \Delta$ be respectively the pullback metric through z of the flat metric, the gradient and the laplacian referred to $|dz|^2$. The Weierstrass representation of X in terms of z gives the following estimates for $\|\nabla R\|, |\Delta R|$ (this length being also computed with respect to $|dz|^2$):

$$\|\nabla R\|^2 \le \tfrac{1}{2}\left(|g|^2 + |g|^{-2} + 2\right), \qquad |\Delta R| \le \frac{1}{R}\left(|g|^2 + |g|^{-2} + 2\right), \tag{3–1}$$

where g is the Gauss map of X. We divide the proof of Theorem 3.1 into four steps.

STEP 1. *The Theorem holds when g eventually omits neighborhoods of $0, \infty \in \overline{\mathbb{C}}$.*

To see this, define $h = R^a + f(x_3) + cx_3 \in C^2(M - \{R^{-1}(0)\})$, with $a, c > 0$ and f a real C^2 function to be made precise later on. This Step 1 will be done by choosing suitable a, c, f so that h is eventually positive, proper and superharmonic on M. We fix $a \in (\alpha, 1)$ and $b \in (1, 2-a)$. Define $\varphi : \mathbb{R} \to \mathbb{R}$ by $\varphi(u) = -\frac{1}{(u^2+1)^{b/2}}$. This is a smooth strictly negative even function on the real line with a global minimum at $\varphi(0) = -1$ and

$$\lim_{|u|\to\infty} |u|^b \varphi(u) = -1. \tag{3–2}$$

Let $f : \mathbb{R} \to \mathbb{R}$ be the only smooth even function satisfying $f''(u) = \varphi(u)$, $f(0) = f'(0) = 0$. f has a global maximum at $f(0) = 0$, is strictly decreasing on the positive numbers and

$$\lim_{|u|\to\infty} \frac{f(u)}{|u|} = -C(b),$$

$C(b)$ being a positive constant depending on b (the exact value of b does not matter for our purposes; we can see the functions φ and f in Figure 6 below). Neither f nor $C(b)$ depend on the omitted neighborhoods of the vertical directions by g, but only on the region W_α (b depends on a, and a depends only on α).

Since $h = R^a + f(x_3) + cx_3$ and $f(x_3)$ diverges linearly to $-\infty$ as $|x_3| \to \infty$, we can choose $c > 0$ depending only on b such that $f(x_3) + cx_3 > 0$ for all $x_3 > 0$ large. In particular, h diverges to $+\infty$ when $x_3 \to +\infty$.

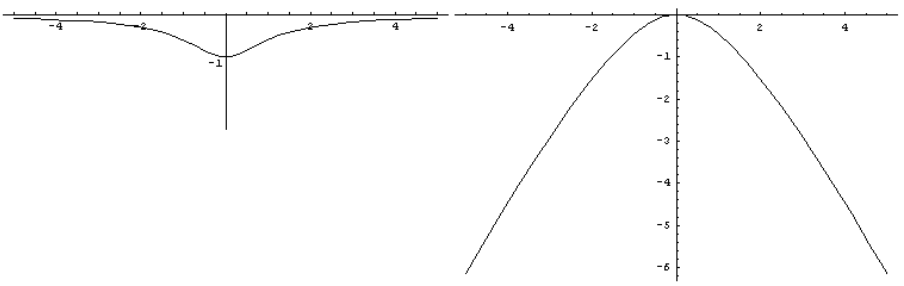


FIGURE 6. The functions φ (left) and f (right) for $b = 1.5$.

When $x_3 < 0$ is large in absolute value, we have $h > R^a - C_1|x_3|$ for a suitable constant $C_1 > 0$ that only depends on b. The hypothesis $X(M) \subset W_\alpha$ implies $|x_3|^{2/\alpha} < x_1^2 + x_2^2$; thus

$$h > (x_1^2 + x_2^2 + x_3^2)^{a/2} - C_1|x_3| > \left(|x_3|^{2/\alpha} + x_3^2\right)^{a/2} - C_1|x_3|.$$

Since $2/\alpha > 2$ and $a > \alpha$, the function appearing in the last right-hand side is positive for $|x_3|$ large enough (depending only on α, a, b) and proper as a function of $|x_3|$.

Finally, in a region of the type $M \cap x_3^{-1}([k,k])$, h is eventually positive because $f(x_3) + cx_3$ is bounded and R^a is proper and positive. From the above arguments, we conclude that h is eventually positive and proper.

Concerning the $|dz|^2$-laplacian of h, firstly note that both $\|\nabla R\|$ and $R|\Delta R|$ are bounded by a positive constant (depending on the omitted neighborhoods of the vertical directions). Therefore, (3–1) allows us to write

(3–3)
$$\Delta h = \Delta(R^a) + f'' \leq |\Delta(R^a)| + \varphi \leq \frac{C}{R^{2-a}} + \varphi \leq \frac{C}{|x_3|^{2-a}} + \varphi \approx \frac{C}{|x_3|^{2-a}} - \frac{1}{|x_3|^b},$$

where we have also used (3–2) in the last approximation, this last one being valid for $|x_3|$ large. Note that C only depends on the missing neighborhoods by the Gauss map. Since $b < 2 - a$, the right-hand side of (3–3) is negative for $|x_3|$ large enough, which gives $\Delta h \leq 0$ in $M - x_3^{-1}([-k,k])$ for a certain $k > 0$ (depending on the omitted neighborhoods by g). In the set $M \cap x_3^{-1}([k,k])$, the inequality $\Delta h \leq \frac{C}{R^{2-a}} + \varphi$ together with the facts that R^{2-a} is proper and φ is strictly negative, imply that Δh is eventually negative. This shows that h is proper, eventually positive and superharmonic, which finishes the proof of Step 1.

STEP 2. *The theorem holds if g eventually omits a neighborhood U of 0 in $\overline{\mathbb{C}}$.*

It is enough to prove this second step for a sufficiently small open neighborhood of zero. Thus $U \subset \overline{\mathbb{C}}$ can be assumed to be open, containing zero, with $\infty \notin \overline{U}$ and $g(M) \cap \overline{U} = \varnothing$. Since $\overline{\mathbb{C}} - \overline{U}$ is open and nondense, it must be hyperbolic and so, the Green's function G in $\overline{\mathbb{C}} - \overline{U}$ with singularity at ∞ exists. Given $k \in \mathbb{N}$, consider the surface $M_k = (G \circ g)^{-1}([0,k]) \subset M$, whose boundary is the disjoint union $\partial M_k = I_k \cup \{G \circ g = k\}$, where $I_k = \partial M \cap \{0 \leq G \circ g < k\}$]; see Figure 7. $X|_{M_k}$ is a proper nonflat minimal immersion with boundary, whose Gauss map omits neighborhoods of zero and infinity, and whose image in $\mathbb{R}^3$ is contained in

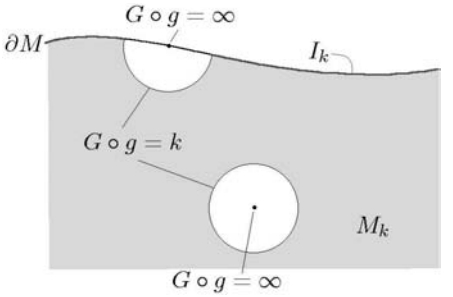


FIGURE 7. The shaded surface is M_k.

W_α. By Step 1, M_k is parabolic and this holds for each k. Fix a point $p \in M$ such that $g(p) \neq \infty$. For all k sufficiently large, p is an interior point of M_k and as $G \circ g$ is a bounded harmonic function on M_k, it must satisfy the mean value property

$$\begin{aligned}(G \circ g)(p) &= \int_{I_k} (G \circ g)\mu_p^k + \int_{\{G \circ g = k\}} (G \circ g)\mu_p^k \\ &= \int_{\partial M \cap \{0 \leq G \circ g < k\}} (G \circ g)\mu_p^k + k \int_{\{G \circ g = k\}} \mu_p^k,\end{aligned}$$

where μ_p^k stands for the harmonic measure of M_k with respect to p. The first integral in the last right-hand side is nonnegative; thus

$$0 \leq \int_{\{G \circ g = k\}} \mu_p^k \leq \frac{1}{k}(G \circ g)(p) \overset{(k \to \infty)}{\longrightarrow} 0. \tag{3–4}$$

Since μ_p^k is full on ∂M_k,

$$1 = \int_{\partial M_k} \mu_p^k = \int_{I_k} \mu_p^k + \int_{\{G \circ g = k\}} \mu_p^k,$$

which together with (3–4) imply that $\int_{I_k} \mu_p^k \to 1$ as k tends to ∞. But this last integral can be computed as $\int_{I_k} \mu_p^k = \mu_p^k(I_k) = h_{I_k}(p)$, where h_{I_k} is the (unique) bounded solution of the Dirichlet problem

$$(\star)_{I_k} \begin{cases} \Delta h_{I_k} = 0 & \text{in } M_k, \\ h_{I_k} = 1 & \text{in } I_k, \\ h_{I_k} = 0 & \text{in } \{G \circ g = k\}, \end{cases}$$

vanishing at the ideal boundary. To finish this second step, we only need to check that the (unique) bounded harmonic function h in M with boundary values $h = 1$ in ∂M and that vanishes at the ideal boundary of M is the constant 1. To see this, recall that such a function satisfies $0 \leq h \leq 1$ in M (because h is constructed as a limit of solutions u_k of Dirichlet problems in relatively compact open subsets $\Omega_k \nearrow M$ and $0 \leq u_k \leq 1$ in Ω_k for each k; see Section 1). This inequality for h implies that for $k \in \mathbb{N}$ fixed, it holds that $h_{I_k} \leq h$ in ∂M_k. Since M_k is parabolic, this implies $h_{I_k} \leq h$ in M_k. Evaluating at p and taking limits as $k \to \infty$, we have $1 = \lim_k h_{I_k}(p) \leq h(p) \leq 1$. Thus h attains an interior maximum at p with value 1, so it must be constant 1. This finishes Step 2.

STEP 3. *The theorem is valid when $g(M)$ omits an open subset $\varnothing \neq U \subset \overline{\mathbb{C}}$.*

If zero is a point of U, then the statement follows directly from Step 2. Thus, we can assume that $0 \notin U$. Without loss of generality we can assume $0 \notin \overline{U}$ and $g(M) \cap \overline{U} = \varnothing$. Since $\overline{\mathbb{C}} - \overline{U}$ is open and nondense, it must be hyperbolic. Since $0 \in \overline{\mathbb{C}} - \overline{U}$, the Green's function G in $\overline{\mathbb{C}} - \overline{U}$ with singularity at 0 exists. Consider the surface $M_k = (G \circ g)^{-1}([0,k]) \subset M$. Since $X(M_k) \subset W_\alpha$ and the Gauss map restricted to M_k omits a neighborhood of zero in $\overline{\mathbb{C}}$, Step 2 guarantees that M_k is parabolic. From this point, the proof of Step 2 can be repeated without changes.

Now we can prove the theorem. Up to removing a compact subset of M, we can assume that $g(M)$ is contained in an open hyperbolic subset $\Omega \subset \overline{\mathbb{C}}$. Fix a point $q \in \Omega$ and let G be the Green's function in Ω with singularity at q. Given $k \in \mathbb{N}$, define $U(k) = G^{-1}((k, +\infty])$, which is an open neighborhood of q. As in previous steps, we consider the surface $M_k = (G \circ g)^{-1}([0,k]) = M - g^{-1}(U(k))$. Since $X(M_k) \subset W_\alpha$ and $g|_{M_k}$ omits $U(k)$, Step 3 gives that M_k is parabolic. Following again the arguments in the proof of Step 2 and recalling that the parabolicity is not affected by adding or removing compact subsets, we conclude the proof. □

References

[1] L. V. Ahlfors and L. Sario, *Riemann surfaces*, Princeton Univ. Press, Princeton, 1960.

[2] P. Collin, R. Kusner, W. H. Meeks III and H. Rosenberg, *The topology, geometry and conformal structure of properly embedded minimal surfaces*, preprint.

[3] H. M. Farkas and I. M. Kra, *Riemann surfaces*, Springer, New York, 1979.

[4] F. J. López, *Some Picard theorems for minimal surfaces*, Trans. Amer. Math. Soc. **356** (2004), 703–733.

[5] F. J. López and J. Pérez, *Parabolicity and Gauss map of minimal surfaces*, to appear in Indiana Univ. Math. Journal.

[6] W. H. Meeks III and H. Rosenberg, *Maximum principles at infinity with applications to minimal and constant mean curvature surfaces*, preprint.

[7] W. H. Meeks III and H. Rosenberg, *The uniqueness of the helicoid and the asymptotic geometry of properly embedded minimal surfaces with finite topology*, preprint.

[8] F. Xavier, *Why no new simply-connected embedded minimal surfaces have been found since 1776*, preprint.

Departamento de Geometria y Topologia, Universidad de Granada, Fuentenueva s/n, 18071, Granada, Spain

E-mail address: jperez@ugr.es

Clay Mathematics Proceedings
Volume 2, 2004

The Isoperimetric Problem

Antonio Ros

Contents

1. Presentation 175
1.1. The isoperimetric problem for a region 176
1.2. Soap bubbles 177
1.3. Euclidean space, slabs and balls 178
1.4. The isoperimetric problem for the Gaussian measure 182
1.5. Cubes and boxes 183
1.6. The periodic isoperimetric problem 187
2. The Isoperimetric Problem for Riemannian 3-Manifolds 188
2.1. The stability condition 188
2.2. The 3-dimensional projective space 191
2.3. Isoperimetry and bending energy 193
2.4. The isoperimetric profile 195
2.5. Lévy–Gromov isoperimetric inequality 197
3. The Isoperimetric Problem for Measures 199
3.1. The Gaussian measure 199
3.2. Symmetrization with respect to a model measure 201
3.3. Isoperimetric problem for product spaces 204
3.4. Sobolev-type inequalities 205
References 207

1. Presentation

The isoperimetric problem is an active field of research in several areas, such as in differential geometry, discrete and convex geometry, probability, Banach spaces theory and PDEs.

In this section we will consider some situations where the problem has been completely solved and some others where it remains open. In both cases I have chosen the simplest examples. We will start with the classical version: the isoperimetric problem in a region, for instance, the whole Euclidean space, the ball or

Research partially supported by a MCYT-FEDER Grant No. BFM2001-3318.

©2004 Clay Mathematics Institute

the cube. Even these concrete examples profit from a broader context. In these notes we will study the isoperimetric problem not only for a region but also for a measure, especially the Gaussian measure, which is an important and fine tool in isoperimetry.

Concerning the literature on the isoperimetric problem we mention the classic texts by Hadwiger [**33**], Osserman [**57**] and Burago & Zalgaller [**18**]. We will not treat here the special features of the 2-dimensional situation; see [**43**] for more information about this case.

I would like to thank F. Morgan, J. Pérez and M. Ritoré for several helpful discussions and comments on these notes.

1.1. The isoperimetric problem for a region. Let M^n be a Riemannian manifold of dimension n with or without boundary. In this paper *volumes* and *areas* in M will mean n-dimensional and $(n-1)$-dimensional Riemannian measures, respectively. The n-dimensional sphere and closed Euclidean ball of radius r will be denoted by $\mathbf{S}^n(r)$ and $B^n(r)$. When $r=1$ we will simply write $\mathbf{S}^n$ and B^n.

Given a positive number $t < V(M)$, where $V(M)$ denotes the volume of M, the *isoperimetric problem* consists in studying, among the compact hypersurfaces $\Sigma \subset M$ enclosing a region Ω of volume $V(\Omega) = t$, those which minimize the area $A(\Sigma)$, see Figure 1. Note that we are not counting the area coming from the boundary of M. If the manifold M has smooth boundary or it is the product of manifolds with smooth boundary (like the cube), there are no differences between the isoperimetric problems for M and for the interior of M. So we will take one manifold or the other at our convenience, depending on the arguments we are going to use.

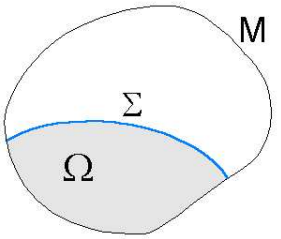


FIGURE 1. Isoperimetric problem in a region.

By combination of the results of Almgren [**2**], Grüter [**32**], Gonzalez, Massari, Tamanini [**26**] we obtain the following fundamental existence and regularity theorem (see also Morgan [**55**]).

THEOREM 1.1. *If M^n is compact, then, for any t, $0 < t < V(M)$, there exists a compact region $\Omega \subset M$ whose boundary $\Sigma = \partial\Omega$ minimizes area among regions of volume t. Moreover, except for a closed singular set of Hausdorff dimension at most $n-8$, the boundary Σ of any minimizing region is a smooth embedded hypersurface with constant mean curvature and, if $\partial M \cap \Sigma \neq \varnothing$, then ∂M and Σ meet orthogonally.*

In particular, if $n \leq 7$, then Σ is smooth. The region Ω and its boundary Σ are called an *isoperimetric region* and an *isoperimetric hypersurface* respectively. The

hypersurface Σ has an area-minimizing tangent cone at each point. If a tangent cone at $p \in \Sigma$ is a hyperplane, then p is a regular point of Σ. The theorem also applies to the case when M itself is noncompact but M/G is compact, G being the isometry group of M.

The *isoperimetric profile* of M is the function $I = I_M : (0, V(M)) \to \mathbb{R}$ defined by

$$I(t) = \min\{A(\partial\Omega) : \Omega \subset M \text{ region with } V(\Omega) = t\}.$$

General properties of this function will be considered in section §2.4. Explicit lower bounds for the profile I are very important in applications and are called *isoperimetric inequalities*. For results of this type see Chavel [**19, 20**], Gallot [**25**] and sections §2 and §3 below.

Another important aspect of the problem (which, in fact, is one of the main matters of this paper) is the study of the geometry and the topology of the solutions and the explicit description of the isoperimetric regions when the ambient space is simple enough. If M is the Euclidean space, the sphere or the hyperbolic space, then isoperimetric domains are metric balls. However, this question remains open in many interesting cases. Concerning that point, there are in the isoperimetry field a certain number of *natural conjectures* which may be useful, at this moment, to fix our ideas about the problem. Let's consider, for instance, the following ones:

(1) The isoperimetric regions on a flat torus $T^n = \mathbf{S}^1 \times \cdots \times \mathbf{S}^1$ are of the type (ball in T^m) $\times\, T^{n-m}$.
(2) Isoperimetric regions in the projective space $\mathbf{P}^n = \mathbf{S}^n/\pm$ are tubular neighborhoods of linear subvarieties.
(3) Isoperimetric regions in the complex hyperbolic space are geodesic balls.

In high dimension many of these conjectures probably fail. In fact, the natural conjecture is already wrong for a very simple space: in the slab $\mathbb{R}^n \times [0,1]$ with $n \geq 9$, unduloids are sometimes better than spheres and cylinders; see Theorem 1.4. However, in the 3-dimensional case these kinds of conjectures are more accurate: Statement (2) is known to be true (see Theorem 2.7) and partial results in §1.5 and §2.1 support (1).

1.2. Soap bubbles. Soap bubble experiments are also useful to improve our intuition about the behavior of the solutions of the isoperimetric problem.

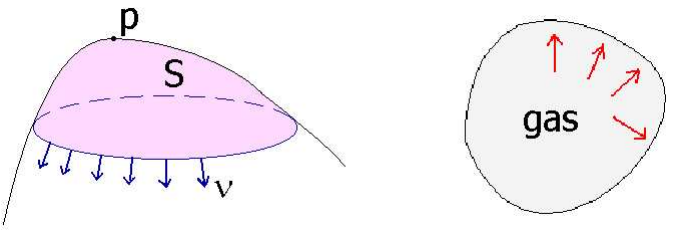


FIGURE 2. Soap bubbles and isoperimetric problem.

From the theoretical point of view, a soap film is a membrane which is also a homogeneous fluid. To deduce how a soap film in equilibrium pushes the space near it, we cut a small piece S in the membrane, see Figure 2. If we want S to remain in equilibrium, we must reproduce in some way the action of the rest of the

membrane over the piece. Because of the fluidity assumption this action is given by a distribution of forces along the boundary of S. Moreover these forces follow the direction of the conormal vector ν: they are tangent to the membrane and normal to the boundary of S. Finally the homogeneity implies that the length of the vectors in this distribution is constant, say of length 1, and thus S pushes the space around it with a force given by

$$F(S) = \int_{\partial S} \nu \, ds.$$

Then, by a simple computation we can obtain the value of *the pressure* at a point p in the film (which is defined as the limit of the mean value of the forces $F(S)$ when S converges to p),

$$\lim_{S \to p} \frac{F(S)}{A(S)} = 2H(p)N(p),$$

where N is the unit normal vector field of S and H is its mean curvature.

Consider now a soap bubble enclosing a volume of gas as in Figure 2. The gas inside pushes the bubble in a homogeneous and isotropic way: it induces on the surface of the bubble a distribution of forces which is normal to the surface and has the same intensity at any point of the bubble. Since the membrane is in equilibrium, the pressure density on the surface must be opposite to this distribution and, thus, the mean curvature of the bubble must be constant.

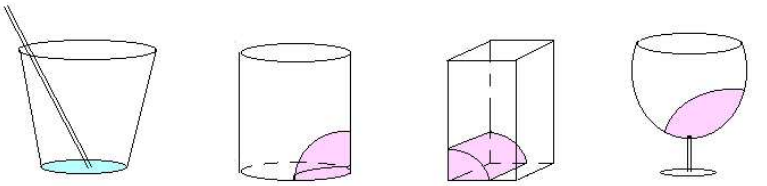


FIGURE 3. Soap bubble experiment.

We can also reason that the energy of the bubble is proportional to its area, so an equilibrium soap bubble minimizes area (at least locally) under a volume constraint. Therefore we can visualize easily (local) solutions of the isoperimetric problem for different kinds of regions in the Euclidean 3-space by reproducing the soap film experiment in Figure 3. The case of a cubic vessel is particularly enlightening: it is possible to obtain the five bubbles which appear in Figure 9.

1.3. Euclidean space, slabs and balls. The symmetry group of the ambient space M can be used to obtain symmetry properties of its isoperimetric regions. These kinds of arguments are called *symmetrizations* and were already used by Schwarz [**70**] and Steiner [**74**] to solve the isoperimetric problem in the Euclidean space (see §3.2 for a generalization of the ideas of these authors).

In this section we will see a different symmetrization argument, depending on the existence and regularity of isoperimetric regions in Theorem 1.1, which was first introduced by Hsiang [**40, 42**]. We will show how the idea works in the Euclidean case.

THEOREM 1.2. *Isoperimetric hypersurfaces in* $\mathbb{R}^n$ *are spheres.*

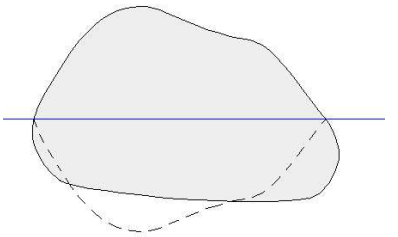


FIGURE 4. If a hyperplane bisects the volume of an isoperimetric region, then the region is symmetric with respect to the hyperplane.

PROOF. Let Ω be an isoperimetric region in $\mathbb{R}^n$ with $\Sigma = \partial\Omega$. First we observe that a stability argument gives that the set $R(\Sigma)$ of regular points of Σ is connected; see Lemma 2.1 in section §2.1 below.

Consider a hyperplane $P \subset \mathbb{R}^n$ so that the regions $\Omega^+ = \Omega \cap P^+$ and $\Omega^- = \Omega \cap P^-$ have the same volume, $P^\pm$ being the half-spaces determined by P, as in Figure 4. Assuming $A(\Sigma \cap P^+) \leq A(\Sigma \cap P^-)$ one can see that the symmetric region $\Omega' = \Omega^+ \cup s(\Omega^+)$, where s denotes the reflection with respect to P, is an isoperimetric region too. In particular,

$$A(\Sigma \cap P^+) = A(\Sigma \cap P^-)$$

and $\Sigma' = \partial\Omega'$ satisfies the regularity properties stated in Theorem 1.1.

If $R(\Sigma) \cap P = \varnothing$, then the regular set of Σ would be either empty or disconnected. Since both options are impossible we conclude that P meets the regular set, it follows from the unique continuation property [**4**] applied to the constant mean curvature equation that $R(\Sigma) = R(\Sigma')$. Then $\Sigma = \Sigma'$ and we get that Ω is symmetric with respect to P.

Since each family of parallel hyperplanes in $\mathbb{R}^n$ contains a hyperplane P which bisects the volume of Ω, we conclude easily that Ω is a ball. □

The preceding arguments also show that the isoperimetric regions in the n-dimensional sphere are metric balls. In the low dimensional case (when the solutions are smooth), Alexandrov reflection technique [**1**] can also be used to solve the isoperimetric problem in the Euclidean space. However, this technique does not work in the three-sphere.

The easy modifications needed to reproduce the argument in the other situations considered in this paper are left to the reader. The only extra fact we need to use in the cases below is that any isoperimetric hypersurface of revolution is necessarily nonsingular. Outside of the axis of revolution this follows from Theorem 1.1 (otherwise the singular set would be too large). Using that the unique minimal cone of revolution in $\mathbb{R}^n$ is a hyperplane it follows that points in the axis are regular too.

THEOREM 1.3. *Isoperimetric hypersurfaces in a half-space are half-spheres.*

PROOF. Hsiang symmetrization gives that any isoperimetric hypersurface has an axis of revolution. To conclude the theorem from that point, we can repeat the argument in the proof of the case 1 of Theorem 1.5 below. □

The slab presents interesting and surprising behavior with respect to the isoperimetric problem. The symmetrization argument in Theorem 1.2 shows that any

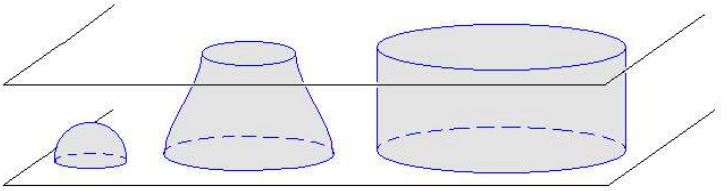


FIGURE 5. Tentative solutions of the isoperimetric problem in the slab.

solution is a hypersurface of revolution. But the complete answer is the following (the case $n = 9$ is still open); see Figure 5.

THEOREM 1.4 (Pedrosa & Ritore, [**58**]). *If $n \leq 8$, isoperimetric surfaces in the slab $\mathbb{R}^{n-1} \times [0,1]$ are half-spheres and cylinders. However, if $n \geq 10$ unduloids solve the isoperimetric problem for certain intermediate values of the volume.*

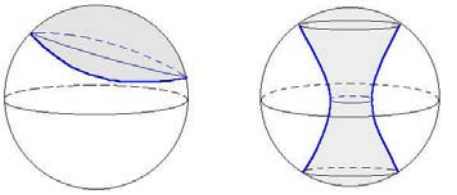


FIGURE 6. Candidates to solve the isoperimetric problem in the ball.

By using the existence and regularity Theorem 1.1, we will give a new proof of the isoperimetric property of spherical caps in the ball $B = \{x \in \mathbb{R}^n : |x| \leq 1\}$.

THEOREM 1.5 (Almgren [**2**]; Bokowski & Sperner [**16**]). *Isoperimetric hypersurfaces in a ball are either hyperplanes through the origin or spherical caps.*

PROOF. Although the arguments below work in any dimension, in order to explain the ideas more intuitively, we will assume that B is a ball in $\mathbb{R}^3$. Let Ω be an isoperimetric region of the ball B and $\Sigma = \partial\Omega$. First observe that $\Sigma \cap \partial B \neq \varnothing$: otherwise by moving Σ until we touch the first time ∂B we obtain a new isoperimetric region which meets the boundary of the ball tangentially and we contradict Theorem 1.1. The symmetrization argument in Theorem 1.2 shows that Σ is connected and has an axis of revolution. We remark that in the ball there are constant mean curvature surfaces satisfying these conditions other than the spherical caps, as in Figure 6. In fact one of the difficulties of the problem, which appears also in most situations, is how to exclude these alternative candidates. We discuss two possibilities:

Σ is a disk. In this case we will show that Σ is spherical or planar by using an argument from [**18**]; see Figure 7.

After changing Ω by its complementary region if necessary, we can consider an auxiliary second ball B' such that $V(B \cap B') = V(\Omega)$ and $\partial B' \cap \partial B = \partial\Sigma$. This is always possible unless Σ encloses the same volume as the plane section through $\partial\Sigma$, a case for which the result is clear.

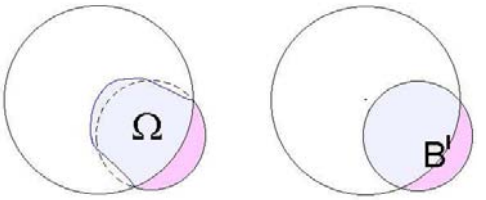


FIGURE 7. Discoidal solutions are spherical.

Define the new regions $B \cap B'$ in the ball B and $W = \Omega \cup (B' - B)$ in $\mathbb{R}^3$. Since $V(B \cap B') = V(\Omega)$, from the minimizing property of Ω we have $A(\Sigma) \leq A(B \cap \partial B')$. Thus $A(\partial W) = A(\Sigma) + A(\partial B' - B) \leq A(\partial B')$ and, using the fact that $V(W) = V(\Omega) + V(B' - B) = V(B')$, the isoperimetric property of the ball in $\mathbb{R}^3$ implies that W is a ball. In particular, since $\Sigma \subset W$, we conclude that Σ is a spherical cap.

Σ *is an annulus.* We will prove that this second case is impossible; see Figure 8. Let's rotate Ω a little bit around an axis of the ball orthogonal to the axis of revolution to get a new region Ω' with boundary $\Sigma' = \partial\Omega'$. Among the regions defined in $\Omega \cup \Omega'$ by $\Sigma \cup \Sigma'$, let's consider the following ones (see Figure 8):

$\Omega_1 = \Omega \cap \Omega'$, which contains the center of the ball,
Ω_3 and Ω_4, the two components of $\Omega - \Omega'$ which meet ∂B, and
Ω_2 and Ω_5, components of $\Omega' - \Omega$ which intersect the boundary of B.

We claim that these five regions are pairwise disjoint. The only nontrivial cases to be considered are the couples Ω_3, Ω_4 and Ω_2, Ω_5. Since Ω is rotationally invariant, Ω' can be also obtained as the symmetric image of Ω with respect to a certain plane P through the center of B. Clearly P intersects Ω_1 and is disjoint from Ω_i, $i = 2, 3, 4, 5$. Since $\Omega_3 \cap \partial B$ and $\Omega_4 \cap \partial B$ lie at different sides of P we conclude that $\Omega_3 \neq \Omega_4$ and, in the same way, $\Omega_2 \neq \Omega_5$.

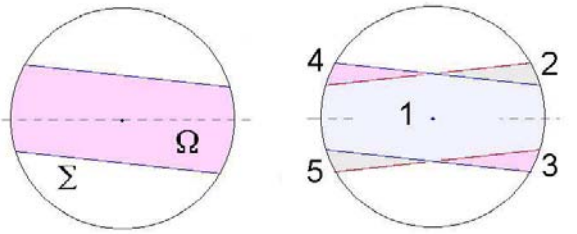


FIGURE 8. Topological annuli cannot be solutions.

Finally define in B the region $\Omega'' = (\Omega - \Omega_3) \cup \Omega_2$. Since $\partial\Omega''$ contains pieces coming from both Σ and Σ', it follows that Ω'' is not smooth. On the other hand, since Ω_2 and Ω_3 are symmetric with respect to P we see that $V(\Omega'') = V(\Omega)$ and $A(\partial\Omega'') = A(\Sigma)$. Thus Ω'' is also an isoperimetric region, which contradicts the regularity of Theorem 1.1. □

Another proof of Theorem 1.5, which essentially uses an infinitesimal version of the argument above, is given in Ros [**66**].

1.4. The isoperimetric problem for the Gaussian measure. We want to extend the family of objects where we are studying the isoperimetric problem. Let M^n be a Riemannian manifold with metric ds^2 and distance d. For any closed set $R \subset M$ and $r > 0$ we can consider the *r-enlargement* $R_r = \{x \in M : d(x, R) \leq r\}$. Given a (positive, Borel) measure μ on M, we define the *boundary measure* (also called *μ-area* or *Minkowski content*) μ^+ as follows:

$$\mu^+(R) = \liminf_{r \to 0} \frac{\mu(R_r) - \mu(R)}{r}.$$

If R is smooth and μ is the Riemannian measure, then $\mu^+(R)$ is just the area of ∂R. Now we can formulate the isoperimetric problem for the measure μ: we want to minimize $\mu^+(R)$ among regions with $\mu(R) = t$. Note that in order to study the isoperimetric problem we only need to have a measure and a distance (not necessarily a Riemannian one). Although we will not follow this idea, this more abstract formulation is useful in some areas such as discrete geometry, probability and Banach space theory; see for instance [**36**] and [**49**].

The *isoperimetric profile* $I_\mu = I_{(ds^2,\mu)}$ of (M, μ) is defined, as above, by

$$I_\mu(t) = \inf\{\mu^+(R) : \mu(R) = t\}, \qquad 0 \leq t \leq \mu(M).$$

If we change the metric or the measure by a positive factor a, it follows easily that

$$I_{a\mu}(t) = aI_\mu\Big(\frac{t}{a}\Big) \quad \text{and} \quad I_{(a^2 ds^2,\mu)} = \frac{1}{a} I_{(ds^2,\mu)}.$$

Among the new examples which appear in our extended setting, we remark the following ones (in these notes we will be interested only in nice new situations and we will only consider absolutely continuous measures).

1. *Smooth measures* of the type $d\mu = f\,dV$, with $f \in C^\infty(M)$, $f > 0$. If R is good enough, then
$$\mu(R) = \int_R f\,dV \qquad \text{and} \quad \mu^+(R) = \int_{\partial R} f\,dA.$$
2. If $V(M) < \infty$ we can consider the *normalized Riemannian measure* (or *Riemannian probability*) $\mu_M = V/V(M)$.
3. The *Gaussian measure* $\gamma = \gamma_n$ in $\mathbb{R}^n$ defined by
$$d\gamma = \exp(-\pi|x|^2)dV.$$

Note that γ is a probability measure, that is, $\gamma(\mathbb{R}^3) = 1$. It is well-known that the Gaussian measure appears as the limit of the binomial distribution. We will see in section §3 that it appears also as a suitable limit of the spherical measure when the dimension goes to infinity. For $n \geq 2$, γ_n is characterized (up to scaling) as the unique probability measure on $\mathbb{R}^n$ which is both rotationally invariant and a product measure.

The isoperimetric problem for the Gaussian measure has remarkably simple solutions; see Theorem 3.3 for a proof.

THEOREM 1.6 ([**17**],[**73**]). *Among regions $R \subset \mathbb{R}^n$ with prefixed $\gamma(R)$, half-spaces minimize the Gaussian area.*

In particular, $I_\gamma(\frac{1}{2}) = 1$ and the corresponding isoperimetric hypersurfaces are hyperplanes passing through the origin.

Next we give a comparison result for the isoperimetric profiles of two spaces which are related by means of a Lipschitz map.

PROPOSITION 1.7. *Let (M,μ) and (M',μ') be Riemannian manifolds with measures and $\phi : M \to M'$ a Lipschitz map with Lipschitz constant $c > 0$. If $\phi(\mu) = \mu'$, then $I_\mu \le cI_{\mu'}$.*

PROOF. Let $R' \subset M'$ be a closed set and $R = \phi^{-1}(R')$. The hypothesis $\phi(\mu) = \mu'$ means that $\mu'(R') = \mu(R)$. The Lipschitz assumption implies that, for any $r > 0$, $\phi(R_r) \subset R'_{cr}$ and therefore $\mu'(R'_{cr}) \ge \mu(\phi^{-1}(\phi(R_r)) \ge \mu(R_r)$. Thus we conclude that

$$\frac{\mu'(R'_{cr}) - \mu'(R')}{cr} \ge \frac{\mu(R_r) - \mu(R)}{cr}$$

and the proposition follows by taking $r \to 0$. □

If we assume that μ and μ' are smooth, ϕ is a diffeomorphism and there exists a μ-isoperimetric hypersurface $\Sigma \subset M$ enclosing a region Ω with $\mu(\Omega) = t$, then one can see easily that the equality $I_\mu(t) = cI_{\mu'}(t)$ is equivalent to the condition $|d\phi(N)| = c$, where N is the unit normal vector along a Σ.

As a first application of the proposition we get, by taking a suitable linear map ϕ and using Theorem 1.5, the following fact.

COROLLARY 1.8. *Let B be the unit ball in $\mathbb{R}^n$ and*

$$E = \left\{(x_1, \cdots, x_n) : \frac{x_1^2}{a_1^2} + \cdots + \frac{x_n^2}{a_n^2} \le 1\right\}$$

an ellipsoid with $0 < a_1 \le \cdots \le a_n$ and $a_1 \cdots a_n = 1$, both spaces endowed with the Lebesgue measure. Then $a_n I_E \ge I_B$ and the section with the hyperplane $\{x_n = 0\}$ minimizes area among hypersurfaces which separate E in two equal volumes.

1.5. Cubes and boxes. Consider the box $W = (0, a_1) \times \cdots \times (0, a_n) \subset \mathbb{R}^n$, with $0 < a_1 \le \cdots \le a_n$. The unit cube corresponds to the case $a_1 = \cdots = a_n = 1$. First we observe that the symmetrization argument of section §1.3 implies that the isoperimetric problem in W is equivalent (after reflection through the faces of the box) to the isoperimetric problem in the rectangular torus $T^n = \mathbb{R}^n/\Gamma$, where Γ is the lattice generated by the vectors $(2a_1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 2a_n)$.

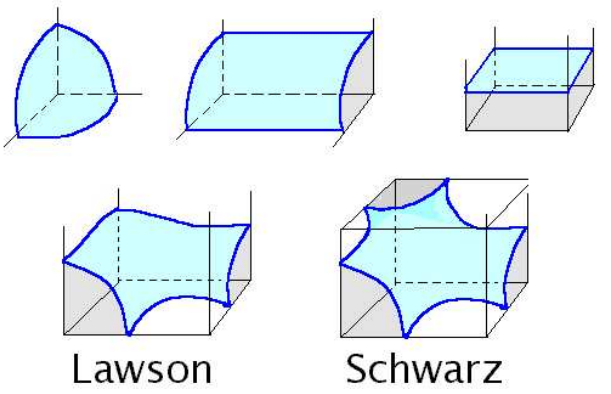


FIGURE 9. Candidates for isoperimetric surfaces in the cube.

For $n = 3$ (from Theorem 1.1 the case $n = 2$ is trivial) Ritoré [**60**] proves that any isoperimetric surface belongs to some of the types described in Figure 9; see §2.1 below (see also Ritoré and Ros [**64**] for other related results). It is natural to hope that isoperimetric surfaces in a 3-dimensional box are spheres, cylinders and planes. However the problem remains open. Lawson type surfaces sometimes come close to beating those simpler candidates. Indeed, by using Brakke's Surface Evolver we can see that in the unit cube there exists a Lawson surface of area 1.017 which encloses a volume $1/\pi$ (as compared to the conjectured spheres-cylinders-planes area which is equal to 1; see Figure 10). In higher dimensions the corresponding conjecture is probably wrong.

The next result is a nice application of the Gaussian isoperimetric inequality (see for instance [**47**]): it gives a sharp isoperimetric inequality in the cube and solves explicitly the isoperimetric problem when the prescribed volume is one half of the total volume.

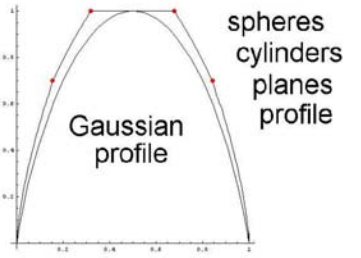


FIGURE 10. Comparison between Gaussian and tentative cubical profiles.

THEOREM 1.9. *Let $W = (0,1)^n$ be the open unit cube in $\mathbb{R}^n$. Then $I_W \geq I_\gamma$.*

PROOF. Consider the diffeomorphism $f : \mathbb{R}^n \to W$ defined by

$$df_{(x_1,\ldots,x_n)} = \begin{pmatrix} e^{-\pi x_1^2} & & \\ & \ddots & \\ & & e^{-\pi x_n^2} \end{pmatrix}.$$

Since f is a 1-Lipschitz map which transforms the Gaussian measure into the Lebesgue one, the result follows directly from Proposition 1.7. □

In Figure 10 we have drawn both the Gaussian profile and the candidate profile of the unit cube in $\mathbb{R}^3$ (which is given by pieces of spheres, cylinders and planes). Note how sharp the estimate in Theorem 1.9 is. Since both curves coincide for $t = \frac{1}{2}$, we can conclude the following fact.

COROLLARY 1.10 (Hadwiger, [**34**]). *Among surfaces which divide the cube in two equal volumes, the least area is given (precisely) by hyperplanes parallel to the faces of the cube.*

The argument used in the proof of Theorem 1.9 gives the following more general result.

THEOREM 1.11. *Given $0 < a_1 \leq \cdots \leq a_n$ with $a_1 \cdots a_n = 1$, consider the box $W = (0, a_1) \times \cdots \times (0, a_n) \subset \mathbb{R}^n$. Then $a_n I_W \geq I_\gamma$ and $\{x_n = a_n/2\}$ has least area among hypersurfaces which divide the box in two equal volumes.*

Since isoperimetric problems in boxes and rectangular tori are equivalent, we can write the following consequence of the last theorem.

COROLLARY 1.12. *Consider the lattice* $\Gamma \subset \mathbb{R}^n$ *spanned by the vectors*

$$(a_1, 0, \dots, 0), \ \dots, \ (0, \dots, 0, a_n),$$

with $0 < a_1 \leq \cdots \leq a_n$. *Among hypersurfaces which divide the torus* $\mathbb{R}^n/\Gamma$ *in two equal volumes, the minimum of the area is given by* $2a_1 \cdots a_{n-1}$. *This area is attained by the pair of parallel* $(n-1)$*-tori* $\{x_n = 0\}$ *and* $\{x_n = a_n/2\}$.

Below we have two applications of the isoperimetric inequality in the cube. The first one concerns the connections between lattices and convex bodies and it is related with the classical Minkowski theorem about lattice points in a symmetric convex body of given volume. We present the result for the case of the cubic lattice, but the proof extends without changes to other orthogonal lattices.

THEOREM 1.13 (Nosarzewska [**56**]; Schmidt [**71**]; Bokowski, Hadwiger, Wills [**15**]). *Let* $\Gamma = \mathbb{Z}^n$ *be the integer cubic lattice and* $K \subset \mathbb{R}^n$ *a convex body. Then*

$$\#(\Gamma \cap K) \geq V(K) - \frac{1}{2} A(\partial K),$$

where $\#(X)$ *denotes the number of points of the set* X.

PROOF. Denote by C a generic unit cube in $\mathbb{R}^n$ of the type

$$C = \{(x_1, \dots, x_n) : k_i - \frac{1}{2} \leq x_i \leq k_i + \frac{1}{2}, \ i = 1, \dots, n\},$$

where $k_1, \dots, k_n$ are arbitrary integer numbers. The centers $p = (k_1, \dots, k_n)$ of the cubes are just the points of the lattice Γ.

If for some of these cubes C we have $V(K \cap C) > \frac{1}{2}$, then the center p of C belongs to K: to see that observe that otherwise, by separation properties of convex sets, there should exist a plane through p such that $K \cap C$ lies at one side of this plane, which would imply that the volume of $K \cap C$ is at most $\frac{1}{2}$ as in Figure 11.

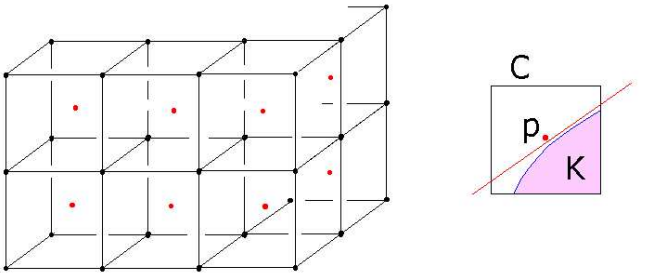


FIGURE 11. Lattices and convex bodies.

On the other hand, if $v = V(K \cap C) \leq \frac{1}{2}$, the isoperimetric inequality for the cube (see Theorem 1.9) gives

$$A(\partial K \cap C) \geq I_\gamma(v) \geq 4v(1-v) \geq 2v = 2V(K \cap C),$$

where we have used that the Gaussian profile satisfies the estimate $I_\gamma(t) \geq 4t(1-t)$; see §3.1 for more details.

Using the properties we have for the above two families of cubes, we conclude the proof as follows:

$$V(K) = \sum_C V(C \cap K) = \sum_{V(C\cap K) > \frac{1}{2}} V(C \cap K) + \sum_{V(C\cap K) \leq \frac{1}{2}} V(C \cap K)$$

$$\leq \#(\Gamma \cap K) + \sum_C A(\partial K \cap C)/2 \leq \#(\Gamma \cap K) + A(\partial K)/2.$$

□

For arbitrary lattices, an extended version of the above result is given in [**72**].

To finish this section we give an estimate for the area of triple periodic minimal surfaces. The result (and the proof) extend to higher dimensional rectangular tori.

THEOREM 1.14. *Let* Γ *be the orthogonal lattice generated by*

$$\{(a,0,0),\ (0,b,0),\ (0,0,c)\}, \qquad 0 < a \leq b \leq c,$$

and $\Sigma \subset \mathbb{R}^3/\Gamma$ *an embedded nonflat closed orientable minimal surface. Then* $A(\Sigma) > 2ab$.

PROOF. First observe that Σ separates the torus $\mathbb{R}^3/\Gamma$: otherwise the pullback image of Σ in $\mathbb{R}^3$ would be disconnected, which contradicts the maximum principle (see [**53**]).

Let Ω be one of the regions in the torus with $\partial\Omega = \Sigma$. Assume that $V(\Omega) \leq abc/2$ and take the enlargement Ω_r, $r \geq 0$, of Ω such that $V(\Omega_r) = abc/2$. Parameterizing $\mathbb{R}^3/\Gamma$ by the geodesics leaving Σ orthogonally we get $dV = (1 - tk_1)(1 - tk_2)\, dt\, dA$ and, if we define $C(r)$ as the set of points in Σ whose distance to $\partial\Omega_r$ is equal to r, we have

$$A(\partial\Omega_r) \leq \int_{C(r)} (1 - rk_1)(1 - rk_2)\, dA.$$

We remark that $(1 - rk_1)(1 - rk_2) \geq 0$ on $C(r)$. From Corollary 1.12 and the integral inequality above, combined with Schwarz's inequality, we obtain

$$2ab \leq A(\partial\Omega_r) \leq \int_\Sigma (1 - rH)^2\, dA = A(\Sigma),$$

where we have used that the mean curvature H of Σ vanishes.

Suppose the equality holds. If $r > 0$ Schwarz's inequality gives that Σ is flat. If $r = 0$, then Σ would be an isoperimetric region, which contradicts Corollary 1.12. Since both options are impossible, we conclude that $A(\Sigma) > 2ab$. □

Note that $\mathbb{R}^3/\Gamma$ contains flat 2-tori of area ab. The area of a P-Schwarz minimal surface in the unit cubic torus is ~ 2.34 (the theorem gives that this area is greater than 2). Note that, since the P-minimal surface is stable [**69**] and (by symmetry arguments) separates the torus in two equal volumes, it follows that in the cubic torus there are local minima of the isoperimetric problem which are not global ones. R. Kusner pointed out to me that by desingularizing the union of the two flat 2-tori induced by the planes $x = 0$ and $y = 0$ in the cubic unit torus (see Traizet [**77**]), we obtain embedded *nonorientable* nonflat compact minimal surfaces with area smaller than 2.

1.6. The periodic isoperimetric problem. The explicit description of the solutions of the isoperimetric problem in flat 3-tori (the so called *periodic isoperimetric problem*) is one of the nicest open problems in classical geometry. The natural conjecture in this case says that any solution is a sphere, a (quotient) of a cylinder or two parallel flat tori. For rectangular 3-tori $T^3 = \mathbb{R}^3/\Gamma$, where Γ is a lattice generated by three pairwise orthogonal vectors, we have a good control on the shape of the best candidates (see §1.5 and §2.1) and the conjecture seems to be well tested. For general flat 3-tori the information we have is weaker and the natural conjecture is not so evident (though we believe it). The problem extends naturally to flat 3-manifolds and to the quotient of $\mathbb{R}^3$ by a given crystallographic group.

A different interesting extension of the problem is the study of stable surfaces (i.e., local minima of the isoperimetric problem; see section §2.1) in flat three-tori. It is known that this family contains some beautiful examples (other than the sphere, the cylinder and the plane) such as the Schwarz P and D triply periodic minimal surfaces and Schoen minimal Gyroid (see Ross [**69**]), but exhaustive work in this direction has never been done. The most important restriction we know at this moment about a stable surface is that its genus is at most 4; see [**63**] and [**62**]. The situation where the lattice is not fixed, but is allowed to be deformed, could be geometrically significant too.

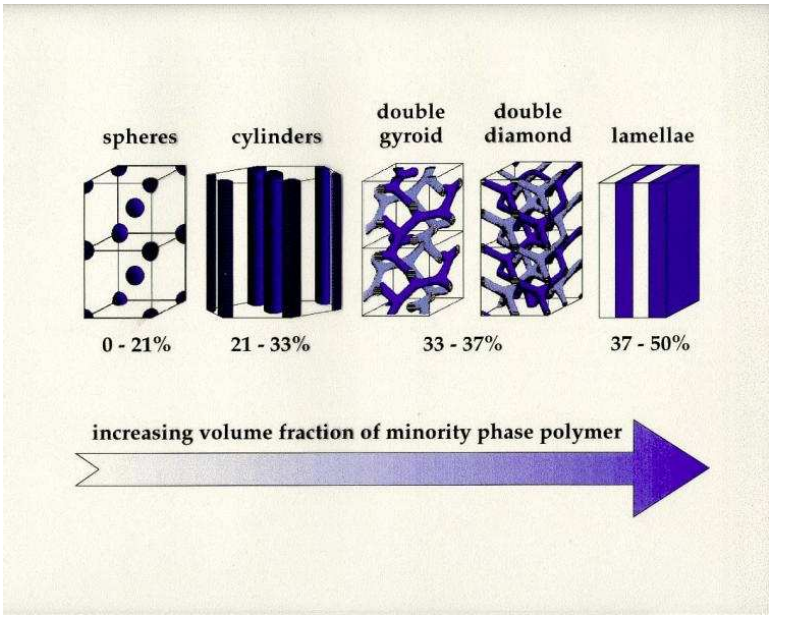


FIGURE 12. From Edwin Thomas' talk on material science, MSRI, Berkeley, 1999.

Nanogeometry. The periodic stability condition is the simplest geometric model for a group of important phenomena appearing in material sciences. Take two liquids which repel each other (like oil and water) and put them together in a vessel. The equilibrium position of the mixture corresponds to a local minimum of the energy of the system, which, assuming suitable ideal conditions, is given by

the area of the interface Σ separating both materials. Since the volume fraction of the liquids is given, we see that Σ is a local solution of the isoperimetric problem (that is, a stable surface). It happens that there are several pairs of substances of this type that (for *chemical reasons*) behave periodically. These periods hold at a very small scale (on the scale of nanometers) and the main shapes which appear in the laboratory, under our conditions, are described in Figure 12. The parallelism between these shapes and the periodic stable surfaces listed above is clear. For more details see [**75**] and [**35**]. Instead of the area, the bending energy (§2.3) is sometimes considered to explain these phenomena. However, from the theoretical point of view, this last functional is harder to study, and it does not seem possible at this moment to classify its extremal surfaces.

2. The Isoperimetric Problem for Riemannian 3-Manifolds

In this section M will be a 3-dimensional orientable compact manifold without boundary. The volume of M will be denoted by $v = V(M)$. Some of the results below are also true in higher dimension. However we will only consider the 3-dimensional case for the sake of clarity.

First we explain several facts about *stable* constant mean curvature surfaces. Some of these stability arguments allow us to solve the isoperimetric problem in the real projective space. Then we will study the *bending energy*, an interesting functional which can be related, in several ways, to the isoperimetric problem. The general properties of the isoperimetric profile and the isoperimetric inequality of Lévy and Gromov will be considered too.

2.1. The stability condition. Stability is a key notion in the study of the isoperimetric problem. In the following we will see several interesting consequences of this idea.

A closed orientable surface $\Sigma \subset M$ is a *stable surface* if it minimizes the area up to second order under a volume constraint. The stability of Σ is equivalent to the following facts (see [**5**]):

(1) Σ has constant mean curvature and
(2) $Q(u,u) \geq 0$ for any function u in the Sobolev space $H^1(\Sigma)$ with $\int_\Sigma u dA = 0$, Q being a quadratic form defined by the second variation formula for the area.

This quadratic form is given by

$$Q(u,u) = \int_\Sigma \left(|\nabla u|^2 - (\operatorname{Ric} N + |\sigma|^2) u^2 \right) \, dA, \tag{2–1}$$

where $\operatorname{Ric} N$ is the Ricci curvature of M in the direction of the unit normal vector to Σ and σ is the second fundamental form of the immersion. The function u corresponds to an infinitesimal normal deformation of the surface, and the zero mean value condition means that the deformation preserves the volume infinitesimally. The *Jacobi operator* $L = \Delta + \operatorname{Ric} N + |\sigma|^2$ is related to Q by the formula $Q(u,v) = -\int_\Sigma uLv \, dA$.

There are several notions of stability in the constant mean curvature context, but in this paper we will only use the one above. The most important class of stable surfaces is given by the isoperimetric surfaces.

If Σ had boundary $\partial\Sigma \subset \partial M$, then the quadratic form Q should be modified to

$$Q(u,u) = \int_\Sigma \left(|\nabla u|^2 - (\mathrm{Ric}\, N + |\sigma|^2)u^2\right)\, dA - \int_{\partial\Sigma} \kappa\, u^2\, ds, \tag{2–2}$$

where κ is the normal curvature of ∂M along N, see for instance [**68**].

If W is a region in $\mathbb{R}^3$ and $\Sigma \subset W$ has constant mean curvature and meets ∂W orthogonally, stability corresponds with soap bubbles which can really be blown in the experiment of Figure 3.

A simple consequence of (2–1) and (2–2) is that if $\mathrm{Ric} > 0$ and ∂M is convex, then any stable surface is connected. In the case $\mathrm{Ric} \geq 0$, Σ is either connected or totally geodesic. In fact, if Σ is not connected we can use as a test function u a locally constant function, and the above assertion follows directly.

Formulae (2–1) and (2–2) are valid in higher dimension and isoperimetric hypersurfaces $\Sigma \subset M$ satisfy the stability condition. Moreover locally constant functions on the regular set $R(\Sigma)$ of Σ belong to the Sobolev space $H^1(\Sigma) = H^1(R(\Sigma))$. The equality of the above spaces follows because the Hausdorff codimension of the singular set is large; see Theorem 1.1. Therefore we have the following result, which was used in the proof of Theorem 1.2.

LEMMA 2.1. *Let M be an n-dimensional Riemannian manifold and Σ an isoperimetric hypersurface. If* $\mathrm{Ric} \geq 0$ *and ∂M is convex, then either the regular set $R(\Sigma)$ of Σ is connected or Σ is totally geodesic.*

Although the idea of stability is naturally connected with the isoperimetric problem, it has not been widely used until recently. The pioneer result in this direction was the following one.

THEOREM 2.2 (Barbosa & do Carmo [**5**]). *Let Σ be a compact orientable surface immersed in $\mathbb{R}^3$ with constant mean curvature. If Σ is stable, then it is a round sphere.*

Consider the 3-dimensional box $W = [0,a_1]\times[0,a_2]\times[0,a_3]$, $0 < a_1 \leq a_2 \leq a_3$. Recall that any isoperimetric surface Σ in W gives an isoperimetric surface Σ' in the torus $T = \mathbb{R}^3/\mathrm{span}\{(2a_1,0,0),\,(0,2a_2,0),\,(0,0,2a_3)\}$ by taking reflections with respect to the faces of W. We remark that the argument below only uses the *stability* of Σ'.

PROPOSITION 2.3 (Ritoré [**60**]). *Let Σ be a nonflat isoperimetric surface in a box W. Then Σ has 1-1 projections over the three faces of W (as in Figure 9).*

PROOF. We work not with Σ but with Σ', which is also connected and orientable. Denote by s_i the symmetry in T induced by the symmetry with respect to the plane $x_i = 0$ in $\mathbb{R}^3$. Note that the set of points fixed by s_i consists of the tori $\{x_i = 0\}$ and $\{x_i = a_i\}$ and so it separates T. In particular, since Σ' is connected, $C_i = \{p \in \Sigma' : s_i(p) = p\}$ is nonempty and separates Σ' too.

Let $L = \Delta + |\sigma|^2$ be the Jacobi operator and $N = (N_1, N_2, N_3)$ the unit normal vector of Σ'. Then $u = N_i$ satisfies $u \not\equiv 0$, $Lu = 0$ and $u \circ s_i = -u$. Hence u vanishes along the curve of fixed points C_i. Assume, reasoning by contradiction, that u vanishes at some point outside C_i. Then u will have at least three nodal domains, that is $\{u \neq 0\}$ has at least three connected components Σ_α, $\alpha = 1,2,3$: to conclude this, use that nonzero solutions of the Jacobi equation $Lu = 0$ change sign near each point p with $u(p) = 0$.

Consider the functions u_α on Σ' defined by $u_\alpha = u$ on Σ_α and $u_\alpha = 0$ on $\Sigma' - \Sigma_\alpha$. Since $Q(u_1, u_2) = 0$, we can construct a nonzero function of the type $v = a_1 u_1 + a_2 u_2$ satisfying $Q(v, v) = 0$ and $\int_{\Sigma'} v\, dA = 0$. The stability of Σ' then implies that $Lv = 0$. Since v vanishes on Σ_3 we contradict the unique continuation property [**4**]. Therefore we have that u has no zeros in the interior of Σ and the proposition follows easily. □

See Proposition 1 in [**66**] for a more general argument of this type.

It can be shown (see [**60, 61**]) that any nonflat constant mean curvature surface in a box W with 1-1 projections into the faces of the box and which meets ∂W orthogonally, is either a piece of a sphere or lies in one of the two other families of nonflat surfaces that appear in Figure 9. The classical Schwarz P-minimal surface in the cube is the simplest hexagonal example. The first example of pentagonal type was constructed by Lawson [**46**]. Moreover Ritoré [**60, 61**] has proved that the moduli space of Schwarz (resp. Lawson) type surfaces is parameterized by the moduli space of hyperbolic hexagons (resp. pentagons) with right angles at each vertex. If we do not consider the restriction on the projections of the surface, then the family of constant mean curvature surfaces which meet ∂W orthogonally is much larger; see Große-Brauckmann [**31**].

Ross [**69**] has shown that the Schwarz P-minimal surface is stable in the cube. Lawson type surfaces can be constructed by soap bubbles experiments as in Figure 3. So, apparently some of them are stable too.

Another important stability argument depends on the existence of conformal spherical maps on compact Riemann surfaces. It was introduced by Hersch [**38**] and extended by Yang and Yau [**79**]. The version below can be found in Ritoré and Ros [**63**].

THEOREM 2.4. *Let M^3 be a compact orientable 3-manifold with $\partial M = \varnothing$ and Ricci curvature* $\mathrm{Ric} \geq 2$. *If Σ is an isoperimetric surface, then Σ is compact, connected and orientable of genus less than or equal to* 3. *Moreover, if genus*$(\Sigma) = 2$ *or* 3, *then* $(1 + H^2)A(\Sigma) \leq 2\pi$.

PROOF. Assume that there is a conformal (i.e., meromorphic) map $\phi : \Sigma \longrightarrow \mathbf{S}^2 \subset \mathbb{R}^3$ whose degree satisfies

$$\deg \phi \leq 1 + \left[\frac{g+1}{2}\right] \quad \text{and} \quad \int_\Sigma \phi\, dA = 0, \tag{2–3}$$

where $[x]$ denotes the integer part of x and $g = \mathrm{genus}\,\Sigma$. We will take ϕ as a test function for the stability condition. The Gauss equation transforms $\mathrm{Ric}\, N + |\sigma|^2$ into $\mathrm{Ric}\, e_1 + \mathrm{Ric}\, e_2 + 4H^2 - 2K$, where e_1, e_2 is an orthonormal basis of the tangent plane to Σ and H and K are the mean curvature and the Gauss curvature, respectively. Using the Gauss–Bonnet theorem and the facts that $|\phi| = 1$ and $|\nabla \phi|^2 = 2\,\mathrm{Jac}\,\phi$, we get

$$\begin{aligned} 0 \leq Q(\phi, \phi) &= \int_\Sigma |\nabla \phi|^2 - (\mathrm{Ric}\, N + |\sigma|^2) \\ &= 8\pi \deg \phi - \int_\Sigma \left(\mathrm{Ric}\, e_1 + \mathrm{Ric}\, e_2 + 4H^2 - 2K\right) \\ &\leq 8\pi(1 + [\tfrac{g+1}{2}]) - 4(1 + H^2)A(\Sigma) + 8\pi(1 - g), \end{aligned}$$

which implies the desired result.

It remains to show that a map ϕ satisfying (2–3) exists. The Brill–Noether theorem (see [**27**], for instance) guarantees the existence of a nonconstant meromorphic map $\phi : \Sigma \to \mathbf{S}^2$ whose degree satisfies (2–3). Now we prove that ϕ can be modified, by composition with a suitable conformal transformation of the sphere, in order to get a map with mean value zero. Let G be the group of conformal transformations of $\mathbf{S}^2$ (which coincides with the conformal group of $O = \{x \in \mathbb{R}^3 : |x| < 1\}$). It is easy to see that there exists a unique smooth map $f : x \mapsto f_x$ from O into G satisfying these conditions:

(i) $f_0 = \mathrm{Id}$ is the identity map.
(ii) For any $x \neq 0$, $f_x(0) = x$ and $d(f_x)_0 = \lambda_x \mathrm{Id}$ for some $\lambda_x > 0$.
(iii) Given $y \in \mathbf{S}^2$, we have that, when $x \to y$, f_x converges almost everywhere to the constant map y.

Define the continuous map $F : O \to O$ by

$$F(x) = \frac{1}{A(\Sigma)} \int_\Sigma f_x \circ \phi \, dA.$$

Property (iii) implies that F extends continuously to the closed balls, $F : B^3 \to B^3$ such that $F(y) = y$ for all y with $|y| = 1$. Then, by a clear topological argument, F is onto and, in particular, there exists $x \in O$ such that

$$\int_\Sigma f_x \circ \phi \, dA = 0.$$

2.2. The 3-dimensional projective space. The isoperimetric problem for 3-dimensional space-forms, that is, complete 3-spaces with constant sectional curvature, is a very interesting particular situation. The simply connected case is solved by symmetrization, but if the fundamental group is nontrivial this method does not work or gives little information. It is natural to hope that the complexity of the solutions will depend closely on the complexity of the fundamental group. However the following theorem gives restrictions which are valid in any 3-space-form.

THEOREM 2.5. *Let Σ be an isoperimetric surface of an orientable 3-dimensional space-form $M(c)$ with constant curvature c.*

(a) *If Σ has genus zero, then Σ is an umbilical sphere.*
(b) *If genus$(\Sigma) = 1$, then Σ is flat.*

PROOF. In the first case, the result follows from the Hopf uniqueness theorem [**39**] for constant mean curvature spheres. Assertion (b) is proved in [**63**] as follows. Assume Σ is nonflat. It is well known that since Σ is of genus 1, its geometry is controlled by the sinh-Gordon equation: there is a flat metric ds_0^2 on Σ and a positive constant a such that the metric on Σ can be written as $ds^2 = a\, e^{2w} ds_0^2$ where w satisfies $\Delta_o w + \sinh w \cosh w = 0$ (Δ_0 is the Laplacian of the metric ds_0^2). Moreover the quadratic form Q, in term of the flat metric is given by

$$Q(u,u) = \int_\Sigma \left(|\nabla_0 u|^2 - (\cosh^2 w + \sinh^2 w) u^2 \right) dA_0.$$

Take $\Omega \subset \Sigma$ a nodal domain of w (i.e., a connected component of $\{w \neq 0\}$) and consider the function $u = \sinh w$ in Ω and $u = 0$ in $\Sigma - \Omega$. Direct computation,

using the sinh-Gordon equation, gives $-u\Delta_0 u = (\cosh^2 w - |\nabla_0 w|^2)\sinh^2 w$ and therefore

$$\begin{aligned} Q(u,u) &= \int_\Omega \left(-u\Delta_0 u - (\cosh^2 w + \sinh^2 w)u^2\right) dA_0 \\ &= -\int_\Omega \left(|\nabla_0 w|^2 + sinh^2 w)sinh^2 w dA_0 < 0. \end{aligned}$$

On the other hand, at every point, the sign of w coincides with the sign of the Gauss curvature of Σ and, since Σ is nonflat, the Gauss–Bonnet theorem gives that w has at least two nodal domains. Thus we have on Σ two functions u_1 and u_2 satisfying $Q(u_1,u_1) < 0$ and $Q(u_2,u_2) < 0$. Since the supports of these functions are disjoint we get $Q(u_1,u_2) = 0$ and so a linear combination of the u_i will contradict the stability of Σ. □

As an application of the last theorem we solve the isoperimetric problem in a slab.

COROLLARY 2.6 ([**63**]). *Isoperimetric surfaces in a slab of* $\mathbb{R}^3$ *and in the flat three manifold* $\mathbf{S}^1 \times \mathbb{R}^2$ *are round spheres and flat cylinders.*

PROOF. The symmetrization argument in §1.3 gives that these two isoperimetric problems are equivalent and that any solution must be a surface of revolution. In particular, any isoperimetric surface in $\mathbf{S}^1 \times \mathbb{R}$ must have genus 0 or 1, and the result follows from Theorem 2.5 above. □

By combining various of the results we have already proved, we can give a complete solution of the isoperimetric problem in the projective space $\mathbf{P}^3 = \mathbf{S}^3/\{\pm\}$, or in other words, we can solve the isoperimetric problem for antipodally symmetric regions on the 3-sphere. As far as I know the only nontrivial 3-space forms where we can solve the problem at this moment are $\mathbf{P}^3$, the Lens space obtained as a quotient of $\mathbf{S}^3$ by the cube roots of the unity (see §2.3) and the quotient of $\mathbb{R}^3$ by a screw motion (see [**63**]). In the first two cases we cannot use symmetrization. The solution in the projective space depends on stability arguments.

THEOREM 2.7 (Ritore & Ros [**63**]). *Isoperimetric surfaces of* $\mathbf{P}^3$ *are geodesic spheres and tubes around geodesics.*

PROOF. Let $\Sigma \subset \mathbf{P}^3$ be an isoperimetric surface. From Theorem 2.4, Σ must be connected, orientable and genus$(\Sigma) \leq 3$.

If genus$(\Sigma) = 0, 1$, then, thanks to Theorem 2.5, Σ must be umbilical or flat. Such surfaces are easily classified and coincide with the ones in the statement of the theorem.

Let us see that the genus$(\Sigma) = 2, 3$ case cannot hold. The last part of Theorem 2.4 gives $(1 + H^2)A(\Sigma) \leq 2\pi$. On the other hand, for any closed surface in the 3-sphere $\widetilde{\Sigma} \subset \mathbf{S}^3$ with mean curvature $\widetilde{H}$ we have the Willmore inequality $\int_{\widetilde{\Sigma}}(1 + \widetilde{H}^2) \geq 4\pi$ with the equality holding only if $\widetilde{\Sigma}$ is an umbilical sphere (see [**78**] and §2.3 below). If we take as $\widetilde{\Sigma}$ the pullback image of Σ to $\mathbf{S}^3$, we obtain

$$4\pi \geq 2(1 + H^2)A(\Sigma) = \int_{\widetilde{\Sigma}}(1 + H^2) \geq 4\pi.$$

Then $\widetilde{\Sigma}$ is a round sphere, which is impossible because $\widetilde{\Sigma}$ must enclose an antipodally symmetric region. □

Direct computations give that for small (resp. large) volumes the solution is a geodesic sphere (resp. the complement of a geodesic sphere). For intermediate values of the volume the isoperimetric solution is a tube around a geodesic. In particular we have the following equivalent results.

COROLLARY 2.8. (a) *Suppose* Σ *divides* $\mathbf{P}^3$ *into two pieces of equal volume. Then* $A(\Sigma) \geq \pi^2$ *and equality holds if and only if* Σ *is the minimal Clifford torus.*

(b) *Suppose* Σ *divides* $\mathbf{S}^3$ *in two antipodally symmetric pieces of equal volume. Then* $A(\Sigma) \geq 2\pi^2$ *and equality holds if and only if* Σ *is the minimal Clifford torus.*

2.3. Isoperimetry and bending energy. Consider the functional which associates to each compact surface Σ in the Euclidean space $\mathbb{R}^3$ the integral of its squared mean curvature, $\int_\Sigma H^2\, dA$. This expression is called the *bending energy* of Σ and has a rich structure. It is natural to ask what are its minima under various topological restrictions. The global minimum is 4π, which is attained only for the round sphere. An interesting open question is to decide if the minimum for tori is given by a certain anchor ring, which has energy equal to $2\pi^2$ (Willmore [**78**]). Since the functional is compatible with conformal geometry, if Σ is viewed as a surface in the sphere $\mathbf{S}^3$ (via the stereographic projection), then the bending energy is given by

$$\int_\Sigma (1 + H^2)\, dA, \tag{2–4}$$

where now H denotes the mean curvature in the sphere. The Willmore conjecture says that the minimum of (2–4) for tori equals $2\pi^2$ and is attained by the minimal Clifford torus in $\mathbf{S}^3$; see [**66**] for more information about the current state of this question. Here we are interested in the connections between bending energy and the isoperimetric problem. We have seen in the proof of Theorem 2.7 how the first one is useful in understanding the second one, and we will see relations in the other direction too.

Given a smooth region Ω on M with $\partial\Omega = \Sigma$, we can parameterize $M - \Omega$ (outside a closed subset of measure zero) by geodesics leaving Σ orthogonally, i.e., by the normal exponential map $\exp_\Sigma$; see [**37**]. Then $dV = g(p,t)\, dt\, dA$, where $g(p,t)$ is the Jacobian of $\exp_\Sigma$, $p \in \Sigma$ and $t \in \mathbb{R}$. The *cut function* $c : \Sigma \to \mathbb{R}$ is a continuous function which associates to each point p the largest $t > 0$ such that $d(\alpha(t), \Sigma) = t$, where $\alpha(t)$ is the geodesic with $\alpha(0) = p$ which leaves Ω orthogonally and d denotes the Riemannian distance in M.

For any $r \geq 0$ we can estimate the area of the boundary of the enlargement $\Omega_r = \{p \in M : d(p,\Omega) \leq r\}$ as follows: if we introduce the set $C(r) = \{p \in \Sigma : d(p, \partial\Omega_r) = r\} \subset \{p \in \Sigma : c(p) \geq r\}$, then we have

$$A(\partial\Omega_r) \leq \int_{C(r)} g(p,r)\, dA. \tag{2–5}$$

PROPOSITION 2.9 ([**65**]). *Let* $\Sigma \subset M$ *be a closed surface which separates an ambient three-manifold* M *with* $\mathrm{Ric} \geq 2$. *Then*

$$\int_\Sigma (1 + H^2)\, dA \geq \max I_M = I_M(v/2).$$

If the equality holds, then either M *is minimal or totally umbilical.*

PROOF. The last equality will be proved in section §2.4. Let $\Omega \subset M$ be one of the regions bounded by Σ and assume that $V(\Omega) \leq v/2$. Consider the enlargement Ω_r, $r \geq 0$, such that $\mathrm{V}(\Omega_r) = v/2$. The comparison theorem of Heintze and Karcher [**37**], combined with Schwarz's inequality, gives that, if $p \in C(r)$, then the Jacobian of the normal exponential map satisfies $g(p,r) \leq (\cos r - H \sin r)^2 \leq 1 + H^2$. Therefore, formula (2–5) gives

$$A(\partial\Omega_r) \leq \int_M (1 + H^2)\, dA.$$

If the equality holds, then Schwarz's inequality implies that M is umbilical (if $r > 0$) or minimal (if $r = 0$). □

PROPOSITION 2.10 ([**66**]). *Let M be a compact orientable 3-manifold with* $\mathrm{Ric} \geq 2$ *whose profile satisfies* $\max I_M > 2\pi$. *Then any isoperimetric surface of M is either a sphere or a torus.*

PROOF. Let Σ be an isoperimetric surface of M with mean curvature H. From Proposition 2.9 we obtain

$$(1 + H^2)A(\Sigma) \geq \max I_M > 2\pi.$$

If $\mathrm{genus}(\Sigma) \geq 2$, then Theorem 2.4 implies $(1 + H^2)A(\Sigma) \leq 2\pi$, which contradicts the last inequality. □

Let $\mathbb{Z}_n$ be the group of the n^{th} roots of unity. This group acts on the unit 3-sphere $\mathbf{S}^3 \subset \mathbb{C}^2$ by complex multiplication $\lambda(z_1, z_2) = (\lambda z_1, \lambda z_2)$, for any $\lambda \in \mathbb{Z}_n$ and $(z_1, z_2) \in \mathbf{S}^3$. The quotient spaces $\mathbf{L}_n = \mathbf{S}^3/\mathbb{Z}_n$ are called *lens spaces*. Note that $\mathbf{L}_2$ is just the projective space. The lens space $\mathbf{L}_3$ is the unique elliptic 3-space-form with fundamental group equal to $\mathbb{Z}_3$.

In section §3.3 we will show certain isoperimetric inequalities for product measures. These results can be extended to fiber bundles; see [**67**], and imply in particular that $\max I_{\mathbf{L}_3} > 2\pi$. Therefore Proposition 2.10 combined with Theorem 2.5 allows us to solve completely the isoperimetric problem for the lens space $\mathbf{L}_3$.

THEOREM 2.11 ([**67**]). *The isoperimetric surfaces of the lens space $\mathbf{L}_3 = \mathbf{S}^3/\mathbb{Z}_3$ are geodesic spheres and tubes around geodesics. In particular, the (quotient) of the minimal Clifford torus has least area among surfaces which separate $\mathbf{L}_3$ in two equal volumes.*

THEOREM 2.12 (Willmore [**78**]). *For any compact surface Σ immersed in the unit 3-sphere we have $\int_\Sigma (1 + H^2)\, dA \geq 4\pi$ and the equality holds if and only if Σ is a metric sphere.*

PROOF. If Σ is not embedded it was proved by Li and Yau [**51**] that the bending energy is larger than 8π. If Σ is embedded, it separates $\mathbf{S}^3$. Since the maximum of the isoperimetric profile of the 3-sphere is 4π, the claim follows from Proposition 2.9. □

Similarly, using the solution of the isoperimetric problem in the projective space, we can prove the Willmore conjecture in the antipodally invariant case.

THEOREM 2.13 (Ros [**65**]). *Let $\Sigma \subset \mathbf{S}^3$ be an antipodally invariant closed surface of odd genus. Then*

$$\int_\Sigma (1 + H^2)\, dA \geq 2\pi^2.$$

If the equality holds, then Σ is the minimal Clifford torus.

PROOF. The odd genus assumption implies that the quotient surface $\Sigma' = \Sigma/\pm$ separates projective space $\mathbf{P}^3$; see [**65**]. Therefore using Proposition 2.9 and Theorem 2.7 we obtain

$$\int_\Sigma (1+H^2)\,dA = 2\int_{\Sigma'} (1+H^2)\,dA \geq 2\max I_{\mathbf{P}^3} = 2\pi^2.$$

The equality case follows from Proposition 2.9 and Corollary 2.8. □

Another proof of this theorem has been given by Topping [**76**]. The Willmore conjecture has been also proved for tori in $\mathbb{R}^3$ that are symmetric with respect to a point; see Ros [**66**].

2.4. The isoperimetric profile. In this section we will consider some general properties of the isoperimetric profile $I = I_M : (0, v) \to \mathbb{R}$ of a compact Riemannian 3-manifold M, $I(t) = \inf\{\mathrm{A}(\partial\Omega) : \Omega \subset M \text{ region, } V(\Omega) = t\}$, where $v = V(M)$. Since the complement of an isoperimetric region is also an isoperimetric region, we get $I(v-2t) = I(t)$, for any t. Items (a) and (b) in the next theorem were first proved by Bavard and Pansu [**9**]. Berard and Meyer [**10**] proved (2–8) by using a different argument. The isoperimetric profile has been also studied by Gallot [**25**] and Hsiang [**41**].

THEOREM 2.14. *Given t, $0 < t < v$, let Ω be any isoperimetric region in M with $V(\Omega) = t$ and $\partial\Omega = \Sigma$. The isoperimetric profile I of M satisfies the following properties.*

(a) *I has left and right derivatives $I'_+(t)$ and $I'_-(t)$, for any $0 < t < v$. Moreover if H is the mean curvature of Σ, then*

$$I'_+(t) \leq 2H \leq I'_-(t). \tag{2–6}$$

(b) *If σ denotes the second fundamental form of Σ, then the second derivative of I satisfies (in the weak sense)*

$$I''(t)I(t)^2 + \int_\Sigma \left(\mathrm{Ric}\, N + |\sigma|^2\right) \leq 0. \tag{2–7}$$

(c) *For small t, any isoperimetric region of volume t is a convex region contained in a small neighborhood of some point of M. In particular*

$$I(t) \sim (36\pi t^2)^{1/3} \quad \textit{when } t \to 0. \tag{2–8}$$

PROOF. We will only show assertion (c). Let Σ_n be a sequence of isoperimetric surfaces enclosing volumes $t_n \to 0$. We discuss the following possibilities.

1. *The curvature of the sequence Σ_n is unbounded.* By homothetically expanding the surfaces Σ_n, we obtain a sequence $\Sigma'_n \subset B(r_n)$ of properly embedded constant mean curvature surfaces in $B(r_n) = \{x \in \mathbb{R}^3 : |x| \leq r_n\}$, endowed with a metric ds_n^2 which converges smoothly to the usual Euclidean metric. Moreover $r_n \to \infty$, $0 \in \Sigma'_n$ for each n and the second fundamental form of the surface satisfies $\max |\sigma_{\Sigma'_n}|^2 = |\sigma_{\Sigma'_n}|^2(0) = 2$.

Since $\sigma_{\Sigma'_n}$ is bounded, then locally the surface Σ'_n consists of a certain number of sheets as in Figure 13. Each one of these sheets is a graph over a planar domain of a function with bounded derivatives. If two sheets become arbitrarily close nearby some point when n goes to infinity, then we can modify easily the surface Σ_n in

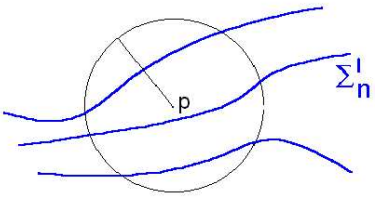


FIGURE 13. The surface Σ'_n is given locally, and independently of n, by graphs of functions with bounded derivatives.

order to get a new surface with least area and enclosing the same volume. In Figure 14 we have a scheme of these modifications in various cases. In the cases (a) and (b), two sheets come together and the enclosed region lies at different sides of the surface. If several sheets of the surface are involved, then we can use (c).

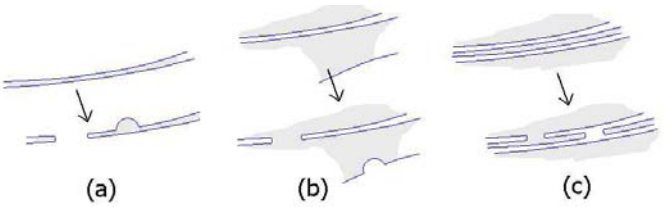


FIGURE 14. If two sheets of Σ'_n become arbitrarily close we can reduce area without modifying the enclosed volume, which contradicts the minimizing property of the surface Σ_n.

Hence, since Σ'_n has constant mean curvature, standard compactness results (see [**59**], for instance), given that (after taking a subsequence) Σ'_n converges smoothly and with multiplicity 1 to a properly embedded surface $\Sigma' \subset \mathbb{R}^3$ of constant mean curvature $H_{\Sigma'}$, where now $\mathbb{R}^3$ is endowed with the Euclidean metric, with $0 \in \Sigma'$ and $|\sigma_{\Sigma'}|^2(0) = 2$.

The fact that the surfaces Σ_n minimize area under volume constraint allow us to conclude that the surface Σ' (which is orientable) satisfies the following form of the stability condition (which applies also to noncompact surfaces):

$$(2\text{–}9) \qquad Q_{\Sigma'}(u,u) \geq 0, \quad \forall\, u \in C^\infty(\Sigma') \text{with compact support and } \int_{\Sigma'} u\, dA = 0,$$

$Q_{\Sigma'}$ being the quadratic form defined in (2–1). This condition is known to imply that Σ' is either a union of parallel planes or a round sphere; see da Silveira [**23**] or López Ros [**52**]. Since the component of Σ' passing through the origin is nonflat, we have that Σ' is a sphere (of radius 1). Therefore, for n large enough, a certain component of Σ_n can be written, after suitable rescaling, as a graph over a round sphere and the function which defines the graph converges to zero. If the surfaces Σ_n were connected, assertion (c) would be proved.

If Σ_n were disconnected we could repeat the argument above and produce another limit sphere by using a different component of Σ_n. Note that the union of these two limit spheres is unstable and this contradiction proves that Σ_n is connected for n large enough.

2. *The curvature of* $\{\Sigma_n\}$ *is bounded.* In this case we rescale the surfaces so that they enclose a volume equal to 1. By taking limits in this situation we will produce a family of pairwise disjoint planes in $\mathbb{R}^3$ enclosing a finite volume and this contradiction finishes the proof. □

We list some direct consequences of the theorem.

(a) I is locally Lipschitz in $(0, v)$ and extends continuously to 0 and v, $I(0) = I(v) = 0$.

(b) If $\mathrm{Ric}_M \geq 0$, then I is concave (use (2–7)). As the profile is symmetric, it follows in particular that I attains its maximum at $t = v/2$.

(c) In 3-space forms, solutions of the isoperimetric problem for small prescribed volumes are round spheres: lift these solutions to the universal covering and either compare area-volume with the ones of the metric spheres or use the Alexandrov theorem [**1**].

(d) Theorem 2.14 extends to 3-manifolds M such that M/G is compact, where G is the group of isometries of M. Using this extension we can prove the isoperimetric property of the spheres in $\mathbb{R}^3$ in a new way: It is clear that in $\mathbb{R}^3$ solutions for different volumes differ by a homothety. Therefore the isoperimetric profile of the Euclidean space given by $I(t) = ct^{\frac{2}{3}}$ for a certain constant $c > 0$. If Σ is an isoperimetric region enclosing a volume t, from (2–6) and Schwarz inequality we have $I'(t)^2 = 4H^2 \leq 2|\sigma|^2$, which combined with (2–7) gives

$$I''(t)I(t)^2 + \tfrac{1}{2}I'(t)^2 I(t) \leq 0. \tag{2–10}$$

But using the explicit expression of I we obtain in fact the equality in (2–10), and this implies that Σ is totally umbilical.

The arguments used in the proof of item (c) in Theorem 2.14 also show the following result.

PROPOSITION 2.15. *If* Σ_n *is a sequence of isoperimetric surfaces of* M *enclosing volumes* $t_n \to t$, $0 < t < v$, *then (up to a subsequence)* Σ_n *converges smoothly and with multiplicity 1 to an isoperimetric surface* Σ *enclosing a volume* t.

As a consequence of this proposition, for each t, $0 < t < v$, there exist isoperimetric surfaces Σ_+ and Σ_- enclosing regions of volume t and whose mean curvatures are given by $I'_+(t)/2$ and $I'_-(t)/2$, respectively. In particular, if for a certain prescribed volume t, there is a unique solution of the isoperimetric problem, then the profile I is differentiable at t.

2.5. Lévy–Gromov isoperimetric inequality. We present here one of the basic isoperimetric inequalities in Riemannian geometry. It was first proved by Lévy for convex hypersurfaces in Euclidean space. The proof below contains, in particular, the solution of the isoperimetric problem in the sphere and therefore, by an easy limit argument, gives the sharp Euclidean isoperimetric inequality too. An improved version, which involves bounds on the diameter of the manifolds and also works partially for the negative curvature case, is given in Berard, Besson and Gallot [**11**]. The argument applies in any dimension, in spite of the possible existence of singularities for solutions of the isoperimetric problem.

THEOREM 2.16 (Lévy [**50**]; Gromov [**28**]). *Let* M^3 *be a compact Riemannian manifold with Ricci curvature greater than or equal to the one of the sphere* $\mathbf{S}^3(r)$

of radius r, $\mathrm{Ric} \geq 2/r^2$. *Denote by μ_M and ϑ the normalized Riemannian measures of M and $\mathbf{S}^3(r)$ respectively. Then $I_{\mu_M} \geq I_\vartheta$. Moreover, $I_{\mu_M}(s) = I_\vartheta(s)$ for some $0 < s < 1$ implies $M = \mathbf{S}^3(r)$.*

PROOF. Normalize so that $r = 1$. Let $\Sigma \subset M$ be an isoperimetric surface enclosing a region Ω with $V(\Omega) = sV(M)$, for some $s \in (0,1)$, and mean curvature H (with respect to the inner normal). Consider a metric sphere $\Sigma_0 \subset \mathbf{S}^3$ bounding a spherical ball Ω_0 with $V(\Omega_0) = sV(\mathbf{S}^3)$ and mean curvature H_0.

Assume $H \geq H_0$ and parameterize both regions Ω, Ω_0 by geodesics leaving the boundary orthogonally. Then $dV = g(p,t)\,dt\,dA$ and $dV_0 = g_0(t)\,dt\,dA_0$, where the subscript 0 denotes the spherical counterpart. In the sphere, the Jacobian of the exponential map is given by $g_0(t) = (\cos t - H_0 \sin t)^2$.

Consider the cut functions c and c_0 of Σ and Σ_0 respectively (note that now we are not parameterizing $M - \Omega$, as in §2.3, but Ω itself). Clearly c_0 is constant. Under our hypothesis, the comparison theorem of Heintze and Karcher [**37**] says that $c(p) \leq c_0$ and $g(p,t) \leq g_0(t)$. Therefore

$$\begin{aligned} V(\Omega) &= \int_\Sigma \int_0^{c(p)} g(p,t)\,dt\,dA \leq \int_\Sigma \int_0^{c_0} g_0(t)\,dt\,dA \\ &= A(\Sigma) \int_0^{c_0} g_0(t)\,dt = \frac{A(\Sigma)V(\Omega_0)}{A(\Sigma_0)}, \end{aligned}$$

which just says that $I_\vartheta(s) \leq I_{\mu_M}(s)$. In the case $H \leq H_0$ the argument above applies to the complementary regions $M - \Omega$ and $\mathbf{S}^3 - \Omega_0$ (whose mean curvatures are $-H$ and $-H_0$, respectively) and gives $I_\vartheta(1-s) \leq I_{\mu_M}(1-s)$. So we obtain the desired inequality by using the symmetry of the profile.

If the equality holds we conclude that $H = H_0$, so the argument applies to both Ω and $M - \Omega$. Moreover we get $c = c_0$ and $g = g_0$. The remaining part of the proof follows from [**37**]. □

If $\mathrm{Ric}_M \geq 2$, the comparison result in [**37**] implies that $\mathrm{V}(M) \leq \mathrm{V}(\mathbf{S}^3)$ and the equality characterizes the round sphere. We can prove that, when the volume of M is large, we have strong control on the topology of isoperimetric surfaces.

COROLLARY 2.17 (Ros [**66**]). *Let M be a compact 3-manifold with* $\mathrm{Ric} \geq 2$. *If $V(M) \geq V(\mathbf{S}^3)/2$, then any isoperimetric surface of M is homeomorphic either to a sphere or to a torus.*

PROOF. Recall that I_M denotes the profile with respect to the (unnormalized) Riemannian measure. Assuming M is not a round sphere, Theorem 2.16 gives for $t = \frac{1}{2}$,

$$\frac{\max I_M}{v} = \frac{I_M(v/2)}{v} > \frac{4\pi}{V(\mathbf{S}^3)} = 2\pi,$$

where $v = V(M)$, and the corollary follows from Proposition 2.10. □

Concerning the topology of isoperimetric surfaces Σ in a positively curved ambient space M, we conjecture that if M is a 3-sphere with a metric of positive Ricci curvature (resp. a strictly convex domain $\mathbb{R}^3$), then Σ is homeomorphic to a sphere (resp. a disc).

3. The Isoperimetric Problem for Measures

In this section we first study the Gaussian measure and then we prove, by extending the symmetrization of Steiner [**74**] and Schwarz [**70**], a comparison isoperimetric inequality for product spaces. This comparison is sharp and allows us to obtain important geometric consequences. In particular, we will see that the isoperimetric problem for a product of two compact Riemannian manifolds always admits some solution of slab type [**8**]. It is well-known that, in a Riemannian manifold, the isoperimetric inequality is equivalent to the Sobolev inequality and gives bounds on the eigenvalues of the Laplacian; see [**19, 25**]. Here we will see that isoperimetric inequalities of Gaussian type imply three fundamental analytic inequalities.

In this section μ will be an absolutely continuous probability measure on a Riemannian manifold M^n: $d\mu = \varphi\, dV$ and $\mu(M) = 1$. Moreover, we will assume that $\mu(D) > 0$ for any nonempty open subset $D \subset M$. We will say that (M, μ) is a *probability space* and we will denote its isoperimetric profile by I_μ. Note that $I_\mu(0) = I_\mu(1) = 0$. The hypotheses about μ could be relaxed in some of the results below, but we have chosen to work in a comfortable context for expository reasons.

If M has finite volume, we can consider its *Riemannian probability* μ_M defined as the normalized Riemannian measure $d\mu_M = dV/V(M)$.

3.1. The Gaussian measure. The Gaussian measure $\gamma = \gamma_n$ on $\mathbb{R}^n$ is the probability measure defined by

$$d\gamma = \exp(-\pi|x|^2)dV.$$

Classically, γ_1 appears in the *central limit theorem*: consider in the discrete n-dimensional cube $\{1, -1\}^n \subset \mathbb{R}^n$ the normalized counting measure μ_n. Then the orthogonal projection of μ_n onto the main diagonal of the cube $\mathbb{R} = \operatorname{span}\{(1, \ldots, 1)\}$ converges to the Gaussian measure on $\mathbb{R}$ (up to suitable renormalization).

Let us see that the same holds if we project the normalized spherical measure in dimension m onto a given n-dimensional subspace and take a limit as m goes to ∞. Let $c_m = V(\mathbf{S}^m)$ be the volume of the m-dimensional unit sphere and $\rho_m = c_{m-1}/c_m$. Among the properties of the the sequence $\{c_m\}$ we mention that

$$(m-1)c_m = 2\pi c_{m-2} \quad \text{and} \quad \rho_m \sim \sqrt{\frac{m}{2\pi}} \quad \text{when } m \to \infty. \tag{3–1}$$

Denote by ϑ_m the Riemannian probability on the sphere $\mathbf{S}^m(\rho_m)$. The radius ρ_m is chosen so that the profile of ϑ_m satisfies $I_{\vartheta_m}(\frac{1}{2}) = 1$. Given $n \leq m$, we consider the projection $p : \mathbf{S}^m(\rho_m) \to B^n(\rho_m)$, $p(x, y) = x$, and the image measure $\sigma_{m,n} = p(\vartheta_m)$ induced on the ball $B^n(\rho_m) = \{x \in \mathbb{R}^n : |x| \leq \rho_m\}$. Then, for any region R in the ball $B^n(\rho_m)$, we have

(3–2)

$$\sigma_{m,n}(R) = \vartheta_m(p^{-1}(R)) = \frac{c_{m-n}}{c_m \rho_m^n} \int_R \left(1 - \frac{|x|^2}{\rho_m^2}\right)^{-(n+1)/2} \left(1 - \frac{|x|^2}{\rho_m^2}\right)^{m/2} dV.$$

Taking limits in the expression (3–2), and using the relation on the right in (3–1), we obtain the following result (sometimes attributed to Poincaré):

PROPOSITION 3.1. *When $m \to \infty$, the projection $\sigma_{m,n} = p(\vartheta_m)$ of the spherical probability ϑ_m converges to the Gaussian measure γ_n on $\mathbb{R}^n$.*

Note that, in the case $n = m - 1$, the integrand in (3–2) is constant. Thus we get the following known result.

LEMMA 3.2. *The projection $p : \mathbf{S}^m(\rho_m) \to B^{m-1}(\rho_m)$ transforms the spherical probability ϑ_m in the Riemannian probability β_{m-1} of the ball $B^{m-1}(\rho_m)$.*

In particular, from Proposition 1.7 we get that the profiles of the sphere and the ball are related by $I_{\beta_{m-1}} \geq I_{\vartheta_m}$. This inequality is sharp because $I_{\beta_{m-1}}(\frac{1}{2}) = I_{\vartheta_m}(\frac{1}{2})$.

Proposition 3.1 allows us to solve the isoperimetric problem in the Gaussian space as a limit of the spherical isoperimetric problem (note that the same holds for the Euclidean space).

THEOREM 3.3 (Borell [**17**]; Sudakov & Tsirel'son [**73**]). *Among regions $R \subset \mathbb{R}^n$ with prefixed $\gamma(R)$, half spaces minimize the Gaussian area.*

PROOF. Let $R \subset \mathbb{R}^n$ be a closed subset and $E = \{x_n \leq a\} \subset \mathbb{R}^n$ a half-space with $\gamma(E) = \gamma(R)$. If we fix $b < a$ and we consider the half-space $F = \{x_n \leq b\}$, then for m large enough we have

$$\vartheta_m(p^{-1}(R)) > \vartheta_m(p^{-1}(F)), \tag{3–3}$$

where $p : \mathbf{S}^m(\rho_m) \to \mathbb{R}^n$ is the orthogonal projection of the sphere to the linear subspace $\mathbb{R}^n$.

Note that $p^{-1}(F)$ is an isoperimetric region of the spherical probability ϑ_m. Moreover, for this measure, the isoperimetric inequality can be written in the following integral form (see §3.2 below): if $Q \subset \mathbf{S}^n(\rho_m)$ is a closed subset and G is a metric ball in the same sphere with $\vartheta_m(Q) \geq \vartheta_m(G)$, then, for any $r > 0$, the r-enlargements of Q and G satisfy

$$\vartheta_m(Q_r) \geq \vartheta_m(G_r). \tag{3–4}$$

Since the projection p reduces the distance we see that the r-enlargement of R and the one of its pullback image are related by

$$p^{-1}(R_r) \supset \{p^{-1}(R)\}_r. \tag{3–5}$$

Therefore, (3–3), (3–4) and (3–5) give

$$\vartheta_m(p^{-1}(R_r)) \geq \vartheta_m(\{p^{-1}(F)\}_r). \tag{3–6}$$

Taking limits in (3–6) when m goes to infinity we obtain, from Proposition 3.1,

$$\gamma_n(R_r) \geq \gamma_n(F_r).$$

The left term of this inequality is clear. To get the right one we have used that, when m is large, the sphere $\mathbf{S}^m(\rho_m)$ becomes arbitrarily flat and the normal vector of the boundary of the ball $p^{-1}(F)$ in $\mathbb{S}^m(\rho_m)$ becomes almost parallel to $\mathbb{R}^n$. Finally, if $b \to a$ we conclude $\gamma_n(R_r) \geq \gamma_n(E_r)$ and that gives directly the isoperimetric inequality we desired. □

If (M, μ) and (M_1, μ_1) are two probability spaces, then it is clear that $I_\mu \geq I_{\mu \otimes \mu_1}$. In particular we have $I_\mu \geq I_{\mu^{\otimes 2}} \geq \cdots I_{\mu^{\otimes n}}$. The theorem above implies that $I_{\gamma_1} = I_{\gamma_2} = \cdots = I_{\gamma_n}$ (note that $\gamma_n = \gamma_1^{\otimes n}$). This is another remarkable property of the Gaussian measure that in fact characterizes γ among probability measures on $\mathbb{R}$; see Bobkov and Houdré [**13**].

Given a probability space (M, μ), its *Gaussian constant* is defined as

$$c = c(\mu) = \inf_t \frac{I_\mu(t)}{I_\gamma(t)}. \tag{3–7}$$

This constant gives the best possible Gaussian isoperimetric inequality in (M, μ). If M is compact and μ_M is the normalized Riemannian measure, then $c(\mu_M) > 0$ and there exists $t_0 \in (0,1)$ such that $I_{\mu_M}(t_0) = cI_\gamma(t_0)$. This follows because the asymptotic behavior of the above functions at $t = 0$ is known to be given by

$$I_{\mu_M}(t) \sim a\, t^{(n-1)/n} \quad \text{and} \quad I_\gamma(t) \sim t(4\pi \log 1/t)^{1/2},$$

for some $a > 0$; see [**10**] and [**6**]. Note the parallelism between the Gaussian constant and the usual Cheeger constant (see for instance [**19**]), which is defined as in (3–7), but replacing I_γ with the piecewise linear function $J(t) = \min\{t, 1-t\}$. As we will see in the next section, in several applications of the isoperimetric inequalities, the Gaussian constant gives sharper results than the Cheeger one.

The isoperimetric profile of the spheres satisfies the following extraordinary sequence of inequalities.

THEOREM 3.4 (Barthe [**6**]).

$$I_{\vartheta_m} \geq I_{\vartheta_{m+1}} \geq I_\gamma \geq (I_{\vartheta_{m+1}})^{(m+1)/m} \geq (I_{\vartheta_m})^{m/(m-1)}. \tag{3–8}$$

Here we list some remarks related with the result above.

1. Specializing the theorem to $m = 2$ we have $2\sqrt{t(1-t)} \geq I_\gamma(t) \geq 4t(1-t)$.
2. Since $I_{\vartheta_m}(\frac{1}{2}) = I_\gamma(\frac{1}{2}) = 1$, we see that the inequalities (3.4) are sharp.
3. $I_{\vartheta_m} \to I_\gamma$ when m goes to ∞.
4. The normalized Riemannian measure β_n on $B^n(\rho_{n+1})$ satisfies $I_{\beta_n} \geq I_\gamma$ and the equality holds at $t = \frac{1}{2}$ (use Lemma 3.2).
5. If M^n is compact with empty boundary, $\mathrm{Ric}_M \geq (n-1)/\rho_n^2$ and μ_M is its Riemannian probability, then $I_{\mu_M} \geq I_\gamma$. This follows from the high dimensional version of Theorem 2.16 which, under our hypothesis gives that $I_{\mu_M} \geq I_{\vartheta_n}$. Note that $(n-1)/\rho_n^2$ converges to 2π when $n \to \infty$.

3.2. Symmetrization with respect to a model measure. We say that a probability measure μ_0 on an n_0-dimensional Riemannian manifold M_0 is a *model measure* if there exists a continuous family (in the sense of the Hausdorff distance on compact subsets) $\mathcal{D} = \{D^t \mid 0 \leq t \leq 1\}$ of closed subsets of M_0 satisfying the following conditions:

(i) $\mu_0(D^t) = t$ and $D^s \subset D^t$, for $0 \leq s \leq t \leq 1$.
(ii) D^t is a smooth isoperimetric domain of μ_0 and $I_{\mu_0}(t) = \mu_0^+(D^t)$ is positive and smooth for $0 < t < 1$.
(iii) The r-enlargement of D^t satisfies $(D^t)_r = D^s$ for some $s = s(t,r)$, $0 \leq t \leq 1$.
(iv) $D^1 = M_0$ and D^0 is either a point or the empty set.

The essential property is that enlargements of isoperimetric regions in $\mathcal{D}$ are also isoperimetric regions. Concerning the enlargements of the empty set, we are using the convention $\varnothing_r = \varnothing$.

Among the examples of model measures we have the Riemannian probability in $\mathbf{S}^n$. A second important group is given by the Gaussian measure $(\mathbb{R}^n, \gamma)$ and symmetric log-concave probability measures on $\mathbb{R}$, that is measures of the type $d\mu = e^{-\tau} dt$, where τ is a convex function with $\tau(-t) = \tau(t)$ for any $t \in \mathbb{R}$; in this last case the isoperimetric regions are the intervals $\{t : t \leq a\}$, $a \in \mathbb{R}$; see Bobkov and Houdré [**14**].

PROPOSITION 3.5. *Let (M,μ) be a probability space and (M_0,μ_0) a model measure. Then $I_\mu \geq I_{\mu_0}$ if and only if for any nonempty closed set $\Omega \subset M$ and for all $r \geq 0$ its r-enlargement satisfies $\mu(\Omega_r) \geq \mu_0(D_r)$, where $D \in \mathcal{D}$ with $\mu_0(D) = \mu(\Omega)$.*

PROOF. Suppose that $I_\mu \geq I_{\mu_0}$ and take $f(r) = \mu(\Omega_r)$ and $h(r) = \mu_0(D_r)$. Observe that f and h are continuous: the right continuity is clear. To prove the continuity to the left use that, for $r > 0$, $\partial\Omega_r$ is a set of measure zero. Moreover $h(r)$ is smooth for $0 < h(r) < 1$ and, in this range, $h'(r) = I_{\mu_0}(h(r))$. Consider the continuous function $F : [0,1] \to [0,\infty]$ defined by the conditions $F(0) = 0$ and $F'(t) = 1/I_{\mu_0}(t)$ for $0 < t < 1$. Then $(F \circ h)'(r) = 1$ whenever $0 < h(r) < 1$.

We want to show that $f(r) \geq h(r)$ for $r > 0$. If either $f(r) = 1$ or D is empty, that is trivial. So, we only need to prove the inequality in the case $D \neq \varnothing$ and $0 < r < r_0 = \min\{\rho : f(\rho) = 1\}$. For these r we have $f(r) > 0$ and the mean value theorem gives

$$\begin{aligned}\liminf_{\rho \to r^+} \frac{F(f(\rho)) - F(f(r))}{\rho - r} &= \liminf F'(t_\rho)\frac{f(\rho) - f(r)}{\rho - r} = F'(f(r))\mu^+(\Omega(r)) \\ &\geq F'(f(r))I_\mu(f(r)) = \frac{I_\mu(f(r))}{I_{\mu_0}(f(r))} \geq 1,\end{aligned}$$

where t_ρ is an intermediate value between $f(r)$ and $f(\rho)$. Since $f(0) = h(0)$ we deduce that $F(f(r)) \geq F(h(r))$ for $0 < r < r_0$. In fact, if that were false, we could find a number $0 < a < r_0$ so that $F(f(a)) - F(h(a)) < 0$ and $0 < h(r) < 1$ for $0 < r < a$. Hence, for some $\varepsilon > 0$, the continuous function

$$g(r) = F(f(r)) - F(h(r)) - \frac{r}{a}\big(F(f(a)) - F(h(a))\big), \quad 0 \leq r \leq a$$

would satisfy $g(0) = g(a) = 0$ and $\liminf_{\rho \to r^+} g(\rho) > \varepsilon$, $0 < r < a$, which is impossible. Finally, F being an increasing function, we get $f(r) \geq h(r)$ as we claimed. □

A modified version of the above result (under weaker hypotheses) can be found in [**14**].

Symmetrization is a well-known and useful construction in Euclidean geometry which, among other things, gives the Euclidean isoperimetric inequality; see Steiner [**74**], Schwarz [**70**] and [**18**]. Steiner–Schwarz symmetrization differs from Hsiang symmetrization introduced in §1.3. In the simple planar case, the idea is as follows: given a region Ω in the plane, as in Figure 15, its symmetrization is the figure Ω^S meeting vertical lines L in closed intervals symmetric with respect to the x-axis and with the same length that $\Omega \cap L$.

Now we will see how this construction, joint with its main properties, can be extended to general product spaces (in fact, the idea works also in the fiber bundles setting; see [**67**]). A proof of the isoperimetric property of half-spaces in the Gaussian space using this symmetrization was given by Ehrhard [**21**].

Consider three probability spaces $(M_0^{n_0},\mu_0)$, $(M_1^{n_1},\mu_1)$ and $(M_2^{n_2},\mu_2)$ and assume that μ_0 is a model measure. Given a closed subset $\Omega \subset M_1 \times M_2$, the section of Ω through the point $x \in M_1$ will be denoted by $\Omega(x) = \Omega \cap (\{x\} \times M_2)$. The same notation will be used to denote the sections of regions in $M_1 \times M_0$.

The *symmetrization with respect to* μ_0 of Ω is the subset $\Omega^S \subset M_1 \times M_0$ defined as follows. Take an arbitrary point $x \in M_1$.

If $\Omega(x) = \varnothing$, then $\Omega^S(x) = \varnothing$.

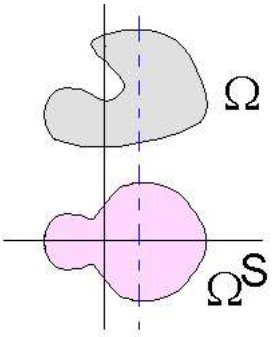


FIGURE 15. The classical symmetrization was used by Steiner and Schwarz to prove the Euclidean isoperimetric inequality.

If $\Omega(x) \neq \varnothing$, then $\Omega^S(x) = \{x\} \times D$, where $D \in \mathcal{D}$ with $\mu_0(D) = \mu_2(\Omega(x))$.

PROPOSITION 3.6. *Assume $I_{\mu_2} \geq I_{\mu_0}$. If $\Omega \subset M_1 \times M_2$ is a closed subset, then*

(a) *Ω^S is a closed set in $M_1 \times M_0$,*
(b) *$\mu_1 \otimes \mu_0(\Omega^S) = \mu_1 \otimes \mu_2(\Omega)$,*
(c) *$\Omega \subset \Omega'$ implies $\Omega^S \subset (\Omega')^S$,*
(d) *$(\Omega_r)^S \supset (\Omega^S)_r$,*
(e) *$(\mu_1 \otimes \mu_2)^+(\Omega) \geq (\mu_1 \otimes \mu_0)^+(\Omega^S)$.*

PROOF. To prove (a) take a sequence $\{(x_k, z_k)\}_k$ in Ω^S which converges to a point $(x, z) \in M_1 \times M_0$. We can also assume that $t_k = \mu_2(\Omega(x_k)) = \mu_0(D^{t_k})$ converges to a number t. Since z_k lies in D^{t_k} we have z belongs to D^t. So we only need to show that $t \leq \mu_2(\Omega(x)) = \mu_0(\Omega^S(x))$.

Consider the characteristic functions $\chi_k, \chi : M_2 \to \mathbb{R}$ of $\Omega(x_k)$ and $\Omega(x)$, respectively. Observe that $\limsup_k \chi_k \leq \chi$: in fact, if $\xi \in M_2$ satisfies $\limsup \chi_k(\xi) = 1$, then $\chi(\xi) = 1$ by the closeness of Ω, and if $\limsup \chi_k(\xi) = 0$, then there is nothing to prove. So, applying Fatou's lemma we conclude (a).

The definition of Ω^S implies clearly assertions (b) and (c).

Now we prove (d). First we remark that, for $r > 0$, the formula $\bigcup_x \Omega^S(x) = \Omega^S$ implies that $(\Omega^S)_r$ is the closure of $\bigcup_x \{\Omega^S(x)\}_r$. On the other hand, since $\Omega(x)_r \subset \Omega_r$, item (c) gives $(\Omega(x)_r)^S \subset (\Omega_r)^S$ and thus, $\bigcup_x (\Omega(x)_r)^S \subset (\Omega_r)^S$. Therefore, to

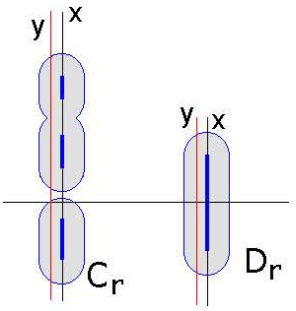


FIGURE 16. The sections through the point $y \in M_1$ of C_r and D_r.

conclude (d) it is enough to check that, if $\Omega(x) \neq \varnothing$, then

$$(\Omega(x)_r)^S \supset (\Omega^S(x))_r.$$

To prove this we use the notation $C = \Omega(x)$ and $D = \Omega^S(x) = C^S$. Thus $C \subset \{x\} \times M_2 \equiv M_2$, $D \subset \{x\} \times M_0 \equiv M_0$ and $\mu_2(C) = \mu_0(D)$. Denote by d_i the Riemannian distance in M_i, $i = 0, 1, 2$ and recall that the Riemannian distance d in the product manifold $M_1 \times M_2$ satisfies the identity $d = \sqrt{d_1^2 + d_2^2}$. Hence, if we take $y \in M_1$ with $d_1(x, y) \leq r$ and we define $r' = \sqrt{r^2 - d_1(x, y)^2}$, we can write $C_r(y) \equiv \{z \in M_2 : d_2(z, C) \leq r'\}$ and $D_r(y) \equiv \{z \in M_0 : d_0(z, D) \leq r'\}$; see Figure 16. So, Proposition 3.5 allows us to obtain $\mu_2(C_r(y)) \geq \mu_0(D_r(y))$, which implies, by definition of symmetrization, $(C_r)^S(y) \supset D_r(y)$. Then $(C_r)^S \supset D_r$, the desired inclusion.

The assertion in (e) follows directly from (b), (c) and (d). □

3.3. Isoperimetric problem for product spaces. Symmetrization with respect to a model measure and its properties in Proposition 3.6 give immediately comparison results for isoperimetric inequalities in product probability spaces.

THEOREM 3.7. *Consider probability spaces (M_0, μ_0), (M_1, μ_1) and (M_2, μ_2). If μ_0 is a model measure and $I_{\mu_2} \geq I_{\mu_0}$, then $I_{\mu_1 \otimes \mu_2} \geq I_{\mu_1 \otimes \mu_0}$.*

Since the n-dimensional Gaussian measure is a product measure $\gamma_n = (\gamma_1)^{\otimes n}$ and its isoperimetric profile does not depend on n, we obtain the interesting consequences of the theorem by taking $\mu_0 = \gamma$.

THEOREM 3.8 (Barthe & Maurey [**8**]; Barthe [**6**]). (a) *Let (M_1, μ_1) and (M_2, μ_2) be two probability spaces with $I_{\mu_i} \geq cI_\gamma$, $i = 1, 2$ and $c > 0$. Then, $I_{\mu_1 \otimes \mu_2} \geq cI_\gamma$. Moreover, if $\Omega \subset M_1$ is an isoperimetric region with $\mu_1(\Omega) = a$ and $I_{\mu_1}(a) = cI_\gamma(a)$, then $\Omega \times M_2$ is an isoperimetric region of $\mu_1 \otimes \mu_2$.*

(b) *If the manifolds M_i, $i = 1, 2$, are compact (possibly with boundary), then the isoperimetric problem in the Riemannian product $M_1 \times M_2$ (endowed with its Riemannian measure) admits for some volume a solution of the type $\Omega \times M_2$ or $M_1 \times \Omega$, where Ω is an isoperimetric region in the corresponding factor.*

(c) *Among hypersurfaces of $\mathbf{S}^{n_1}(r_1) \times \cdots \times \mathbf{S}^{n_k}(r_k)$ dividing the space in two equal volumes, the area is minimized by $\mathbf{S}^{n_1 - 1}(r_1) \times \mathbf{S}^{n_2}(r_2) \times \cdots \times \mathbf{S}^{n_k}(r_k)$ (up to order).*

(d) *Among hypersurfaces in a product of Euclidean balls $B^{n_1}(r_1) \times \cdots \times B^{n_k}(r_k)$ dividing the space in two equal volumes, the area is minimized by $B^{n_1 - 1}(r_1) \times B^{n_2}(r_2) \times \cdots \times B^{n_k}(r_k)$ (up to order); see Figure 17.*

(e) *If (M, μ) is a probability space with $I_\mu \geq I_\gamma$, then $I_{\mu \otimes \gamma} = I_\gamma$.*

PROOF. Given $c > 0$, the profile of $(M, ds^2/c^2, \mu)$ is c times the profile of (M, ds^2, μ). So, by rescaling, it is enough to prove the assertions above for $c = 1$. Items (a) and (e) follow from Theorem 3.7 and the isoperimetric properties of γ. To prove (b), use that under these hypotheses the Gaussian constants of the normalized Riemannian measures are positive and take c the minimum of both constants. Items (c) and (d) follow from the sharp estimates for the profile of the sphere and the ball given in the comments after Theorem 3.4. □

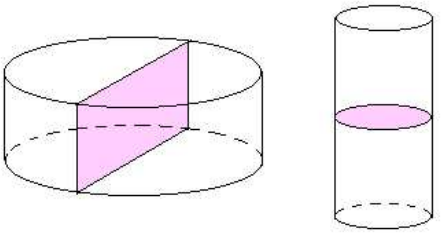


FIGURE 17. Surfaces of least area among the ones which divide a can $B^2(r) \times [0,1]$ in two equal volumes.

The above symmetrization arguments work, without changes, when the spaces have infinite measure. Among the consequences which can be obtained in this case (using as model measure the Lebesgue one) we mention the following result; see [**29**, p. 335].

THEOREM 3.9. *Let $(M_1^{n_1}, \mu_1)$ and $(M_2^{n_2}, \mu_2)$ be measure spaces with infinite total measure and denote by ω_m the Lebesgue measure in $\mathbb{R}^m$.*

(a) *If $I_{\mu_2} \geq I_{\omega_m}$, then $I_{\mu_1 \otimes \mu_2} \geq I_{\mu_1 \otimes \omega_m}$.*
(b) *If $I_{\mu_1} \geq I_{\omega_{m_1}}$ and $I_{\mu_2} \geq I_{\omega_{m_2}}$, then $I_{\mu_1 \otimes \mu_2} \geq I_{\omega_{m_1+m_2}}$.*

After the first version of these notes was written, we received the paper Barthe [**7**] which contains a proof of Theorem 3.7 in the case that μ_0 is a symmetric log-concave measure on $\mathbb{R}$.

3.4. Sobolev-type inequalities. In this section we give three analytic inequalities for a probability space (M, μ) which admits a Gaussian isoperimetric inequality, that is, a space with positive Gaussian constant $c(\mu)$. We know that this family includes any compact manifold (with or without boundary) endowed with its Riemannian probability; see Theorem 2.14(c) and [**10**].

First we will see a functional inequality which, for a given probability space, is equivalent to the Gaussian isoperimetric inequality. In fact, the geometric properties in Theorem 3.8 were first proved in [**8**] by following this analytic formulation, instead of using symmetrization. Earlier versions of the next result were proved by Ehrhard [**22**] and Bobkov [**12**]. The general version below is due to Barthe and Maurey [**8**].

THEOREM 3.10. *Let (M, μ) be a probability space and $c > 0$. Then, $I_\mu \geq cI_\gamma$ if and only if for any locally Lipschitz function $f : M \to [0,1]$,*

$$I_\gamma\left(\int_M f\,d\mu\right) \leq \int_M \sqrt{I_\gamma(f)^2 + \frac{1}{c^2}|\nabla f|^2}\,d\mu. \tag{3–9}$$

PROOF. Assume $c = 1$. If (3–9) holds, then for any closed subset $R \subset M$ we define $f_\varepsilon(x) = (1 - d(x,R)/\varepsilon)_+$, where d denotes the Riemannian distance in M and $h_+ = \max\{h, 0\}$. Putting $f = f_\varepsilon$ in (3–9) and taking $\varepsilon \to 0$ we obtain $\mu^+(\partial R) \geq I_\gamma(\mu(R))$.

Conversely, if $I_\mu \geq I_\gamma$ consider the probability measure $\mu \otimes \gamma$ in $M \times \mathbb{R}$ and the closed subset $R = R(f) = \{(x,t) \in M \times \mathbb{R} : \gamma(t) \leq f(x)\}$.

Then $(\mu\otimes\gamma)(R)=\int_M f d\mu$ and

$$(\mu\otimes\gamma)^+(\partial R)=\int_M \sqrt{I_\gamma(f)^2+|\nabla f|^2}\,d\mu.$$

We can conclude the proof by using that $I_{\mu\otimes\gamma}\geq I_\gamma$; see Theorem 3.8(e). □

The next inequality was first proved by Gross [**30**] for the Gaussian measure itself. It is an important tool in analysis with applications, among others, in evolution problems; see for instance [**24, 44**]. The proofs below have been taken from Ledoux [**48**].

THEOREM 3.11 (Logarithmic Sobolev inequality). *Let (M,μ) be a probability space with $I_\mu\geq cI_\gamma$ for some $c>0$. Then, for any smooth bounded positive function $g:M\to\mathbb{R}$,*

$$\int_M g\log g\,d\mu\leq\int_M g\,d\mu\log\left(\int_M g\,d\mu\right)+\int_M\frac{|\nabla g|^2}{4\pi c^2 g}\,d\mu. \tag{3–10}$$

PROOF. We normalize so that $c=1/\sqrt{2\pi}$ and we use the asymptotic behavior of the Gaussian profile at $t=0$, which is $cI_\gamma(t)\sim t\sqrt{2\log t^{-1}}$.

Put $f=e^{-\lambda}g$ in (3–9), with $\lambda>0$ large enough. Then

$$e^{-\lambda}\int g\sqrt{2\log\frac{e^\lambda}{\int_M g d\mu}}d\mu\leq e^{-\lambda}\int\sqrt{g^2(2\log\frac{e^\lambda}{g})+|\nabla g|^2}\,d\mu+\text{error term},$$

which can be written as

$$0\leq\int\sqrt{2g^2(\lambda+\log\frac{1}{g}+\frac{|\nabla g|^2}{2g^2})}-\sqrt{2g^2(\lambda+\log\frac{1}{\int g})}\,d\mu+\text{error term}.$$

Multiplying by $\sqrt{\lambda}$, taking $\lambda\to\infty$ and using the formula

$$\lim_{\lambda\to\infty}\sqrt{\lambda}\left(\sqrt{a(\lambda+b_1)}-\sqrt{a(\lambda+b_2)}\right)=\sqrt{a}(b_1-b_2),$$

we obtain,

$$0\leq\sqrt{2}\int\left(g\log\frac{1}{g}+\frac{|\nabla g|^2}{2g}-g\log\frac{1}{\int g}\right)d\mu.$$

□

THEOREM 3.12 (Poincaré inequality). *Assume the probability space (M,μ) satisfies $I_\mu\geq cI_\gamma$, for some $c>0$. Then, for any $h:M\to\mathbb{R}$ with $\int_M h d\mu=0$,*

$$\int_M h^2\,d\mu\leq\frac{1}{2\pi c^2}\int_M|\nabla h|^2\,d\mu. \tag{3–11}$$

PROOF. Put $g=(1+\varepsilon h)^2$ in the log-Sobolev inequality (3–10) and take $\varepsilon\to 0$. □

To convince the reader of the sharpness of the above results, we consider a couple of examples. For the sphere $\mathbf{S}^n(\rho_n)$, Theorem 3.12 gives that the first eigenvalue of the Laplacian λ_1 satisfies $\lambda_1\geq 2\pi$. The correct value is $\lambda_1=n/\rho_n^2$, which converges to 2π when n goes to infinity.

For the flat torus $T^n=\mathbf{S}^1(1/\pi)\times\cdots\times\mathbf{S}^1(1/\pi)$ Theorem 3.12 gives $\lambda_1\geq 2\pi$ (note that $\rho_1=1/\pi$ and so the Gaussian constant of T^n is $c=1$). The exact value

is $\lambda_1 = \pi^2$. The Cheeger constant of T^n is $h = 1$ (by Theorem 1.9) and so the Cheeger theorem (see [**19**]) gives the estimate $\lambda_1 \geq h^2/4 = 0.25$.

References

[1] A.D. Alexandrov, *Uniqueness theorems for surfaces in the large I*, Vestnik Leningrad Univ. Math. **11** (1956), 5–17.

[2] F. J. Almgren, *Existence and regularity almost everywhere of solutions to elliptic variational problems with constraints*, Mem. AMS **165** (1976).

[3] F. J. Almgren, Spherical symmetrization, Proc. International workshop on integral functions in the calculus of variations, Trieste 1985, Red. Circ. Mat. Palermo (2) Supple. (1987), 11–25.

[4] N. Aronszajn, *A unique continuation theorem for solution of elliptic partial differential equations or inequalities of second order*, J. Math. Pures Appl. **36** (1957), 235–249.

[5] J. L. Barbosa and M. do Carmo, *Stability of hypersurfaces with constant mean curvature*, Math. Z. **185** (1984), 339–353.

[6] F. Barthe, *Extremal properties of central half-spaces for product measures*, J. Funct. Anal. **182** (2001), 81–107.

[7] ———, *Log-concave and spherical models in isoperimetry*, preprint.

[8] F. Barthe and B. Maurey, *Some remarks on isoperimetry of Gaussian type*, Ann. Inst. H. Poincaré Probab. Statist. **36** (2000), 419–434.

[9] C. Bavard and P. Pansu, *Sur le volume minimal de* $\mathbf{R}^2$, Ann. Sci. École Norm. Sup. (4) **19** (1986), 479–490.

[10] P. Bérard and D. Meyer, *Inégalités isopérimétriques et applications*, Ann. Sci. École Norm. Sup. (4) **15** (1982), 513–541.

[11] P. Berard, G. Besson and S. Gallot, *Sur une inegalité isopérimetrique qui généralise celle de Paul Lévy–Gromov*, Invent. Math. **80** (1985), 295–308.

[12] S. G. Bobkov, *An isoperimetric inequality on the discrete cube and an elementary proof of the isoperimetric inequality in Gauss space*, Ann. Probab. **24** (1) (1997), 206–214.

[13] Bobkov, S. G. and Udre, K (Houdre, C.), *Characterization of Gaussian measures in terms of the isoperimetric property of half-spaces* (Russian), Zap. Nauchn. Sem. S.-Petersburg. Otdel. Mat. Inst. Steklov. (POMI) 228 (1996), Veroyatn. i Stat. 1, 31–38, 356; translation in J. Math. Sci. (New York) **93** (1999), 270–275.

[14] S. G. Bobkov and C. Houdré, *Some connections between isoperimetric and Sobolev-type inequalities*, Mem. Amer. Math. Soc. **129** (1997).

[15] J. Bokowski, H. Hadwiger and J. M. Wills, *Eine Ungleichung zwischen Volumen, Oberfläche und Gitterpunkzhal konvexer Körper in n-dimensionalen euklidischen Raum*, Math. Z. **127** (1972), 363–364.

[16] J. Bokowski and E. Sperner, *Zerlegung konvexer Körper durch minimale Trennflächen*, J. Reine Angew. Math. **311/312** (1979), 80–100.

[17] C. Borell, *The Brunn–Minkowski inequality in Gauss space*, Invent. Math. **30** (1975), 207–216.

[18] Y. D. Burago and V. A. Zalgaller, *Geometric inequalities*, Springer, Berlin (1988).

[19] I. Chavel, *Eigenvalues in Riemannian geometry*, Academic Press, 1984.

[20] ———, *Isoperimetric Inequalities: Differential Geometric and Analytic Perspectives*, Cambridge Tracts in Mathematics 145, Cambridge University Press, Cambridge, 2001.

[21] A. Ehrhard, *Symetrisation dans l'espace de Gauss*, Math. Scand. **53** (1983), 281–301.

[22] ———, *Inégalités isopérimétriques et intégrales de Dirichlet gaussienes*, Ann. Sci. Ecole Norm. Sup. **17** (1984), 317–332.

[23] A. da Silveira, *Stability of complete noncompact surfaces with constant mean curvature*, Math. Ann. **277** (1987), 629–638.

[24] E. B. Davies, *Heat kernels and spectral theory*, Cambridge Univ. Press, (1989).

[25] S. Gallot, *Inégalités isopérimetriques et analytiques sur les variétés Riemannienes*, Asterisque **163–164** (1988), 31–91.

[26] E. Gonzalez, U. Massari and I. Tamanini, *On the regularity of boundaries of sets minimizing perimeter with a volume constraint*, Indiana Univ. Math. J. **32** (1983), 25–37.

[27] P. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley, New York, (1978).

[28] M. Gromov, *Paul Levy's isoperimetric inequality*, prepublications I.H.E.S. (1980).

[29] ———, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhäuser, Boston (1999).
[30] L. Gross, *Logarithmic Sobolev inequalities* Amer. J. Math. **97** (1976), 1061–1083.
[31] K. Große-Brauckmann, *New surfaces of constant mean curvature* Math. Z. **214** (1993), 527–565.
[32] M. Grüter, *Boundary regularity for solutions of a partioning problem*, Arch. Rat. Mech. Anal. **97** (1987), 261–270.
[33] H. Hadwiger, *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie*, Springer, Berlin (1957).
[34] ———, *Gitterperiodische Punktmengen und Isoperimetrie*, Monatsh. Math. **76** (1972), 410–418.
[35] D.A. Hajduk, P.E. Harper, S.M. Gruner, C.C. Honecker, G. Kim, E. L. Thomas and L. J. Fetters, *The giroid: A new equilibrium morphology in weakly segregated diblock copolymers*, Macromolecules **27** (1994), 4063–4075.
[36] L. H. Harper, *Optimal numberings and isoperimetric problems on graphs*, J. Combinatorial Theory **1** (1966), 385–393.
[37] E. Heintze and H. Karcher, *A general comparison theorem with applications to volume estimates for submanifolds*, Ann. Sci. Ec. Norm. Super. **11** (1978), 451–470.
[38] J. Hersch, *Quatre propertiés isopérimétriques de membranes sphériques homogenes*, C. R. Acad. Sci. Paris Ser. A-B **270** (1970), A1645-A1648.
[39] H. Hopf, *Differential geometry in the large*, Lecture Notes in Mathematics, **1000** (1989), Springer.
[40] W. Y. Hsiang, *A symmetry theorem on isoperimetric regions*, PAM-409 (1988), UC Berkeley.
[41] ———, *On soap bubbles and isoperimetric regions in noncompact symmetric spaces. I*, Tohoku Math. J. (2) **44** (1992), 151–175.
[42] ———, *Isoperimetric regions and soap bubbles*, Proceedings conference in honor to Manfredo do Carmo, Pitman survey in pure and. appl. math . **52** (1991), 229–240.
[43] H. Howards, M. Hutchings and F. Morgan, *The isoperimetric problem on surfaces*, Am. Math. Monthly **106** (1999), 430–439.
[44] G. Huisken, *Local and global behavior of hypersurfaces moving by mean curvature*, Differential geometry: partial differential equations on manifolds, Proc. Symp. Pure Math. **54** Part 1 (1993), 175–191.
[45] M. Hutchings, F. Morgan, M. Ritoré and A. Ros, *Proof of the double bubble conjecture* Ann. Math. (2) **155** (2002), 459–489.
[46] H. B. Lawson, *Complete minimal surfaces in* $\mathbf{S}^3$, Ann. of Math. (2) **92** (1970), 335–374.
[47] M. Ledoux, *Concentration of measure and logarithmic Sobolev inequalities. Séminaire de Probabilités.* XXXIII, 120–216, Lecture Notes in Math., **1709**, Springer, Berlin, 1999.
[48] ———, *The geometry of the Markov diffusion generators*, Probability theory, Ann. Fac. Sci. Toulouse Math. (6) **9** (2000), 305–366.
[49] M. Ledoux and M. Talagrand, *Probability in Banach spaces. Isoperimetry and processes*, Results in Mathematics and Related Areas (3), **23**, Springer, Berlin, 1991.
[50] P. Lévy, *Leçons dánalyse fonctionelle*, Gauthier–Villars, Paris 1922.
[51] P. Li and S. T. Yau, *A new conformal invariant and its applications to the Willmore conjecture and first eigenvalue of compact surfaces*, Invent. Math. **69**(1982), 269–291.
[52] F. J. López and A. Ros, *Complete minimal surfaces with index one and stable constant mean curvature surfaces*, Comment. Math. Helvet. **64** (1989), 34–43.
[53] W. H. Meeks, *The theory of triply periodic minimal surfaces*, Indiana Univ. Math. J. **39** (1990), 877–936.
[54] F. Morgan, *Geometric measure theory: a beginner's guide*, 2nd ed., Academic Press, 1998.
[55] ———, *Regularity of isoperimetric hypersurfaces in Riemannian manifolds*, preprint.
[56] M. Nosarzewska, *Évaluation de la différence entre l'aire d'une région plane convexe et le nombre des points avec coordonnées entiéres couvert par elle*, Colloq. Math. **1** (1948), 305–311.
[57] R. Osserman, *The isoperimetric inequality*, Bull. Amer. Math. Soc. **84** (1978), 1182–1238.
[58] R. Pedrosa and M. Ritoré, *Isoperimetric domains in the Riemannian product of a circle with a simply connected space form and applications to free boundary problems*, Indiana Univ. Math. J. **48** (1999), 1357–1394.

[59] J. Perez and A. Ros, *Properly embedded minimal surfaces with finite total curvature*, Proceedings of the CIME course "Minimal surfaces in flat three-manifolds" at Martina-Franca, Italy, summer 1999 (to appear).
[60] M. Ritoré, *Superficies con curvatura media constante*, PhD Thesis, Granada, 1994, available at `http://www.ugr.es/~ritore`.
[61] ———, *Examples of constant mean curvature surfaces obtained from harmonic maps to the two sphere*, Math. Z. **226** (1997), 127–146.
[62] ———, *Index one minimal surfaces in flat three-space forms*, Indiana Univ. Math. J. **46**, (1997), 1137–1153.
[63] M. Ritoré and A. Ros, *Stable Constant Mean Curvature Tori and the Isoperimetric Problem in Three Space Forms*, Comment. Math. Helvet. **67** (1992), 293–305.
[64] ———, *The spaces of index one minimal surfaces and stable constant mean curvature surfaces embedded in flat three manifolds*, Trans. Amer. Math. Soc. **348** (1996), 391–410.
[65] A. Ros, *The Willmore conjecture in the real projective space*, Math. Research Letters **6** (1999), 487–494.
[66] ———, *The isoperimetric and Willmore problems*, Proc. congress in Memory of A. Gray, Contemporary Math. (to appear).
[67] ———, *The isoperimetric problem for lens spaces* (preprint).
[68] A. Ros and E. Vergasta, *Stability for hypersurfaces of constant mean curvature with free boundary*, Geom. Dedicata **56** (1995), 19–33.
[69] M. Ross, *Schwarz' P and D surfaces are stable*, Differential Geom. Appl. **2** (1992), 179–195.
[70] H. A. Schwarz, *Gesammelte Mathematische Abhandlungen*, Springer Verlag, Berlin (1890).
[71] W. M. Schmidt, *Volume, surface area and the number of integer points covered by a convex set*, Arch. Math. (Basel) **23** (1972), 537–543.
[72] U. Schnell, *Lattice inequalities for convex bodies and arbitrary lattices*, Monatsh. Math. **116** (1993), 331–337.
[73] V. N. Sudakov and B. S. Tsirel'son, *Extremal properties of half-spaces for spherically invariant measures*, J. Soviet Math. (1978), 9–18.
[74] J. Steiner, *Sur le maximum et le minimum des figures dans le plan, sur la sphére et dans l'espace en géneral* J. Reine Angew. Math. **24** (1842), 93–152.
[75] E. L Thomas, D. M. Anderson, C.S. Henkee and D. Hoffman, *Periodic area-minimizing surfaces in block copolymers*, Nature **334** (1988), 598–602.
[76] P. Topping, *Towards the Willmore conjecture*, Calc. Var. and PDE, **11** (2000), 361–393.
[77] Traizet, *Construction of tripy periodic minimal surfaces*, preprint.
[78] T. J. Willmore, *Note on embedded surfaces*, An. Sti. Univ. Al. I. Cuza Iasi Sect. I a Mat. **11** (1965), 493–496.
[79] P. Yang and S. T. Yau, *Eigenvalues of the Laplacian of compact Riemann surfaces and minimal submanifolds*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **7** (1980), 55–63.

Departamento de Geometria y Topologia, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain
E-mail address: aros@ugr.es

Clay Mathematics Proceedings
Volume 2, 2004

Flat Structures, Teichmüller Theory and Handle Addition for Minimal Surfaces

Michael Wolf

Abstract. The basic obstruction to using the Weierstrass representation to create minimal surfaces of complicated topology but few symmetries is the period problem: the one-forms obtained by complexifying the coordinate differentials must have purely imaginary periods before the Weierstrass representation is well-defined. In this paper, we discuss a procedure (developed with Matthias Weber) for describing the period problem in terms of Euclidean domains, and then solving it using Teichmüller theory. A novel feature of the proof is that it is inductive: complicated surfaces arise by the addition of handles or ends to less complicated surfaces.

1. Introduction

The Weierstrass representation is a powerful tool bringing methods of complex analysis to bear on problems of existence in minimal surface theory. It is particularly powerful in constructing minimal surfaces of low genus and few ends, as the problem of ensuring that the resulting representation is well-defined (the so-called "period problem") naturally translates into a problem in a low-dimensional moduli space, which can then be solved either explicitly or by the use of techniques adapted to low-dimensional spaces, like the intermediate value theorem (for one-dimensional spaces).

These latter methods see particular success in the handle-addition theorems of Chen–Gackstatter, Karcher and others ([**CG81**], [**CG82**], [**HMI90**], [**HW**], [**Kar88**], [**Kar89**], [**Kar91**]). These results beg the question as to how extensively and systematically one can continue to add handles to existing minimal surfaces.

We describe here work with Matthias Weber in which we show an inductive method that adds successive handles (and ends) to known minimal surfaces, producing families of minimal surfaces of increasing complexity. We might very informally state the result thus:

Theorem 1.0.1 (Handle addition: Weber–Wolf). *For a number of classes of geometric descriptions of surfaces, there exist families of minimal examples indexed by genus or by genus and number of ends.*

Partially supported by NSF grant DMS 9971563.

©2004 Clay Mathematics Institute

The generality of the result is that the architecture of the proof and many of the components apply to a number of widely varying classes of surfaces, while the particular class of surfaces only affects one (global) choice within the proof.

We now describe the results more carefully, beginning with four sample theorems of the type of Theorem 1.0.1.

THEOREM 1.0.2 (Adding handles to Enneper [**WW98**]). *For each $g \geq 0$, there is a complete minimal surface E_g in $\mathbb{E}^3$ of genus $g \geq 0$ and one Enneper-type end with total curvature $-2\pi(g+1)$ and eight symmetries.*

This theorem has something of a long history: the original surface is due to Enneper [**Enn69**], and then one and two handles were added by Chen–Gackstatter [**CG82**]. A surface with an Enneper-type end and genus three was found by d'Espirito-Santo [**San94**], and then shortly afterwards, Thayer displayed convincing numerical evidence of the existence of surfaces of this type for all genera $g \leq 35$ [**Tha94**]. For arbitrary genus, the result above was independently obtained by Sato [**Sat96**].

Even though these surfaces are not embedded, they are perhaps still interesting: the surface E_g minimizes, among all complete immersed minimal surfaces of genus g, the degree of the Gauss map; in other words, it maximizes the total curvature (by the Jorge–Meeks inequality; see [**JI83**]). An example of these surfaces is drawn in Figure 4.

THEOREM 1.0.3 (Adding handles and ends to Costa [**WW**]). *For all odd genera g, there is a complete minimal surface $CT_g \subset \mathbb{E}^3$ of genus g which is embedded outside a compact set, with g parallel (horizontal) planar ends and two catenoid ends. The symmetry group (of order eight) of CT_g is generated by reflective symmetries about a pair of orthogonal vertical planes and a rotational symmetry about a horizontal line.*

These surfaces look as shown in Figure 1. (Many of the figures in this paper, including all the images of minimal surfaces, were made by Matthias Weber.)

Here one might imagine that the original Costa surface results from desingularizing the intersection of a horizontal surface with a (vertical) catenoid; analogously, these surfaces might result from desingularizing the intersection of an odd number of planes with a catenoid.

We generalize these CT_g surfaces as follows, imagining **D**rilling additional **H**oles to obtain surfaces $DH_{m,n}$.

THEOREM 1.0.4 (Adding handles to CT_g surfaces).
(i) *For every pair of integers $n \geq m \geq 1$, there exists a complete minimal surface $DH_{m,n} \subset \mathbb{E}^3$ of genus $m+n+1$ which is embedded outside a compact set with the following properties: it has $2n+1$ vertical normals, $2m+1$ (parallel) planar ends, and two catenoid ends. The symmetry group is as in Theorem 1.0.3.*
(ii) *For $n < m$, there is no complete minimal surface with those symmetries of the type $DH_{m,n}$ (and $2n+1$ vertical normals, $2m+1$ planar ends, and two catenoid ends).*

What is interesting is how this theorem is consistent with the Hoffman–Meeks conjecture (see [**HK97**]) that if there exists an embedded minimal surface of genus g with e ends, then $g \geq e-2$. The information in Theorem 1.0.4 can be encoded in a table:

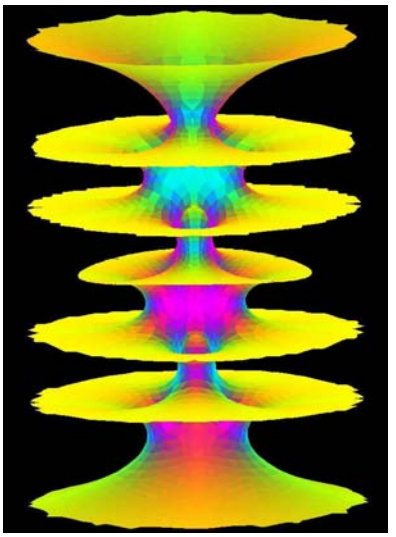


FIGURE 1. The surface CT_5 of genus 5 with five planar ends and two catenoid ends. One can imagine this as being a desingularization of the intersection of five horizontal planes with a catenoid.

$m\backslash n$	0	1	2	3	4	5	6
0	Costa	Horgan	M–W	$*$	$*$	$*$	$*$
1	$-$	B–W	$+$	$+$	$+$	$+$	$+$
2	$-$	$-$	$+$	$+$	$+$	$+$	$+$
3	$-$	$-$	$-$	$+$	$+$	$+$	$+$
4	$-$	$-$	$-$	$-$	$+$	$+$	$+$
5	$-$	$-$	$-$	$-$	$-$	$+$	$+$

TABLE 1. Existence ($+$) and nonexistence ($-$) for surfaces of type (m, n). Costa refers to the surface announced in [**Cos84**], Horgan refers to the surface discussed in [**Web98**], M–W and $*$ refer to the surfaces proved not to exist in [**MW01**], and B–W refers to the surface of Boix and Wohlgemuth (see [**BW**]).

The techniques of Theorem 1.0.1 also apply to periodic surfaces; for instance, we can inductively add handles to Scherk's doubly periodic surface (see Figure 8).

THEOREM 1.0.5 (Adding handles to Scherk's doubly periodic surface [**WW01**]). *For each $g \geq 0$, there is a complete doubly-periodic minimal surface S_g in $\mathbb{E}^3$, whose quotient (by the maximal group of translational symmetries) has genus $g \geq 0$ and four planar parallel ends, parallel to the ends of Scherk's doubly periodic surface S_0.*

Weber [**Web00**] has added handles to the surfaces [**CHMI89**] of Callahan–Hoffman–Meeks; again the architecture of the proof is identical.

THEOREM 1.0.6 (Adding handles to Callahan–Hoffman–Meeks [**Web00**]). *For each pair of integers $1 \leq m \leq n$, there is a singly periodic complete minimal surface*

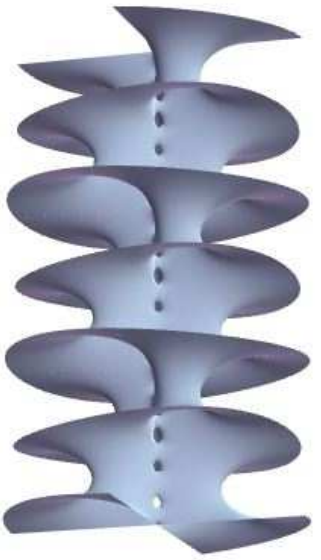


FIGURE 2. Weber's handle addition to the Callahan–Hoffman–Meeks surface.

$CHM_{m,n}$ in $\mathbb{E}^3$ invariant under a translation whose quotient has $2m$ planar ends, genus $m+n+1$ and eight symmetries.

In this paper, we describe the proofs of these theorems, focusing for the sake of concreteness almost exclusively on the case of the Enneper-ended examples of Theorem 1.0.2. After defining our notation and recalling some background information in that notation in the next section, we give an overview of the method in the third and fourth sections. In these final sections, we focus on each of the steps of the proof, giving some, but not all, of the details of each of the steps of the proofs. Complete proofs may be found in the referenced articles.

2. Background and Notation

2.1. The Weierstrass representation. Our arguments are founded upon the Weierstrass representation. To set notation, recall that for a minimal surface on whose underlying Riemann surface $\mathcal{R}$ we have a holomorphic function G and a holomorphic form dh (not necessarily exact, despite the notation), the minimal immersion may be represented via a map $\mathcal{R} \to \mathbb{E}^3$ by

$$z \mapsto \mathrm{Re}\int_{\rho_0}^{z} \left(\frac{1}{2}\Big(G - \frac{1}{G}\Big)\, dh,\ \frac{i}{2}\Big(G + \frac{1}{G}\Big) dh,\ dh\right).$$

For this surface, G will be the Gauss map (postcomposed with stereographic projection) and dh will be the complexified differential of the third coordinate in $\mathbb{E}^3$.

As basic examples, we mention:

(1) *Enneper's surface* (1869), where we may take $G = z$ and $dh = z\,dz$ on $\mathcal{R} = \mathbb{C}$ (see [**Web03**, Figure 4, p. 4] in this volume); the surfaces for Theorem 1.0.2 have one end asymptotic to the end of this surface.
(2) *The catenoid*, for which we may take $G = z$, $dh = dz/z$ on $\mathbb{C} - \{0\}$ (see [**Web03**, Figure 3, p. 26] in this volume); the top and bottom ends of the family for 1.0.3 are asymptotically catenoids.

There are, of course, some restrictions on the possible Weierstrass data. To begin, we compute that the metric on $F(\mathcal{R})$ is

$$ds_{F(R)} = \left(|G| + \frac{1}{|G|}\right)|dh|; \tag{2–1}$$

thus, to get a regular metric, we need compatible divisors for G and dh at regular points of the surface. This is a pretty mild restriction, as it is all local.

The global problem for producing minimal surfaces is that of well-definedness: continuation around a cycle must leave the map unchanged. Thus we require

$$\operatorname{Re}\int_\gamma \frac{1}{2}\Big(G - \frac{1}{G}\Big)\,dh = \operatorname{Re}\int_\gamma \frac{i}{2}\Big(G + \frac{1}{G}\Big)\,dh = \operatorname{Re}\int_\gamma dh = 0$$

for every cycle $\gamma \subset \mathcal{R}$.

Rephrasing: We prefer to work with the meromorphic data of two forms $G\,dh$ and $(1/G)\,dh$. Then we can phrase the **period problem** as

$$\int_\gamma G\,dh = \overline{\int_\gamma \frac{1}{G}\,dh}, \qquad \operatorname{Re}\int_\gamma dh = 0$$

for every cycle $\gamma \subset \mathcal{R}$ that we wish to be mapped by F to a closed curve. (For periodic surfaces, we may wish to allow some cycles γ to not satisfy the period problem in a specified way.)

3. The Geometry of Orthodisks

How should one approach the period problem? More generally, how does one picture a period of a one-form on a Riemann surface? For the type of problems we are considering, where the fundamental domain of the desired minimal surface (with respect to the desired symmetry group) is a planar domain, we adopt the perspective that a one-form on a Riemann surface provides for a singular Euclidean structure on the surface.

This is but a global summary of some straightforward local computations. Locally, if $\alpha = f(z)dz$ is a one-form on a Riemann surface $\mathcal{R}$, then we can define a line element ds_α by

$$ds_\alpha = |\alpha| = |f(z)|\,|dz|$$

Then, away from the zeroes of α, the metric ds_α has curvature

$$K = \frac{-2}{f(z)}\,\partial\overline{\partial}\log f(z) = 0$$

since f is meromorphic.

At a zero or pole p of α, by comparing the radius and circumference of small circle about the singularity p, we find that if $\alpha = (z^k + \text{higher-order terms})\,dz$, then ds_α is isometric to a Euclidean cone with cone angle $2\pi(k+1)$ at p.

3.1. Flat structure triples. Here's a motivating example (W. Thurston): Consider a parallelogram in the plane with the usual toral identifications, and its two complementary regions Ω_0 and Ω_∞ in $\hat{\mathbb{C}}$. Consider the one-form dz on each of these regions.

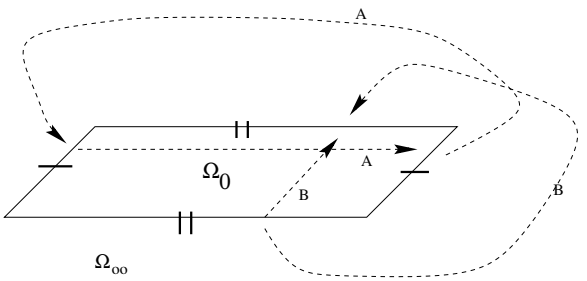


FIGURE 3. A parallelogram with identifications defines two tori, each with a one-form.

With the identifications, each domain is a torus, say T_0 and T_∞, respectively. The one-form dz restricts to be the familiar nonvanishing holomorphic form ω_0 on T_0, but on T_∞, the one-form dz has a second order pole at ∞, and is regular elsewhere. Thus, this form, say ω_∞, corresponds to a Weierstrass $\mathfrak{P}$-function (multiplied by the nonvanishing holomorphic one-form) on T_∞. The zero of this form occurs at the vertex of the parallelogram — note the total cone angle there of 6π.

Most significantly, note that here we have two (typically different) Riemann surfaces, T_0 and T_∞, each equipped with a one-form so that under a natural correspondence of cycles, the periods of the one-forms along those cycles either agree or are negatives of one another. Here the periods are found using that the developing map $z \to \int^z \omega_i$ of the surface T_i to $\widehat{\mathbb{C}}$ has image Ω_i for $i = 0, \infty$. Thus the periods, say $\int_A \omega_i = \int_A dz$ are given by differences of endpoints as shown above. In particular, the A-periods of ω_i are negatives of each other, while the B-periods agree.

This device of pairs of flat metrics representing pairs of Riemann surfaces and one-forms where the Euclidean geometry of the pairs reflects a relationship between the periods of the one-forms drives all of our constructions.

As we noted in the introduction, when we discuss details in this paper, we will focus on the case of the Enneper-ended surfaces of Theorem 1.0.2. So before going much farther, we now deduce the general forms of the flat structures required for these Enneper-ended surfaces E_g. Since we have the benefit of a computer simulation of some of these surfaces, we can use these images; in general, we would have some image suggested at least by intuition.

We begin our analysis using the figure. First, note that this surface has reflections in two vertical planes (as well as one additional symmetry we will discuss somewhat later). Thus, its quotient by the group of reflections will be a planar domain which itself has a symmetry. Next, observe that there are nine points, say $H_1, H_2, \ldots, H_9$ on the central axis where the tangent plane is horizontal. Thus the form dh, which is a component of the Weierstrass data, will have a zero at those points: we need to decide on the order of the zero. The simulation Figure 4 leads us to imagine that each zero is simple, and so we will assume that for the rest of

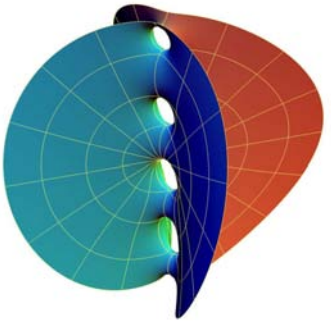


FIGURE 4. The Chen–Gackstatter Surface with four handles.

the discussion — remember that at this stage, we are trying to set up a candidate space for proving existence, and so we are free to assume anything we want about that surface we are trying to prove exists. (The peril of this, of course, is that we can restrict our space of candidates too much, and so exclude the actual minimal surface from consideration.) Since these points H_i are on the surface, we will need the induced metric to be regular, and so, from formula 2–1, we require that the Gauss map G have either a simple zero or a simple pole at those points.

We look once more at our simulation, and do not observe any other points with a horizontal tangent plane, so we declare that the divisor of the surface is supported on the points H_i and the single end, say E, where our surface is asymptotic to Enneper's original surface. Thus we can read off the relevant Weierstrass data at that point E from the corresponding data at the end of Enneper's surface: we declare that dh should have a third order pole at E, while G should have a simple zero or pole, depending on how we orient the surface.

With these considerations in mind, we can tabulate the Weierstrass data on the divisor of G (or dh) as follows:

	ord dh	ord G
H_1	1	-1
H_2	1	1
H_3	1	-1
H_4	1	1
H_5	1	-1
H_6	1	1
H_7	1	-1
H_8	1	1
H_9	1	-1
E	-3	-1

TABLE 2. Divisor data for the surface E_4.

All of these derivations of divisor data for a Weierstrass representation are quite standard in the theory of minimal surfaces: our next goal is to simplify the

appearance of the period problem by converting this data into candidates for the flat structures of the forms $G\,dh$, $(1/G)\,dh$ and dh. To do this, we go back to the simulation and make one further observation: the surface meets the four vertical (symmetry) coordinate planes in curves along which G is either purely real or purely imaginary, and for which dh is purely real (this also involves a choice of stereographic projection for the Gauss map). Thus the forms $G\,dh$ and $(1/G)\,dh$ are either purely real or purely imaginary along those arcs. Thus the image of the one quarter of the surface bounded by the vertical planes will develop onto a region bounded by a polygonal arc with either purely real or purely imaginary line segments.

What are the angles that these segments meet at? Since the segments meet at one of the points H_i or E, this is already determined by the divisor data in the table above. We convert this complex analytic information to geometric data in the next table, using that the cone corresponding to a zero of order k has an angle of $2\pi(k+1)$.

	$\operatorname{ord} G\,dh$	$\operatorname{ord}(1/G)\,dh$	$\angle G\,dh$	$\angle \frac{1}{G}\,dh$	$\angle dh$
H_1	0	2	$\pi/2$	$3\pi/2$	π
H_2	2	0	$3\pi/2$	$\pi/2$	π
H_3	0	2	$\pi/2$	$3\pi/2$	π
H_4	2	0	$3\pi/2$	$\pi/2$	π
H_5	0	2	$\pi/2$	$3\pi/2$	π
H_6	2	0	$3\pi/2$	$\pi/2$	π
H_7	0	2	$\pi/2$	$3\pi/2$	π
H_8	2	0	$3\pi/2$	$\pi/2$	π
H_9	0	2	$\pi/2$	$3\pi/2$	π
E	-2	-4	$-\pi/2$	$-3\pi/2$	$-\pi$

TABLE 3. Divisors and cone angles for the forms $G\,dh$ and $(1/G)\,dh$.

Here the last three columns give one fourth of the total cone angle at the indicated point: we display just one fourth of the angle as the one fourth of the surface we display includes just one fourth of a neighborhood of each of these points.

In every one of the situations we treat, we may assume that the surface is sufficiently symmetric so as to guarantee that a fundamental domain is planar. Indeed, a drawback of our method is that it is not immediately clear how to extend this method to the case where the fundamental domain of the minimal surface is a nonplanar surface, or even a topologically nontrivial planar surface. So, we focus for the rest of the time on topological disks with flat structures.

DEFINITION 3.1.1. *An orthodisk is a closed topological disk with distinguished points on the boundary equipped with both a Euclidean metric and an isometric developing into $\mathbb{E}^2$. The Euclidean metric renders the intervals between the distinguished points as nonsingular geodesics; we label these intervals by their endpoints.*

We typically blur the distinction between the orthodisk and its developed image in $\mathbb{E}^2$.

We are almost in a position from which we can display the possible flat structures for the forms $G\,dh$, $(1/G)\,dh$ and dh; we will of course now refer to these structures as the orthodisks for $G\,dh$, $(1/G)\,dh$ and dh. Before doing that however, we note that in this particular situation, we find one more restriction from our simulation which is particularly helpful: we see a straight line on the surface connecting the middle handle H_5 with the end E. We will thus require all of our forms to respect the symmetry about this line. With this in mind, we claim that the developed image of the form $G\,dh$ looks like the domain to the northwest in the figure below.

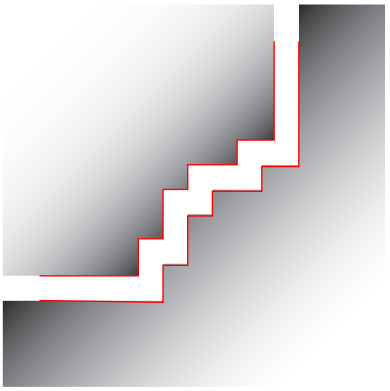


FIGURE 5. Flat structures for $G\,dh$ for E_4.

Here we should imagine that when we label the vertices counter-clockwise, the first finite vertex is H_1, the final finite vertex is H_9, and the end E is at infinity. We see from the table just above that the angles at the H_i are correct, and it is not much harder to decide that the angle at the infinite end E is also correctly represented. We have also imposed the symmetry about the line on the surface from H_5 to E by requiring a reflective symmetry about the diagonal line in the diagram from H_5 to E.

Finally, we remark that the process of deriving a domain, say $\Omega_{G\,dh}$, from a Riemann surface and the one-form $G\,dh$, is reversible. One begins with such a domain, and then takes a branched double cover (branched over the vertices) over the double of the domain. The one-form dz pulls back to give the desired form: $G\,dh$ in this case.

3.2. Representing the period problem. We now take up the problem of representing periods in these flat structures. The basic observation is that if we develop via a formula like

$$z(\zeta) = \int^{\zeta} G\,dh$$

to obtain a domain, say $\Omega_{G\,dh}$ in the z-plane as in Figure 5, then the Fundamental Theorem of Calculus implies that we may represent the form $G\,dh$ in that domain $\Omega_{G\,dh}$ as dz. But then, for an arc γ on the surface that develops into the plane as $z(\gamma)$, we find that

$$\int_{\gamma} G\,dh = \int_{z(\gamma)} dz$$

and this last integral is simply the difference of the z-values of the endpoints of $z(\gamma)$.

With this in mind, we observe from the topology of the simulation in Figure 4 that any cycle on the minimal surface E_4 projects to be homologous to a sum of edges in the boundary $\partial\Omega_{G\,dh}$. Thus, since from the previous paragraph we have on $\Omega_{G\,dh}$ the representation of $G\,dh = dz$, we see that the periods of $G\,dh$ are expressible as sums of the (complex) lengths of the sides (i.e., the difference of the endpoints of the segments as complex numbers).

Thus, the period condition, expressed analytically as

$$\int_\gamma G\,dh = \overline{\int_\gamma \frac{1}{G}\,dh} \tag{3–1}$$

translates to a geometric condition that the complex length of an edge H_iH_{i+1} in $\Omega_{G\,dh}$ should be the conjugate of the complex length of the corresponding edge H_iH_{i+1} in $\Omega_{(1/G)\,dh}$.

This brings us to an idiosyncrasy of our construction. We find it easier to construe our vertical planes as being vertical coordinate planes so that $\Omega_{G\,dh}$ has horizontal and vertical boundary edges. Also, then the analogous domain $\Omega_{1/G\,dh}$ will also have horizontal and vertical boundary edges. But, in order to force (3–1), we will then have to construe the real line in the figure Figure 5 as being diagonal, parallel to the line $\{y = x\}$.

From these considerations, we can use the domain $\Omega_{G\,dh}$ and the table of cone angles of Table 3 above to force a precise description of the domain $\Omega_{(1/G)\,dh}$. Indeed, this is the domain in Figure 5 which is complementary to $\Omega_{G\,dh}$; here since we keep the domain to the left of the boundary, we label the points in a manner *opposite* to how we labeled them for $\Omega_{G\,dh}$, i.e., the first finite point H_1 is the corner at the top right, while the last finite point, H_9 is the corner at the bottom left.

DEFINITION 3.2.1. *A pair of similarly labeled orthodisks Ω_1 and Ω_2 are called conjugate if, after lifting the canonical form dz from $\mathbb{E}^2$ to the surfaces S_1 and S_2 which (branched) doubly covers Ω_1 and Ω_2, respectively, the periods of the forms are conjugate.*

We observe that this condition can be verified by considering arcs which connect corresponding boundary edges of Ω_1 and Ω_2, as long as we keep in mind that these arcs are only one piece of an entire circle on the surfaces S_i.

It is now evident why the symmetry we imposed about the diagonal line from H_5 to E is so useful: because of that symmetry, the segments $\overline{H_iH_{i+1}}$ and $\overline{H_{9-i}H_{10-i}}$ have the same length and opposite orientation in both domains $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$, and these orientations are reversed for the "complementary" domain. The upshot is that these two orthodisks, $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$, when drawn as in Figure 5 precisely fit together so that the closure of their union is exactly $\mathbb{C}$.

Where has the (horizontal) period problem (3–1) gone? We really haven't done anything — developing one-forms into $\mathbb{C}$ cannot be anything beyond a psychological convenience — but we have displayed domains where (3–1) is satisfied. However, when we drafted these domains using the condition (3–1), we ignored the conformal structure of the resulting domains. What we have accomplished in this process of describing a pair of conjugate orthodisks $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ is the transformation of the period problem into a:

Conformal Problem: Find two such conjugate domains $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ which are conformal by a map which takes vertices H_i and E to corresponding vertices H_i and E.

Solving this problem then solves the (horizontal) period problem, as we can begin with the conformal structure common to both $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$, and then take the branched cover of the double of either domain to obtain a Riemann surface. Then the canonical one-forms dz on each domain lift to forms $G\,dh$ and $(1/G)\,dh$ on that Riemann surface which, by construction, satisfy (3–1).

REMARK 3.2.2. *We can also follow this process to produce an orthodisk, say Ω_{dh} corresponding to the one-form dh. A surprising feature of the proofs of all the theorems in the introduction is that finding a domain Ω_{dh} appropriately conformal to $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ which satisfies the period condition* $\operatorname{Re}\int_\gamma dh = 0$ *is nearly automatic from the geometric data. In particular, in the case of the Enneper-ended surfaces E_g, this reflects that such forms dh are exact.*

A more thorough discussion of orthodisks and Weierstrass data can be found in [**WW**].

3.2.1. *Orthodisks for other minimal surfaces.* We pause for a moment from our treatment of the orthodisks relevant to the Enneper-ended surfaces E_g to display the geometries of orthodisks relevant for the study of other surfaces.

We begin by noting that when we compute the orthodisks for the "higher genus" Costa surfaces CT_g, we get surfaces with more sheets and more branch points, as in the Figures 6 and 7. Here the fatter dots at some of the points P_i indicate that the point P_i is a (finite) branch point. For example, on the left side of Figure 10 below, the surface is a two-sheeted branched cover (with boundary) over $\widehat{\mathbb{C}}$, with branch points (at ∞) at P_1 and P_3 and a finite branch point at P_2.

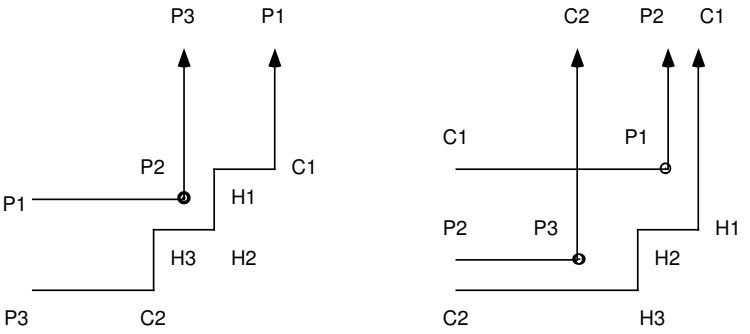


FIGURE 6. Orthodisks for a genus three Costa surface.

The doubly periodic Scherk surfaces with handles (as in Figure 8) have orthodisks like those pictured in Figure 9.

3.3. Moduli spaces of conjugate flat structures. How are we to solve the conformal problem of the last subsection? The basic approach is to regard that problem as a problem in the moduli spaces of pairs of conjugate orthodisks $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$. In particular, observe that the domain $\Omega_{G\,dh}$, because of the symmetry about the line H_5E is determined by the edge lengths of the segments $\overline{H_iH_{i+1}}$ for $i = 1, 2, 3, 4$. The conformal structure (here we *always* mean

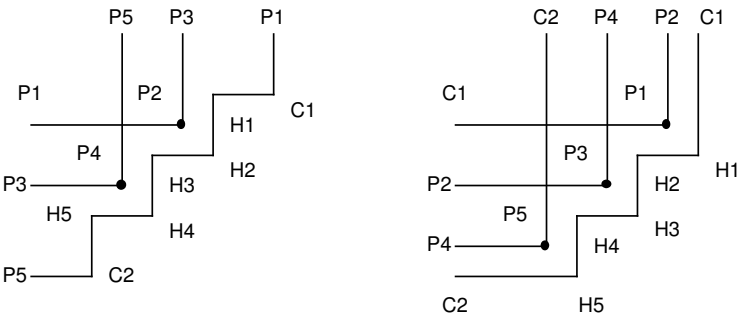


FIGURE 7. Orthodisks for a genus five Costa surface.

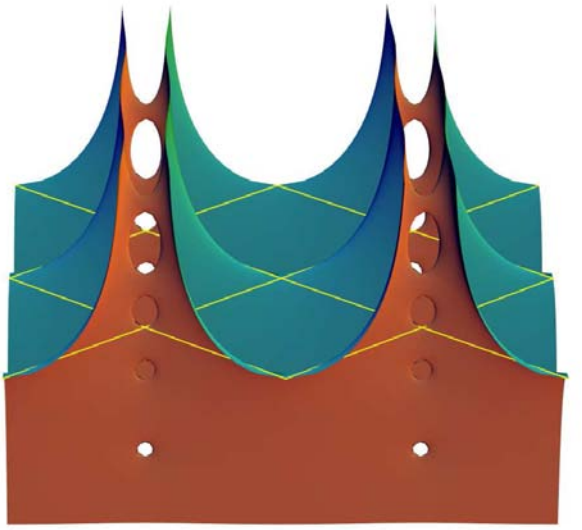


FIGURE 8. The genus 4 doubly periodic Scherk.

by "conformal structure" that the points H_i and E are distinguished) is also given by these numbers, up to scaling. Thus the moduli space $\mathcal{M}_4$ of possible conformal structures of domains $\Omega_{G\,dh}$ for the surface E_4 is given as the simplex $\mathcal{M}_4 = \{(x_1, x_2, x_3, x_4) | \sum x_i = 1, x_i > 0\}$. As such domains immediately define $\Omega_{(1/G)\,dh}$ by construction (in this case, by taking the complementary domain in $\mathbb{C}$), we see that solving the conformal problem in $\mathcal{M}_4$ is equivalent to producing the minimal surface E_4.

Before outlining how we solve that conformal problem, we pause briefly to explore these moduli spaces.

3.3.1. *A singleton moduli space: The Costa surface* $DH_{0,0}$ $(= CT_0)$. The Costa surface is the basic example of a complete finite total curvature embedded minimal surface of nontrivial topology. It is also the starting point for the investigations for the Theorem 1.0.3. Here, we seek a torus with two catenoid ends (C_1 and C_2), one planar end P_1 and one finite point H_1 with vertical normal. We also assume our standard eight symmetries: reflections about two orthogonal vertical planes and a

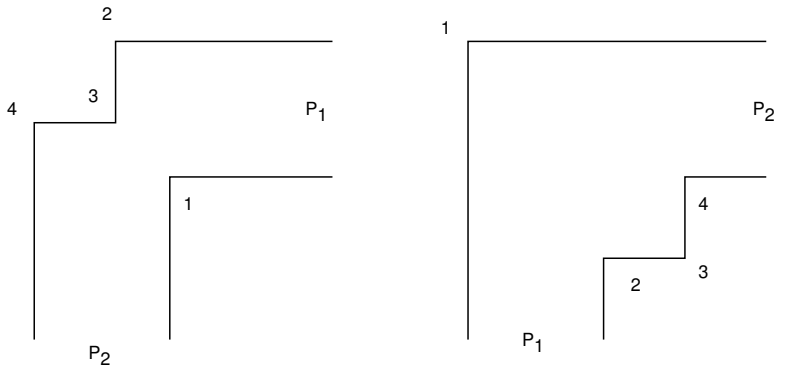


FIGURE 9. Orthodisks for genus 4 doubly periodic Scherk.

rotation about a horizontal line. (Previous existence and uniqueness proofs may be found in [**Cos84**] and [**HMI85**])

As we did for the surface E_4 in subsection 3.1, we can deduce the divisor data for the Weierstrass data for this surface, assuming that the points occur on the boundary of a fundamental quarter of the surface in the order $C_1 - P_1 - C_2 - H_1$. (Below we reproduce a computer image of a fundamental domain of Costa's surface for the group generated by reflections in the two vertical planes. This makes apparent the conformal polygon, and the order of the special points on its boundary.)

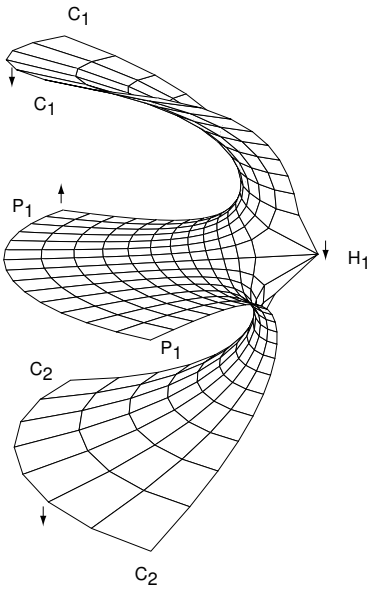


FIGURE 10. One quarter of Costa's surface.

Table 4 corresponds to Table 3; it computes the cone angles for the forms $G\,dh$ and $(1/G)\,dh$ on the putative conformal polygon for Costa's surface. These flat

structures are drawn below, assuming again an additional symmetry about the line $\{y = -x\}$.

	$\operatorname{ord} G$	$\operatorname{ord} dh$	$\operatorname{ord} G\,dh$	$\operatorname{ord} \frac{1}{G}\,dh$	$\angle G\,dh$	$\angle \frac{1}{G}\,dh$	$\angle dh$
$C_1 \downarrow$	1	-1	0	-2	$\pi/2$	$-\pi/2$	0
$P_1 \uparrow$	-3	1	-2	4	$-\pi/2$	$5\pi/2$	π
$C_2 \downarrow$	1	-1	0	-2	$\pi/2$	$-\pi/2$	0
$H_1 \downarrow$	1	1	2	0	$3\pi/2$	$\pi/2$	π

TABLE 4. Divisors and (quarter) cone angles for Costa's surface.

The crucial observation to make at this point is that any conformal quadrilateral with a symmetry across a diagonal is conformally equivalent to a square.

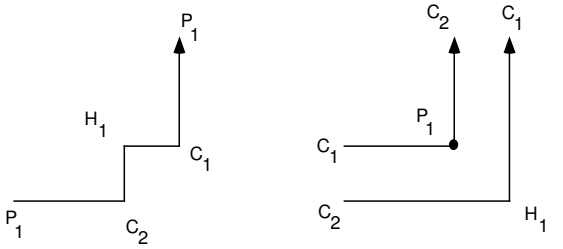


FIGURE 11. Orthodisks for $G\,dh$ and $(1/G)\,dh$ for Costa's surface.

Thus we conclude that the moduli space of possible examples consists only of the singleton of a pair of square tori, so the only element in the moduli space is a reflexive pair. This establishes the existence of this surface, by a proof that is somewhat distinct from the other proofs of existence of this surface ([**Cos84**], [**HK97**]).

3.3.2. *Empty moduli spaces: Catenoid with one handle.* In this example, we try to construct a minimal surface with two catenoid ends and one handle. Now, a theorem of Schoen [**Sch83**] implies that any minimal surface of genus $g \geq 1$ with but two embedded catenoid ends cannot exist; here we will require in addition the eightfold symmetry present in all of our examples. Yet, it is clear from an analysis of the orthodisks pictured below that such a surface cannot exist with 4-fold symmetry, because there is no symmetric and conjugate pair: the periods from $\overline{C_1C_2}$ to $\overline{H_1H_2}$ are conjugate about the line $\{y = -x\}$, while the periods from $\overline{C_2H_2}$ to $\overline{H_1C_1}$ are conjugate across the line $\{y = +x\}$.

A crucial part of Theorem 1.0.4 is the proof of the nonexistence of a whole class of possible configurations, consistent with the Hoffman–Meeks conjecture. The technique we use there is mostly just an iterative use of the technique we exhibit here for the nonexistence of the one-handled catenoid.

3.3.3. *The Horgan surface $DH_{0,1}$.* The second example of nonexistence is called the Horgan surface (see [**HK97**]): to visualize it, start with one plane and two handles, one growing upwards and another growing downwards — the necks of the handles should be perpendicular to one another. Both handles connect to catenoid ends. One imagines the surface to be almost as in Figure 13.

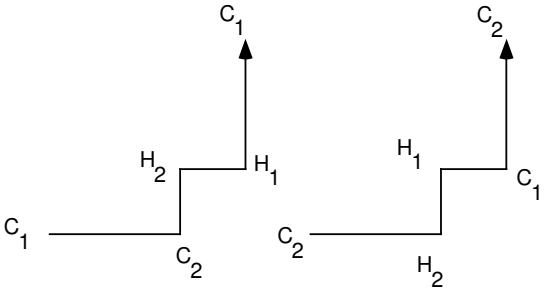


FIGURE 12. Catenoid with a handle.

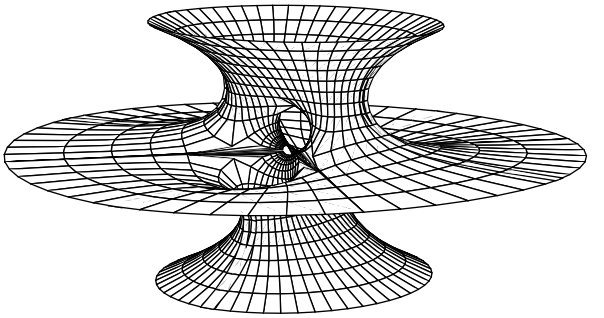


FIGURE 13. The hypothetical Horgan surface.

This pattern leads to a sequence of distinguished points on the boundary in the order $C_1PC_2H_2MH_1C_1$, where M denotes a (new type of) regular point where the symmetry lines cross. The orthodisks are as follows:

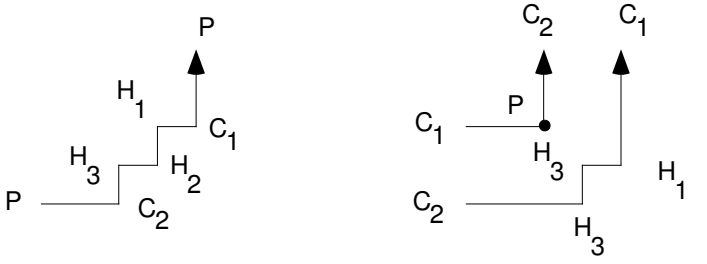


FIGURE 14. Orthodisks corresponding to the Horgan surface.

Recently, Weber [**Web98**] showed that this surface cannot exist by a method that is different from that used in our previous nonexistence proofs. In the case of this surface, there is a nontrivial moduli space, but our general existence proof fails at precisely one place for the case of this class of surface (we are unable to prove "Properness of the Height Function").

4. Architecture of the Proof

4.1. Converting the period problem into a conformal problem. Our goal in this section is to outline the approach we use to prove the theorems stated at the outset of this note.

While the details of the arguments are sometimes quite involved, the basic logic of the approach and the ideas of the proofs are quite simple. Here, we outline the argument, as a step-by-step recipe. (For the sake of making this section somewhat self-contained, we repeat some of the background we developed in computing the orthodisks for E_4.)

Step 1: Draw the surface. The first step in proving the existence of a minimal surface is to work out a detailed proposal. This can either be done numerically, as in the work of Thayer [**Tha94**] for the Chen–Gackstatter surfaces E_g or in Boix and Wohlgemuth ([**BW**]) for the low genus surfaces of the classes CT_g, or it can be schematic, showing how various portions of the surface might fit together. Often, as in the case of E_4 above we need to narrow the list of possibilities, and so additional symmetries might be assumed. Sometimes, there are quite a few combinatorial possibilities, and so, for each such class, we might attempt to follow the recipe and see if any of the rest of the recipe fails.

Step 2. Compute the divisors for the forms $G\,dh$ and $(1/G)\,dh$. From the model that we drew in Step 1, we can compute the divisors for the Weierstrass data, which we earlier defined to be the Gauss map G and the "height" form dh. (Note here how important it is that the Weierstrass representation be given in terms of geometrically defined quantities — for us, this gives the passage between the extrinsic geometry of the minimal surface as defined in Step 1 and the conformal geometry and Teichmüller theory of the later steps.) Thus we can also compute the divisors for the meromorphic forms $G\,dh$ and $(1/G)\,dh$ on the Riemann surface (so far undetermined, but assumed to exist) underlying the minimal surface. Of course the divisors for a form determine the form up to a constant, so the divisor information nearly determines the Weierstrass data for our surface. For the surface E_4, this step was completed in Table 2.

Step 3: Compute the orthodisks for the forms $G\,dh$ and $(1/G)\,dh$ required by the period conditions. A meromorphic form on a Riemann surface defines a flat singular (conformal) metric on that surface: for example, from the form $G\,dh$ on our putative Riemann surface, we determine a line element $ds_{G\,dh} = |G\,dh|$. This metric is locally Euclidean away from the support of the divisor of the form and has a complete Euclidean cone structure in a neighborhood of a zero or pole of the form. Thus we can develop the universal cover of the surface into the Euclidean plane.

The orthodisks for the forms $G\,dh$ and $(1/G)\,dh$ are not completely arbitrary: because the periods for the pair of forms must be conjugate by formula (3–1), the orthodisks must develop into domains which have a particular Euclidean geometric relationship to one another. This relationship is crucial to our approach, so we will dwell on it somewhat (generalizing the discussion of section 3.2). If the map $D : \Omega \longrightarrow \mathbb{E}^2$ is the map which develops the flat structure of a form, say α, on a domain Ω into $\mathbb{E}^2$, then the map D pulls back the canonical form dz on $\mathbb{C} \cong \mathbb{E}^2$ to the form α on Ω. Thus the periods of α on the Riemann surface are given by integrals of dz along the developed image of paths in $\mathbb{C}$, i.e., by differences of the complex numbers representing endpoints of those paths in $\mathbb{C}$.

We construe all of this as requiring that the flat structures develop into domains that are "conjugate": if we collect all of the differences in positions of parallel sides for the developed image of the form $G\,dh$ into a large complex-valued n-tuple $V_{G\,dh}$, and we collect all of the differences in positions of corresponding parallel sides for the developed image of the form $(1/G)\,dh$ into a large complex-valued n-tuple $V_{(1/G)\,dh}$, then these two complex-valued vectors $V_{G\,dh}$ and $V_{(1/G)\,dh}$ should be conjugate. This is the flat structure implication of the period condition formula (3–1), here using that our situation allows that the periods of all cycles can be found from differences of positions of parallel sides in a flat structure. Thus, we translate the "period problem" into a statement about the Euclidean geometry of the developed flat structures. This we already did for the surface E_4 in section 3.

As we remarked earlier, the vertical period problem ($\mathrm{Re}\int_\gamma dh = 0$) will be trivially solved for the surfaces we treat here.

Step 4: Define the moduli space of pairs of conjugate flat domains. Now we work backwards. We know the general form of the developed images (called $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$, respectively) of orthodisks associated to the forms $G\,dh$ and $(1/G)\,dh$, but in general, there are quite a few parameters of the orthodisks left undetermined, even after we have assumed symmetries, determined the Weierstrass divisor data for the models and used the period conditions (3–1) to restrict the relative Euclidean geometries of the pair $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$. Thus, there is a moduli space Δ of possible candidates of pairs $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$: our period problem (3–1) is now a conformal problem of finding such a pair which are conformally equivalent by a map which preserves the corresponding cone points. (Solving this problem means that there is a well-defined Riemann surface which can be developed into $\mathbb{E}^2$ in two ways, so that the pair of pullbacks of the form dz give a pair of forms $G\,dh$ and $(1/G)\,dh$ with conjugate periods.)

(The condition of conjugacy of the domains $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ often dictates some restrictions on the moduli space, and even a collection of geometrically defined coordinates. We do not meet these in the considerations of the surface E_g, but they are crucially relevant in the discussion of the surfaces CT_g, where they enter into the proof of nonexistence of some candidates. We saw a brief glimmer of this in our discussion of the nonexistence of the catenoid with a handle.)

Step 5: Solve the conformal problem using Teichmüller theory. At this juncture, our minimal surface problem has become a problem in finding a special point in a product Δ of moduli spaces of complex domains: we will have no further references to minimal surface theory. The plan is straightforward. We will define a height function $\mathcal{H} : \Delta \longrightarrow \mathbb{R}$ with these properties:

1. Reflexivity: The height $\mathcal{H}$ equals 0 only at a solution to the conformal problem.
2. Properness: The height $\mathcal{H}$ is proper on Δ. This ensures the existence of a critical point.
3. Noncritical Flow: If the height $\mathcal{H}$ at a pair $(\Omega_{G\,dh}, \Omega_{(1/G)\,dh})$ does not vanish, then the height $\mathcal{H}$ is not critical at that pair $(\Omega_{G\,dh}, \Omega_{(1/G)\,dh})$.

The proof of the solution to the conformal problem then summarizes as follows: we restrict to a locus $\mathcal{Y}$ on which we have the Noncritical Flow Property. By Properness, the height function $\mathcal{H}$ is proper on $\mathcal{Y}$; thus there is a critical point X on $\mathcal{Y}$ for $\mathcal{H}$. The Noncritical Flow then forces $\mathcal{H}(X) = 0$, so by Reflexivity, the

surface represented by X is a solution to our conformal problem, and hence also defines a solution to the minimal surface problem.

We will see that Reflexivity will be relatively straightforward, but Properness and Noncritical Flow will require some care. This is because in these steps, we are forced to compare two flat structures of a domain (and their underlying conformal structures.) In the customary treatment of the period problem, the difficulty was in understanding periods of two one-forms on a single surface — here the periods of the one-forms have been transmuted into a flat structure on the Riemann surface, and so the basic difficulty of the original global problem emerges in relating the pair of orthodisks.

How should we define this height function? We need a finite number of conformal invariants of $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ which satisfy these conditions:

1. Reflexivity: They form a complete set of conformal invariants, i.e., the two domains $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ are conformally equivalent if and only if the given set of conformal invariants agree.
2. Properness: They have computable/estimable asymptotics.
3. Noncritical Flow: They have computable derivatives.

The natural candidates for these invariants are the "Extremal Lengths" of curve systems, first defined by Beurling and Ahlfors in the 1940's. (We discuss the rudiments of this subject below: see [**Ahl66**] or [**Oht70**] for more details.)

Before plunging into a discussion of extremal lengths, we provide a context by just producing a height function for the Enneper-ended surfaces E_g. In this discussion, the curves Γ_i refer to the set of arcs in the domain which have one endpoint on $H_{i-1}H_i$ and another endpoint on $H_{i+1}H_{i+2}$ (where here we construe $H_0 = E$).

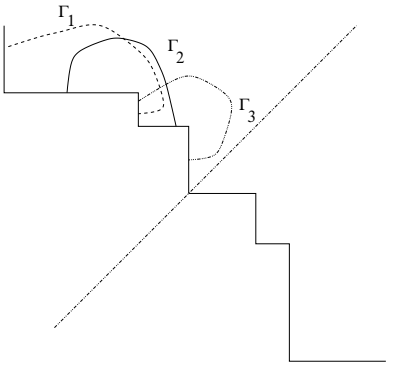


FIGURE 15. Curve systems for E_4.

REMARK 4.1.1. *In general, the decision about which curve systems to use is the one aspect of the proof which is not yet automatic — presently, we do not have a good procedure for deciding from a pair of orthodisks how to choose a system of curves for the height function which will allow for all three aspects of the proof to work at once.*

We define the height as

$$\mathcal{H} = \sum_{i=1}^{n} \mathcal{H}(\Gamma_i)$$

with

$$\mathcal{H}(\Gamma_i) = \left(\mathrm{Ext}_{\Gamma_i}\, \Omega_{G\,dh} - \mathrm{Ext}_{\Gamma_i}\, \Omega_{(1/G)\,dh}\right)^2 + \left(\exp(1/\mathrm{Ext}_{\Gamma_i}(\Omega_{G\,dh})^{-1}) - \exp(1/\mathrm{Ext}_{\Gamma_i}(\Omega_{(1/G)\,dh}))\right)^2.$$

The idea is to compare corresponding extremal lengths on $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$. We will say more about the form of each term when we discuss Properness.

4.2. Background on extremal length. Before resuming our outline, we pause to recall some aspects of the notion of Extremal Length. The definition of Extremal Length is very very flexible. Consider a *set* Γ of curves on a Riemann surface $\mathcal{R}$, and the set $\mathcal{M} = \{\rho \geq 0\}$ of (Borel measurable) conformal metrics on that Riemann surface. We define the *extremal length* of Γ on $\mathcal{R}$ to be

$$\mathrm{Ext}_{\mathcal{R}}(\Gamma) = \sup_{\rho \in \mathcal{M}} \frac{\inf_{\gamma \in \Gamma} \left(l_\rho(\gamma)\right)^2}{\mathrm{area}_\rho(\mathcal{R})},$$

where $l_\rho(\gamma)$ is the ρ-length of γ and $area_\rho$ is the ρ-area of the Riemann surface $\mathcal{R}$.

The idea is that this gives some notion of length that depends only on the underlying Riemann surface $\mathcal{R}$, and not on a choice of conformal metric on $\mathcal{R}$.

Here are three useful observations that guide many of our choices — we will not refer to it explicitly in the sequel, but it is crucial for some of the details of the arguments that we are omitting in this treatment. First, as $\mathrm{Ext}_{\mathcal{R}}(\Gamma)$ is a supremum, we immediately obtain a lower bound for this invariant of extremal length from any conformal metric in which we can estimate shortest lengths.

On the other hand, since Γ is usually taken to be a homotopy class of simple curves, we can characterize the solution metric and thus obtain a different definition of Extremal length. In particular, an easy length-area argument (see [**Ahl66**]) shows that the solution metric for the curves in an $h \times l$ rectangle connecting opposite sides is given by the standard flat metric on that rectangle.

We can thus equivalently define

$$\mathrm{Ext}_{\mathcal{R}}(\Gamma) = \inf_{R_{h\times l} \subset \mathcal{R}} \frac{l}{h},$$

where the infimum is over all $h \times l$ rectangles $R_{h\times l}$ embedded in $\mathcal{R}$ with horizontal curves in Γ.

This new definition, now in terms of an infimum, allows us to obtain *upper* bounds for $\mathrm{Ext}_{\mathcal{R}}(\Gamma)$ by mapping in appropriate rectangles (or annuli if the curves in Γ are closed).

Finally, as we can explicitly solve the extremal length problems for points on curves encircling boundary points of a disk, we know estimates for how extremal length changes with infinitesimal changes of data, as well as under degeneration. Also, the formulae are invertible: given the extremal lengths of $n-3$ cycles encircling consecutive pairs of points, we can locate the n points.

This last fact allows us to conclude that for these pairs, say Z, of planar domains, we have $\mathcal{H}(Z) = 0$ if and only if Z represents a conformal pair of conjugate domains.

4.3. Analysis of the height function. We now return to our outline of the proof of the theorem.

Step 5.1: Reflexivity. Having defined the height function via the extremal lengths of the curves Γ_i, for a good choice of Γ_i, this step of reflexivity is always immediate, at least for the case of a pair of domains $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ which are topological disks. (It is at present far from clear how to extend this architecture to the case of topologically nontrivial orthodisks.) We simply recall from the previous subsection that to determine the position of n boundary points of a disk, we require the extremal lengths of $n-3$ distinct arcs encircling pairs of those boundary points. Our height function is constructed so that it vanishes only when such a set of extremal lengths for both domains agree.

Step 5.2: Properness. We need to show that as one of the conformal structures, (underlying one of the orthodisks $\Omega_{G\,dh}$ or $\Omega_{(1/G)\,dh}$) degenerates, then the height of the pair goes to infinity. The subtlety here is that the asymptotic differences in the extremal lengths of corresponding surfaces can be quite minute: our height function is specifically created to blow up these small differences.

Our height function will measure differences in the extremal lengths $\mathrm{Ext}_{\Omega_{G\,dh}}\gamma_i$ and $\mathrm{Ext}_{\Omega_{(1/G)\,dh}}\gamma_i$. Often, but not always, a geometric degeneration of the orthodisk of either $\Omega_{G\,dh}$ or $\Omega_{(1/G)\,dh}$ will force one of the extremal lengths $\mathrm{Ext}_\bullet(\gamma_i)$ to tend to zero or infinity, while the other extremal length stays finite and bounded away from zero. This is a straightforward situation where it will be obvious that the height function will blow up. A more subtle case arises when a geometric degeneration of the orthodisk forces *both* of the extremal lengths $\mathrm{Ext}_{\Omega_{G\,dh}}(\gamma_i)$ and $\mathrm{Ext}_{\Omega_{(1/G)\,dh}}(\gamma_i)$ to *simultaneously* decay (or explode). In that case, we begin by observing that there is a natural map between the vector $\langle \mathrm{Ext}_{\Omega_{G\,dh}}(\gamma_i)\rangle$ and the vector $\langle \mathrm{Ext}_{\Omega_{(1/G)\,dh}}(\gamma_i)\rangle$. This pair of vectors is reminiscent of pairs of solutions to a hypergeometric differential equation, and we show, by a monodromy argument analogous to that used in the study of those equations, that it is not possible for corresponding components of that vector to vanish or blow up at identical rates. In particular, we show that the logarithmic terms in the asymptotic expansion of the extremal lengths near zero have a different sign, and this sign difference forces a difference in the rates of decay that is detected by the height function, forcing it to blow up in this case. In the next paragraph, we give some of the crucial details of this monodromy argument.

The Monodromy Argument. Let Δ be a moduli space of dimension at least two parametrizing pairs of orthodisks X_1 and X_2 corresponding to given formal Weierstrass data as usual. Suppose γ is a cycle in the underlying conformal polygon which joins edges P_1P_2 and Q_1Q_2 which are parallel (and hence nonadjacent) in the flat structures. (In our applications, γ will be one of the cycles used in the height function). Denote by R_1 the vertex before Q_1 and by R_2 the vertex after Q_2 and observe that by assumption, $R_2 \neq P_1$ and $P_2 \neq R_1$. Introduce a second cycle β which connects R_1Q_1 with Q_2R_2.

We formulate our properness claim more precisely in the following two lemmas:

LEMMA 4.3.1. *Suppose that for a sequence $p_n \in \Delta$ with $p_n \to p_0 \in \partial\Delta$ we have that $\mathrm{Ext}_{X_1(p_n)}(\gamma) \to 0$ and $\mathrm{Ext}_{X_2(p_n)}(\gamma) \to 0$. Suppose furthermore that γ is a cycle encircling a single edge which degenerates geometrically to 0 as $n \to \infty$ in $X_1(p_n)$. Then*

$$\left|\exp(1/\mathrm{Ext}_{X_1(p_n)}(\gamma)) - \exp(1/\mathrm{Ext}_{X_2(p_n)}(\gamma))\right|^2 \to \infty.$$

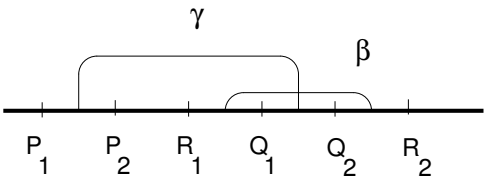


FIGURE 16. Local picture for monodromy argument.

LEMMA 4.3.2. *Suppose that for a sequence* $p_n \in \Delta$ *with* $p_n \to p_0 \in \partial\Delta$ *we have that* $\mathrm{Ext}_{X_1(p_n)}(\gamma) \to \infty$ *and* $\mathrm{Ext}_{X_2(p_n)}(\gamma) \to \infty$. *Suppose furthermore that* γ *is a cycle with an endpoint on an edge which degenerates geometrically to* 0 *as* $n \to \infty$ *in* $X_1(p_n)$. *Then*

$$\left|\exp(\mathrm{Ext}_{X_1(p_n)}(\gamma)) - \exp(\mathrm{Ext}_{X_2(p_n)}(\gamma))\right|^2 \to \infty.$$

PROOF. We first prove Lemma 4.3.1, assuming for now the truth of this next:

COROLLARY 4.3.3. *Either* $\|F_1(\gamma)\| - \frac{\log\delta}{\pi}\|F_1(\beta)\|$ *or* $\|F_1(\gamma)\| + \frac{\log\delta}{\pi}\|F_1(\beta)\|$ *is real analytic in* δ *for* $\delta = 0$. *In the first case,* $\|F_2(\gamma)\| + \frac{\log\delta}{\pi}\|F_2(\beta)\|$ *is real analytic in* δ, *in the second* $\|F_2(\gamma)\| - \frac{\log\delta}{\pi}\|F_1(\beta)\|$.

Consider the conformal polygons corresponding to the pair of orthodisks. Normalize the punctures by Möbius transformations so that

$$P_1 = -\infty, \qquad P_2 = 0, \qquad Q_1 = \varepsilon, \qquad Q_2 = 1$$

for X_1 and

$$P_1 = -\infty, \qquad P_2 = 0, \qquad Q_1 = \varepsilon', \qquad Q_2 = 1$$

for X_2. By the assumption of 4.3.1, we know that $\varepsilon, \varepsilon' \to 0$ as $n \to \infty$. We now apply the monodromy Corollary 4.3.3 to the curve $\varepsilon_0 e^{it}$ and conclude that either

$$\frac{|\mathrm{Per}\,\beta(X_1)|}{|\mathrm{Per}\,\gamma(X_1)|} + \frac{1}{\pi}\log\varepsilon \tag{4–1}$$

is single-valued in ε while

$$\frac{|\mathrm{Per}\,\beta(X_2)|}{|\mathrm{Per}\,\gamma(X_2)|} - \frac{1}{\pi}\log\varepsilon' \tag{4–2}$$

is single-valued in ε', or the same statement holds for the analogous quantities with opposite signs. Without loss of generality, we can treat the first case.

Now suppose that ε' is real analytic (and hence single-valued) in ε near $\varepsilon = 0$. Then using that X_1 and X_2 are conjugate implies that the absolute lengths of β in X_1 and X_2 are equal, as are those of γ; hence

$$\frac{|\mathrm{Per}\,\beta(X_1)|}{|\mathrm{Per}\,\gamma(X_1)|} = \frac{|\mathrm{Per}\,\beta(X_2)|}{|\mathrm{Per}\,\gamma(X_2)|}.$$

Thus we see, after subtracting (4–2) from (4–1), that

$$\log(\varepsilon\varepsilon'(\varepsilon))$$

is single-valued in ε near $\varepsilon = 0$ which contradicts that $\varepsilon, \varepsilon' \to 0$.

Now Ohtsuka's [**Oht70**] extremal length formula states that for the current normalization of $X_1(p_n)$ we have

$$\operatorname{Ext}\gamma = O\left(|\log\varepsilon|^{-1}\right)$$

(see [**WW98**, Lemma 4.5.3] and [**Oht70**]). We conclude that

$$\left|\exp(1/\operatorname{Ext}_{X_1(p_n)}(\gamma)) - \exp(1/\operatorname{Ext}_{X_2(p_n)}(\gamma))\right| = O\Big(\frac{1}{\varepsilon} - \frac{1}{\varepsilon'}\Big),$$

which goes to infinity, since we have shown that ε and ε' tend to zero at different rates. This proves Lemma 4.3.1.

The proof of Lemma 4.3.2 is very similar: for convenience, we normalize the points of the punctured disks such that

$$P_1 = -\infty, \qquad P_2 = 0, \qquad Q_1 = 1, \qquad Q_2 = 1+\varepsilon$$

for X_1 and

$$P_1 = -\infty, \qquad P_2 = 0, \qquad Q_1 = 1, \qquad Q_2 = 1+\varepsilon'$$

for X_2.

By the assumption of Lemma 4.3.2, we know that $\varepsilon, \varepsilon' \to 0$ as $n \to \infty$. We now apply the monodromy Corollary 4.3.3 to the curve $1+\varepsilon_0 e^{it}$ and conclude that

$$\frac{\operatorname{Per}\gamma(X_1)}{\operatorname{Per}\beta(X_1)} + \frac{1}{\pi}\log\varepsilon$$

is single-valued in ε while

$$\frac{\operatorname{Per}\gamma(X_2)}{\operatorname{Per}\beta(X_2)} - \frac{1}{\pi}\log\varepsilon'$$

is single-valued in ε'. The rest of the proof is identical to the proof of Lemma 4.3.1.

To prove the important Corollary 4.3.3, we will need asymptotic expansions of the extremal length in terms of the Euclidean geometric invariants of the orthodisks. Though not much is known explicitly about extremal lengths in general, for the cycles we always choose we can reduce this problem to an asymptotic control of Schwarz–Christoffel integrals. Their monodromy properties allow us to distinguish their asymptotic behavior by the sign of logarithmic terms.

We introduce some notation: Suppose we have an orthodisk such that the angles at the vertices alternate between $\pi/2$ and $-\pi/2$ modulo 2π. Consider the Schwarz–Christoffel map

$$F : z \mapsto \int_i^z (t-t_1)^{a_1/2} \cdot \ldots \cdot (t-t_k)^{a_k/2} \tag{4–3}$$

from a conformal polygon with vertices at t_i to this orthodisk. Here we take each a_i to be an odd integer. Choose four distinct vertices $t_i, t_{i+1}, t_j, t_{j+1}$ so that $j \equiv i \pmod 2$, ensuring that the edges $t_i t_{i+1}$ and $t_j t_{j+1}$ are parallel in the orthodisk geometry. (See Figure 17.) Introduce a cycle γ in the upper half plane connecting edge (t_i, t_{i+1}) with edge (t_j, t_{j+1}) and denote by $\bar\gamma$ the closed cycle obtained from γ and its mirror image across the real axis. Similarly, denote by β the cycle connecting $(t_{j-1}t_j)$ with $(t_{j+1}t_{j+2})$ and by $\bar\beta$ the cycle together with its mirror image.

Now consider the Schwarz–Christoffel period integrals

$$F(\gamma) = \frac{1}{2}\int_\gamma (t-t_1)^{a_1/2}\cdots(t-t_k)^{a_k/2}, \qquad F(\beta) = \frac{1}{2}\int_\beta (t-t_1)^{a_1/2}\cdots(t-t_k)^{a_k/2},$$

as multivalued functions depending on the *now complex* parameters t_i.

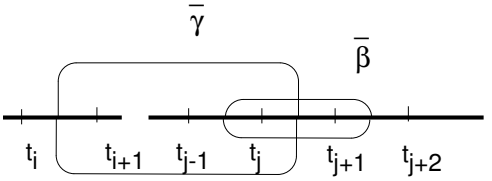


FIGURE 17. Cycles for analytic continuation.

THEOREM 4.3.4. *Under analytic continuation of t_{j+1} around t_j the periods change their values like*

$$F(\gamma) \to F(\gamma) + 2F(\beta), \qquad F(\beta) \to F(\beta).$$

To see this note that the path of analytic continuation of t_{j+1} around t_j gives rise to an isotopy of $\mathbb{C}$ which moves t_{j+1} along this path. This isotopy drags β and γ to new cycles we will call β' and γ', respectively.

Because the curve β is defined to surround t_j and t_{j+1}, the analytic continuation of t_j around t_{j+1} merely returns β to β'. Thus, because β' equals β, their periods are also equal. On the other hand, the curve γ is not equal to the new "dragged" curve γ'. To see this, note that the period of γ' is obtained by developing the flat structure of the doubled orthodisk along γ'. To compute this flat structure, observe the crucial fact that because the exponents in (4–3) are halves of odd integers, the angles at the orthodisk vertices are either $\pi/2$ or $-\pi/2$, modulo 2π; thus the angles of the doubled flat structure equal π, modulo 2π. Thus the arc γ' develops into the union of the arc γ *along with* the arcs β and $\bar{\beta}$; in particular, we see that the period of γ' equals the period of γ plus twice the period of β. □

Now denote by $\delta := t_{j+1} - t_j$ and fix all t_i other than t_{j+1}: we regard t_{j+1} as the independent variable.

THEOREM 4.3.5. *The function $F(\gamma) - \dfrac{\log \delta}{\pi i} F(\beta)$ is single-valued and holomorphic near $\delta = 0$.*

PROOF. By definition, the function is locally holomorphic near $\delta = 0$. By Theorem 4.3.4, it is single-valued. □

Now for the properness argument, we are interested in the geometric coordinates for the moduli space Δ — these are just the absolute values of the the periods. More precisely we are interested in

$$\|F(\gamma)\| := |\operatorname{Re} F(\gamma)| + |\operatorname{Im} F(\gamma)|.$$

We translate the preceding statement about periods into a statement about their respective absolute values.

We consider two conjugate orthodisks parameterized by Schwarz–Christoffel maps F_1 and F_2 defined on the same conformal polygon. Recall that we have constructed β and γ to be purely horizontal or vertical.

With this preparation, we can now prove the technical result Corollary 4.3.3 about degeneration that was the foundation for our properness argument.

PROOF OF COROLLARY 4.3.3. Recall that the above periods are linear combinations of the geometric coordinates where the coefficients are just elements of

$\{1, -1, i, -i\}$. Now by construction, $F_j(\beta)$ is purely real or purely imaginary; moreover, the direction of $F_j(\beta)$ is $\pm i$ times the direction of $F_j(\gamma)$. This, together with Theorem 4.3.5, implies the first claim. Next, note that if we turn left at a vertex in X_1, we will turn right at the corresponding vertex in the conjugate orthodisk X_2, and vice versa. Thus, if the directions of the corresponding edges for γ in corresponding orthodisks differ by $+i$, then the directions of the corresponding edges for β will differ by a $-i$, and vice versa. This implies the second claim. □

This last corollary is really the crux of the properness argument: it is the critical ingredient in the proof of Lemma 4.3.1 and Lemma 4.3.2 that says that the extremal lengths of corresponding cycles on a pair of degenerating orthodisk domains cannot be real-analytically related. We have constructed the height function to exploit the precise differences in the orders of decay.

REMARK 4.3.6. *Properness does require a pair of interlocking simple cycles as above. In very low genus, such pairs may not exist. This is exactly the (only) point at which a proof by our method fails to prove the existence of the Horgan surface.*

Steps 5.3 and 5.4: Noncritical flow and regeneration. We wish to show that if $Z_0 \in \Delta$ is a pair of conjugate orthodisks in the moduli space Δ, then if $\mathcal{H}(Z_0) \neq 0$, then we can embed Z_0 in a family $Z_t \subset \Delta$ so that $\frac{d}{dt}\mathcal{H}(Z_t) \neq 0$. There are two steps to this argument: first we observe that if we find ourselves on a one-dimensional locus $\mathcal{Y} = Z_t$ on which $\mathrm{Ext}_{\Omega_{G\,dh}}(\gamma_i) = \mathrm{Ext}_{\Omega_{(1/G)\,dh}}(\gamma_i)$ for every γ_i with but one exception, then we can differentiate along that locus and find $(d/dt)\mathcal{H}(Z_t) \neq 0$. In the second step, we produce the locus $\mathcal{Y}$ by induction: this is the essence of the "handle addition" in the process. We begin in the next paragraph with the first step, and then conclude the Noncritical Flow step with a discussion of the inductive "regeneration" in the following paragraph.

Noncritical flow along good loci. The domains $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ have a remarkable property: for many choices of cycles γ_i, if $\mathrm{Ext}_{\Omega_{G\,dh}}(\gamma_i) > \mathrm{Ext}_{\Omega_{(1/G)\,dh}}(\gamma_i)$, then when we deform $\Omega_{G\,dh}$ so as to decrease $\mathrm{Ext}_{\Omega_{G\,dh}}(\gamma_i)$, the conjugacy condition forces us to deform $\Omega_{(1/G)\,dh}$ so as to increase $\mathrm{Ext}_{\Omega_{(1/G)\,dh}}(\gamma_i)$. We can thus always deform $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ so as to reduce one term of the height function $\mathcal{H}$.

To see this, suppose that we are so lucky as to find ourselves at a Z where only the one "height" $\mathcal{H}(\Gamma_1) \neq 0$, i.e., we have equality of extremal lengths $\mathrm{Ext}_{\Omega_{G\,dh}}(\Gamma_i) = \mathrm{Ext}_{\Omega_{(1/G)\,dh}}(\Gamma_i)$ of the "coordinate" cycles Γ_i for $i \neq 1$. As extremal length is given by conformally mapping the domain to a rectangle, we can imagine each half of an orthodisk being foliated by the image (under the conformal map) of the horizontal foliation of an appropriate rectangle. This is illustrated in Figure 18.

Note that the foliations on $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ extend to a (singular) foliation on $\mathbf{C}$. If $\mathrm{Ext}_{\Gamma_1}(\Omega_{G\,dh}) > \mathrm{Ext}_{\Gamma_1}(\Omega_{(1/G)\,dh})$, we deform as follows, noting that we can simultaneously decrease $\mathrm{Ext}_{\Gamma_1}(\Omega_{G\,dh})$ while increasing $\mathrm{Ext}_{\Gamma_1}(\Omega_{(1/G)\,dh})$.

In fact, this situation only has to do with alternating left-right turns. (Formally, we compute $(d/dt)\,\mathrm{Ext}_{\Gamma_1}\,\Omega_{G\,dh}(t)$, using well-known formulas [**Gar87**] about derivatives of extremal length functions on Teichmüller space and some of the elements of infinitesimal Teichmüller theory. In fact, when we do this carefully, we see that the heuristic above often does not quite work as drawn — typically we push on a side far away from the images of the corners of the rectangles.)

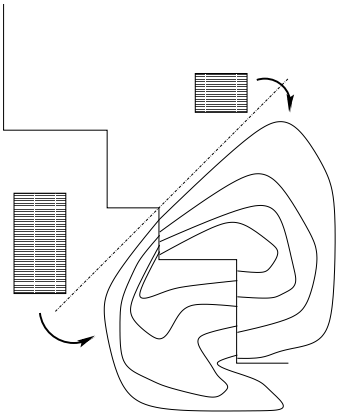


FIGURE 18. The image of the horizontal foliation of a rectangle foliates one of the halves of an orthodisk under a mapping by an appropriate conformal map. Typically, the rectangles required for $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ will be conformally distinct.

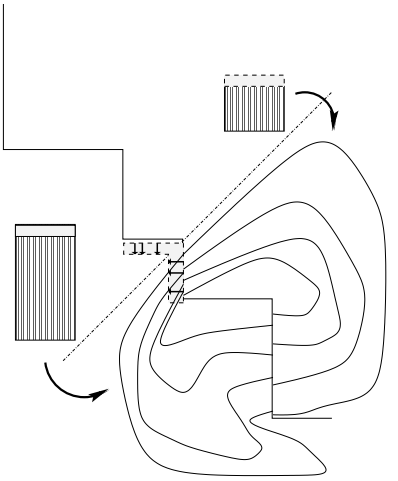


FIGURE 19. Pushing a side decreases one extremal length while increasing the other.

REMARK 4.3.7. *While this process seems quite robust, in very low genus (and in cases where the "good locus" $\mathcal{Y}$ is two-dimensional instead of one-dimensional), the proofs of this step become more involved. See for example* [**WW**].

Regeneration. In the process described in the previous step, an issue arises: we might be able to reduce one term of the height function via a deformation, but this might affect the other terms, so as not to provide for an overall decrease in height. We thus seek a locus $\mathcal{Y}$ in our moduli space where the height function has but a single nonvanishing term, and all the other terms vanish to at least second order. If we can find such a locus $\mathcal{Y}$, we can flow (as described in the previous paragraph) along that locus to a solution.

We find this desired locus by first considering the boundary $\partial\overline{\Delta}$ of the closure $\overline{\Delta}$ of the moduli space Δ: this boundary has strata of moduli spaces Δ' for minimal surface problems of lower complexity. Looking ahead to a proof by induction, we assume that there is a solution X' of those lower complexity (either genus or number of ends or both) problems represented on such a boundary strata Δ' (with all of the corresponding extremal lengths in agreement). Then we prove that there is a locus $\mathcal{Y} \subset \Delta$ inside the larger moduli space Δ (with $\mathcal{Y}$ limiting on X') which has the analogues of those same extremal lengths in agreement. As a corollary of that condition, the height function on $\mathcal{Y}$ has the desired simple properties.

This inductive step is a crucial part of the "handle addition" program. We are unable to see our way through the combinatorics of deforming the orthodisks at an arbitrary point $Z \in \Delta_g$, and so we restrict ourselves to a special sublocus built out of a lower genus solution.

Let us now explore some of the details of this step in the case of our example of the genus four surface E_4 with an end asymptotic to Enneper's surface. Here Δ_4 is a three-dimensional simplex bounded by four two-dimensional faces. Looking back at the figure Figure 5, we recognize that each of the boundary faces is represented by limits of orthodisks degenerating in a simple way: associate to each such boundary face, say $\mathcal{F}_i$, a finite edge, say E_i in the Figure 5. Then, consider a family of orthodisks in which the length of E_i is tending to zero, but the lengths of the other edges are converging to finite and positive numbers. This family converges to a point on $\mathcal{F}_i$, and we see that it is natural to parametrize $\mathcal{F}_i$ by the projectivization of the (positive and finite) lengths of the remaining edges. (We could then proceed to the faces of higher codimension in a similar manner, but this is not necessary for our purposes here.) We focus on the face $\mathcal{F}_1$ obtained by (symmetrically) collapsing the edges H_4H_5 and H_5H_6.

This face $\mathcal{F}_1$ parametrizes conjugate pairs for the problem of finding a genus three Enneper-ended surface E_3. We assume that there is such a solution, say X'.

At this less complex solution X' in the moduli space $\mathcal{F}_1 \subset \overline{\Delta}$, we aim to undo the degeneration which collapsed the two central edges to a point. In particular, we seek a path $\mathcal{Y} \subset \overline{\Delta}$ along which we retain the essential properties of X' that

$$\mathrm{Ext}_{\Gamma_2}(\Omega_{G\,dh}) = \mathrm{Ext}_{\Gamma_2}(\Omega_{(1/G)\,dh})$$

and

$$\mathrm{Ext}_{\Gamma_3}(\Omega_{G\,dh}) = \mathrm{Ext}_{\Gamma_3}(\Omega_{(1/G)\,dh}).$$

Geometrically, we aim to "cut a small" corner into a lower genus pair of orthodisks, and then readjust the lengths of the larger sides so as to satisfy the conditions above. See Figure 20.

By the implicit function theorem, to prove this, it is basically enough to show that X' is locally unique in the (finite-dimensional) lower genus moduli space $\mathcal{F}_1$. But consider a deforming X' by, say, pushing an edge "into" $\Omega_{G\,dh}$. Thus we also push the corresponding edge "out of" $\Omega_{(1/G)\,dh}$.

Once again, we are comparing the flat geometries of $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ to their underlying conformal geometries; this time our interest is in how a prescribed infinitesimal change in flat geometry (corresponding to a prescribed change in periods) affects the underlying conformal geometry. We study this using the standard calculus from Teichmüller theory [**Ahl61**]: tangent vectors to Teichmüller space are represented by Beltrami differentials, which are tensors of type $d\overline{z} \otimes \frac{\partial}{\partial z}$ that can

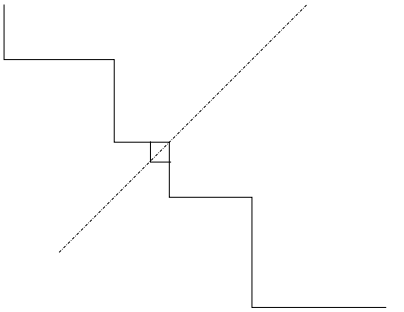


FIGURE 20. Orthodisks for $G\,dh$ for E_4.

be readily computed from a prescribed deformation of structure like the "pushing into" or "pushing out of" an edge. (See [**WW**], for example, for full details.)

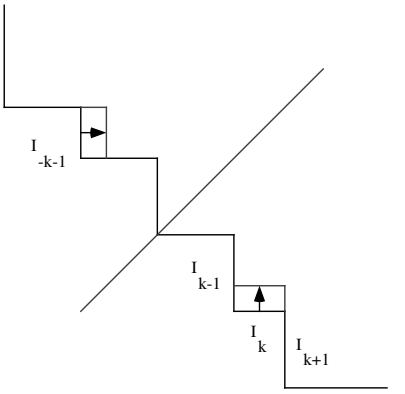


FIGURE 21. We push the k-th edge I_k and its reflection I_{-k-1} into $\Omega_{G\,dh}$.

Formally then, the first operation of "pushing in" corresponds to an infinitesimal Beltrami differential $\dot{\mu}_G$ on $\Omega_{G\,dh}$, and the second operation of "pushing out" corresponds to an infinitesimal Beltrami differential $\dot{\mu}_{1/G}$ on $\Omega_{(1/G)\,dh}$. But as X' is a solution, we know that there is a *conformal* map $w : \Omega_{G\,dh} \to \Omega_{(1/G)\,dh}$, and we can compute, using the tensor type of a Beltrami differential together with the "staircase-like" geometry of $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$, that

$$w^*(\dot{\mu}_{1/G}) = -\dot{\mu}_G \tag{4–4}$$

Thus, any such deformation deforms the conformal structures of $\Omega_{G\,dh}$ and $\Omega_{(1/G)\,dh}$ in *opposite* directions, where here by directions we mean as tangent vectors to appropriate Teichmüller spaces.

Remarks:

(1) The Noncritical Flow/Regeneration inductive step seems quite general. Thus, in some sense, the most difficult period problem to solve is the low genus one. It is therefore significant that Horgan does not exist, as its existence might then have yielded existence of "k-handled Horgan".

(2) We produced the face $\mathcal{F}_1$ by a process of allowing the orthodisks to degenerate: we (symmetrically) collapsed the middle edges H_4H_5 and H_5H_6 of the structures in

Figure 5. On the surfaces obtained by taking a branched cover over the double of one of these structures, we would see lifts of the points H_4, H_5, and H_6 approaching one another (as measured in the flat metric) along such a degenerating family. Formally however, in terms of the underlying conformal structures, it does not make sense to speak of points coalescing on Riemann surfaces; it is only proper to speak of curves pinching off to form noded surfaces. (Noded surfaces are complex spaces where each neighborhood is either biholomorphic to a disk or biholomorphic to a domain $\{zw = t : |z| < \varepsilon, |w| < \varepsilon\}$ in $\mathbb{C}^2$. The complements of the nodes (where $z = w = 0$ above) are punctured Riemann surfaces, occasionally referred to as the regular components of the noded surface.) In the case of the Enneper-ended surfaces, since the limit of a degenerating family of structures $\Omega_{G\,dh}$ was a lower genus flat structure of the type $\Omega'_{G\,dh}$, we "see" only one regular component of the noded surface, as the limiting one-form on the other component is zero. This does not always hold in handle addition; for example in the Costa-type surfaces of [**WW**], the degenerate structures appear to be of the form of less complex structures, along with other very simple components. (See Figure 22 for the limit of a degenerating family of the form Figure 6.) The regeneration step is then more complex than the relatively simple "cutting of the corner" of Figure 20.

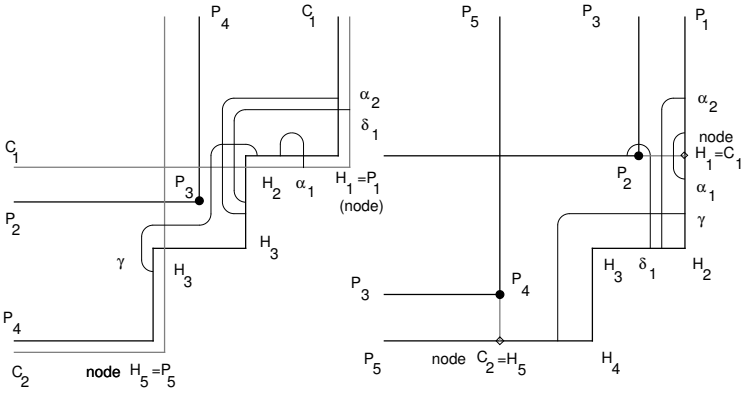


FIGURE 22. The limit of a family of degenerating surfaces of the type shown in Figure 6.

Here, on the right hand side, the regular components of the noded surface appear either as lower complexity orthodisks or as points, while on the left-hand side, the regular components which are not lower complexity orthodisks appear as nontrivial bigons.

5. Beyond Enneper-Ended Surfaces

We end with a few comments about the adjustments required to adapt the outline above to the surfaces of the types in Theorems 1.0.3, 1.0.4 and 1.0.5, as well as deformations of those surfaces in continuous families. Basically, the above argument goes through relatively unchanged for the surfaces of Costa type. Of course, we need to choose the appropriate cycles, but then the Reflexivity and Properness proofs are identical. The discussion of Noncritical Flow and Regeneration are a bit

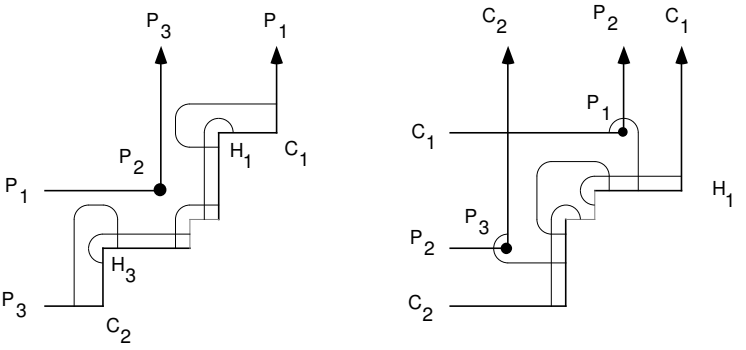


FIGURE 23. Regenerating orthodisks for $DH_{m,n}$, case $m < n$.

different however, as the degenerations of these orthodisks are different. In particular, when we regenerate, there are two parameters to add, and we need to always solve a two-dimensional flow problem.

An interesting feature of this is that, given a surface of type (m, n), we can choose to regenerate in either this two-parameter way, or in the way we discussed with respect to the Enneper ends (Figure 23).

This second method leads to surfaces like $DH(1, 2)$, shown in Figure 24.

FIGURE 24. The surface $DH(1, 2)$.

This represents how we fill in the superdiagonal entries in Table 1 (page 213).

Finally, we note that it is possible to deform the planar ends of the surfaces described in Theorem 1.0.4 into catenoid ends with specified growth rates near zero. A new type of special point is introduced into the orthodisk, but only mild modifications are required for the proofs.

References

[Ahl61] L. Ahlfors. Some remarks on Teichmüller's space of Riemann surfaces. *Ann. Math.*, 74:171–191, 1961.

[Ahl66] L. Ahlfors. *Lectures on quasiconformal mappings.* Van Nostrand, New York, 1966.

[BW] E. Boix and M. Wohlgemuth. Numerical results on embedded minimal surfaces of finite total curvature. In preparation.

[CG81] C. C. Chen and F. Gackstatter. Elliptic and hyperelliptic functions and complete minimal surfaces with handles. *Instituto de Matemática e Estatística, Universidade de São Paulo*, 27, 1981.

[CG82] C. C. Chen and F. Gackstatter. Elliptische und Hyperelliptische Functionen und vollständige Minimalflächen von Enneperschen Typ. *Math. Ann.*, 259:359–369, 1982.

[CHMI89] M. Callahan, D. Hoffman, and W. H. Meeks III. Embedded minimal surfaces with an infinite number of ends. *Inventiones Math.*, 96:459–505, 1989.

[Cos84] C. Costa. Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends. *Bull. Soc. Bras. Mat.*, 15:47–54, 1984.

[Enn69] A. Enneper. Die cyklischen Flächen. *Z. Math. und Phys.*, 14:393–421, 1869.

[Gar87] F. Gardiner. *Teichmuller Theory and Quadratic Differentials.* Wiley Interscience, New York, 1987.

[HK97] D. Hoffman and H. Karcher. Complete embedded minimal surfaces of finite total curvature. In *Encyclopedia of Mathematics*, pages 5–93, 1997. R. Osserman, editor, Springer Verlag.

[HMI85] D. Hoffman and W. H. Meeks III. A complete embedded minimal surface in $\mathbb{R}^3$ with genus one and three ends. *Journal of Differential Geometry*, 21:109–127, 1985.

[HMI90] D. Hoffman and W. H. Meeks III. Embedded minimal surfaces of finite topology. *Annals of Mathematics*, 131:1–34, 1990.

[HW] D. Hoffman and F. Wei. Construction of a helicoid with a handle. Preprint.

[JI83] L. Jorge and W.H. Meeks III. The topology of complete minimal surfaces of finite total curvature. *Topology*, 22:203–221, 1983.

[Kar88] H. Karcher. Embedded minimal surfaces derived from Scherk's examples. *Manuscripta Math.*, 62:83–114, 1988.

[Kar89] H. Karcher. Construction of minimal surfaces. *Surveys in Geometry*, pages 1–96, 1989. University of Tokyo, 1989, and Lecture Notes No. 12, SFB256, Bonn, 1989.

[Kar91] H. Karcher. Construction of higher genus embedded minimal surfaces. In *Geometry and Topology of Submanifolds III (Leeds, 1990)*, pages 174–191, River Edge, NJ, 1991. World Sci. Publishing.

[MW01] F. Martin and M. Weber. On properly embedded minimal surfaces with three ends. *Duke Math. Journal*, 107:533–559, 2001.

[Oht70] M. Ohtsuka. *Dirichlet Problem, Extremal Length, and Prime Ends.* Van Nostrand Reinhold, New York, 1970.

[San94] N. Do Espirito Santo. Complete minimal surfaces with type enneper end. *Ann. Inst. Fourier (Grenoble)*, 44:525–577, 1994.

[Sat96] K. Sato. Existence proof of one-ended minimal surfaces with finite total curvature. *Tohoku Math. J.*, 48:229–246, 1996.

[Sch83] R. Schoen. Uniqueness, symmetry, and embeddedness of minimal surfaces. *Journal of Differential Geometry*, 18:791–809, 1983.

[Tha94] E. Thayer. *Complete Minimal Surfaces in Euclidean 3-Space.* PhD thesis, University of Massachusetts at Amherst, 1994.

[Web98] M. Weber. On the Horgan minimal non-surface. *Calc. Var.*, 7:373–379, 1998.

[Web00] M. Weber. On singly periodic minimal surfaces invariant under a translation. *Manuscripta Mathematica*, 101:125–142, 2000.

[Web03] M. Weber. Classical minimal surfaces in euclidean space by examples: geometric and computational aspects of the Weierstrass representation, This volume, pp. 19–63.

[WW] M. Weber and M. Wolf. Teichmüller theory and handle addition for minimal surfaces. *Annals of Math.*, 156:713–795, 2002.

[WW98] M. Weber and M. Wolf. Minimal surfaces of least total curvature and moduli spaces of plane polygonal arcs. *Geom. and Funct. Anal.*, 8:1129–1170, 1998.

[WW01] M. Weber and M. Wolf. *Minimal Surfaces and Teichmueller Theory*. 2001. In preparation.

Department of Mathematics, Rice University, Houston, TX 77005
E-mail address: mwolf@math.rice.edu

Clay Mathematics Proceedings
Volume 2, 2004

The Genus-One Helicoid as a Limit of Screw-Motion Invariant Helicoids with Handles

Matthias Weber
(Joint work with D. Hoffman and M. Wolf)

Contents

1. Introduction 243
2. The Translation-Invariant Helicoid with Handles 244
3. The Screw-Motion Invariant Helicoids with Handles 251
4. The Genus-One Helicoid 255
References 257

1. Introduction

Besides the notorious plane, the only known complete embedded minimal surfaces of finite topology (that is, of finite genus and with finitely many ends) in euclidean space in the nineteenth century were the catenoid and the helicoid.

By a theorem of Collin [**Col97**], any such surface has either *finite total curvature* or just *one end*. The finite total curvature case is rather well developed. There are many examples (the Costa surface and its descendants) and many interesting conjectures (see besides others [**Cos82, Cos84, Cos89, HMI85, HMI87, HMI89, HMI90, Kap97, Kar89, Tra98, Tra01, WW98, WW02, Woh91, Woh93**]).

In the 1-ended case, there is only one new example so far, the Genus-One Helicoid [**HKW93a, HKW93b**]. Its existence proof (Hoffman, Karcher, Wei) was very complicated, and for the embeddedness one had to believe the computer graphics by Jim Hoffman.

However, Hoffman and Wei [**HW**] outlined another, more conceptual approach. They proposed the existence of a family of screw motion invariant minimal surfaces with a helicoidal end and periodic handles. This family should contain both the translation-invariant helicoid with handles, a surface known to exist and to be embedded [**HKW99**], and, as a limit, the genus-one helicoid. Using the maximum principle and known results about helicoidal ends [**HM99, HPR99**], the embeddedness of the genus-one helicoid would follow.

©2004 Clay Mathematics Institute

They derived the Weierstrass data for this family and provided numerical and graphical evidence that this family indeed exists. However, they were not able to solve the period problem at this time.

We will here present joint work of Hoffman, Wolf and myself [**WHW01**] which is based on the original idea of Hoffman and Wei, the cone metric technique developed by Wolf and myself, and my Habilitation thesis [**Web00**] which proves both the existence of the proposed family and the embeddedness of the genus-one helicoid.

In the first part, we will give a short complete proof of the existence of the translation-invariant helicoid with handles, as this serves as the model case for the general case, which will be sketched in the second part.

2. The Translation-Invariant Helicoid with Handles

THEOREM 2.1. *There exists a complete embedded minimal surface $\mathcal{H}_1$ in $\mathbb{R}^3$ with the following properties:*

(1) *The surface is invariant under a vertical translation.*
(2) *The quotient of the surface by the translation has genus 1.*
(3) *The quotient surface has two helicoidal ends.*
(4) *The vertical coordinate axis lies on the surface and is a symmetry line.*
(5) *The quotient surface has two parallel horizontal lines crossing the vertical axis which are lines of (the same) symmetry.*

2A. The Weierstrass data. The quotient surface can be represented by a torus $\mathcal{T}_1$ punctured at two points corresponding to the two helicoidal ends. This torus is necessarily *rhombic*, as the required symmetries have connected fixed point sets.

At the ends, the height differential must have simple poles with *purely imaginary residues* at the ends. To locate the zeros, note that these will be permuted by rotations around the vertical or a horizontal axis. Since there are only two and they can't lie on the vertical axis, they must both lie on one of the horizontal axes. At these points, the Gauss map is vertical.

Normalize a fundamental domain for $\mathcal{T}_1$ to have vertices at $\pm 1, \pm\tau i$ so that the vertical diagonal represents the vertical straight line and the horizontal diagonal represents the horizontal straight lines. Using symmetry, we can then write

$$V_1 = -a, \qquad V_2 = a, \qquad E_1 = -b, \qquad E_2 = b,$$

with $a < b < 1$. Abel's theorem applied to G forces

$$a + b = 1.$$

Here is the divisor table for $\mathcal{H}_1$:

	E_1	V_1	V_2	E_2
dh	∞	0	0	∞
G	∞	∞	0	0
$G\,dh$	∞^2	$*$	0^2	$*$
$(1/G)\,dh$	$*$	0^2	$*$	∞^2

The horizontal diagonal of $\mathcal{T}_1$ corresponds to the two horizontal straight lines and the vertical diagonal to the vertical line of $\mathcal{H}_1$. We will encounter a slightly

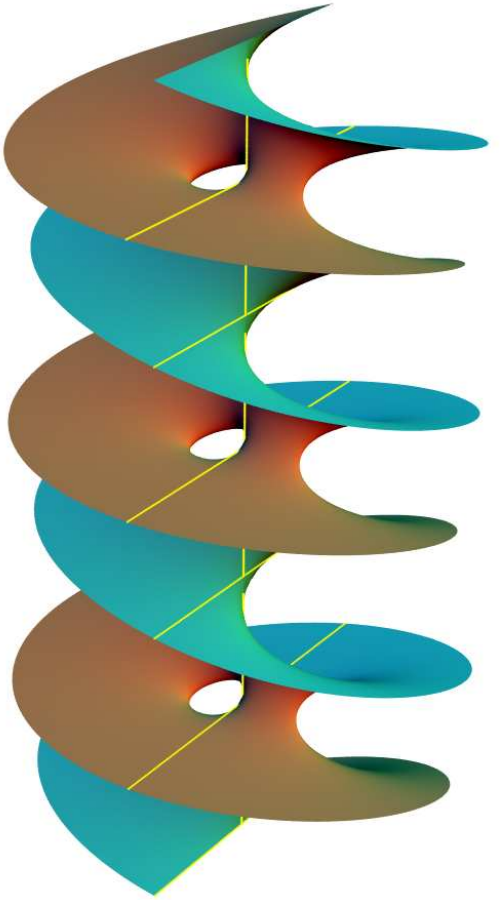


FIGURE 1. The translation-invariant helicoid with handles.

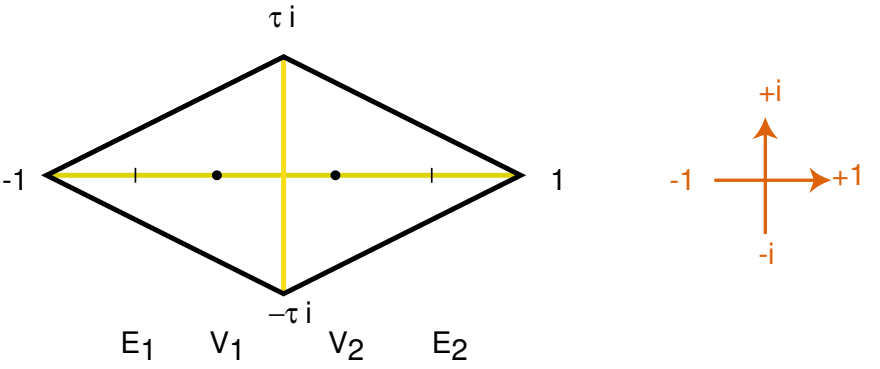


FIGURE 2. Divisor of the Weierstrass data.

confusing technical problem in these figures: The imaginary direction in the complex plane does not always correspond to the vertical direction in the 3-dimensional

euclidean space where the surface is sitting. To avoid any misunderstandings, we have included compasses in each planar figure in this section.

2B. The horizontal period condition. The *horizontal* period condition requires that

$$\int_\alpha G\,dh = \overline{\int_\alpha \frac{1}{G}\,dh}$$

for all closed cycles α on the torus $\mathcal{T}_1$.

This is equivalent to the condition that the periods of the first two complex coordinate differentials are purely imaginary; see the author's companion article starting on page 19 of this volume.

Consider the involution $\iota : z \mapsto -z$. It carries the divisor of $G\,dh$ to the divisor of $(1/G)\,dh$, and as the differential at a fixed point is -1 we have

$$\iota^* \frac{1}{G}\,dh = -G\,dh.$$

Using the period condition it follows that

$$\int_\alpha G\,dh = -\int_\alpha \iota^* \frac{1}{G}\,dh = -\int_{\iota(\alpha)} \frac{1}{G}\,dh = \int_\alpha \frac{1}{G}\,dh = \overline{\int_\alpha G\,dh},$$

so that all the periods of $G\,dh$ need to be *purely real*.

Slit the plane from -1 to 1, and glue the $[-1, 0]$ portion of the upper part of the slit to the $[0, 1]$ portion of the lower part of the slit, and vice versa.

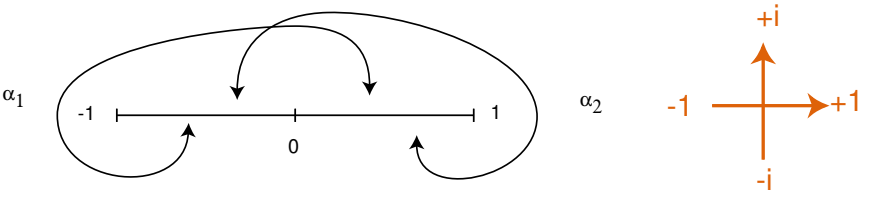


FIGURE 3. Cone metric construction of the rhombic torus and the meromorphic 1-form.

This defines a *cone metric* on a rhombic torus with trivial holonomy and two cone points. At the point $\{-1, 0, 1\}$, the cone angle is 6π, and at ∞, the cone angle is -2π. Hence the cone metric comes from a 1-form with double order pole and zero and only real periods.

We identify this cone metric torus with a rhombic lattice torus $\mathcal{T}_1$ so that the imaginary axis of the slit plane becomes the horizontal diagonal of the rhombus. We translate the cone metric along the horizontal diagonal so that the 6π cone point coincides with $V_2 = a$. This defines $G\,dh$. Finally, define $(1/G)\,dh = -\iota^* G dh$.

It is conceivable that there are other tori and forms $G\,dh$ with the desired conditions. On a rhombic torus one can show that this is in fact the only choice, see [**Web02**].

Using the *jigsaw puzzle* in Figure 4, we can decompose the cone metric into two planar domains. In Figure 5 the cycles α_1 and α_2 from Figure 3 are included.

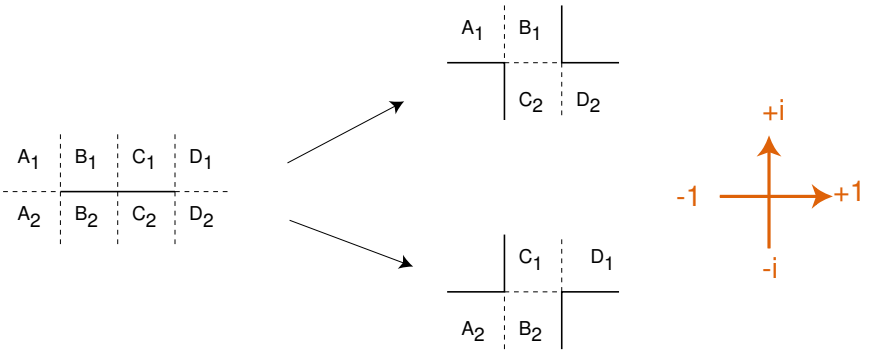


FIGURE 4. A jigsaw puzzle.

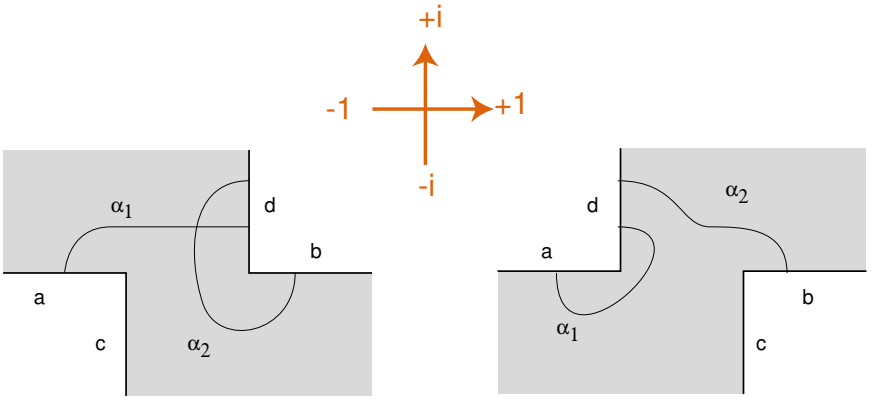


FIGURE 5. Alternative description of the cone metric.

2C. The vertical period condition. Since we have defined $G\,dh$ and $\frac{1}{G}\,dh$ on a fixed torus $\mathcal{T}_1$, dh is also defined up to sign. The only freedom we are left with is to vary the parameter a.

We first check that the residues of dh at E_1 and E_2 are *purely imaginary*. This is necessary so that the ends at E_1 and E_2 are helicoidal. Consider a positive real tangent vector V at the center 0 of the rhombic fundamental domain. By definition, $Gdh(V)$ is positive imaginary, and $1/G\,dh(V) = -\iota^* G\,dh(V) = Gdh(V)$ is as well. Hence

$$dh(V) = \sqrt{G\,dh(V) \cdot 1/G\,dh(V)} \in i\mathbb{R}$$

at 0. Now note that dh is odd, as the divisor is invariant under ι and it can't be even, as it then would descend to the quotient sphere. Therefore this condition holds on the whole horizontal diagonal. Hence the residue of dh at the ends E_1 and E_2 is purely imaginary, as desired.

We now choose the sign of dh so that at 0,

$$dh(V) \in i\mathbb{R}^-.$$

The vertical period condition requires, for the cycle γ in Figure 6, that

$$\mathrm{Re} \int_\gamma dh = 0.$$

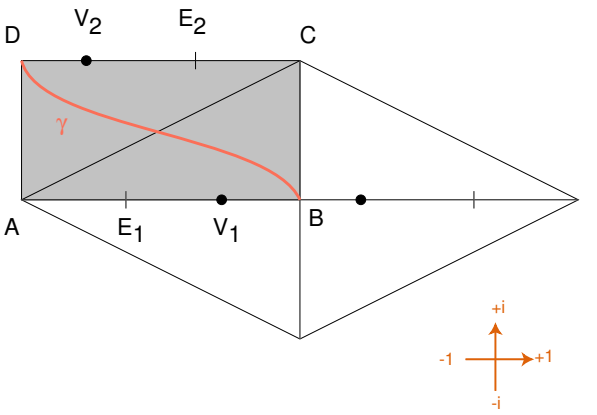


FIGURE 6. Vertical period condition.

We call the shaded rectangle the *fundamental rectangle*. Its boundary edges are the straight lines on the surface. The fundamental rectangle is mapped by the surface parameterization to a surface patch which looks as in Figure 7. The path γ is mapped to a closed curve connecting B with D around the handle.

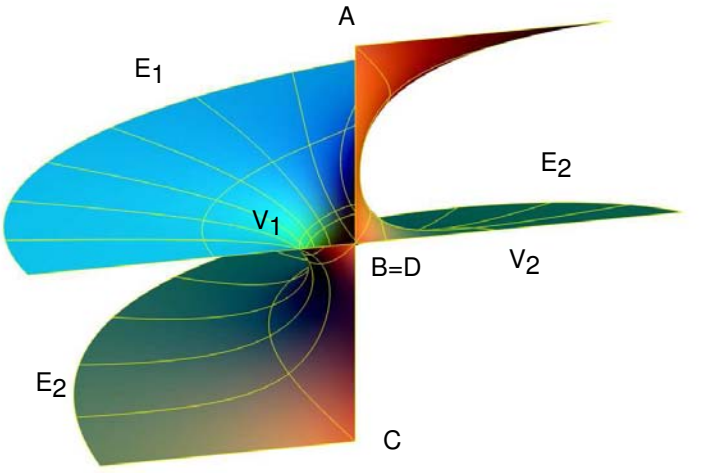


FIGURE 7. Image of the fundamental rectangle.

PROPOSITION 2.2. *The periods of dh are continuous in a for* $a \in [0, \frac{1}{2}]$.

PROOF. The argument requires that we spend a few lines on the meaning of the word continuous. For the spaces of meromorphic 1-forms and functions on a fixed torus, we will use the topology of locally uniform convergence. This topology has the property that translations of forms, products, other standard arithmetic operations and integrals over cycles avoiding singularities are continuous. It is by no means the only topology with this property, but it is quite convenient.

Now, the forms $G\,dh$ and $(1/G)\,dh$ depend on a only by a translation and hence make sense for all a. By the definition of dh, since

$$dh = \sqrt{G\,dh \cdot \frac{1}{G}\,dh},$$

it follows that also dh depends continuously on a for all a, and therefore also the periods of dh depend continuously on a. □

PROPOSITION 2.3. *For $0 < a < \frac{1}{2}$, the map $\int^z dh$ maps the fundamental rectangle to a polygonal domain. The positive real axis points up in the dh figures so that up in the figure means up in space. The edges emanating from V_1 and V_2 are slits, and the other horizontal edges extend to infinity to the left and right, so that the region has two infinite half-strips as ends.*

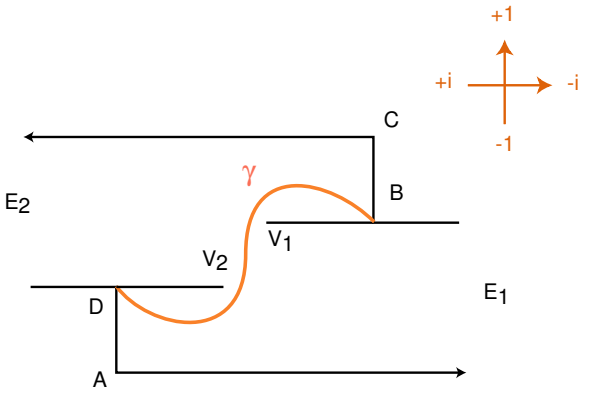


FIGURE 8. Image of the fundamental rectangle under $\int dh$.

PROPOSITION 2.4. *For $a = \frac{1}{2}$, $\int^z dh = \lambda i z + c$ with $\lambda \in \mathbb{R}^-$.*

PROOF. In this case, $a = b = \frac{1}{2}$ so that $G\,dh$ and $(1/G)\,dh$ cancel to $dh = \lambda i\,dz$ (recall that we defined $(1/G)\,dh = -\iota^* G\,dh$). □

Figure 9 shows the image of $\int^z dh$ for $a = 0.49$.

PROPOSITION 2.5. *For $a = 0$, $\int^z dh$ maps the fundamental rectangle to the domain in Figure 10.*

PROOF. In this case, $b = 1$ so that $G\,dh = -(1/G)\,dh$ and hence $dh = iG\,dh$. At this point the referee complained that one can deduce $G \equiv -i$ in the limit case. This is indeed the case, but doesn't help at all — we need to determine boundary values for period integrals to be able to apply the intermediate value theorem.

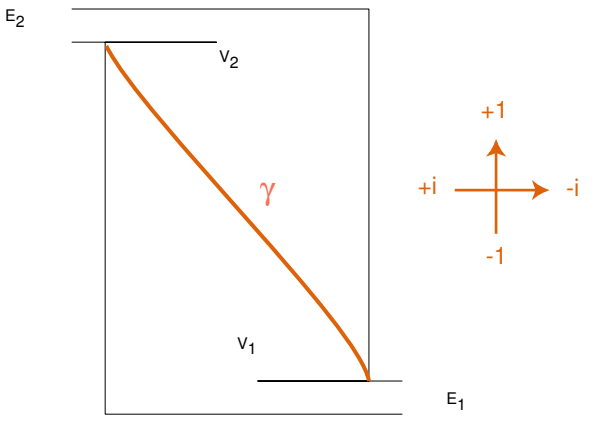


FIGURE 9. Vertical period condition.

In this case, the form $iG\,dh$ is symmetric with respect to the rhombus diagonals, and hence the fundamental rectangle is mapped to one of the (rotated) pieces of the jigsaw puzzle (Figure 4), because it decomposes into these two pieces under the same symmetries. It must be the piece in Figure 10, because, by construction, the edge E_1V_1 must point to the left, and this rules out the other jigsaw piece. □

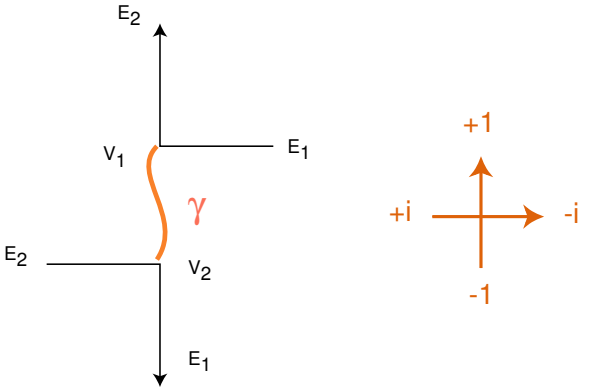


FIGURE 10. Limit domain for dh at $a = 0$.

From the intermediate value theorem we get

THEOREM 2.6. *There is a value* $0 < a < \frac{1}{2}$ *which solves the vertical period problem.*

This concludes the existence proof of the translation-invariant helicoid with handles.

REMARK 2.7. The embeddedness of this surface was proved first in [**HKW99**]. There is now an alternative proof by F. Martín which uses the above deformation of the fundamental piece in the parameter a.

3. The Screw-Motion Invariant Helicoids with Handles

In this section, we modify the proof for the existence of the translation-invariant helicoid with handles to screw-motion invariant helicoids with handles. One of the key advantages of the cone metric approach is that multivalued forms can be treated easily.

The main additional difficulty in this section comes from the fact that we have to allow the underlying torus to vary as well as the 1-forms. This makes it necessary to be very careful with the topology on the parameter spaces. We will be rather sketchy in the arguments, however, and refer the reader to [**WHW01**] for the full details.

We begin by reviewing conditions on the Weierstrass data for screw-motion invariant minimal surfaces in general.

3A. Singly periodic minimal surfaces invariant under a screw motion. Suppose Σ is a complete minimal surface in $\mathbb{R}^3$ invariant under a *vertical screw motion* with rotation angle ϕ and translation t. Let X be the quotient surface. Cutting Σ with a horizontal plane defines a homology class $\eta \in H_1(X, \mathbb{Z})$.

PROPOSITION 3.1. *The Weierstrass data $(dG/G, dh)$ are well-defined on X and satisfy*

(1) $\int_\gamma dG/G = \#(\gamma, \eta) \cdot \phi$,
(2) $\int_\gamma dh = \#(\gamma, \eta) \cdot t$,
(3) $\int_\gamma G\, dh = \overline{\int_\gamma 1/G\, dh} \qquad \forall \#(\gamma, \eta) = 0.$

If, on the other hand, Weierstrass data $(dG/G, dh)$ satisfy these conditions, they define a screw motion invariant surface as stated.

Here is a list of the known examples of screw-motion invariant minimal surfaces:

(1) Karcher [**Kar88, Web00**]: twisted Scherk surfaces
(2) Hoffman, Karcher and Wohlrab [**Lyn93**]: twisted Fischer–Koch surfaces
(3) Callahan, Hoffman and Karcher [**CHK93**]: twisted Callahan–Hoffman–Meeks surfaces
(4) Hoffman, Karcher and Wei [**HKW93a, HKW93b, HKW99, HW**] and Weber, Hoffman and Wolf [**Web00, WHW01**]: twisted helicoids with handles
(5) Traizet and Weber [**TW01**]: many new examples, among them helicoids with arbitrarily many handles in the quotient

3B. The screw-motion invariant helicoids with handles.

THEOREM 3.2. *For any angle $\pi < \phi < \infty$, there exists a complete embedded minimal surface $\mathcal{H}_k$ in $\mathbb{R}^3$ with the following properties:*

(1) *The surface is invariant under a vertical screw motion with angle $\phi = 2\pi k$ and translation t.*
(2) *The quotient of the surface by the screw motion has genus 1 and two helicoidal ends.*
(3) *The vertical coordinate axis lies on the surface.*
(4) *The quotient surface has two horizontal lines crossing the vertical axis.*
(5) *For the twist angle we have $\phi = \int_\nu dG/G$, where ν is a segment of the vertical coordinate axis of length t.*

To prove this theorem, we begin by analyzing the divisors of the Weierstrass data. The multivaluedness of the Gauss map is visible as the real exponent k at the ends.

The divisors of G and dh are given by

	E_1	V_1	V_2	E_2
dh	∞	0	0	∞
G	∞^k	∞	0	0^k
$G\,dh$	∞^{k+1}	$*$	0^2	0^{k-1}
$(1/G)\,dh$	0^{k-1}	0^2	$*$	∞^{k+1}

Instead of using dG/G to define G, we construct a cone metric for $|G\,dh|$ as follows: Take the slit torus domain as for the translation-invariant helicoid with handles, choose a point hi on the positive imaginary axis, slit the positive imaginary axis above this point to $i\infty$ and sew in a cone with cone angle $2\pi(k-1)$ so that the total cone angle at hi becomes $2\pi k$. This cone metric defines a multivalued 1-form $G\,dh$.

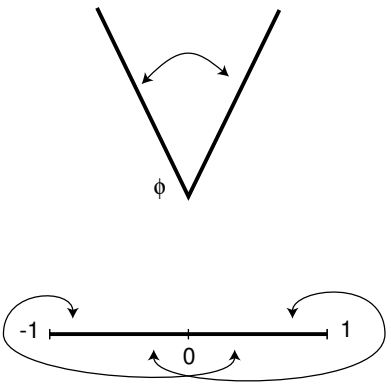


FIGURE 11. Cone metric for $G\,dh$.

For each $(k,h) \in K = (\frac{1}{2},\infty] \times [0,\infty]$, we have defined a cone metric $|G\,dh|$. Note that for $h = 0$ the construction still makes sense even though cone points have coalesced. For $h = \infty$, the whole cone has disappeared. It will be of great importance to ensure that further constructions are continuous in (k,h) on the full set K.

We identify the cone metric torus with a rhombic torus $\mathcal{T}_{k,h}$ given by a rhombus with corners at $\pm 1, \pm i\tau$ so that the vertical axis becomes the horizontal rhombus diagonal and so that the divisor becomes symmetric at the vertical rhombic diagonal. We can arrange so that

$$V_1 = -a, \qquad V_2 = a, \qquad E_1 = -b, \qquad E_2 = b.$$

Then $(1/G)\,dh$ is defines as $-\iota^* G\,dh$ with $\iota(z) = -z$.

For $k = \infty$, the same construction works as long as we realize that a cone with infinite cone angle represents only a single cone point. This means $E_1 = E_2$ and hence $b = 1$ in the construction. We will come back to this later.

In any case, this pair of cone metrics defines a pair of of multivalued 1-forms with the right monodromy so that the horizontal period condition is satisfied.

In this screw-motion invariant situation, the solution to the horizontal period problem does *not* define the underlying torus uniquely. We find a 1-parameter family of candidates $\mathcal{T}_{k,h}$ where $h \geq 0$. However, the parameter k is the only free parameter left, because it specifies the position of the cone point E_2 in addition to V_2 and E_1. Hence the divisor of dh is completely determined (by symmetry).

3C. A theorem of Abel.

PROPOSITION 3.3. *Suppose there is a family of screw-motion invariant surfaces $\mathcal{H}_k$ as in Theorem 3.2 so that $\mathcal{H}_1$ is the translation-invariant helicoid with handles. Then necessarily*

$$a + kb = k.$$

PROOF. Apply the reciprocity formula to the forms dz and $\frac{dG}{G}$: Let R be a rhombic fundamental domain of $\mathcal{T}_{k,h}$ with corners at ± 1 and $\pm i\tau$. Then, by the residue theorem,

$$\int_{\partial R} z \cdot \frac{dG}{G} = 4\pi i(ka + b).$$

On the other hand, by direct computation

$$\int_{\partial R} z \cdot \frac{dG}{G} = (1 + i\tau)2\pi k - (1 - i\tau)2\pi k = 4\pi ik.$$

Here we use that, along the boundary edges, G changes its value by $e^{\pm 2\pi ik}$. This is done deliberately, but this is the only choice which is *continuous* in k and holds with the translation invariant case ($k = 1$). □

3D. Continuity of the vertical period. Let γ be the cycle connecting 0 with $-1 + \tau i$ diagonally in the fundamental rectangle. Let $K = (\frac{1}{2}, \infty] \times [0, \infty]$ be the *parameter rectangle*. Then define

$$H : K \to \mathbb{R}, \qquad H(k, h) = \operatorname{Re} \int_\gamma dh.$$

PROPOSITION 3.4. *H is continuous in K and real analytic in $(\frac{1}{2}, \infty) \times (0, \infty)$.*

PROOF. The real analyticity in the interior of K is clear as H is given by the real part of a composition of holomorphic period maps and inverse period maps.

Hence we need to understand continuity only at the boundary. We distinguish two cases.

Suppose first that h is large. Decompose the $|G\,dh|$ cone metric into two domains, as shown in Figure 12. These domains correspond to two domains in $\mathcal{T}_{k,h}$, as shown in Figure 13.

The white region with the $|G\,dh|$ cone metric is isometric to the complement $\Omega = \Omega(k, h)$ of a neighborhood of 0 in the complex plane with a cone metric given *explicitly* by

$$ds = \left| d\left(1 + \frac{z}{kh}\right)^k \cdot h \right| = \left| \left(1 + \frac{z}{kh}\right)^{k-1} dz \right|.$$

This metric has cone points at $-kh$ and ∞, and the point $-kh$ is at distance h from 0, as

$$\int_{-kh}^{0} d\left(1 + \frac{z}{kh}\right)^k \cdot h = \left(1 + \frac{z}{kh}\right)^k \cdot h \Bigg|_{-kh}^{0} = h.$$

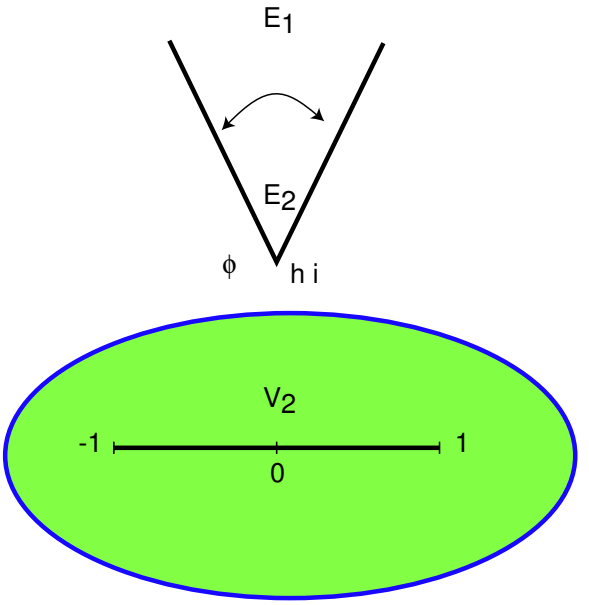


FIGURE 12. Domain decomposition.

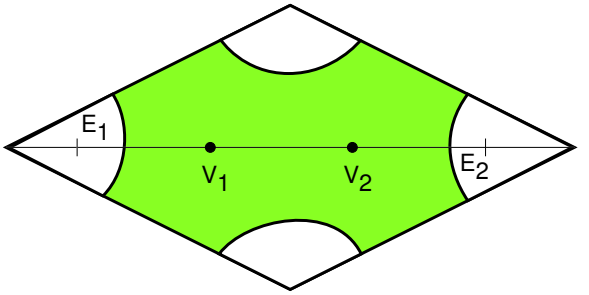


FIGURE 13. Domain decomposition.

To understand the convergence of the cone metric $|G\,dh|$, it suffices to understand the convergence of the metric ds and of the gluing maps between the cone metric in Ω and the shaded region in $\mathcal{T}_{k,h}$. The gluing maps are defined in an overlap region of the shaded and white domains.

Now for $h \to \infty$, $ds \to |dz|$ uniformly in k and locally uniformly in z. Similarly, for $k \to \infty$, $ds \to |e^{z/h} dz|$.

In particular, the convergence is uniform in the compact gluing region. These statements suffice to conclude that the 1-forms $G dh$ (and hence $(1/G)\,dh$, dh and $H(k,h)$) converge locally uniformly for (k,h) converging to the upper and right boundary of K.

For $h \to 0$, one uses another decomposition of the cone metric. This is topologically more complicated, but the convergence is simpler as only finite cone points coalesce. Details can be found in [**WHW01**].

□

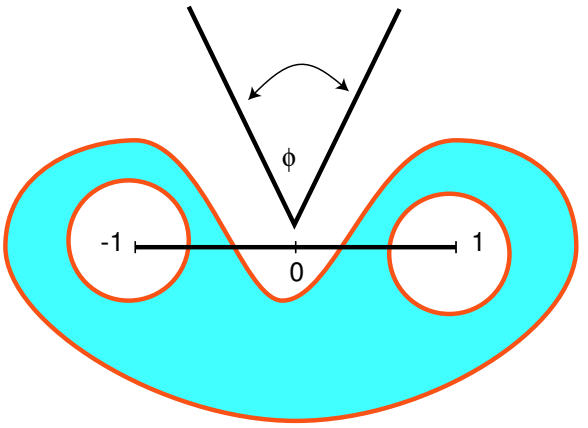


FIGURE 14. Domain decomposition.

Using the continuity of H, one proves the existence of an $\mathcal{H}_k$ exactly in the same way as for the translation-invariant $\mathcal{H}_1$:

The following two propositions are proved in exactly the same way as Propositions 2.4 and 2.5 in the translation invariant case.

PROPOSITION 3.5. *For $h = 0$, $\int^z dh = \lambda i z + c$.*

This gives by the intermediate value theorem:

PROPOSITION 3.6. *For $h = \infty$, the cone metric $|G\,dh|_\infty$ is the same as for $\mathcal{H}_1$ and hence $\int^z dh = \int G dh_\infty$.*

THEOREM 3.7. *For any $k > \frac{1}{2}$, there is a value $0 < h < \infty$ (or $\frac{k}{k+1} < b < 1$) such that the vertical period problem for the twisted genus-1 helicoid (resp. the genus-one helicoid, if $k = \infty$) is solved.*

REMARK 3.8. The method breaks down for $k \leq \frac{1}{2}$, as then the limit case $h = 0$ cannot be realized anymore. There is strong theoretical [**TW01**] and numerical evidence that the $k = \frac{1}{2}$ surface does not exist.

4. The Genus-One Helicoid

4A. Existence.

THEOREM 4.1. *There exists a complete embedded minimal surface in $\mathbb{R}^3$ with the following properties:*

(1) *The surface has genus 1 and infinite total curvature.*
(2) *The surface has one helicoidal end.*
(3) *The vertical coordinate axis lies on the surface and is a symmetry line.*
(4) *The Gauss map has an essential singularity at the end.*
(5) *The surface has one parallel horizontal line crossing the vertical axis.*

The existence part of the proof is very similar to the one for the twisted helicoids: For the $|G\,dh|$ cone metric, one uses a cone with an infinite cone angle. The height differential has a double order pole at the end and two simple zeros at the finite points where the Gauss map is vertical. These zeros can coalesce either at the

end or at the center of the torus. These limit cases correspond again to explicitly known non-solutions of the vertical period problem with opposite signs. Details can again be found in [**WHW01**].

4B. Continuity of the family.

THEOREM 4.2. *There is a continuous family of twisted helicoids $\mathcal{H}_k$ for $\frac{1}{2} < k_0 \leq k \leq \infty$.*

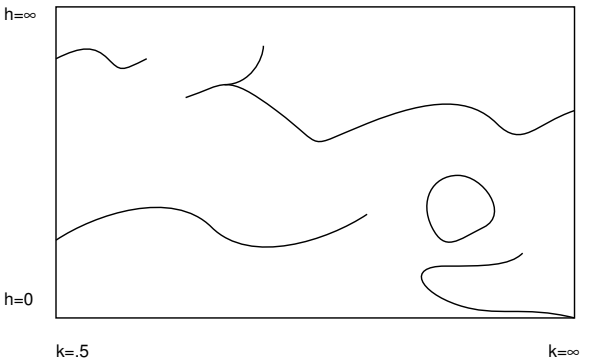


FIGURE 15. The solution locus?

PROOF. Fix $k_0 > \frac{1}{2}$, and denote the *solution locus* by

$$Z = \{(k,h) \in [k_0, \infty] \times [0, \infty] : H(k,h) = 0\}.$$

For every k, there is an $h(k)$ such that $(k, h(k)) \in Z$. Suppose there is no connected component of Z containing points with $k = k_0$ and $k = \infty$. Then Z could be *separated* by an arc connecting the horizontal edges $h = 0$ and $h = \infty$ which are *disjoint* from Z. However, H is continuous along that arc and takes positive and negative values at the endpoints by Propositions 3.5 and 3.6. Hence there is a connected component Z_0 of Z connecting a solution at k_0 with a solution at $k = \infty$. Since H is *real analytic*, Z is locally path connected and hence Z_0 is path connected. □

REMARK 4.3. We conjecture that this family is smooth and unique. Note however that numerical computations of Bobenko [**Bob**] indicate that there is a non-embedded genus-one helicoid.

4C. Embeddedness of the genus-one helicoid.

THEOREM 4.4. *The twisted helicoids $\mathcal{H}_k$ of the continuous family are embedded.*

PROOF. We know that any $\mathcal{H}_1$ is *embedded*. Scale and arrange the $\mathcal{H}_k$ in $\mathbb{R}^3$ so as to satisfy these properties:

(1) All $\mathcal{H}_k$ are *asymptotic* to the same helicoid as $\mathcal{H}_\infty$.
(2) The *straight line* containing V_1 and V_2 is the x-axis.
(3) The centers of the rhombi are mapped to the origin.

Since the helicoidal end is embedded [**HM99, HPR99**], the maximum principle implies that all $\mathcal{H}_k$ are embedded. □

REMARK 4.5. One can easily arrange the $\mathcal{H}_k$ so that they converge to the ordinary helicoid. We avoid this by choosing a common (geometric) reference point of *all* surfaces. Since we have continuity of the family including infinity, this gives convergence of the geometric surfaces in $\mathbb{R}^3$.

References

[Bob] A. I. Bobenko. Helicoids with handles and Baker–Akhiezer spinors. SFB preprint 264, Berlin.

[CHK93] M. Callahan, D. Hoffman, and H. Karcher. A family of singly-periodic minimal surfaces invariant under a screw motion. *Experimental Mathematics* 2:157–182, 1993.

[Col97] P. Collin. Topologie et courbure des surface minimales proprement plongées de $\mathbb{R}^3$. *Ann. Math.*, 145:1–31, 1997.

[Cos82] C. Costa. *Imersões mínimas em $\mathbb{R}^3$ de gênero um e curvatura total finita.* Ph.D. thesis, IMPA, Rio de Janeiro, Brasil, 1982.

[Cos84] C. Costa. Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends. *Bull. Soc. Bras. Mat.*, 15:47–54, 1984.

[Cos89] C. Costa. Uniqueness of minimal surfaces embedded in $\mathbb{R}^3$ with total curvature 12π. *Journal of Differential Geometry*, 30(3):597–618, 1989.

[HKW93a] D. Hoffman, H. Karcher, and F. Wei. Adding handles to the helicoid. *Bulletin of the AMS, New Series*, 29(1):77–84, 1993.

[HKW93b] D. Hoffman, H. Karcher, and F. Wei. The genus one helicoid and the minimal surfaces that led to its discovery. In *Global Analysis and Modern Mathematics*. Publish or Perish Press, 1993. K. Uhlenbeck, editor, p. 119–170.

[HKW99] D. Hoffman, H. Karcher, and F. Wei. The singly periodic genus one helicoid. *Commentarii Math. Helv.*, pages 248–279, 1999.

[HM99] D. Hoffman and J. McCuan. Embedded minimal annular ends asymptotic to the helicoid. Manuscript, 1999.

[HMI85] D. Hoffman and W. H. Meeks III. A complete embedded minimal surface in $\mathbb{R}^3$ with genus one and three ends. *Journal of Differential Geometry*, 21:109–127, 1985.

[HMI87] D. Hoffman and W. H. Meeks III. Properties of properly embedded minimal surfaces of finite total curvature. *Bulletin of the AMS*, 17(2):296–300, 1987.

[HMI89] D. Hoffman and W. H. Meeks III. The asymptotic behavior of properly embedded minimal surfaces of finite topology. *Journal of the AMS*, 2(4):667–681, 1989.

[HMI90] D. Hoffman and W. H. Meeks III. The strong halfspace theorem for minimal surfaces. *Inventiones Math.*, 101:373–377, 1990.

[HPR99] L. Hauswirth, J. Pérez, and P. Romon. Embedded minimals ends of finite type. Preprint, 1999.

[HW] D. Hoffman and F. Wei. Construction of a helicoid with a handle. Preprint.

[Kap97] N. Kapouleas. Complete embedded minimal surfaces of finite total curvature. *J. Diff. Geom.*, 47:95–169, 1997.

[Kar88] H. Karcher. Embedded minimal surfaces derived from Scherk's examples. *Manuscripta Math.*, 62:83–114, 1988.

[Kar89] H. Karcher. Construction of minimal surfaces. *Surveys in Geometry*, pages 1–96, 1989. University of Tokyo, 1989, and Lecture Notes No. 12, SFB256, Bonn, 1989.

[Lyn93] A. Lynker. Einfach periodische elliptische Minimalflächen. Diplomarbeit Bonn, 1993.

[Tra98] M. Traizet. Gluing minimal surfaces with implicit function theorem. Preprint 171 Tours, 1998.

[Tra01] M. Traizet. Construction of an asymmetric minimal surface. Preprint, 2001.

[TW01] M. Traizet and M. Weber. Embedded minimal surfaces with helicoidal ends. Preprint, 2001.

[Web00] M. Weber. The genus one helicoid is embedded. Habilitationsschrift Bonn, 2000.

[Web02] M. Weber. Period quotient maps of meromorphic 1-forms and minimal surfaces on tori. *Journal of Geometric Analysis*, 12, 2002.

[WHW01] M. Weber, D. Hoffman, and M. Wolf. An embedded helicoid with a handle. Preprint, 2001.

[Woh91] M. Wohlgemuth. Higher genus minimal surfaces by growing handles out of a catenoid. *Manuscripta Math.*, 70:397–428, 1991.

[Woh93] M. Wohlgemuth. *Vollständige Minimalflächen höheren Geschlechts und endlicher Totalkrümmung*. Ph.D. thesis, University of Bonn, April 1993.

[WW98] M. Weber and M. Wolf. Minimal surfaces of least total curvature and moduli spaces of plane polygonal arcs. *Geom. and Funct. Anal.*, 8:1129–1170, 1998.

[WW02] M. Weber and M. Wolf. Teichmüller theory and handle addition for minimal surfaces. *Annals of Math.*, 2002.

Department of Mathematics, Indiana University, Rawles Hall, Bloomington, IN 47405

E-mail address: matweber@indiana.edu

Clay Mathematics Proceedings
Volume 2, 2004

Computing Minimal Surfaces

David Hoffman

1. Introduction

This paper is a version of a lecture I gave at the CMI Summer School, with some parts deleted and others added. My reason for doing this is to keep the focus mainly on one theme, one that is at turns controversial and conventional, obvious and elusive: the use of computational simulation in minimal surface research. This is not meant to be an historical survey, rather a personal one about how computation has been used in the subject. By computation, I mean numerical calculations made using computers, and I should add that these calculations include in almost all cases computations for the production of images to aid in understanding what they mean. Indeed, anyone who has looked at the collected works of Schwarz [**38**], or the historical sections of [**32**] or [**11**], understands that it is not a twentieth-century innovation to calculate examples of minimal surfaces, and to sketch them based on these numerical estimates. Using computers has made it possible to make harder computations and to look at more complicated examples.

There are good reasons to look at more complicated examples, and these reasons are not produced by the existence of the computer. The "we compute them because we can" and the "looking for the keys under the lamp post" rationales for mathematical visualization leave me cold. The desire to explore new examples is theory-driven and is often done as part of an effort to answer qualitative questions. There has been no shortage of disagreement on this question in the past decade. But in my view, even what computations can be made at any given time is determined by the the way surfaces are conceived, and that always has a theoretical basis. Theoretical considerations and questions of general understanding are there from the beginning and cannot be easily separated in a meaningful way from the computing. It is this intermingling of abstraction and numerics — theory and computation if you must — that most interests me here.[1] At the end of this paper, I will

Supported by research grant DE-FG03-95ER25250/A007 of the Applied Mathematical Science subprogram of the Office of Energy Research, U.S. Department of Energy, and National Science Foundation, Division of Mathematical Sciences research grant DMS-0139410.

[1]I do not mean to downplay or disparage the enormous utility of computer-aided visualization of minimal surfaces for educational purposes. Also, there were classically known minimal surfaces whose appearance was not well known to experts in the subject in the early 1980's. For example Riemann's one-parameter family of minimal surfaces — except for the catenoid, the only minimal surfaces fibred by circles and straight lines — were virtually unknown to modern researchers. What

©2004 Clay Mathematics Institute

discuss an example where a theoretical advance was necessary even to "see" what was quite evident in the images made from calculations done originally for another purpose.

I intend to stay very close to my area of expertise, and this will naturally limit the scope of the paper. I will focus on noncompact, properly immersed minimal surfaces, and I will not discuss in any detail problems that involve specification or restriction of boundary values for compact surfaces. Nor will I be concerned with minimal surfaces with singularities. In fact, I will for the most part stick to periodic embedded minimal surfaces and their relationship to $\mathcal{H}e_1$, a properly immersed minimal surface of genus one with one end, which is asymptotic to the helicoid. Such a surface was discovered and investigated by Karcher, Wei and me. There was overwhelming evidence that this surface was embedded, but we could not prove it. Some eight years later, its embeddedness was proved by Weber, Wolf and myself.

Strictly speaking, the last sentence of the previous paragraph is not true. Weber, Wolf and I proved that there exists a properly embedded minimal surface with all the defining properties of the $\mathcal{H}e_1$. It is almost certainly the same surface, but as of this writing, that is a conjecture. This situation has some similarity to the discovery of the gyroid, and that is where I will begin in the next section.

Acknowledgments. I thank Jim Hoffman of MSRI, who made the computer-generated images in this paper, as well as the software to generate them. Thanks also to Hermann Karcher, Martin Traizet, Matthias Weber, Mike Wolf and Fusheng Wei whose work is discussed here.

2. The Gyroid and the Chen–Gackstatter Surface

Before narrowing the focus to very recent research, there are two earlier discoveries — investigations done without the use of computers — that I want to mention. They have not received as much attention as they should and therefore are not so generally recognized as being important to the intellectual landscape of the subject of computation and minimal surfaces.

2.1. The gyroid. In the late sixties, Alan Schoen, working for the NASA Electronics Research Laboratory in Cambridge, MA, was interested in triply periodic, space-dividing surfaces, defined either by restricting their symmetry group to one of the crystallographic groups or by specifying a "skeletal graph." By a skeletal graph of $\mathcal{S}$, I mean in this context two connected spatial graphs that lie on either side of a surface $\mathcal{S}$ and onto which the each domain of $\mathbb{R}^3 - \mathcal{S}$ can be retracted through surfaces preserving the symmetry group of $\mathcal{S}$. Schoen was able to construct a number of new examples of triply periodic embedded minimal surfaces, all of which were embedded and all of which contained either straight lines or possessed planes of symmetry [**36**]. These properties allowed him to use the Schwarz reflection principle to establish their existence as minimal surfaces and to make models of them. The period problem in some cases was solved experimentally.

---

they really looked like was not clear even to those researchers who knew about them. I do not know of any published images of this surface before the late 1980's. Also, the first indication for me of what Scherk's singly periodic surface (conjugate to Scherk's doubly periodic surface, which is discussed in Section 3) looks like came from a drawing by J. Pitts in his thesis, published in 1981 [**27**]. Scherk discovered this surface in the 1830's [**35**].

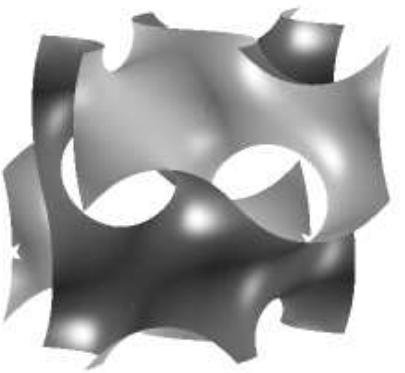


FIGURE 1. The gyroid

One example, the one he was to name the *gyroid*, ("gyroidal", Schoen noted, meant spiral or gyratory in crystallographers' lingo) had no symmetries that related to Schwarz reflection. Even though he was able to build a model of the surface (to compute it) he was not able to prove that it was in fact a minimal surface. One of the models he sent to Robert Osserman at Stanford. I saw it in Osserman's office when I was a graduate student in the late 1960's and was writing my thesis under his supervision. A picture of this model is on the frontispiece of Osserman's very influential monograph *A survey of minimal surfaces* [**34**], published in 1969. Schoen discussed his example with many mathematicians, but it was Blaine Lawson — then also a graduate student working under Osserman's direction — who was the most helpful in pointing Schoen in the right direction. Schoen realized that in fact the gyroid was a member of the associate family that contained the P-surface and the D-surface.

Discovered by Schwarz in the Nineteenth Century, these two surfaces are locally conjugate (their coordinate functions are conjugate harmonic functions). See Figure 2. A fundamental domain of the D-surface is bounded by a polygon consisting of four edges of a tetrahedron. It can be extended by Schwarz reflection to produce the entire surface. The conjugate minimal surface to this piece is bounded by planar curves whose plane the surface meets orthogonally. This piece extends by Schwarz reflection to produce the P-surface. The conjugacy relation identifies a basic piece of the D- with the P-surface.

For any angle ϕ,

$$\cos\phi D + \sin\phi P$$

is another minimal surface. Addition in this formula is addition in $\mathbb{R}^3$ of points that correspond by the conjugacy relation. The gyroid, the G-surface, turns out to be given by

$$G = \cos\phi_0 D + \sin\phi_0 P,$$

with $\phi_0 \sim 38.015°$. This value of ϕ, modulo π, is the unique value for which the extended surface closes up to give a triply periodic surface.

The G-surface was the first example of a triply periodic and embedded minimal surface to be found that did not contain lines or possess planes of reflective symmetry.

The correspondence between the models that Schoen made and the G-surface as described mathematically in the formula above was generally accepted, although without proof. There was no proof of embeddedness except by reference to this correspondence. There was not a proof in any traditional mathematical sense. It was only in 1996 that Meinhard Wolgemuth and Karsten Große-Brauckmann published a paper that proved that the gyroid is embedded [**29**].

The desire to prove that a minimal surface is embedded was how I got involved with computing minimal surfaces. In the case of Costa's surface, the computation was made to see whether that surface was even a candidate [**9, 13, 14, 15, 19, 20**]. In the case of the gyroid, the computation was made (although not by computer) before it was known that it could be minimal. That the associate surface, G, was accurately modeled by the plastic gyroid of Schoen's construction — and was therefore embedded — was "evident". I must admit that I did not realize that there was a mathematical gap between G and the gyroid until it was pointed out to me by Hermann Karcher.

2.1.1. *Intermaterial dividing surfaces.* The gyroid has had a second career as the poster surface for material scientists. A short digression is needed to explain why.

A variety of immiscible liquids, in the presence of a soap or some other surfactant, can self-assemble into a rich variety of regular mesophases. Characterized by their "inter-material dividing surfaces" — where the different substances touch — these structures also occur in microphase-separated block copolymers. The understanding of the interface is key to prediction of material properties, but the full relationship between the curvature of the dividing surfaces and the relevant molecular and macromolecular physics is not well understood. Moreover, there is only a partial theoretical understanding of the range of possible periodic surfaces that might occur as interfaces. Considering the surface to be an interface between two regions of fixed — but possibly unequal — volume, it is natural to impose constant mean curvature, $H = c$, as a property of the dividing surface. In the case of mean curvature $H = 0$, various classical minimal surfaces have been conjectured as being appropriate models for observed interfaces (the observation being done principally by transmission-electron-microscope methods), first among them the D-surface of Schwarz. In di-block copolymers, interfaces with the gyroidal symmetries were observed. However, these surfaces were not minimal. It has been conjectured, and there is some theoretical evidence and a good deal of computational evidence, that the gyroid sits in a one-parameter family of embedded triply periodic surfaces with nonzero constant mean curvature, which begins at the gyroid and degenerates to its skeletal graph [**1, 39, 12, 24, 23, 28, 2**].

2.2. The Chen–Gackstatter surface. The second discovery whose importance I wish to emphasize was made by C. C. Chen and Fritz Gackstatter. Around 1980, they constructed a complete, immersed minimal surface of genus one and finite total curvature [**6, 7**]. The surface that now bears their names is, in my opinion, one of the most beautiful minimal surfaces. A sculpture of the surface by Ralph Carlson is used by MSRI as a gift of appreciation to generous supporters (Figure 3). It has also been presented to others who have participated in some of the more successful of MSRI's outreach events: Tom Stoppard and Steve Martin are both in possession of bronze Chen–Gackstatters. Its discovery — really its construction — was a landmark in the subject, although it was not recognized at the time

and is even today not well appreciated. That minimal surfaces could be represented by an integral formula involving analytic data was a nineteenth-century discovery, which we associate with the names of Weierstrass, Enneper, Riemann and Schwarz.

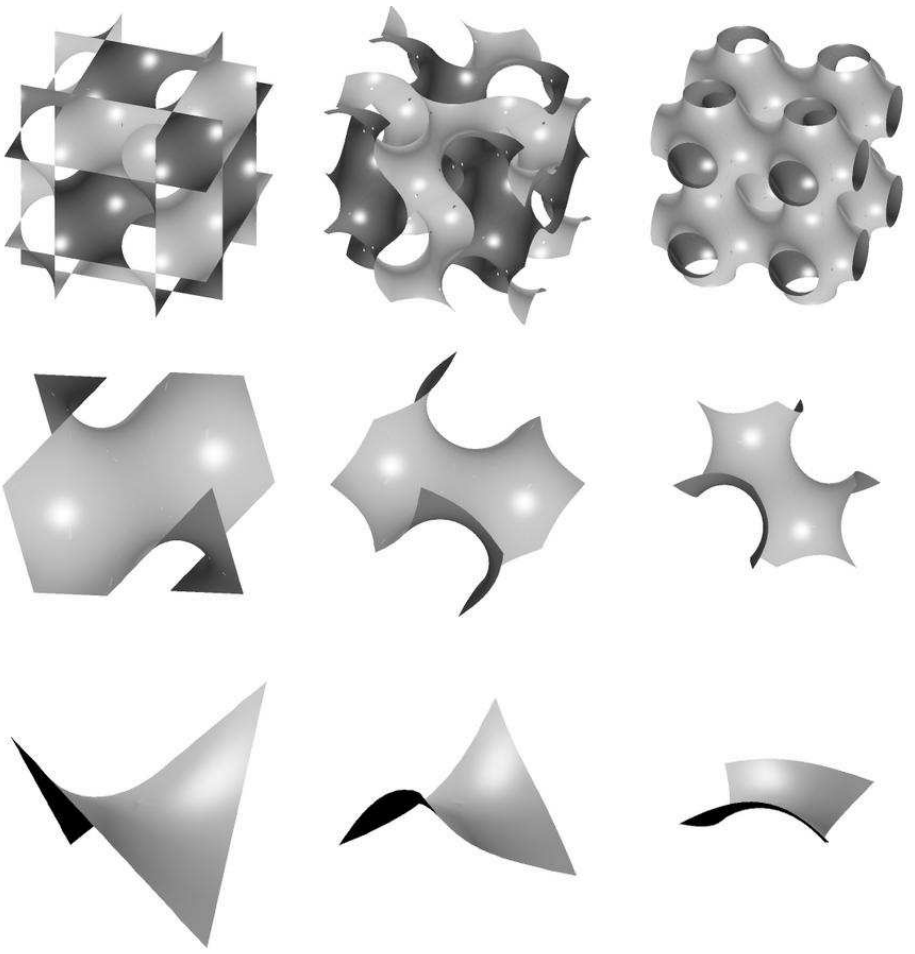


FIGURE 2. The Schwarz D- and P-surfaces, and the Gyroid of A. Schoen. The left and right columns illustrate the D- and P-surfaces, respectively. The center column illustrates the gyroid. The surface at bottom left is Schwarz's solution to the Plateau Problem for a quadrilateral made up of edges of a tetrahedron. It extends to the D-surface. At bottom right is its conjugate surface, a basic part of the P-surface, and in the bottom middle is the associate surface that is a piece of the gyroid. The middle row consists of six of the corresponding pieces in the bottom row; the extension is by Schwarz reflection. The top row consists of regions of these three surfaces inside of cubes. Each extends by translation to R^3. (In this illustration the number of fundamental units in each image is not the same.)

It was this representation that Schwarz used to write down the first solution to a nontrivial Plateau Problem for a nontrivial boundary. Schwarz found an integral parameterization for the minimal surface bounded by a quadrilateral made up of four edges of a tetrahedron. See Figure 2. This surface is a basic building block for the D-surface mentioned above. However, the idea that the entire D-surface could be described in terms of algebraic data on a compact Riemann surface of genus three (providing a global representation for D, modulo translations) was not made clear until the work of Osserman [**34, 33**]. Osserman proved that a complete and regular minimal surface in $\mathbb{R}^3$ whose total curvature was finite has the following properties:

i. The underlying Riemann surface determined by the metric induced from $\mathbb{R}^3$ is a compact Riemann surface $\mathcal{R}$, punctured in a finite number of points $\{p_1, \ldots, p_n\}$, one for each topological end;
ii. The stereographic projection (from $(0,0,1)$), g, of the Gauss map, well defined on the minimal surface, extends to a meromorphic function on $\mathcal{R}$;
iii. The holomorphic one-form, η, defined as the exterior derivative of $h := x_3 + ix_3^*$, where '*' denotes harmonic conjugation, extends to a meromorphic one-form on $\mathcal{R}$. The zeros of η on the Riemann surface occur precisely at the zeros and poles of g on the surface and have the same order. At the punctures, $\eta = dh$ must have a pole wherever the extension of g is not equal ∞.

Moreover, any "data" $\{\mathcal{R}, \{p_1, \ldots, p_n\}, g, \eta\}$ satisfying these conditions produce a conformal minimal immersion of (a covering of) $\mathcal{R} \setminus \{p_1, \ldots, p_n\}$ given by the "Weierstrass representation formula:"

$$X(p) = \mathrm{Re} \int_{\mathbf{p}}^{p} (\frac{1}{2}(\frac{1}{g} - g), \frac{i}{2}(\frac{1}{g} + g), 1)\eta. \tag{2–1}$$

The first uses of this characterization were to prove general results about complete minimal surfaces of finite total curvature: their total curvature was a multiple of -4π; their Gauss map could not omit more than three points; if the total curvature of such a surface was equal to -4π, then the surface was either the catenoid or Enneper's surface. See [**34**].

Remember that at this time there were very few well-known complete minimal surfaces of finite total curvature: the plane, the catenoid and Enneper's surface. (It is quite possible that the "generalized Enneper surfaces", $g = z^k$, $\eta = z^k\, dz$ on $\mathbb{C}$ were known — when $k = 1$ this is Enneper's — but I do not know a reference.) In particular, there were no known examples with genus greater than zero. Klotz and Sario [**30**] constructed examples of genus zero with arbitrary numbers of ends, and examples of genus $p > 0$ with $n \geq 4$ ends. However, many of the examples were essentially multiple coverings and some of them had branch points.

The first successful attempt (and for all I know the first attempt) to use the Weierstrass representations to construct a complete minimal surface of finite type with specific geometric and topological properties (genus, behavior at infinity) was made around 1980 by Chen and Gackstatter. They wanted to construct a minimal surface of genus one with (in the light of the aforementioned result of Klotz and Sario) fewer than four ends. They chose as the underlying Riemann surface the simplest candidate: the *Gaussian* or *square* torus. This is the Riemann surface of

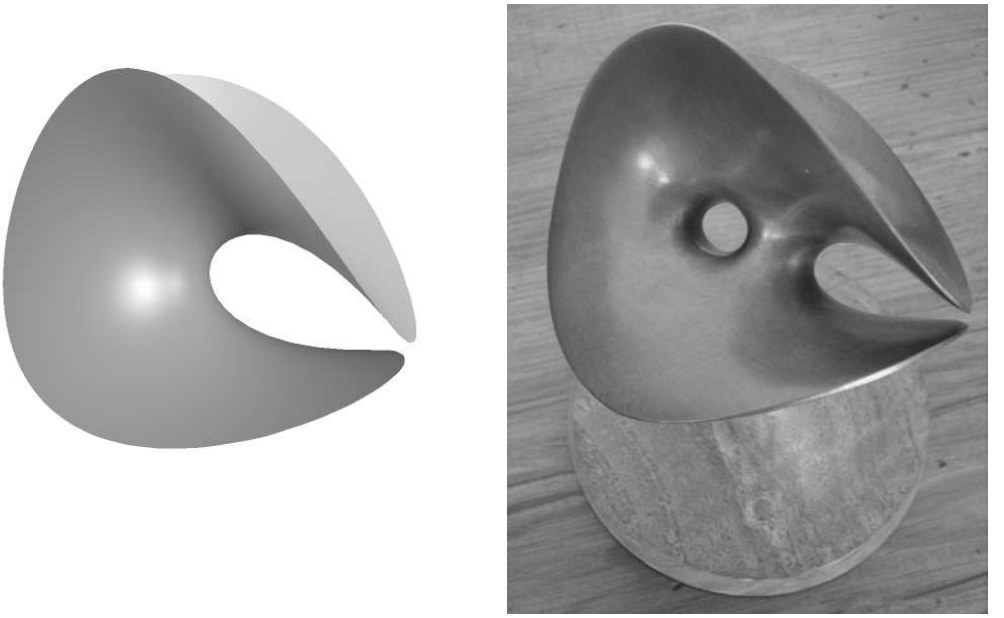


FIGURE 3. Left: Enneper's surface. Right: A bronze sculpture of the Chen–Gackstatter surface made by Ralph Carlson.

$\mathbb{C}$ modulo the integer lattice. It is also the Riemann surface of the equation

$$w^2 = z(z^2 - 1). \tag{2–2}$$

It seems clear that they were looking for the simplest new example, and therefore specified that it have just one end. Having done that, they had a very limited choice of models among examples of finite total curvature. An end like the catenoid or the plane would force the surface to sit in a half-space, a contradiction of the maximum principle. This left them with the end of Enneper's surface as an model. The Gauss map, g, would then have to be single-valued at the end, and η would have to have a triple pole. This is achieved by specifying the Weierstrass data $g = \frac{\lambda w}{z}$ and $\eta = c\,dz$, where w and z are the elliptic functions on the square torus given in (2–2), and c and λ are real constants to be determined. Chen and Gackstatter showed that the period problem was solvable and therefore the proper choice of constants provided a single-valued Weierstrass representation. (For more details of the Weierstrass-data construction of the Chen–Gackstatter surface see the original paper [**6**] or the first sections of [**17**].) They understood that what they had constructed was the simplest possible genus-one example of a complete minimal surface of finite total curvature. They even knew qualitatively what it looked like.

Chen and Gackstatter posed as a "great question" the existence of a genus-one example with two ends and drew by hand a picture of how such a surface might look: a catenoid with a handle, possessing the maximum amount of symmetry possible. Several years later, Rick Schoen proved that such a surface did not exist, even without the symmetry assumptions [**37**].

In my opinion, the construction of the Chen–Gackstatter surface, a surface that is not embedded, was a key early step in the modern study of properly embedded

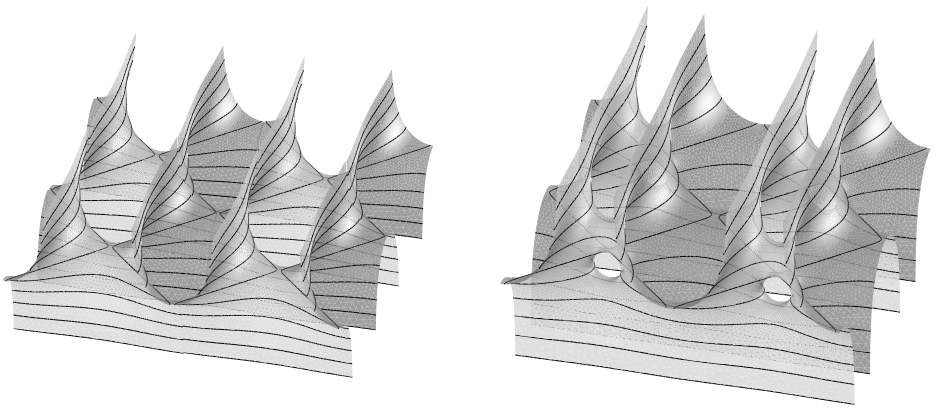


FIGURE 4. Left: Scherk's doubly periodic surface. Right: Scherk's doubly periodic surface with a handle.

minimal surfaces of finite topology. It was the first attempt to build a minimal surface to desired geometric specifications, and it was successful. Together with the work of Osserman, the aforementioned paper of Schoen, the important results about embedded ends by L. Jorge and W. Meeks [**26**] (both published in 1983) and the thesis of Costa [**9**] (published in 1982), it was part of the foundation of the modern study of properly embedded minimal surfaces of finite topology.

3. Symmetry Breaking and Families of Minimal Surfaces

The catenoid and the helicoid were the first minimal surfaces to be discovered because they are the only nonplanar minimal surfaces invariant under a one-parameter group of symmetries. From an analytic point of view, the second-order differential equation of minimality becomes an ordinary differential equation — for a rotational or a ruled minimal surface, respectively. The discovery of Costa's surface over 200 years later, and the proof that it is embedded, depends on the high degree of symmetry — albeit finite — of the surface. The construction of the Costa–Hoffman–Meeks surfaces — generalizations of the surface of Costa to all genera greater than one — also relied on symmetry constructions [**19, 20**]. The Weierstrass representation (under simplifying symmetry assumptions) was used to construct many new minimal surfaces of finite topology and new singly and triply periodic minimal surfaces. However, the doubly periodic minimal surfaces seemed to resist modification. It was Fusheng Wei who made a breakthrough by realizing that one could succeed in modifying some doubly periodic minimal tori by breaking symmetry in the following sense: add only half the number of handles necessary to produce the most symmetric generalization [**47**]. Wei produced, among other things, doubly periodic, embedded, minimal surfaces with genus equal to two in the quotient.

Hermann Karcher understood immediately the importance of this construction. Very soon after he heard about Wei's work, Karcher produced what is now known as "Scherk's doubly periodic surface with a handle".

The classical Scherk doubly periodic minimal surface can be described qualitatively as follows. Consider over a square in the plane a saddle-like surface of infinite height, bounded by four vertical lines passing through the corners. Extend the surface by Schwarz reflection around these vertical edges, producing a doubly periodic surface that is — except for the vertical lines — a graph over a region of the plane. The domain of the graph can be considered to correspond to the interior of the black squares on a checkerboard. Karcher modified Scherk's doubly periodic surface by adding handles not to every saddle but only to every other saddle. It should be made clear that by adding fewer handles and not insisting on the maximally symmetric situation, Karcher was actually *simplifying* the situation; the genus of the quotient surface was equal to one, not two as it would be if handles were inserted on each saddle. The remaining symmetry forced the quotient surface to be the square torus.

3.1. Families. It was well known that Scherk's doubly periodic surface is a member of a smooth one-parameter family of embedded minimal surfaces. This family is easy to describe. Deform the unit square through rhombi in such a manner that the length of one of the diagonals remains equal to $\sqrt{2}$, and the length, say l, of the other diagonal varies from $\sqrt{2}$ to ∞. Over each rhombus, follow the construction (over the square) of the saddle surface bounded by four verticals at the vertices. This produces a minimal graph over the interior of the rhombus, which is bounded by the four verticals, and which can be extended by Schwarz reflection in the verticals to a doubly periodic surface (see Figure 5). Producing this deformation on a computer using the Weierstrass representation is not hard. Fix θ, $0 \leq \theta < \pi/4$, and take as the domain the closed unit disk minus four boundary points: $D - \{\pm e^{i\theta}, \pm ie^{-i\theta}\}$. The Weierstrass data

$$g = z, \qquad \eta = \frac{2\,dz}{z(z^2 - e^{2i\theta}(z^2 + e^{-2i\theta})},$$

produces in the interior of D a saddle-like minimal graph over a rhombus whose diagonals (after a rescaling if necessary) have lengths equal to $\sqrt{2}$ and $l(\theta)$, with $l \to \infty$ as $\theta \to \pi/4$. The four arcs of $\partial D - \{\pm e^{i\theta}, \pm ie^{-i\theta}\}$ are mapped into the four vertical lines over the vertices of the rhombus, and they bound the minimal saddle. When $\theta = 0$, one gets the standard Scherk's surface whose fundamental domain is a graph over the interior of a square.

What was not well known at the time to researchers in the field is that the limit as $\theta \to \frac{\pi}{4}$ of this deformation family of singly periodic surfaces is the helicoid, a singly periodic minimal surface. This was "discovered" by Meinhard Wohlgemuth and me in 1987 while looking at computer generated images of the Scherk family near the limiting value of $\theta = \frac{\pi}{4}$. The fixed-length diagonal of the rhombus becomes a part of the axis of the helicoid; the other diagonal diverges to one of the lines that rule the helicoid. Once we "saw" this, it was not hard to verify it.

In principle, this relationship in the abstract was known to Scherk; it was at the very least implicit in his work. However, it is likely that Scherk had no clear idea what these surfaces looked like as doubly periodic surfaces. The important discovery that lay nascent in this observation was that families of doubly periodic, embedded minimal surfaces could limit to embedded, singly periodic minimal surfaces. This came out in subsequent research in two ways.

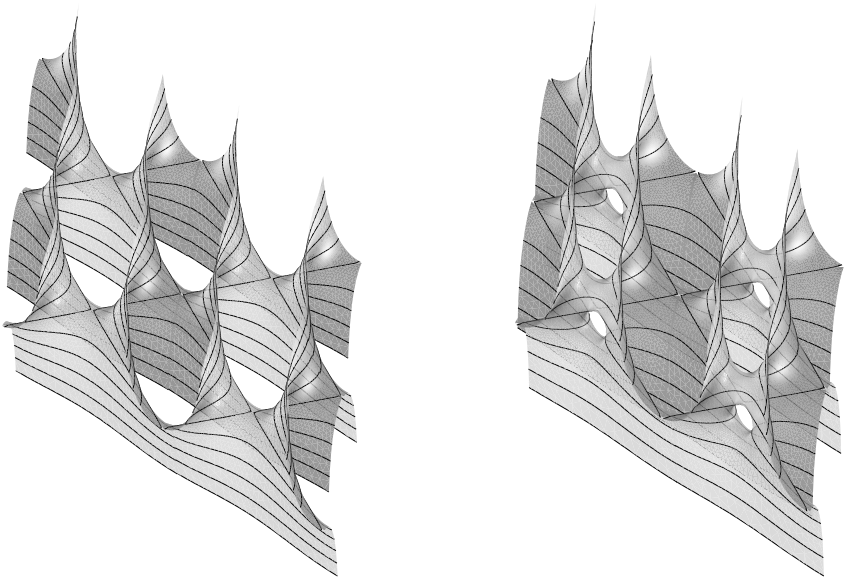


FIGURE 5. Deformations of Scherk's doubly periodic surface (left) and Scherk's doubly periodic surface with a handle (right).

The first one is relatively mundane: look at known families of periodic minimal surfaces and show that the helicoid is a limit in a natural fashion. For example, Wayne Rossman and I did this for the classical Riemann minimal surfaces [**21**]. Modulo translation, these singly periodic surfaces are conformally rectangular tori:

$$w^2 = z(z-1)(z-\lambda), \tag{3–1}$$

for $\lambda > 0$ real. Degeneration occurs as $\lambda \to 0$ or $\lambda \to \infty$. In one case, the limit (suitably normalized) is a helicoid, in the other it is a catenoid. Since the tori associated with λ and $\frac{1}{\lambda}$ are conformally the same, it is not surprising that the corresponding Riemann examples are in fact conjugate surfaces, as are the helicoid and the catenoid.

The second way, and the subject of the next section, is the conception of new, and previously unknown, examples by the construction of deformation families of periodic minimal surfaces and then the thought experiment of imagining the limit surface.

3.2. The limit surface of a deformation family. We asked ourselves: Can the handle added to Scherk's surface survive a deformation similar to that of Scherk's doubly periodic surface? The interest in doing so was two-fold. First, it would be an interesting test to find out whether it was reasonable to expect, in general, that an embedded minimal surface could be deformed through a family of like surfaces. Second, if the surface could be deformed, would it have a limit surface?

With knowledge of the Scherk doubly periodic family, it was natural to try to generalize the Weierstrass construction of the Scherk family to rhombic tori. This

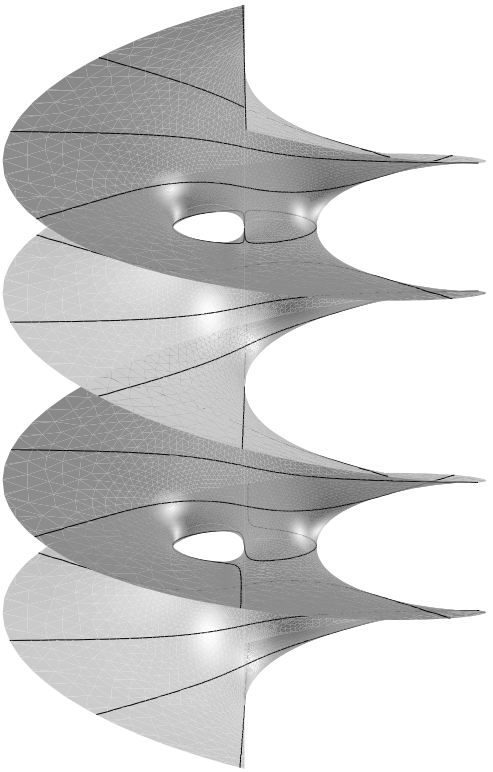


FIGURE 6. $\mathcal{H}_1$, the genus-one helicoid with translational symmetry.

is what Wei and I set out to do. We hoped that for each such torus — or at least for those close to the square torus — the complex parameter for the position of the puncture could be chosen to solve a two-dimensional period problem. Very few such period problems had been solved up to that time. If the new deformation family did not degenerate too much, one could hope for a limit surface that would not be among those already known.

Wei succeeded in working out the family (see Figure 5). Computer-generated images showed, as we hoped, that the family appeared to be converging to a surface asymptotic, a helicoid but with one with one additional handle in each fundamental domain (in the surface modulo translation). See Figure 6. More details about the calculation and the Weierstrass representations can be found in [**10**] and [**17**].

It is important to restate carefully what actually was done. Nothing at all was proved in the traditional sense. In fact, the existence of this family of doubly periodic surfaces has never been formally established. The period problem — finding a zero of a mapping from a domain in the plane to $\mathbb{R}^2$ — was solved numerically. Moreover, this was done for a large but finite number of rhombic tori, starting with the square and moving toward degeneracy. Not only was there no traditional proof of the existence of each surface, there was nothing remotely close to a proof

of continuity of the solutions as a function of the input data (the conformal type and the location of the punctures corresponding to the ends). Nonetheless, the computation together with the images and the animations produced from them contained a huge amount of mathematical information. First, the conjecture that the deformation family existed was, with a high degree of certainty, correct. Second, the surfaces, if they existed, had the expected qualitative behavior. Third, and most important, the suspicion that there should be a singly periodic limit surface, with one helicoidal end and genus equal to one in the quotient, went from a reasonable guess to a conviction. Karcher, Wei and I set out to prove directly that there was such a surface.

What we had was the firm belief that we had a qualitatively correct picture of the surface and its symmetries. From such a starting point we already had a lot of experience deducing Weierstrass data. This was (not unexpectedly) a two-parameter problem, again, one real parameter for the underlying rhombic torus and another real parameter for the symmetric placement of the helicoidal ends in the quotient. We solved the period problem numerically and produced clear and convincing images of the surface. It then took some months to give qualitative estimates that showed in a manner entirely independent of computation that the period problem had a solution. At that point, we knew that that there must exist a minimal surface, $\mathcal{H}_1$, with the following properties:

The quotient of $\mathcal{H}_1$ by vertical translation, T, by 2π is a minimal surface in $\mathbb{R}^3/T$ with the following properties:

i. $\mathcal{H}_1/T$ has genus one and two ends;
ii. $\mathcal{H}_1/T$ contains a vertical axis and two horizontal lines;
iii. $\mathcal{H}_1/T$ is asymptotic to portion of the helicoid that has twisted through an angle of 2π.

However, embeddedness was a problem. Even though we knew that it was embedded outside of some cylinder around its axis, and we had convincing pictures of the part of the surface in the cylinder showing that it was embedded there, too, we did not have a proof of embeddedness. This took another, longer, period to establish. In discovering the proof of embeddedness, some subtle qualities of the surface were used (in particular the behavior of the set of points where the stereographic projection of the Gauss map was unitary). These properties were first observed in the images of the surface, then proved to be the case. More details can be found in papers [**16, 10**].

I am not sure that we would have been able to provide a demonstration of embeddedness along these lines if we did not have the images *together with the knowledge of how the images were computed.* I am reasonably confident we would have succeeded eventually, but we would have found another proof, very likely more complicated. Also, we were not able to prove uniqueness of this surface (i.e., that it was the only minimal surface satisfying the properties enumerated above). We suspected this to be the case, and it is true, but the proof of this had to wait approximately seven years, and it is a byproduct of the effort to prove that the genus-one helicoid, $\mathcal{H}e_1$ was embedded.

4. The Genus-One Helicoid

Before describing the genus-one helicoid, let me set the stage by discussing the helicoid. The first nonplanar minimal surface to be discovered was the catenoid,

and this is usually attributed to Euler, and the discovery dated 1744. At that time, however, the currently accepted definition of minimality was not yet formulated. That was to be done by Lagrange, who made the definition for graphs in 1755. Euler had shown that the catenoid was a critical point for area among surfaces of rotation with the same endpoints. In 1776, Meusnier showed that the helicoid was also a minimal surface.

The helicoid is a simply connected, embedded and ruled minimal surface (the only nonplanar, ruled minimal surface). See Figure 7. It is now known that not only is the helicoid not deformable through minimal surfaces, but also that it is the unique simply connected and embedded minimal surface [**31, 25**]. This is a strong affirmative answer to the Generalized Bernstein Question asked by Osserman some 45 years ago. It is a perfect complement to the result of Collin [**8**] showing that the catenoid, the conjugate minimal surface to the helicoid, is the unique properly embedded minimal annulus.

After the discovery of an infinite number of properly embedded minimal surfaces of genus k, $k \geq 1$, in the 1980's, it gradually occurred to various researchers that the same question could be asked about the helicoid. There are two ways to pose the question:

1. What are the properly embedded, nonplanar minimal surfaces of finite topology with one topological end?
2. What are the properly embedded minimal surfaces with infinite total curvature?

Because of the work of Meeks and Rosenberg and Collin cited above, we know that these questions are equivalent.

There was only one known example: the helicoid.

In a conversation that is recounted in [**17**], Harold Rosenberg suggested to Hermann Karcher that Hermann and I add a handle to the helicoid. That is, he suggested that we construct a properly embedded minimal surface of genus one with one helicoidal end. However, there is a problem with doing such a thing concretely. The Weierstrass representation for a helicoidal end is easy to write down in the genus-zero case, but not so easy to specify for a surface of higher genus. To explain why this is so, we need to say more about the helicoid.

The helicoid is singly periodic. In fact, it is invariant under *vertical screw motions*. A vertical screw motion σ_k of $\mathbb{R}^3$, $k \in \mathbb{R}$, is defined to be rotation by an angle of $2\pi k$ around the vertical axis, followed by a vertical translation by $2\pi k$. The quotient of $\mathcal{H}$ by σ_k is a minimal surface in the flat three-manifold $\mathbb{R}^3/\sigma_k$ with the following properties:

i. $\mathcal{H}/\sigma_k$ has genus zero and two ends;
ii. $\mathcal{H}/\sigma_k$ contains a vertical axis and is fibred by horizontal lines.

It is easy to write down the Weierstrass data for $\mathcal{H}/\sigma_k$:

$$g = iz^k, \qquad \eta = \frac{ki\,dz}{z}. \tag{4–1}$$

When $k = 1$, we have the familiar textbook representation. (The factor of i in the definition of g is there for reasons of normalization; it corresponds to a rotation about the vertical axis in $\mathbb{R}^3$.) There is a period in the third coordinate of the representation (2–1) using the data (4–1), giving rise to the helicoid as a singly periodic surface.

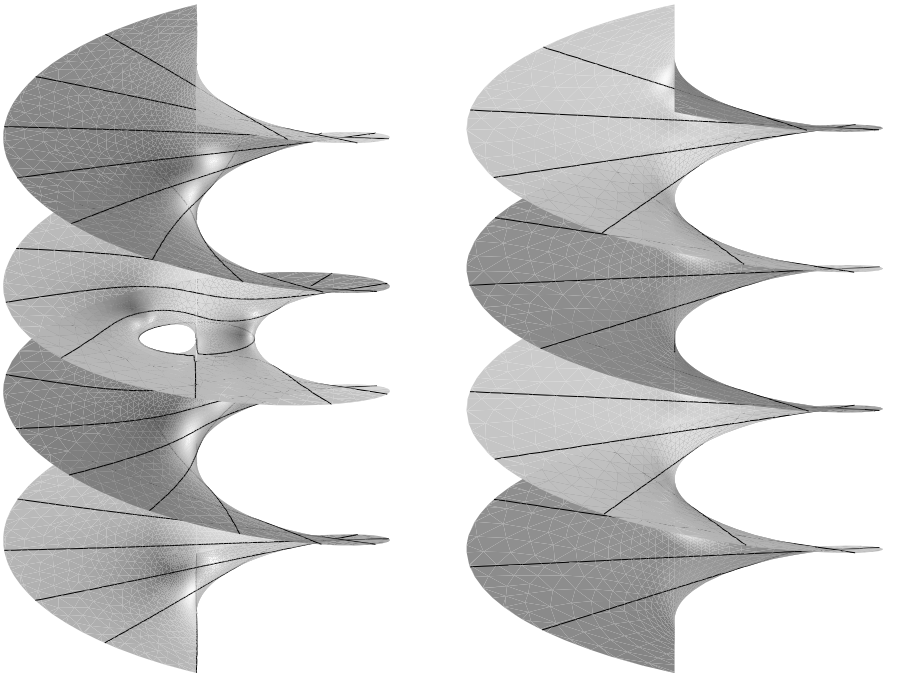


FIGURE 7. The helicoid $\mathcal{H}$ (right) and a genus-one helicoid $\mathcal{H}e_1$ (left).

Note that there are *two* ends in the quotient. If we want look at the entire helicoid — a genus-zero surface with one end — we must take periodicity into account. A change of variables gives

$$g = ie^w, \qquad \eta = i\,dw, \tag{4–2}$$

with a single end at infinity, a point where g has an essential singularity. This is important: to write down Weierstrass data for a possible higher-genus helicoid, it is necessary to specify a function on a higher-genus Riemann surface with a single essential singularity that has the same behavior as e^w at $w = \infty$. If one then wants to use numerical techniques to approximate the solution to a period problem, then the specification of this function has to be efficiently computable in the Weierstrass representation. The way to do this was not obvious.

Karcher had the key insight that enabled us to get around this problem. (See [**17**] for more details.) Not surprisingly, the idea for solution came from the study of singly periodic minimal surfaces invariant under screw motions σ_k. If k is not an integer, the Gauss map of such a surface does not descend to the quotient surface. For example, in the Weierstrass data for the helicoid given in (4–1), the Gauss map g is multivalued when k is not integral. However, the logarithmic derivative, dg/g, is a one-form that is well defined in the quotient. The function g can be reconstructed by integration and exponentiation. If one wants to specify a Gauss map on a Riemann surface with a singularity like that of $g = e^z$ at ∞ on the complex plane, one could specify $dg/g = dz$, then integrate and exponentiate. The

point is that this can be done analogously on a Riemann surface by specifying $\frac{dg}{g}$ to have a divisor with a *double pole* at the puncture point corresponding to the end, simple poles at the vertical points (where one wants $g = 0$ or $g = \infty$) of the desired surface — with appropriate residues — and zeros at the desired branch points. One is then specifying *meromorphic data* for the Gauss map, and this is something with which we had a good deal of experience by the early 1990's.

Of course, integrating the one-form dg/g on a Riemann surface produces a period problem in addition to the already existing period problems for the Weierstrass integral, and this is actually a compound problem because one then needs g to compute the Weierstrass integral. In a situation like this, our experience indicated that we should use symmetry as much as possible in order to simplify the period problem. With a helicoidal end, it is not possible to have any symmetry planes, and any symmetry lines of the surface must also be symmetry lines of the end of a helicoid. This restricts us to an axis and to horizontal lines that meet the axis. But if we have more than one such horizontal line, then the surface would be invariant under a translation (the translation that results from successive 180°-degree rotation around non-intersecting lines). Such a surface either has genus zero or infinite genus, and we are looking for an example with finite topology and positive genus. Therefore, the maximally symmetric situation will happen when there is a vertical axis and *one* horizontal line on the surface, a line that meets the axis. With the additional assumption that the genus equals one, we knew for symmetry reasons that the underlying Riemann surface had to be a rhombic torus punctured in one point. If we took as a model for such a surface a rhombus in the plane with the standard edge identifications, we could assume without loss of generality that the center of the rhombus corresponded to the point where the lines cross on the surface, the diagonals of the rhombus correspond the two lines on the surface, and that the vertex of the rhombus was the puncture corresponding to the helicoidal end.

So we were looking for an example, which we denote by $\mathcal{H}e_1$, with the following defining properties:

(i) $\mathcal{H}e_1$ is a properly immersed minimal surface;
(ii) $\mathcal{H}e_1$ has genus one and one end asymptotic to the helicoid;
(iii) $\mathcal{H}e_1$ contains a single vertical line (*the axis*) and a single horizontal line.

Karcher, Wei and I [**17**] found an $\mathcal{H}e_1$ by specifying the Weierstrass representation according to this prescription and then solving the period problem(s) for this data as a function of parameters that specified the conformal type and the locations of the vertical points and the branch points. The images of the surface we computed left little doubt that we had found an $\mathcal{H}e_1$,..., and that it was embedded.

However, we could not prove embeddedness. Of course, we had an easy path to establish that the surface was embedded outside of a compact set (it was designed to be asymptotic to a helicoid after all). John McCuan and I proved that any end with the symmetry and asymptotics of the end of this $\mathcal{H}e_1$ had to be embedded outside of a compact set [**18**]. Moreover, we had clear pictures of the compact part of the surface, and they indicated clearly that the surface was embedded! But we did not have a proof of embeddedness. We also did not have a proof of uniqueness, although this did not bother us too much at the time. It was the embeddedness that was the real problem.

4.1. Deformations of $\mathcal{H}_1$: the $\mathcal{H}_k$ family. At the same time we were writing up the results on the existence of an $\mathcal{H}e_1$, we were also discovering that the singly periodic genus-one helicoid, $\mathcal{H}_1$, could be perturbed through a family of embedded minimal surfaces. There were sufficient precedents to make us expect that this was possible. In [**3**], Michael Callahan, Karcher and I found, numerically, a deformation of an important class of surface — the embedded, singly periodic minimal surfaces with an infinite number of flat ends, the ones described in [**4**] and [**5**] — through surfaces invariant under a screw motion. This interesting deformation stops after a finite time; a limit Weierstrass representation consistent with the expected symmetries of the surface does not exist. Some years earlier, Karcher has considered deformations of generalized singly periodic Scherk surfaces through screw-motion-invariant families.

Wei succeeded in getting a representation of the desired surfaces that was not only qualitatively correct but (and this was a clever innovation by Fusheng) could be stably computed after making some simple (but not obvious) transformations of the variables.

The screw-motion-invariant minimal surfaces, $\mathcal{H}_k$, which we wanted to find, are described by the following conditions. The quotient of $\mathcal{H}_k$ by σ_k:

(i) has genus one and two ends;
(ii) contains both a vertical axis, and two horizontal lines making an angle of πk with one another;
(iii) is asymptotic to a portion of the helicoid that has twisted through an angle of $2\pi k$.

We computed a large number of them and made an animation of the family.

4.1.1. *Animations of the $\mathcal{H}_k$ family.* The animations showed clearly that these surfaces formed a continuous family[2]. We therefore had strong evidence for the existence of screw-motion-invariant, embedded minimal surfaces all with genus one in the quotient. Equally important, the animation of this family of surfaces was decisive in motivating the next important step in the subject, although it took many years for this to come about.

In the animation we produced, we had to decide on a normalization of the surfaces. On the helicoid, all the points on the vertical axis have the same Gauss curvature because the axis is an orbit of the group of screw-motion isometries of the helicoid. This is certainly not the case for the surfaces $\mathcal{H}_k$, so the best we could do was to choose a distinguished point on the axis of each $\mathcal{H}_k$ and normalize the curvature there to be equal to -1. We specified this point to be one of the two points on the axis (modulo σ_k) where the horizontal lines on $\mathcal{H}_k$ cross. (To be precise, we chose the point on the distinguished horizontal line that contains branch points of the Gauss map.)

The animation provided convincing evidence that the family was continuous in k, and that it rapidly stabilized. That is, if one looked at a slab of height $2n\pi$, then once $k \geq n$, there was little perceptible change in the image. Moreover, the surface to which it appeared to converge was identical in appearance to the images of $\mathcal{H}e_1$. Those images were independently computed from the Weierstrass data for that surface.

[2]These animations can be found at www.msri.org/publications/sgp/jim/geom/minimal/ and in library/helicoidg1/indexc.html and supplement/helicg1plimit/indexc.html

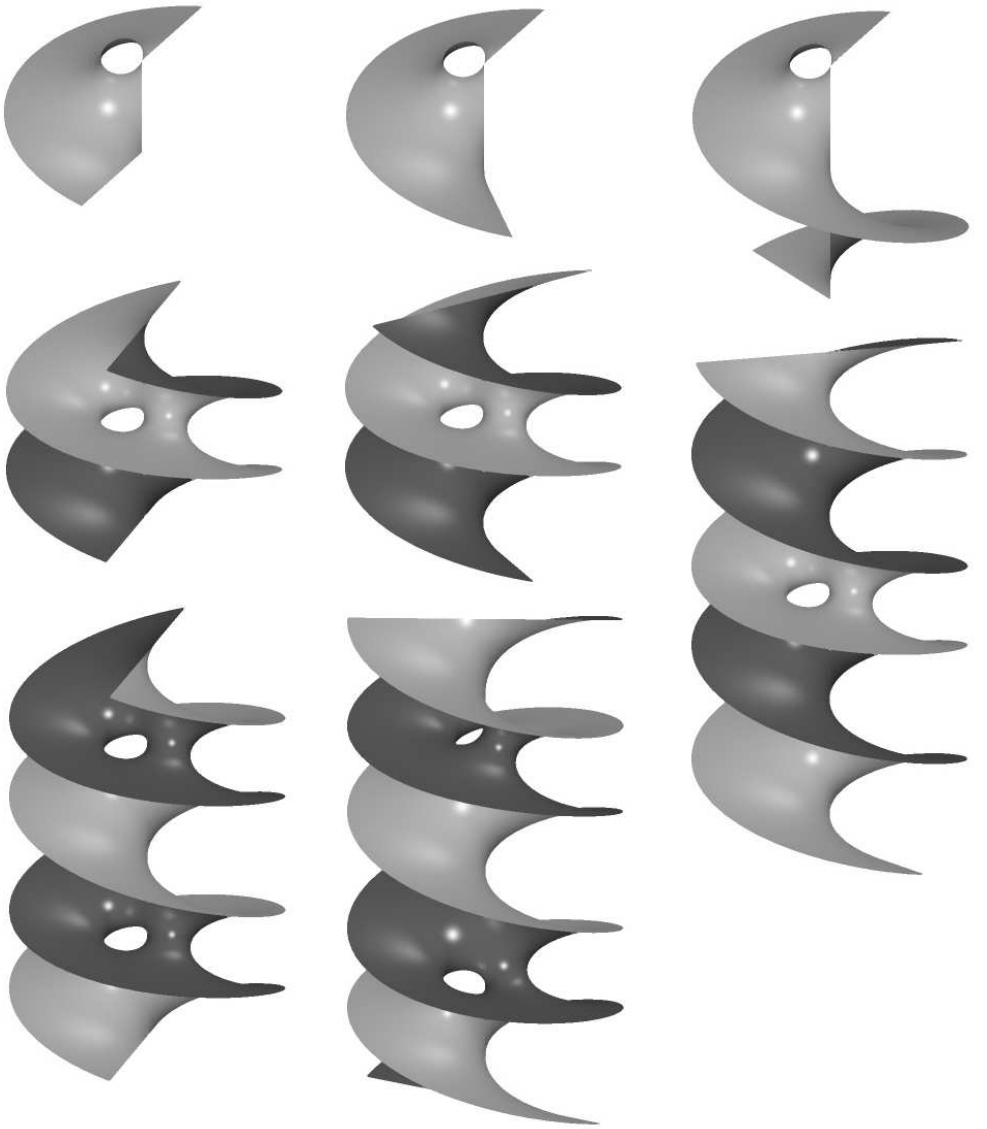


FIGURE 8. Three members of the deformation family $\mathcal{H}_k$. On the left is the singly periodic genus-one helicoid $\mathcal{H}_1$. On top is $\frac{1}{4}$ of $\mathcal{H}_1/T$. The axis and the horizontal lines are clearly visible. Two fundamental domains (copies of $\mathcal{H}_1/T$) are illustrated at the bottom. On the right is $\mathcal{H}_k$ for $k = 1.66$. On the top is $\frac{1}{4}$ of $\mathcal{H}_k/\sigma_k$ On the bottom is one full copy of $\mathcal{H}_k/\sigma_k$.

While we had little doubt that this family existed and that it was smooth in k, we had no way to prove that this was the case. We were not only not able to prove existence for any given k (except of course $k = 1$, which is $\mathcal{H}_1$), we could not prove that the surfaces depended continuously on k.

4.2. Flat structures and embeddedness of $\mathcal{H}e_1$. Various attempts to prove that the $\mathcal{H}e_1$ we found in [**17**] was embedded did not work out, and after a few years it became more attractive to try to prove the existence of the $\mathcal{H}_k$ family and its continuous dependence on k in order to establish the embeddedness of $\mathcal{H}e_1$.[3] While this seemed to be a daunting task, the failure of other attempts to prove the embeddedness of $\mathcal{H}e_1$ suggested to me that this method of proof had to be tried. However, it was clear that a new technique was needed.

The new technique was provided by ideas introduced into minimal surface theory by Matthias Weber and Michael Wolf. They developed a method of construction of Weierstrass data by means of flat structures that could be described either by the absolute value of a one-form on a subset of a Riemann surface or by the application of this metric to the flat plane, considered as $\mathbb{C}$, by means of integration of that one-form. By considering the one-form as essentially dz on a region of the plane, its periods were given by measurement in the region in the plane, and the region was most often a polygonal one. Therefore, the periods could be determined by inspection. A further breakthrough idea — prescribing the twist angle by sewing in a cone to a fixed Riemann surface — was made by Weber. A detailed description of this process is given in the papers of Weber and Wolf in this volume [**48, 45**], in [**44**] and in [**46**].

5. The Limit of $\mathcal{H}_k$ as $k \to \frac{1}{2}$

I claimed at the beginning of this paper that the main research use of computation in minimal surface theory is often not well understood. In my discussion of the surfaces $\mathcal{H}_1$, $\mathcal{H}e_1$ and the family $\mathcal{H}_k$, I have tried to show how theoretical ideas and computational results are interdependent. Both activities can produce mathematical problems not easily answered without reliance on the other. In this final section, I want to recount the story of an unexpected result of the computation of the $\mathcal{H}_k$ family. It led to results that no one expected and it required a theoretical advance even to recognize what had actually been computed.

While the helicoid is invariant under the screw motion σ_k for any $k \neq 0$, our computations indicated that the condition $k > \frac{1}{2}$ is necessary for the period conditions necessary for a surface $\mathcal{H}_k$ to exist and to satisfy the conditions described in Section 4.1. This was a mild surprise, but since we were interested in what happens when $k \to \infty$, we did not pursue it at the time.

Extracting information from pictures requires some knowledge of how the pictures have been computed, and some understanding of the theoretical questions that motivated the computation. In some cases, often the most interesting, it requires some time before the theory catches up with the computation. We have seen

[3]The argument would go like this. Assume that the $\mathcal{H}_k$ form a family of minimal surfaces that vary continuously in k. Since $\mathcal{H}_1$ is embedded and all the $\mathcal{H}_k/\sigma_k$ are embedded outside of a compact set, it follows from a maximum-principle argument that all the $\mathcal{H}_k$ are embedded. A separate argument has to be made that a limit surface exists as $k \to \infty$, that it has genus one, not genus zero, and that the $\mathcal{H}_k$ do not limit to a covering of the limit surface. Then this limit is embedded and can easily be shown to be an $\mathcal{H}e_1$, although perhaps not *the* $\mathcal{H}e_1$ found in [**17**].

several instances of that in the last 15 years.[4] The computation of the $\mathcal{H}_k$ family is an instance where computation was very much in advance of theory. In fact, the behavior of the $\mathcal{H}_k$ family near $k = \frac{1}{2}$ illustrates the point dramatically.

5.1. The computation. At the time of computation of the $\mathcal{H}_k$ family, we had clearly in mind the idea that the family, if it existed, had as its limit, as $k \to \infty$, an $\mathcal{H}e_1$. That the computation gave screw-motion-invariant solutions for twist angles $2\pi k$, $k < 1$, was noted, as was the fact that the computation began to behave poorly as the twist angle approached π, that is, as k approached $\frac{1}{2}$ from above. In order to be explicit and to explain the calculations represented by the graphs in Figure 10, it is necessary to describe briefly how we represented the minimal surfaces.

The torus presented by the equation

$$w^2 = P(z) := z(z - e^{i\theta})(z - e^{-i\theta})(z - d), \tag{5–1}$$

where $0 < \theta < \pi$, and $d < 0$ is a rhombic torus: the cross ratio of the roots of $P(z)$ is unitary, and this property characterizes rhombic tori. The orientation-reversing involutions $(z, w) \to (\bar{z}, \bar{w})$ and $(z, w) \to (\bar{z}, -\bar{w})$ have the property that their fixed-point sets are closed curves that are the lifts of intervals on the extended real axis in the z-plane: the lift of $[d, 0]$ for $(z, w) \to (\bar{z}, -\bar{w})$; the lift of $[\infty, d] \cup [0, \infty]$ for $(z, w) \to (\bar{z}, \bar{w})$. These fixed-point sets can be considered to be the diagonals of a rhombus. They cross at the points $(d, 0)$ and $(0, 0)$. The Weierstrass representation of the $\mathcal{H}_k$ will map one of these diagonals into the vertical axis, and the other into the horizontal lines on $\mathcal{H}_k/\sigma_k$. (Two points are removed from this other diagonal — corresponding to the ends in the quotient — thus creating two segments out of this diagonal.) The involutions above correspond to Schwarz reflection across these lines.

For each value of $d < 0$, we have a one-parameter family of representations, indexed by θ, of rhombic tori. We may define the position of the ends (symmetric with respect to the center of the rhombus) by our choice of θ in (5–1). Details are given in [**22**]. The twist angle $2\pi k$ is presented in Figure 10 as a function of d, with the sign of d changed from negative to positive for convenience. For each value of d, the value of θ is chosen to solve a part of the period problem.

It is important to emphasize again that the input variable to the program was the value of d, and the output was Weierstrass data for which the periods were killed, as well as $k = k(d)$, which determines the twist angle. Because of this, we got values of k less than 1 without having to be clever enough to look for them ourselves. Of course, it is likely that we would have explored this natural possibility, likely but not certain.

The program to kill the periods of the Weierstrass representation performed poorly as d approached values for which $k(d)$ was close to (but always greater than) $\frac{1}{2}$, and could not be pushed below $\frac{1}{2}$. For values k near $\frac{1}{2}$, the program MESH that does an automatic triangulation of the surface as it is being computed, slowed down significantly. We thought that it was likely that this was a numerical

[4]For most of the history of the subject, going back to the eighteenth century, it was computation that lagged behind the theory. To give just one example, Schwarz's solution, using the Weierstrass representation, of the Plateau problem for a four-sided polygon made up of edges of the tetrahedron was the first explicit solution of the Plateau problem for a boundary curve that did not lie on a known minimal surface. This may sound strange because we are used to thinking of the Weierstrass representation as theoretical. In fact it is the first general algorithm for computing minimal surfaces.

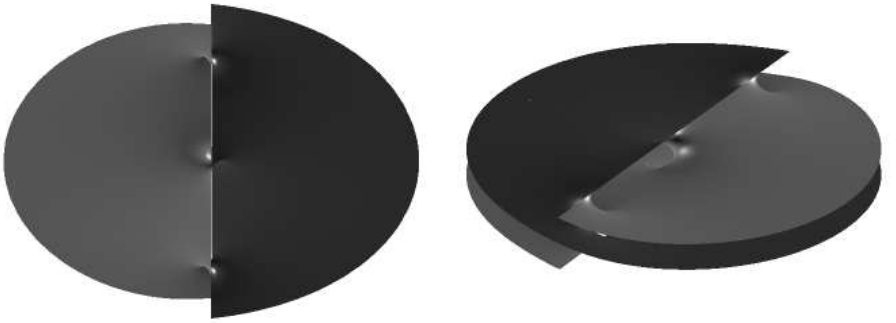


FIGURE 9. Near the limit value $k = \frac{1}{2}$. Screw-motion-invariant genus-one helicoid with twist angle just slightly greater than π. The images here were produced with MESH using a representation of the surfaces under consideration that involves theta functions. This method was found by Weber [**44**] and implemented for use here by Traizet. Introduced for theoretical reasons, it also has computational advantages; it allows one to get much closer to $k = \frac{1}{2}$. The surfaces shown here have $k = 0.501$.

problem, that somewhere in the code we had implicitly assumed that $k \geq 1$. In retrospect, this was not very likely since — as explained above — the value of k is an output of the computation. Still, some sort of poor parameterization could lead to errors like this, and we had seen behavior like this before. For example, at these extreme values, perhaps a branch point moved outside of a small circle used to compute a period. Also, the images of the surfaces for k near $\frac{1}{2}$ did not have any evident errors or telltale signs of impending degeneracy. The holes were not getting smaller, nor were they drifting away from the central axis and dilating.

We were not prepared by any previous experience to take seriously the possibility that the family would stop at some surface that did not appear to be degenerate. In fact, we hypothesized that it was possible for the twist angle to go to zero and imagined the holes getting smaller and smaller, closer together and clustering around the axis of the surface. The normalization of the surfaces (with $K = -1$ at a distinguished axis point) does not imply that the Gauss curvature is bounded as k goes to 0. Near $k = 0$, we expected to have handles lining up along the axis and shrinking. In the limit we expected these handles to converge to points on the axis and to see the helicoid as the limit set.

In retrospect, this point of view turned out not to be justified on the basis of what we saw. It was clear that the purely numerical approach was not able to give a reasonable answer to the question. In this case, without a reasonable theoretical framework, we were not able to use the substantial amount of information that was available to us. Without additional theoretical insight, we could not see what we had computed.

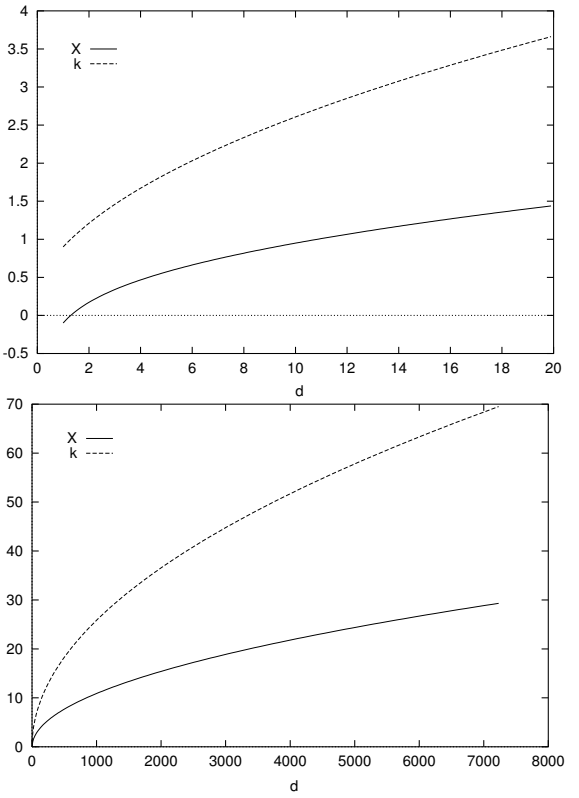


FIGURE 10. The twist angle k and the parameter X (a computed parameter in the Weierstrass data) as functions of d. In these graphs the sign of d has been changed from negative to positive.

5.2. The theory necessary to understand the computation. It turns out that the numerics and the graphics were correct as far as they went, but to realize this took an additional development, a development that was decidedly non-computational in nature. Martin Traizet, now at Tours, was a visitor to the GANG Laboratory in Amherst at the time we were doing these computations beginning in 1993. He was aware of the phenomena described in the previous paragraph. Some six years later, he developed a method of construction of minimal surfaces by a singular perturbation method, in which the Weierstrass representation and the implicit function theorem are used to solve the period problem. His first success in this area was to prove the existence of higher-genus Riemann examples, the originals of which had been computed by Wei [**42**]. He then extended this to construction of embedded minimal surfaces of finite total curvature. The idea was to consider the degenerate limit of known (computationally or theoretically) deformation families as an algebraic object that contained information about the position of degenerating catenoids in the family. He showed that one could perturb slightly off the singular limit to produce Weierstrass representations, which — using an implicit-function theoretical argument — had all periods equal to zero [**40, 41**].

Traizet visited MSRI in the fall of 2000. Matthias Weber (now at Indiana) was at MSRI during that period, and we discussed the possible limits of this family for

small k. Weber noted that there were some reasons to expect that his alternate construction of the deformation family might fail at $k = \frac{1}{2}$. He and Traizet discussed the possibility of extending Traizet's ideas to this case. However, there were no degenerating catenoids in sight. Using Weber's theta-function parameterization of these surfaces they were able to compute surfaces with twist angle closer to to π than we were. That made the difference. The images produced by MESH were quite revealing, and unanticipated. With the aid of images like that of Figure 9, Traizet and Weber had the startling insight that this limit could be considered to be degenerating into properly placed "half helicoids". This at first was hard to credit but in fact you can sort of "see" it in the pictures. Armed with this insight, they have been able to prove the existence near a singular limit (that is, for k near $\frac{1}{2}$) of screw-motion-invariant helicoids of arbitrary genus) [**43**].

Images such as those in Figure 9 were crucial in obtaining this insight. But to be able to "see" what was happening required the theoretical advances of Traizet and Weber. These were in turn partially motivated by numerical deformation experiments and partially by theoretical questions. They then provided a better means of computing these surfaces very close to the critical twist angle of π and a mindset to understand the meaning of the visual output.[5]

It is not possible to understand what a picture means without a knowledge not only of how it was computed but why it was computed, and under what theoretical assumptions. Sometimes, the understanding of subtle information in an image requires theoretical advances that happen years later.

References

[1] D. Anderson, C. Henke, D. Hoffman, and E. L. Thomas. Periodic area-minimizing surfaces in block copolymers. *Nature*, 334(6184):598–601, 1988. August 18 issue.

[2] M. Callahan, D. Hoffman, and J. Hoffman. Computer graphics tools for the study of minimal surfaces. *Communications of the ACM*, 31(6):648–661, 1988.

[3] M. Callahan, D. Hoffman, and H. Karcher. A family of singly-periodic minimal surfaces invariant under a screw motion. *J. Experimental Math.*, 2(3):157–182.

[4] M. Callahan, D. Hoffman, and W. H. Meeks III. Embedded minimal surfaces with an infinite number of ends. *Inventiones Math.*, 96:459–505, 1989.

[5] M. Callahan, D. Hoffman, and W. H. Meeks III. The structure of singly-periodic minimal surfaces. *Inventiones Math.*, 99:455–481, 1990.

[6] C. C. Chen and F. Gackstatter. Elliptic and hyperelliptic functions and complete minimal surfaces with handles. *Instituto de Matemática e Estatistica, Universidade de São Paulo*, 27, 1981.

[7] C. C. Chen and F. Gackstatter. Elliptische und Hyperelliptische Functionen und vollständige Minimalflächen von Enneperschen Typ. *Math. Ann.*, 259:359–369, 1982.

[8] P. Collin. Topologie et courboure des surfaces minimales proprement plongées de $\mathbf{R}^3$. *Ann. of Math.*, 145:1–31, 1997.

[5] It is important to note, however, that Weber and Traizet do not show that there is degeneration as $k(d)$ approaches $\frac{1}{2}$. They show that there is a family with the same properties as the ones constructed here which does degenerate to a foliation by horizontal planes with three vertical lines of singularity. After rescaling, neighborhoods around these lines are close to helicoids. There is no uniqueness result: in particular, we do not know whether or not there is a family that extends past the value $d_{1/2}$ for which $k(d)$ approaches $\frac{1}{2}$. What happens, if anything, on the other side of $d_{1/2}$ is not known. The "numerical approach" used here is not going to tell us much about uniqueness.

[9] C. Costa. *Imersöes minimas en* $\mathbf{R}^3$ *de gênero un e curvatura total finita.* PhD thesis, IMPA, Rio de Janeiro, Brasil, 1982.

[10] F. Wei, D. Hoffman, H. Karcher. The singly periodic genus-one helicoid. *Comment. Math. Helv.*, 74:248–279, 1999.

[11] U. Dierkes, S. Hildebrandt, A. Küster, and O. Wohlrab. *Minimal Surfaces I.* Grundlehren der mathematischen Wissenschaften 296. Springer, 1992.

[12] D. Hadjuk, P. Harper, S. Gruner, C. Honeker, G. Kim, and E. Thomas andL. Fetters. The gyroid: a new equilibrium morphology in weakly segregated diblock copolymers. *Macromolecules*, 27:4063–4075, 1994.

[13] D. Hoffman. The discovery of new embedded minimal surfaces: Elliptic functions, symmetry; computer graphics. In *Proceedings of the Berlin Conference on Global Differential Geometry*, Springer Lecture Series 1748, 1985.

[14] D. Hoffman. The computer-aided discovery of new embedded minimal surfaces. *Mathematical Intelligencer*, 9(3):8–21, 1987.

[15] D. Hoffman. Some basic facts, old and new, about triply periodic embedded minimal surfaces. In *Colloque de Physique*, 1990. Colloque No. 7, Supplément au Journal de Physique, Tome 51, December 1990, 197–208.

[16] D. Hoffman, H. Karcher, and F. Wei. Adding handles to the helicoid. *Bulletin of the AMS, New Series*, 29(1):77–84, 1993.

[17] D. Hoffman, H. Karcher, and F. Wei. The genus one helicoid and the minimal surfaces that led to its discovery. In *Global Analysis and Modern Mathematics.* Publish or Perish Press, 1993. K. Uhlenbeck, editor, p. 119–170.

[18] D. Hoffman and J. McCuan. Embedded minimal annular ends asymptotic to the helicoid. *Comm. Analysis and Geometry*, 11(4), 2003.

[19] D. Hoffman and W. H. Meeks III. A complete embedded minimal surface in $\mathbf{R}^3$ with genus one and three ends. *Journal of Differential Geometry*, 21:109–127, 1985.

[20] D. Hoffman and W. H. Meeks III. Embedded minimal surfaces of finite topology. *Annals of Mathematics*, 131:1–34, 1990.

[21] D. Hoffman and W. Rossman. Limit surfaces of Riemann examples. *J. Geom. Analysis*, 7(1):161–71.

[22] D. Hoffman and F. Wei. Deforming the periodic genus-one helicoid. *Experimental Mathematics*, 11(2):207–218, 2002.

[23] J. Hoffman, E. L. Thomas, et al. Ordered bicontinuous nanoporous and nanorelief ceramic films from self assembling polymers. *Science*, pages 1716–1719, 1999.

[24] J. Hoffman, E.L. Thomas, M. Wohlgemuth, and N. Yufa. Triply periodic bicontinuous cubic microdomain morphologies by symmetries. *Macromolecules*, 34:6083–6089.

[25] W. Meeks III and H. Rosenberg. Minimal surfaces of finite topology. Pages 471–488 in this volume.

[26] L. Jorge and W. H. Meeks III. The topology of complete minimal surfaces of finite total Gaussian curvature. *Topology*, 22(2):203–221, 1983.

[27] J.T.Pitts. *Existence and regularity of minimal surfaces on Riemannian manifolds.* Princeton University Press, Princeton, N.J., 1981.

[28] K.Große-Brauckmann. Gyroids of constant mean curvature. *Experimental Math.*, 6(1):33–50, 1997.

[29] K.Große-Brauckmann and M. Wohlgemuth. The gyroid is embedded and has constant mean curvature companions. *Calc.Var.Partial Differential Equations*, 4(6):499–523, 1996.

[30] T. Klotz and L.Sario. Existence of complete minimal surfaces of arbitrary connectivity and genus. *Proc.Nat.Acad.Sci U.S.A.*, 54:42–44, 1965.

[31] W. H. Meeks III and H. Rosenberg. The uniqueness of the helicoid and the asymptotic geometry of properly embedded minimal surfaces with finite topology. preprint.

[32] J. C. C. Nitsche. *Lectures on Minimal Surfaces*, volume 1. Cambridge University Press, 1989.

[33] R. Osserman. Global properties of minimal surfaces in E^3 and E^n. *Annals of Math.*, 80(2):340–364, 1964.

[34] R. Osserman. *A Survey of Minimal Surfaces.* Dover Publications, New York, 2nd edition, 1986.

[35] H. F. Scherk. Bemerkungen über die kleinste Fläche Innerhalb gegebener Grenzen. *J. R. Angew. Math.*, 13:185–208, 1835.

[36] A. Schoen. Infinite periodic minimal surfaces without self-intersections. Technical Note D-5541, NASA, Cambridge, Mass., May 1970.
[37] R. Schoen. Uniqueness, symmetry, and embeddedness of minimal surfaces. *Journal of Differential Geometry*, 18:791–809, 1983.
[38] H. A. Schwarz. *Gesammelte Mathematische Abhandlungen*, volume 1. Springer, Berlin, 1890.
[39] L. E. Scriven. Equilibrium bicontinuous structures. *Nature*, 263:123–125, 1976.
[40] M. Traizet. An embedded minimal surface with no symmetries. *J. Diff. Geometry*, 60:103.
[41] M. Traizet. Construction of minimal surfaces by gluing Weierstrass representations. Pages 439–451 in this volume.
[42] M. Traizet. Adding handles to the Riemann examples. *Journal of the Inst. Math.Jussieu*, 1(1):145–174, 2002.
[43] M. Traizet and M. Weber. Hermite polynomials and minimal surfaces. Preprint, 2003.
[44] M. Weber. On the embeddedness of the genus one helicoid. Habilitationschrift, University of Bonn, 2000.
[45] M. Weber. The genus one helicoid as a limit of screw-motion invariant helicoids with handles. Pages 243–258 in this volume.
[46] M. Weber, D. Hoffman, and M. Wolf. An embedded genus-one helicoid. preprint, 2002.
[47] F. Wei. Some existence and uniqueness theorems for doubly periodic minimal surfaces. *Invent. Math.*, 109:113–136, 1992.
[48] M. Wolf. Flat Structures, Teichm§ller theory and handle addition for minimal surfaces. Pages 211–241 in this volume.

Mathematical Sciences Research Institute, 17 Gauss Way, Berkeley, CA 94720
E-mail address: david@msri.org

Clay Mathematics Proceedings
Volume 2, 2004

Geometric Aspects of the Theory of Fully Nonlinear Elliptic Equations

Joel Spruck

1. Introduction

We discuss aspects of the theory of fully nonlinear elliptic equations as they pertain to global differential geometry. Advances in this theory in the last twenty years have opened the possibility of tackling extremely complicated existence questions. However, fully nonlinear elliptic equations arise naturally in many areas of geometry and we will also try to illustrate many diverse applications and tools.

For geometric applications, the most important class of fully nonlinear elliptic equations are implicitly defined equations of the form

$$F(A) = f(\kappa_1, \ldots, \kappa_n) = \psi$$

where A, for example, is the second fundamental form of a hypersurface, $f(\lambda)$ is a symmetric function of the eigenvalues of A and ψ is a function of position and the unit normal. Thus in the geometric setting we are studying functions of the principle curvatures of a hypersurface S. We will call a surface S an elliptic Weingarten surface if it satisfies $f(\kappa) = \text{constant} > 0$, and this equation in local coordinates is a nonlinear elliptic equation. In recent years, the study of hypersurfaces with curvature of higher order,

$$f(\kappa) = H_r(\kappa) = \frac{S_r(\kappa)}{S_r(1, \ldots, 1)} = 1$$

has received considerable attention. The linear case ($r = 1$) of mean curvature H is classical, while Gauss curvature K ($r = n$) is the prototypical fully nonlinear case. The case $r = 2$ of scalar curvature has been much much less studied, but is clearly of great geometric interest. In the next section, we show that hypersurfaces with curvature quotient $H_r(\kappa)/H_s(\kappa) = 1$ constant also fall into the general theory.

Before we begin to study the properties of implicitly defined elliptic equations, we will first give a few more (nonstandard) examples.

EXAMPLE 1.1 (Special lagrangian graphs). Let Ω be a bounded domain in $\mathbb{R}^n$ and consider the graph $\nabla u : \Omega \to \mathbb{R}^{2n}$ as an n-dimensional submanifold of $\mathbb{R}^{2n}$. The work of Harvey and Lawson [**21**] shows that such a graph is special Lagrangian (and

Research of the author was partially supported by NSF Grant DMS-0072242.

©2004 Clay Mathematics Institute

in particular is absolutely area minimizing) with respect to the standard symplectic structure if u satisfies

$$f(\lambda) = \operatorname{Im}(\delta_{ij} + iu_{ij}) = \sum_{k=0}^{[(n-1)/2]} (-1)^k S_{2k+1}(\lambda(u_{ij})) = 0 \quad \text{in } \Omega.$$

This fully nonlinear equation although elliptic does not fit directly into the fully nonlinear theory because of the concavity requirement. However in [**5**] it is shown that at least one of the level sets of $f(\lambda) = 0$ is concave and this allows the definition of a new $g(\lambda)$ which is elliptic and concave on the admissible class of functions where the eigenvalues of the hessian lie inside this level set. For recent work, see Yuan [**40**].

EXAMPLE 1.2 (Conformal equations on S^n). Let (S^n, g_0) be the n-sphere with the standard metric and let g be a metric conformal to g_0. Viaclovsky [**39**], Chang–Gursky–Yang [**9**] and Li–Li [**26**] have studied the (global) fully nonlinear equation

$$S_k\Big(R_{ij} - \frac{R}{2(n-1)} g_{ij}\Big) = 1$$

where R_{ij}, R are respectively the Ricci tensor and scalar curvature of the metric g. Let $x = (x_1, \dots, x_n)$ be coordinates obtained from stereographic projection from the North pole and write

$$g = u(x)^{-2}\, dx^2 = u(x)^{-2}(1 + |x|^2)^2 g_0, \qquad u > 0.$$

Then the preceding equation is equivalent to

$$S_k\Big(uu_{ij} - \frac{|\nabla u|^2}{2}\delta_{ij}\Big) = 1.$$

This equation is conformally invariant, that is, for any conformal transformation $T : \mathbb{R}^n \to \mathbb{R}^n$,

$$v(x) = |J(x)|^{-1/n} u(Tx)$$

is again a solution.

EXAMPLE 1.3 (Global problems associated to Curvature flows). Let S_0 be a compact hypersurface of $\mathbb{R}^n$ and consider a flow with normal velocity given by a function of curvature $V^N = f(\kappa)$. For example, flow by mean or Gauss curvature, $f = H, K$ or perhaps the more exotic harmonic curvature (an example of a curvature quotient)

$$H_{n,n-1} = \frac{H_n}{H_{n-1}} = \frac{n}{\sum_{i=1}^n \kappa_i^{-1}};$$

see Huisken–Ilmanen [**24**].

Associated to such a flow we have a global problem in Differential Geometry: Let S be a compact hypersurface satisfying $f(\kappa) = X \cdot \nu \geq 0$. Is S necessarily a sphere? This is the problem of asymptotic shape for the flow, at least for S_0 convex. If we start with such an S, then it flows by homothety. The problem is somewhat subtle; if S is an ellipsoid in $\mathbb{R}^3$, then its Gauss curvature always satisfies $K = (X \cdot \nu)^4$ so for $f(\kappa) = K^{1/4}$ ellipsoids are solutions. See Andrews [**1, 2**].

Another global problem associated to the flow is to characterize translating solitons. These are solutions which are entire graphs $x_{n+1} = u(x)$ satisfying $f(\kappa) = \nu^{n+1}$. Assuming for example that u is convex, is u necessarily radially symmetric with respect to some origin?

For example, if $f = S_1(\kappa)$, then u satisfies

$$\left(\delta_{ij} - \frac{u_i u_j}{1+|\nabla u|^2} u_{ij}\right) = 1.$$

If u is convex, is u necessarily radial? The difficulty is to establish that u is asymptotically quadratic.

If S is the graph of a function $x_{n+1} = u(x)$, defined in a domain $\Omega \subset \mathbb{R}^n$, then the Gauss curvature of the graph may be expressed as

$$K(S) = \frac{\det u_{ij}}{(1+|\nabla u|^2)^{(n+2)/2}}.$$

So if $K(S) = K > 0$ is constant and we want to prescribe boundary values, we arrive at the Monge–Ampère boundary value problem

$$\begin{aligned} \det u_{ij} &= K(1+|\nabla u|^2)^{(n+2)/2} && \text{in } \Omega, \\ u &= \phi && \text{on } \partial\Omega. \end{aligned}$$

It is not difficult to see that the Monge–Ampère operator is elliptic only for the admissible class of strictly convex functions, corresponding to the class of hypersurfaces S with principle curvatures which are strictly positive. The general existence and regularity theory for fully nonlinear equations $F(D^2u, Du) = \psi(x, u, Du)$ is well developed for equations which are uniformly elliptic and concave (that is, $L = F^{ij}\partial_i\partial_j$ is uniformly elliptic, where $F^{ij} = \partial F/\partial u_{ij}$ and F is also concave in D^2u) because of the fundamental Evans–Krylov interior $C^{2+\alpha}$ regularity results and their extensions up to the boundary (see [**14**] and [**4**]). In the case of the Monge–Ampère equation, the operator $F(D^2u) = (\det u_{ij})^{1/n}$ is elliptic (but not uniformly elliptic) and concave on the admissible class of strictly convex functions. This reduces existence and higher regularity questions to the question of obtaining $C^{1,1}$ a priori estimates.

In the geometric setting, we associate to the equation $F(A) = f(\kappa_1, \ldots, \kappa_n)$ the linearized operator $L = F^{ij}(A)\nabla_i\nabla_j$ where $F^{ij} = \partial F/\partial a_{ij}$ and ∇_i is a covariant derivative. Thus L is elliptic if F^{ij} if (say) positive definite.

We now show that the ellipticity and concavity of the nonlinear operator

$$F(A) = f(\lambda)$$

where the λ are the eigenvalues of a matrix function A, can be completely understood from the properties of $f(\lambda)$. We assume that $f(\lambda)$ is a symmetric function defined in an open convex cone Γ symmetric and with vertex at the origin. The symmetry of f implies that if f is a smooth function, so is $F(A)$.

THEOREM 1.4. *Assume that $f_{\lambda_i} > 0$ for all i and that $f(\lambda)$ is concave.*

(i) *When A is diagonal, $F^{ij} = f_i\delta_{ij}$ so $L = F^{ij}\nabla_i\nabla_j$ is elliptic.*

(ii) *$F(A)$ is concave.*

(iii) *When A is diagonal with simple eigenvalues,*

$$F^{ij,rs} = f_{ir}\delta_{ij}\delta_{rs} + \frac{f_i - f_j}{\lambda_i - \lambda_j}(1-\delta_{ij})\delta^{ij}_{rs},$$

where

$$F^{ij,rs} = \frac{\partial^2 F}{\partial a_{ij}\partial a_{rs}}$$

and $\delta^{ij}_{rs} = 1$ if $\{ij\} = \{rs\}$ and 0 otherwise.

PROOF. The computation depends on computing the first and second variations of the eigenvalues of a symmetric matrix when it is diagonal. We first give a quick proof of the concavity of $F(A)$, assuming we have already demonstrated part (i) by making use of the following well-known lemma from linear algebra:

LEMMA 1.5. *Let $(F^{ij}), (B_{ij})$ be symmetric $n \times n$ matrices with eigenvalues*

$$\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_n \geq 0, \qquad \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n.$$

Then $\sum F^{ij} B_{ij} \geq \gamma_1 \mu_1 + \cdots + \gamma_n \mu_n$.

Now let A, B be symmetric matrices with $\lambda(A), \lambda(B) \in \Gamma$. To demonstrate the concavity of $F(A)$ we want to show that

$$F^{ij}(A)(B - A)_{ij} \geq F(B) - F(A).$$

Let B have eigenvalues $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ and let A have eigenvalues $\{\lambda_i\}_{i=1}^n$ arranged so that (the eigenvalues of $G^{ij}(A)$)

$$f_{\lambda_1} \geq f_{\lambda_2} \geq \cdots \geq f_{\lambda_n} > 0.$$

By the lemma, $F^{ij} B_{ij} \geq f_{\lambda_1} \mu_1 + f_{\lambda_1} + \cdots + f_{\lambda_n} \mu_n$, so that

$$F^{ij}(B_{ij} - A_{ij}) = F^{ij} B_{ij} - \sum \lambda_i f_{\lambda_i} \geq \sum f_{\lambda_i}(\lambda)(\mu_i - \lambda_i) \geq f(\mu) - f(\lambda)$$

by the concavity of f, proving (ii).

To prove part i, we write $F^{ij} = \sum_k f_{\lambda_k} \partial \lambda_k / \partial A_{ij}$. We compute $\partial \lambda_k / \partial A_{ij}$ at a generic point where the eigenvalues are simple by making a variation $\tilde{A}_{ij} = A_{ij} + \epsilon$. If $j < i$, then

$$\det(\tilde{A} - \lambda I) = \prod_{k \neq i,j} (\lambda_k - \lambda)(\lambda^2 - (\lambda_i + \lambda_j) + \lambda_i \lambda_j - \epsilon^2).$$

Hence $\tilde{\lambda}_k = \lambda_k$ if $k \neq i, j$ and

$$\tilde{\lambda}_i = \frac{\lambda_i + \lambda_j}{2} + \sqrt{\left(\frac{\lambda_i - \lambda_j}{2}\right)^2 + \epsilon^2}, \qquad \tilde{\lambda}_j = \frac{\lambda_i + \lambda_j}{2} - \sqrt{\left(\frac{\lambda_i - \lambda_j}{2}\right)^2 + \epsilon^2}.$$

It follows that $\tilde{\lambda}_i = \lambda_i + O(\epsilon^2)$, and similarly $\tilde{\lambda}_j = \lambda_j + O(\epsilon^2)$. Hence $\partial \lambda_k / \partial A_{ij} = 0$ for $k \neq i, j$. If now $i = j$, then $\tilde{\lambda}_k = \lambda_k$ if $k \neq i$ and $\tilde{\lambda}_i = \lambda_i + \epsilon$. Hence in all cases, $\partial \lambda_k / \partial A_{ij} = \delta_{ki} \delta_{ij}$. It follows that

$$F^{ij} = \sum_k f_{\lambda_k} \frac{\partial \lambda_k}{\partial A_{ij}} = f_{\lambda_i} \delta_{ij},$$

proving (i).

To prove (iii) we use

$$F^{ij,rs} = \sum_k f_k \frac{\partial^2 \lambda_k}{\partial A_{ij} \partial A_{rs}} + \sum_{k,l} f_{kl} \frac{\partial \lambda_k}{\partial A_{ij}} \frac{\partial \lambda_l}{\partial A_{rs}}.$$

From our previous calculations we see that

$$\sum_{k,l} f_{kl} \frac{\partial \lambda_k}{\partial A_{ij}} \frac{\partial \lambda_l}{\partial A_{rs}} = f_{ir} \delta_{ij} \delta_{rs}.$$

Similarly, $\partial^2 \lambda_k / \partial A_{ij} \partial A_{rs}$ is nonzero only when $\{i, j\} = \{r, s\}$ and $i \neq j$ and

$$\frac{\partial^2 \lambda_i}{\partial A_{ij}^2} = \frac{1}{\lambda_i - \lambda_j}, \quad \frac{\partial^2 \lambda_j}{\partial A_{ij}^2} = -\frac{1}{\lambda_i - \lambda_j} \quad \text{if } \lambda_i > \lambda_j.$$

The formula follows. Note also that $(f_i - f_j)/(\lambda_i - \lambda_j) \leq 0$. To see this, suppose $\lambda_i > \lambda_j$ and note that by the symmetry of the convex cone Γ, the ray

$$\lambda^* + t(\lambda_i - \lambda_j)(e_i - e_j), \quad 0 \leq t \leq 1,$$

is in Γ, where λ^* is obtained from λ by interchanging λ_i and λ_j. Since f is symmetric and concave, the graph of $t \to f(\lambda^* + t(\lambda_i - \lambda_j)(e_i - e_j))$ is concave and symmetric about its maximum which occurs at $t = \frac{1}{2}$. Hence

$$(\lambda_i - \lambda_j)(f_i(\lambda) - f_j(\lambda)) \leq 0$$

from which the result follows. □

As a nice application of the theorem, we can give purely algebraic necessary and sufficient conditions on $f(\kappa)$ which guarantee that the linearized operator is a divergence operator.

THEOREM 1.6. *$L = F^{ij}\nabla_i\nabla_j$ is of divergence form if and only if*

$$f_{jj} = 0 \quad \textit{and} \quad \sum_{r \neq j}\Big(f_{rj} + \frac{f_r - f_j}{\lambda_r - \lambda_j}\Big) = 0 \qquad \textit{for all } j. \tag{1.1}$$

PROOF. We use $\nabla_k A_{ij} = \nabla_i A_{kj}$ (Codazzi equations) for a hypersurface in $\mathbb{R}^{n+1}$ with second fundamental form A_{ij}. Hence $L = \nabla_i(F^{ij}\nabla_j) - F^{ij,rs}\nabla_i A_{rs}$ is a divergence if and only if $F^{ij,rs}\nabla_i A_{rs} = 0$. We may suppose that A_{ij} is diagonal with simple eigenvalues; then

$$\begin{aligned} F^{ij,rs}\nabla_i A_{rs} &= \sum_r f_{jr}\nabla_j A_{rr} + \sum_{r \neq j}\frac{f_r - f_j}{\lambda_r - \lambda_j}\nabla_r A_{rj} \\ &= f_{jj}\nabla_j A_{jj} + \sum_{r \neq j}\Big(f_{rj} + \frac{f_r - f_j}{\lambda_r - \lambda_j}\Big)\nabla_r A_{rj}. \end{aligned}$$

Since this must hold for all possible values of $\nabla_j A_{jj}$ and $\nabla_r A_{rj}$ the result follows. □

It is well-known (see for example [**31**]) that the linearized operator for the higher order mean curvatures are of divergence form. This is usually proven through the use of the so-called Newton tensors. Instead we give a simple proof using Theorem 1.6.

COROLLARY 1.7. *$f = S_r$ satisfies* (1.1) *and so its associated operator L_r is of divergence form.*

PROOF. Let $f = S_r(\kappa)$; then (see Lemma 2.14 below) $f_i = S_{m-1}(\kappa_i')$, and so $f_r = S_{m-1}(\kappa_r') = \kappa_j S_{m-2}(\kappa_{rj}') + S_{m-1}(\kappa_{rj}')$. Therefore,

$$\sum_{r \neq j}\frac{f_r - f_j}{\kappa_r - \kappa_j} = -\sum_{r \neq j} S_{m-2}(\kappa_{rj}').$$

On the other hand,

$$f_{jj} = 0 \quad \text{and} \quad \sum_{r \neq j} f_{rj} = \sum_{r \neq j} S_{m-2}(\kappa_{rj}') = \sum_{r \neq j}\frac{f_r - f_j}{\kappa_r - \kappa_j}$$

as required. □

An outline of the content of the next three sections is as follows. In Section 2, we present many basic results (concavity, ellipticity) on elementary symmetric functions and curvature quotients that are used. The presentation is essentially self-contained. In Section 3, we derive the fundamental identities on Weingarten surfaces extending the well-known method for minimal surfaces. As an application we present a new variant [**10**] of Alexandrov reflection, the "method of moving spheres". In Section 4, we discuss the Monge–Ampère equation and applications to the existence of hypersurfaces of constant positive Gauss curvature (K-hypersurfaces for short) [**18, 34**]. For very recent work on the existence of immersed K-hypersurfaces in $\mathbb{R}^{n+1}$ see [**19, 38**].

2. Hyperbolic Polynomials, Elementary Symmetric Functions and Convexity

Gårding's work on hyperbolic polynomials [**11, 23**] is important to the study of curvature functions.

DEFINITION 2.1. A homogeneous polynomial $p(x)$ of degree m in $\mathbb{R}^n$ is said to be hyperbolic with respect to $a \in \mathbb{R}^n$, and we write "p(x) is hyp(a)", if the equation $p(x+ta)=0$ of degree m (in t) has exactly m real roots for every $x \in \mathbb{R}^n$.

Necessarily $p(a) \neq 0$ and we will assume $p(a) > 0$. It is also easy to see that $p(x)$ has real coefficients and that $Q = \sum_j a_j \partial P/\partial \lambda_j$ is hyp(a) by Rolle's theorem.

THEOREM 2.2. *Suppose $p(x)$ is* hyp(a). *Then the component Γ of $\{x \in \mathbb{R}^n : p(x) \neq 0\}$ containing a is a convex cone, $p(x)$ is* hyp(b) *for any $y \in \Gamma$, and $(p(x))^{1/m}$ is concave.*

PROOF. For simplicity assume $p(a) = 1$ (as we will in applications). Set

$$\Gamma_a = \{x \in \mathbb{R}^n : p(x+ta) \neq 0, t \geq 0\}.$$

Since $p(x+ta) = p(a)\prod(t-t_i(x))$, $x \in \Gamma_a$ if and only if all the t_i are negative. Now Γ_a is open and $a \in \gamma_a$ since $p(a+ta) = (1+t)^m$ has only the real root -1. Suppose $x \in \overline{\Gamma_a}$. Then $p(x+ta) \neq 0$ for $t > 0$ so $x \in \Gamma_a$ if $p(x) \neq 0$. Hence Γ_a is open and closed in $\{x | p(x) \neq 0\}$. But Γ_a is star-shaped with respect to a ($x \in \Gamma_a \Rightarrow \alpha x$, $\alpha x + \beta a \in \Gamma_a$ for $\alpha, \beta > 0$), so $\Gamma_a = \Gamma$. □

For $y \in \Gamma$ and $\varepsilon > 0$ fixed, set

$$E_{y,\varepsilon} = \{x \in \mathbb{R}^n : p(x + i\varepsilon a + isy) \neq 0, \operatorname{Re} s \geq 0\}.$$

Then $E_{y,\varepsilon}$ is open and contains 0 since $p(i\varepsilon a + isy) = (is)^m p(\varepsilon a/s + y) = 0$ implies $s < 0$ (since $y \in \Gamma = \Gamma_a$). If $x \in \underline{E}_{y,\varepsilon}$, then $p(x + i\varepsilon a + isy) \neq 0$ by a theorem of Hurwitz if $\operatorname{Re} s > 0$, while if $\operatorname{Re} s = 0$, $z = x + isy \in \mathbb{R}^n$, $p(z + i\varepsilon a) = \prod(i\varepsilon - t_i(z)) \neq 0$. Hence $E_{y,\varepsilon}$ is both open and closed, so $E_{y,\varepsilon} = \mathbb{R}^n$.

In particular (put $s = 1$), we have shown that

$$p(x + i(\varepsilon a + y)) \neq 0 \quad \text{if } x \in \mathbb{R}^n,\ y \in \Gamma,\ \varepsilon > 0.$$

Since Γ is open, this remains true for $\varepsilon = 0$, i.e., $p(x+iy) \neq 0$ for all $x \in \mathbb{R}^n$. It follows that the equation $p(x+ty) = 0$ has only real roots, for if $t = t_1 + it_2$ is a root with $t_2 \neq 0$, this would mean $p((x+t_1 y)/t_2 + iy) = 0$, a contradiction. Thus we have shown that $p(x)$ is hyp(y) for any $y \in \Gamma$. Hence y can play the role of a so Γ is star-shaped with respect to every point in Γ so is convex.

Finally, we will prove the concavity of $(p(x))^{1/m}$ and the strict monotonicity of $p(x)$ in each argument.

PROPOSITION 2.3. (i) *For any $y \in \Gamma$ and $x \in \mathbb{R}^n$, the function $\phi(s) = (p(sx + y))^{1/m}$ is concave in s when $sx + y \in \Gamma$. In particular, take $x = z - y$ with $z \in \Gamma$ and $0 \le s \le 1$; then*

$$(p(sz + (1 - s)y))^{1/m} \ge s(p(z))^{1/m} + (1 - s)(p(y))^{1/m}.$$

(ii) *$p_{x_i} > 0$ in Γ.*

PROOF. $p(x + ty) = p(y) \prod (t - t_i)$, so

$$p(sx + y) = s^m p\Big(x + \frac{y}{s}\Big) = p(y) \prod (1 - st_i).$$

Since $sx + y \in \Gamma, 1 - st_i > 0$ for all i. Set $f(s) = \log \phi(s)$; then

$$f'(s) = -\sum \frac{t_i}{1 - st_i}, f''(s) = -\sum \frac{t_i^2}{(1 - st_i)^2}.$$

Hence,

$$m^2 e^{-f(s)/m} \frac{d^2}{ds^2} e^{f(s)/m} = f'(s)^2 + mf''(s) = \sum \Big(\frac{t_i}{1 - st_i}\Big)^2 - m \sum \frac{t_i^2}{(1 - st_i)^2} \le 0$$

by Cauchy–Schwartz, completing the proof. □

COROLLARY 2.4. *Let $p(x)$ be a symmetric hyperbolic polynomial of degree m with respect to $a = (1, 1, \ldots, 1)$. Then:*

(i) *Γ is an open (proper)convex cone containing the positive cone Γ_+ and contained in the half-space $\sum x_i > 0$.*

(ii) *$f(x) = p(x)^{1/r}$ is concave and $f_{x_i} > 0$ for all i.*

The last condition is the ellipticity (we explained this in the introduction) and follows easily from the concavity, positivity and analyticity of $f(x)$ in Γ. Starting with $S_n(x) = \prod_{i=1}^n x_i$, $a = (1, \ldots, 1)$, which is obviously hyperbolic, we conclude successively by applying the differential operator $\sum_j \partial/\partial x_j$ that all the elementary symmetric functions $S_r = \sum_{i_1 < \cdots < i_r} x_{i_1} \cdots x_{i_r}$ are also hyp(a) and that $\Gamma(S_r) \subset \Gamma(S_{r-1})$.

Thus the elementary symmetric functions $S_r(\kappa)$ of curvature are the most important examples for study. It is useful to normalize S_r and make the following definitions:

DEFINITION 2.5. Set $H_r(x) = 1/\binom{n}{r}$. The curvature functions $H_r(\kappa)$ of a hypersurface are called the higher-order mean curvatures, with the following particular cases:

$$H_1(\kappa) = \frac{1}{n} \sum \kappa_i \qquad \text{is the mean curvature,}$$

$$H_2(\kappa) = \frac{2}{n(n-1)} \sum_{i<j} \kappa_i \kappa_j \qquad \text{is the scalar curvature,}$$

$$H_n(\kappa) = \prod k_i \qquad \text{is the Gaussian curvature.}$$

The S_r and H_r and the cones Γ_r are surprisingly complex and we will need additional properties. Besides being interesting and important in their own right, they will allow us to introduce other curvature functions that are worthy of further study.

We first state the following elementary but useful lemma which follows from repeated applications of Rolle's theorem.

LEMMA 2.6. [**20**] *Let $P(x,y)$ be a homogeneous polynomial in x and y with real coefficients such that all of its roots $\frac{x}{y}$ are real. Then all polynomials derived by differentiation with respect to x and y also have only real roots.*

Let's apply this to

$$P(x,y) = \prod(x + \lambda_i y) = \sum_{r=0}^{n} S_r(\lambda) x^{n-r} y^r.$$

We obtain

$$\frac{\partial^{n-2} P}{\partial^{n-r-2} x \, \partial^r y} = \frac{n!}{2}(H_r(\lambda)x^2 + 2H_{r+1}(\lambda)\, xy + H_{r+2}(\lambda)y^2).$$

Hence the discriminant must be nonnegative and we have shown:

PROPOSITION 2.7 (Newton inequalities). *For any $\lambda \in \mathbb{R}^n$,*

$$H_{r-1}H_{r+1} < H_r^2 \quad \text{for } 1 \le r \le n,$$

with equality only if all the components of λ are equal.

COROLLARY 2.8 (Maclaurin inequalities). *If $\lambda \in \Gamma_r$, then*

$$H_r^{1/r} \le H_{r-1}^{1/(r-1)}.$$

PROOF. We have

$$(H_0H_2)(H_1H_3)^2 \cdots (H_{r-2}H_r)^r \le H_1^2 H_2^4 \cdots H_{r-1}^{2r},$$

so $H_r^{r-1} \le H_{r-1}^r$. □

We now consider another interesting class of curvature functions.

DEFINITION 2.9. For $\lambda \in \Gamma_r$ and $r > s$ define the curvature quotient

$$H_{r,s}(\lambda) = \frac{H_r(\lambda)}{H_s(\lambda)}.$$

Using the Newton inequalities, we can also prove:

LEMMA 2.10 (Generalized Newton–Maclaurin inequalities). *If $\lambda \in \Gamma_r, 0 \le s < r$ and $0 \le l < k$, then*

$$(H_{r,s})^{1/(r-s)}) \le (H_{k,l})^{1/(k-l)}$$

provided $r \ge k$ and $s \ge l$.

PROOF. Note that the Newton inequalities may be expressed as $H_{i+1,i} \le H_{i,i-1}$ for $1 \le i \le r-1$, so that $H_{r,r-1} \le H_{k,l}^{1/(k-l)}$ for $k \le r$. We proceed by induction on r starting from the trivial case $r = 1$. Then, for $k \le r$,

$$\begin{aligned}(H_{r+1,s})^{1/(r+1-s)} &= (H_{r+1,r}H_{r,s})^{1/(r+1-s)} \le ((H_{k,l})^{1/(k-l)}(H_{k,l})^{r-s/(k-l)})^{1/(r+1-s)} \\ &= (H_{k,l})^{1/(k-l)},\end{aligned}$$

completing the induction. On the other hand if $k = r + 1$ then

$$(H_{r+1,s})^{1/(r+1-s)} \leq (H_{r+1,l})^{1/(r+1-l)}$$

is equivalent to

$$(H_{r+1,l})^{1/(r+1-l)}) \leq (H_{s,l})^{1/(s-l)}.$$

Since $s \leq r$ the previous induction applies and the proof is complete. □

We next prove the crucial concavity of $H_{r,s}$.

THEOREM 2.11. *$(H_{r,s})^{1/(r-s)}$ is concave on Γ_r.*

We reduce to the case $s = r - 1$ using that

$$H_{r,s} = \prod_{i=s+1}^{r} H_{i,i-1}$$

is a product of $r - s \geq 1$ positive terms of homogeneity 1. Suppose we have proved that $H_{i,i-1}$ is concave. Then the concavity of $H_{r,s}$ follows from the following useful lemma of independent interest.

LEMMA 2.12. *Let $R_i(x)$, $i = 1, \ldots, N$ be positive and concave on a convex cone Γ in R^n. Then $R = (\prod_{i=1}^{N} R_i)^{1/N}$ is concave on Γ.*

PROOF. By the concavity of each R_i,

$$2^N \prod_{i=1}^{N} R_i\Big(\frac{x_1 + x_2}{2}\Big) \geq \prod_{i=1}^{N}(R_i(x_1) + R_i(x_2)). \tag{2.1}$$

Write $a_i = R_i(x_1) > 0$, $b_i = R_i(x_2) > 0$, $r_i = a_i/b_i$, $r = (r_1, \ldots, r_N)$. Then using the Maclaurin inequalities in the positive cone of $\mathbb{R}^N$ gives

$$\prod_{i=1}^{N}(r_i+1) = \sum_{i=0}^{N} \binom{N}{i} H_i(r) \geq \sum_{i=1}^{N}\left(1+\binom{N}{i}\left(\prod_{1}^{N} r_j\right)^{i/N}\right) = \left(1+\left(\prod_{1}^{N} r_j\right)^{1/N}\right)^N.$$

In terms of the variables a_i, b_i we have shown that

$$\prod(a_i + b_i) \geq \left((\prod a_i)^{1/N} + (\prod b_i)^{1/N}\right)^N.$$

REMARK 2.13. This is just the well-known result that the nth root of the determinant function on positive definite matrices is concave. More generally, we will soon show that if $f(\lambda)$ is concave on the convex cone Γ and the λ_i are the eigenvalues of a matrix A, then $F(A) = f(\lambda)$ is a concave function of A. We just demonstrated this for $f = S_n(\lambda)^{1/n}$.

Recalling equation (2.1) and the definition of a_i, b_i, this is the concavity of R. □

Thus in order to prove the theorem, it remains to show the concavity of $H_{i,i-1}$ or equivalently of $S_{i,i-1} = S_i/S_{i-1}$. To do so we need some elementary identities, which we collect here for reference:

LEMMA 2.14. *Write $x_i' \in \mathbb{R}^{n-1}$ for the vector obtained from $x \in \mathbb{R}^n$ by omitting the i-th coordinate. Then*

$$\begin{aligned} S_r(x) &= x_i S_{r-1}(x_i') + S_r(x_i'), \\ rS_r(x) &= \sum_{i=1}^n x_i S_{r-1}(x_i'), \\ (n-r)S_r(x) &= \sum_{i=1}^n S_r(x_i'), \\ rS_r(x) &= S_1 S_{r-1} - \sum_{i=1}^n x_i^2 S_{r-2}(x_i'). \end{aligned}$$

Using the lemma, we find

$$\frac{\lambda_i S_{r-1}(\lambda_i')}{S_{r-1}(\lambda)} = \frac{\lambda_i S_{r-1}(\lambda_i')}{\lambda_i S_{r-2}(\lambda_i') + S_{r-1}(\lambda_i')} = \frac{\lambda_i S_{r-1,r-2}(\lambda_i')}{\lambda_i + S_{r-1,r-2}(\lambda_i')}.$$

Hence

$$S_1(\lambda) - rS_{r,r-1}(\lambda) = \sum_{i=1}^n \frac{\lambda_i^2}{\lambda_i + S_{r-1,r-2}(\lambda_i')}.$$

Since $S_{1,0} = S_1$ is concave, we can start an induction. Assume we have shown that $S_{i,i-1}$ is concave for $1 \le i \le r-1$. Restricting to $\{\lambda_i = 0\} \cap \Gamma$, we have that $h(\lambda_i') = S_{r-1,r-2}(\lambda_i')$ is concave and we complete the induction using the following

PROPOSITION 2.15. *Let $\phi^i(\lambda) = \lambda_i^2/(\lambda_i + h(\lambda_i'))$ for $1 \le i \le n$, where h is concave and $\lambda_n + h$ is positive. Then ϕ^i is convex in Γ.*

PROOF. It suffices to consider the case $i = n$, and write $\phi = \phi^n$. By a straightforward computation, we find that $(D^2\phi) = (\phi_{ij})$ is given by

$$\begin{aligned} \phi_{\alpha\beta} &= -\frac{\lambda_n^2}{(\lambda_n + h)^2}\Big(h_{\alpha\beta} - \frac{2}{\lambda_n + h} h_\alpha h_\beta\Big) \qquad \text{for } \alpha, \beta < n, \\ \phi_{\alpha n} &= -\frac{\lambda_n h}{(\lambda_n + h)^3} h_\alpha \qquad \text{for } \alpha < n, \\ \phi_{nn} &= \frac{2h^2}{(\lambda_n + h)^3}, \end{aligned}$$

and

$$\phi_{ij}\xi_i\xi_j = \frac{2}{(\lambda_n + h)^3}(h\xi_n - \lambda_n \sum_{\alpha<n} \xi_\alpha h_\alpha)^2.$$

Therefore $D^2\phi$ is positive semidefinite so ϕ is convex. □

THEOREM 2.16. *Let $f(\lambda) = (H_{r,s}(\lambda)^{1/(r-s)}$ for $1 \le s < r \le n$ and $\lambda \in \Gamma_r$. Then $f_{\lambda_i} > 0$ for all i and f is concave in Γ.*

PROOF. It remains only to show the strict monotonicity of f in each variable. This is equivalent to showing that $S_{r,s}$ is strictly monotone in each variable. To

simplify the notation, we write $S_k(\lambda_i') = S_{k;i}$. Then

$$\begin{aligned}\frac{\partial S_{r,s}}{\partial \lambda_i} &= \frac{S_s S_{r-1;i} - S_r S_{s-1;i}}{S_s^2} \\ &= \frac{(S_{s;i} + \lambda_i S_{s-1;i}) S_{r-1;i} - (S_{r;i} + \lambda_i S_{r-1;i}) S_{s-1;i}}{S_s^2} \\ &= \frac{S_{s;i} S_{r-1;i} - S_{r;i} S_{s-1;i}}{S_s^2} \\ &\geq \frac{n(r-s)}{r(n-s)} \frac{S_{s;i} S_{r-1;i}}{S_s^2} > 0,\end{aligned}$$

where we have used Lemma 2.14 and the generalized Newton–Maclaurin inequalities for the last step. □

3. Calculus on the Hypersurface S; Fundamental Identities

The curvature function $f(\kappa)$ implicitly defines a nonlinear function $G(b_{ij})$ of the second fundamental form by the relation $G(b_{ij}) = f(\kappa)$. Therefore, we can apply the theory of such operators developed in [**5**]. Naturally associated to this operator is the linear elliptic differential operator

$$L = \sum_{i,j=1}^{n} G^{ij} \nabla_i \nabla_j$$

where $G = G(b_i j), G^{ij} = \partial / \partial b_{ij}$. For example, if $f(\kappa) = \sum \kappa_i$, then $L = \Delta$ as is well known.

Instead of working intrinsically using covariant differentiation on S, it is often much easier to let

$$\delta = \nabla - \nu (\nu \cdot \nabla)$$

denote the tangential gradient operator on S, where ∇ is the gradient operator in $\mathbb{R}^{n+1}$. Let $e_1, \cdots, e_{n+1}$ denote the orthonormal coordinate frame of $\mathbb{R}^{n+1}$, and set

$$\delta_i = e_i \cdot \delta, \qquad \nu^i = e_i \cdot \nu, \qquad 1 \leq i \leq n+1.$$

LEMMA 3.1. [**27**] *The curvature matrix* $[\delta_i \nu^j]$ *is symmetric with eigenvalues* $(-\kappa_1, \ldots, -\kappa_n, 0)$ *on* S, *where* $\kappa_1, \ldots, \kappa_n$ *are the principal curvatures of* S. *Moreover, we also have the important commutator formula*

$$\delta_i \delta_j - \delta_j \delta_i = (\nu^i \delta_j \nu - \nu^j \delta_i \nu) \cdot \delta. \tag{3.1}$$

Following [**17**], we extend the symmetric function f to some open, symmetric subset of $\mathbb{R}^{n+1}$ in a canonical way. By a local version of a theorem of Glaeser [**15**, p. 108], we can find a smooth function h that satisfies

$$f(\kappa) = h(\sigma^{(1)}(\kappa), \ldots, \sigma^{(n)}(\kappa)) \qquad \text{for } \kappa \in \Gamma.$$

On the other hand, the elementary symmetric functions $\sigma^{(1)}, \cdots, \sigma^{(n)}$ are naturally defined on $\mathbb{R}^{n+1}$. Set

$$\tilde{\Gamma} = \{\lambda \in \mathbb{R}^{n+1} : (\sigma^{(1)}(\lambda), \ldots, \sigma^{(n)}(\lambda)) \in \mathbb{R}^n \text{ is in the domain of } h\}.$$

Then $\tilde{\Gamma}$ is a symmetric open subset of $\mathbb{R}^{n+1}$. Now define a smooth symmetric function $\tilde{f}$ on $\tilde{\Gamma}$ by

$$\tilde{f}(\lambda) = h(\sigma^{(1)}(\lambda), \ldots, \sigma^{(n)}(\lambda)) \qquad \text{for } \lambda \in \tilde{\Gamma}.$$

We note that $\Gamma \times \{0\} \subset \tilde{\Gamma}$, and

$$\tilde{f}(\kappa, 0) = f(\kappa) \qquad \text{for } \kappa \in \Gamma.$$

It is obvious that at a point $(\kappa, 0) \in \Gamma \times \{0\}$,

$$\tilde{f}_i(\kappa, 0) = f_i(\kappa) \qquad \text{for } 1 \leq i \leq n.$$

Let G denote the function on the linear space of real $(n+1) \times (n+1)$ symmetric matrices given by

$$G(A) = \tilde{f}(\lambda_1, \ldots, \lambda_{n+1}),$$

where $\lambda_1, \ldots, \lambda_{n+1}$ are the eigenvalues of the symmetric matrix A. Then we may write

$$G(-\delta\nu) = f(\kappa). \tag{3.2}$$

Note that if $\tilde{e}_1, \ldots, \tilde{e}_{n+1}$ form another orthonormal coordinate frame of $\mathbb{R}^{n+1}$ and if we set

$$\tilde{\nu}^i = \nu \cdot \tilde{e}_i, \qquad \tilde{\delta}_i = \tilde{e}_i \cdot \delta \qquad \text{for } 1 \leq i \leq n+1,$$

then

$$\nu^i = (e_i \cdot \tilde{e}_j)\tilde{\nu}^j, \qquad \delta_i = (e_i \cdot \tilde{e}_j)\tilde{\delta}_j, \qquad \delta_i \nu^j = (e_i \cdot \tilde{e}_k)(e_j \cdot \tilde{e}_l)\tilde{\delta}_k \tilde{\nu}^l$$

and

$$G^{ij}(-[\delta_k \nu^l]) = G^{ms}([-\tilde{\delta}_k \tilde{\nu}^l])(e_i \cdot \tilde{e}_m)(e_j \cdot \tilde{e}_s),$$

where

$$G^{ij}(A) = \frac{\partial G}{\partial A_{ij}}(A), \qquad A = [A_{ij}].$$

In particular, if we choose at a fixed point on the graph S, $\tilde{e}_{n+1} = \nu$ (which implies $\tilde{\delta}_{n+1} = 0$) and $\tilde{e}_1, \ldots, \tilde{e}_n$ such that the matrix $[\tilde{\delta}_k \tilde{\nu}^l] = [-\kappa_1, \ldots, -\kappa_n, 0]$ is diagonal, then as verified in [**5**],

$$G^{ij}([\tilde{\delta}_k \tilde{\nu}^l]) = f_i \delta_{ij}$$

(here δ_{ij} is the standard Kronecker symbol and is not related to the tangential operators δ_i, δ_j). We also note that $\tilde{\nu}^i = 0$ for $1 \leq i \leq n$, while $\tilde{\nu}^{n+1} = 1$, and $\nu^i = e_i \cdot \tilde{e}_{n+1}$ for $1 \leq i \leq n+1$.

From this, it is not difficult to derive the following formulas:

PROPOSITION 3.2.

$$\begin{aligned} G^{ij}\nu^j\delta_i &= 0, \\ G^{ij}(\delta_{ij} - \nu^i\nu^j) &= \sum f_i, \\ G^{ij}\delta_i\nu^j &= -\sum \kappa_i f_i, \\ G^{ij}\delta_i\nu^k\delta_j\nu^k &= \sum \kappa_i^2 f_i. \end{aligned}$$

We now can define the elliptic operator on S naturally associated to $f(\kappa)$ by

$$L = G^{ij}\delta_i\delta_j.$$

By the commutator formula (3.1) and the first formula in Proposition 3.2, we have

$$(G^{ij}\delta_i\delta_j - G^{ij}\delta_j\delta_i) = G^{ij}(\nu^i\delta_j\nu - \nu^j\delta_i\nu) \cdot \delta = 0.$$

REMARK 3.3. Set $\bar{e}_i = e_i - \nu^i \nu$, the projection of e_i on the tangent space to S at a point $p \in S$. Then $\delta_i = \nabla_{\bar{e}_i}$ is the covariant derivative with respect to the tangent vector $\bar{e}_i$ at p for $i = 1, \ldots, n$, and $\delta_{n+1} = 0$. Note also that $\delta_i \nu^j = 0$ and $G^{ij} = 0$ if i or j is $n+1$. In particular, if at $p \in S$ we choose $e_1, \ldots, e_n$ to be an orthonormal basis of the tangent space and $e_{n+1} = \nu$, then

$$L = \sum_{i,j=1}^{n} G^{ij} \nabla_i \nabla_j$$

where $G = G(b_{ij})$, $G^{ij} = \partial / \partial b_{ij}$, and b_{ij} is the second fundamental form of S.

We are now ready for the main result of the section.

THEOREM 3.4 (Fundamental identities). *For any hypersurface S with position vector X and unit normal ν, we have:*

(1) $LX = (\sum_{i=1}^{n} \kappa_i f_i)\, \nu$;

(2) $L\nu + \sum \kappa_i^2 f_i \nu = -\delta f$;

(3) $LS_1 + \sum k_i^2 f_i S_1 = \sum \kappa_i f_i \, |A|^2 - G^{ij,rs}(\delta_k \delta_r \nu^s)(\delta_k \delta_i \nu^j) + \Delta f$.

PROOF. We compute

$$LX = G^{ij} \delta_i (\delta_j k - \nu^j \nu^k) e_k = -(G^{ij} \delta_i \nu^j)\nu - (G^{ij} \nu^j \delta_i)\nu = \sum \kappa_i f_i,$$

proving the first equality.

To prove the second, we differentiate (3.2) on S with respect to δ_k and use the commutator formula and Proposition 3.2 to obtain

$$\begin{aligned} -\delta_k f = G^{ij} \delta_k \delta_i \nu^j &= G^{ij} \delta_i \delta_k \nu^j + G^{ij}(\nu^k \delta_i \nu^r - \nu^i \delta_k \nu^r) \delta_r \nu^j \\ &= G^{ij} \delta_i \delta_j \nu^k + G^{ij} \delta_i \nu^r \delta_j \nu^r \nu^k. \end{aligned}$$

To prove equality (3) of the theorem's statement we differentiate equality (2) with respect to δ_k and use the commutator formula several times:

$$-\Delta f = G^{ij} \delta_k \delta_i \delta_j \nu^k - G^{ij,rs}(\delta_k \delta_r \nu^s)(\delta_k \delta_i \nu^j) - S_1 \sum \kappa_i^2 f_i. \tag{3.3}$$

Now,

$$\begin{aligned} \delta_k \delta_i (\delta_j \nu^k) &= \delta_i \delta_k \delta_j \nu^k + (\nu^k \delta_i \nu^r - \nu^i \delta_k \nu^r)(\delta_r \delta_j \nu^k), \\ \delta_k \delta_j \nu^k &= \delta_j \delta_k \nu^k + (\nu^k \delta_j \nu^r - \nu^j \delta_k \nu^r) \delta_r \nu^k = \delta_j \delta_k \nu^k - \nu^j \delta_k \nu^r \delta_r \nu^k, \\ \delta_i \delta_k \delta_j \nu^k &= -\delta_i \delta_j S_1 - \delta_i \{\nu^j \delta_k \nu^r \delta_r \nu^k\}. \end{aligned}$$

Using $G^{ij} \nu^i \delta_j = 0$ and $\nu^k \delta_k = 0$ and Remark 3.3, we obtain

$$\begin{aligned} G^{ij} \delta_k \delta_i \delta_j \nu^k &= G^{ij} \delta_i \delta_k \delta_j \nu^k + S_1 \sum {\kappa_i}^2 f_i, \\ G^{ij} \delta_i \delta_k \delta_j \nu^k &= -LS_1 + \sum \kappa_i f_i \, |A|^2. \end{aligned}$$

Combining these with (3.3) proves part (3) of Theorem 3.4. □

COROLLARY 3.5. *Let S be a graph with respect to the e_{n+1} direction (so that $\nu^{n+1} > 0$) satisfying $f(\kappa) = c > 0$. Then $h = S_1 / \nu^{n+1}$ satisfies*

$$Lh + 2G^{ij} \delta_i (\log \nu^{n+1}) \delta_j h \geq 0.$$

In particular, h achieves its maximum on the boundary.

The proof is an elementary computation using Theorem 3.4, which we leave to the reader.

COROLLARY 3.6 (Minkowski integral formulas). *Let S be a compact embedded hypersurface and let $f = H_r(\kappa)$ with associated linearized operator $L = L_r$. Then $L_r(\frac{1}{2}|X|^2) = r(H_r(\kappa)X \cdot \nu + H_{r-1}(\kappa))$. In particular,*

$$\int_S (H_r(\kappa)X \cdot \nu + H_{r-1}(\kappa))\, dA = 0. \tag{3.4}$$

PROOF. Note that $L(|X|^2) = 2(rf(\kappa)X \cdot \nu + \sum f_i) = 2r(H_r(\kappa)X \cdot \nu + H_{r-1}(\kappa)$ by Lemma 2.15. According to Corollary 1.7, L_r is divergence free so formula (3.4) follows by integration (no ellipticity of f is needed or assumed). □

3.1. The method of moving spheres for elliptic Weingarten surfaces. As an application of the preceding calculus, we will outline some recent work [**10**] which develops for Weingarten hypersurfaces "the method of moving spheres", a variant of Alexandrov's famous method of moving planes, where reflection in a family of planes is replaced by inversion in a family of spheres. The method of moving spheres in the geometric setting was discovered by McCuan [**29, 28**] for surfaces of constant mean curvature. Here we will extend McCuan's work to a large class of elliptic Weingarten surfaces S, defined by the relation $f(\kappa) = c$. The fact that this is remotely possible is surprising since the inverted surface no longer satisfies a nice equation. Nevertheless, we shall see that because of Theorem 3.4, enough structure is preserved.

As discussed previously, the function f is a smooth positive symmetric function, positive homogeneous of degree one, defined in an open convex symmetric cone $\Gamma \subset \mathbb{R}^n$, with vertex at the origin, and containing the positive cone $\Gamma^+ \equiv \{\kappa \in \mathbb{R}^n : \text{all } \kappa_i > 0\}$. The hypersurface S is assumed to be "elliptic", that is,

$$f_i \equiv \frac{\partial f}{\partial \kappa_i} > 0 \quad \text{in } \Gamma \qquad \text{for } 1 \le i \le n,$$

and we assume that f is concave in Γ.

Without loss of generality, we will assume that f is normalized by the condition

$$f(1, \dots, 1) = 1.$$

We will also need the technical assumption

$$\sum {\kappa_i}^2 f_i \ge f(\kappa)^2. \tag{3.5}$$

We next show that the curvature quotients $f(\kappa) = H_{r,s}^{1/(r-s)}$, as discussed in Section 2, satisfy this technical condition. We first treat the special case $f = H_r^{1/r}$ of higher order mean curvatures normalized to be of homogeneity one.

LEMMA 3.7. *Let P be homogeneous, hyperbolic with respect to $a = (1, \dots, 1)$, of degree m, with positive coefficients, normalized by $P(a) = 1$. Then $f = P^{1/m}$ satisfies (3.5).*

PROOF. Using that f is homogeneous of degree 1 and concave,

$$f(\mu) \le f(\kappa) + \sum (\mu_i - \kappa_i) f_i(\kappa) = \sum \mu_i f_i(\kappa).$$

Choosing $\mu = \kappa^2 = ({\kappa_1}^2, \dots, {\kappa_n}^2) \in \Gamma$ proves that $\sum \kappa_i^2 f_i(\kappa) \geq f(\kappa^2)$. It remains to prove the inequality

$$f(\kappa^2) \geq f(\kappa)^2.$$

Applying Schwarz's inequality term by term to P gives

$$(P(\kappa))^2 \leq P(\kappa^2)P(a) = P(\kappa^2).$$

Taking m-th roots gives (3.5). □

The argument for the curvature quotients uses some results from Section 2. We first show that it suffices to show that $H_{r,r-1}^{1/(r-1)}$ satisfies (3.5).

LEMMA 3.8. *Let $f_1(\kappa), \dots, f_N(\kappa)$ be admissible curvature functions all satisfying (3.5). Then $f = (\prod_{k=1}^N f_k)^{1/N}$ is admissible and also satisfies (3.5).*

PROOF. We have already shown in Lemma 2.12 that f is concave so we need only demonstrate (3.5). By a routine calculation and the arithmetic geometric mean inequality

$$\sum_{i=1}^n {\kappa_i}^2 \partial_i f = \frac{f}{N} \sum_i \sum_k \frac{{\kappa_i}^2 \partial_i f_k}{f_k} \geq \frac{f}{N} \sum_k f_k \geq f(\prod f_k)^{1/N} = f^2. \qquad \square$$

PROPOSITION 3.9. *$H_{r,r-1}^{1/(r-1)}$ satisfies (3.5).*

PROOF. Using formula (2.5) of Lemma 2.14 we find

$$\sum {\kappa_i}^2 (S_{r-1} \partial_i S_r - S_r \partial_i S_{r-1}) = r S_r^2 - (r+1) S_{r-1} S_{r+1}. \tag{3.6}$$

Using the equality

$$S_r = \binom{n}{r} H_r, \tag{3.7}$$

the Newton inequalities (Proposition 2.7) may be rewritten as

$$(r+1) S_{r-1} S_{r+1} \leq r \frac{n-r}{n-r+1} S_r^2.$$

Using this latter equation in (3.7) we find

$$\sum {\kappa_i}^2 \partial_i \Big(\frac{S_r}{S_{r-1}} \Big) \geq \frac{r}{n-r+1} \Big(\frac{S_r}{S_{r-1}} \Big)^2.$$

Again using (3.7), we see that this is equivalent to

$$\sum {\kappa_i}^2 \partial_i H_{r,r-1} \geq H_{r,r-1}^2, \tag{3.8}$$

and the proof is complete. □

An important step in the method of moving spheres is the following maximum principle.

PROPOSITION 3.10. *Let M be an elliptic Weingarten hypersurface with position vector X and unit normal ν oriented so that $f(\kappa) = c > 0$ and set $h = (|X|^2 + (2/c) X \cdot \nu)$. Then*

$$Lh = 2 \Big(\sum f_i - 1 \Big) - \frac{2}{c} \Big(\sum \kappa_i^2 f_i - c^2 \Big) X \cdot \nu. \tag{3.9}$$

In particular, $Lh \geq 0$ in $\{X \cdot \nu \leq 0\}$.

PROOF. By Theorem 3.4,

$$L(|X|^2) = 2c\, X \cdot \nu + 2\sum f_i \quad \text{and} \quad L(X \cdot \nu) = -c - \sum \kappa_i{}^2 f_i\, X \cdot \nu.$$

Multiplying the second equation by $2/c$ and adding gives (3.9). □

To fix the ideas, we will sketch a proof, using spherical reflection, of Alexandrov's theorem that an embedded closed Weingarten hypersurface in $\mathbb{R}^{n+1}$ is a sphere. For complete details and other applications, see [**10**]. Let M be a closed embedded Weingarten hypersurface with position vector X and let S_ρ of radius ρ and center at the origin (assume the origin lies in the unbounded component of the complement of S). For ρ large, $M \subset S_\rho$ and there is a first value ρ_0 where M is tangent to S_ρ. We decrease ρ and cut off a cap M_ρ and set $\tilde{M}_\rho = I(M_\rho)$.

LEMMA 3.11. [**7**] *The directions of principal curvatures of M_ρ map into the directions of principal curvature for $\tilde{M}_\rho$ and if $\tilde{\kappa}$ denotes the corresponding set of principal curvatures to κ,*

$$\tilde{\kappa} = \frac{1}{\rho^2}\big(|X|^2\kappa + 2X \cdot \nu \vec{1}\big). \tag{3.10}$$

For $\rho < \rho_0$ but close to ρ_0 evidently $X \cdot \nu < 0$ and $\tilde{M}_\rho$ is contained inside M. We then let ρ_1 be the infimum of the values of ρ such that this property of $\tilde{M}_\rho$ holds. Just as in the standard case of Alexandrov reflection, we have to consider the possibilities that $\tilde{M}_{\rho_1}$ is tangent to M at an interior point (where $|X| > \rho_1$ and $X \cdot \nu < 0$) or at a boundary point (where $|X| = \rho_1$ and $X \cdot \nu = 0$). In either case, we want to show that $M \cap \{|X| < \rho_1\} = \tilde{M}_{\rho_1}$.

The following lemma will enable us to compare $f(\tilde{\kappa})$ and $f(\kappa)$ and prove a maximum principle.

LEMMA 3.12. *Suppose $\kappa \in \Gamma, \tilde{\kappa} \in \Gamma$. Then*

$$f(\tilde{\kappa}) \le \frac{|X|^2}{\rho^2} f(\kappa) + \frac{2X \cdot \nu}{\rho^2}. \tag{3.11}$$

PROOF. By the concavity and homogeneity of f,

$$\frac{|X|^2}{2\rho^2} f(\kappa) = f\left(\frac{1}{2}\tilde{\kappa} + \frac{1}{2}\Big(\frac{-2X \cdot \nu}{\rho^2}\Big)\vec{1}\right) \ge \frac{1}{2} f(\tilde{\kappa}) + \frac{1}{2}\Big(\frac{-2X \cdot \nu}{\rho^2}\Big),$$

which is equivalent to (3.11). □

COROLLARY 3.13. *Let M satisfy $f(\kappa) = c$.*

(i) *Suppose M^- is a component of $M \cap \{|X| > \rho\}$ on which $X \cdot \nu < 0$ and set $\tilde{M} = I(M^-)$. Then $f(\tilde{\kappa}$ lies outside $T = \{\lambda \in \Gamma : f(\lambda) \ge c\}$ unless M is a sphere of radius $1/c$.*

(ii) *On $\partial\tilde{M}$, $f(\tilde{\kappa}) \le c$.*

PROOF. From Lemma 3.12 and formula (3.11), if $\tilde{\kappa} \in T$ we have

$$f(\tilde{\kappa}) \le \frac{c}{\rho^2}\Big(|X|^2 + \frac{2}{c} X \cdot \nu\Big) = \frac{c}{\rho^2} h.$$

By Proposition 3.10, h achieves its maximum on $S \cap \{|X| = \rho\}$ so $f(\tilde{\kappa}) \le c$. If the inequality is strict on the interior of $\tilde{M}$, we have a contradiction. Otherwise, h has an interior maximum and so $h \equiv \rho^2$. This implies that all the principle

curvatures of M^- have the value c and so M^- is a sphere of radius $1/c$. By unique continuation, this holds for M and so the first part of the lemma holds. The second part of the lemma is now clear by the preceding argument. □

Now suppose we are in the case of internal tangency. Then by a standard argument $\tilde{\kappa} \geq \kappa$ hence $f(\tilde{\kappa}) \geq c$. But $X \cdot \nu < 0$ so we are in the equality case of Corollary 3.13 so M is a sphere. In the case of boundary tangency, we have $|X| = \rho$ and $X \cdot \nu = 0$ at the point of tangency of M and $\tilde{M}_{\rho_1}$. In particular, $\tilde{\kappa} = \kappa$ by formula (3.10) so $\tilde{\kappa} \in \Gamma$ and so $f(\tilde{\kappa}) \leq c$ by Corollary 3.13. We can write M and $\tilde{M}_{\rho_1}$ locally as graphs over their common tangent plane at the point of tangency with $\tilde{M}_{\rho_1}$ lying above M. Now we can apply the Hopf boundary point lemma to conclude that M is invariant under spherical reflection.

In all cases we have shown that for arbitrary center, M is invariant by reflection in some sphere with that center. By moving the center to infinity along a direction $\vec{n}$ we conclude in the limit that M is invariant by a hyperplane with normal $\vec{n}$ and so M is a sphere. For an interesting discussion of spherical symmetries see [**28**, Section 3].

4. Monge–Ampère Boundary Value Problems and Applications

In this section we discuss boundary value problems in $\mathbb{R}^n$ and S^n for Monge–Ampère equations and discuss applications to existence questions for K-hypersurfaces.

Let $\Omega \subset \mathbb{R}^n$ be a smooth domain and consider the Monge–Ampère equation

$$\begin{aligned} \det(u_{ij}) &= \psi(x, u, \nabla u) && \text{in } \Omega, \\ u &= \phi && \text{on } \partial\Omega, \end{aligned} \tag{4.1}$$

where $\psi > 0$ is smooth, $\phi \in C^\infty(\partial\Omega)$.

REMARK 4.1. The choice $\psi = K(1 + |\nabla u|^2)^{(n+2)/2}$ describes a graph $x_{n+1} = u(x)$ of constant Gauss curvature.

The classical PDE existence theorem is the following:

THEOREM 4.2. [**3, 25**] *Suppose Ω is strictly convex and Ω, ψ, ϕ are smooth. Assume also that for the boundary data ϕ there is a strictly convex subsolution $\underline{u}$; that is,*

$$\begin{aligned} \det(\underline{u}_{ij}) &\geq \psi(x, \underline{u}, \nabla \underline{u}) && \text{in } \Omega, \\ \underline{u} &= \phi && \text{on } \partial\Omega. \end{aligned} \tag{4.2}$$

Then there exists a strictly convex solution $u \in C^\infty(\bar{\Omega})$ a solution to (4.1). If $\psi_u \geq 0$ the solution is unique.

From the point of view of PDE, this result is essentially optimal, but is it geometrically useful? The following examples show that the answer is not really.

EXAMPLE 4.3. Let Γ_1, Γ_0 be strictly convex smooth, closed codimension 2 hypersurfaces in parallel planes, say the planes $x_{n+1} = 1, 0$ respectively. We ask if there is a K-hypersurface solution for K sufficiently small. Intuitively, the answer is clearly yes. Let's specialize further and suppose that the parallel projection of Γ_1, call this projection γ_1, contains Γ_0. It is then not difficult to see that if a solution exists, it must be a graph over the annulus Ω with outer boundary γ_1 and inner

boundary Γ_0 and so satisfies (4.1) with $\phi = 1$ on γ_1 and $\phi = 0$ on Γ_0. However since Ω is not convex, Theorem 4.2 does not apply! However, as shown in [**22**] there is a unique smooth solution as expected.

EXAMPLE 4.4. Let S be an ovaloid in $\mathbb{R}^{n+1}$, that is the boundary of a strictly convex body, and let D be a smooth domain on S with $\partial D = \Gamma = (\Gamma_1, \dots, \Gamma_m)$. We think of D as a strictly convex hypersurface with boundary Γ and ask if we can deform D to a K_0-hypersurface for $0 < K_0 \leq \inf_{P \in S} K(P)$. If yes, we expect that the solution should be a radial graph $X(x) = \rho(x)x$, where x ranges over a domain Ω contained in S^n obtained by projecting D radially (we choose an origin inside S) onto S^n. Moreover, $\rho(x) = \phi(x) > 0$ on $\partial\Omega$, where Γ is the radial graph of ϕ.

The Gauss curvature of X is given by (see [**18**])

$$K[X(x)] = \frac{\det(b_{ij})}{\det(g_{ij})} = \frac{\det(\rho^2\sigma_{ij} + 2\nabla_i\rho\nabla_j\rho - \rho\nabla_{ij}\rho)}{\sigma\,\rho^{2n-2}\,(\rho^2 + |\nabla\rho|^2)^{(n+2)/2}}, \tag{4.3}$$

where σ_{ij} is the standard metric on S^n and $\sigma = \det\sigma_{ij}$.

This expression for the Gauss curvature simplifies considerably if we instead consider the "dual" radial graph

$$u = \frac{1}{\rho}, \qquad \varphi = \frac{1}{\phi}.$$

Then u is a solution to the Monge–Ampère type boundary value problem

$$\begin{aligned} \sigma^{-1}\det(\nabla_{ij}u + u\sigma_{ij}) &= K_0\frac{(u^2 + |\nabla u|^2)^{(n+2)/2}}{u^{n+2}} && \text{in } \Omega, \\ u &= \varphi && \text{on } \partial\Omega. \end{aligned} \tag{4.4}$$

Again the point is that we cannot control the geometry of Ω so we must allow arbitrary geometry. Of course there are obstructions and we want to remove them by assuming the existence of an admissible subsolution. We will present two results from [**18**] for graphs over $\mathbb{R}^n$ and S^n respectively.

THEOREM 4.5. [**18, 11**] *Suppose Ω, ϕ, ψ are smooth and assume there is a locally strictly convex subsolution $\underline{u} \in C^\infty(\bar{\Omega})$, i.e.,*

$$\begin{aligned} \det\underline{u}_{ij} &\geq \psi(x, \underline{u}, \nabla\underline{u}) && \text{in } \Omega, \\ \underline{u} &= \phi && \text{on } \partial\Omega. \end{aligned}$$

Then there exists $u \in C^\infty(\bar{\Omega})$ a solution to (4.1). (If $\psi_u \geq 0$, there is uniqueness.) Moreover, any admissible solution satisfies the estimate $|u|_{C^{2+\alpha}(\Omega)} \leq C$ for a controlled constant C.

THEOREM 4.6. [**18, 11**] *Let $\Omega \subset S^n$ be a smooth domain that does not contain any hemisphere. Assume there is a locally strictly convex (i.e., $\nabla_{ij}\underline{u} + \underline{u}\,\sigma_{ij} > 0$) subsolution $\underline{u}$ to (4.4). Then there exists $u \in C^\infty(\bar{\Omega})$ a solution to (4.4). Moreover, any admissible solution satisfies the estimate $|u|_{C^{2+\alpha}(\Omega)} \leq C$ for a controlled constant C.*

COROLLARY 4.7. *Example 4.4 has a smooth solution inside S as conjectured.*

PROOF. Choose an origin inside the convex hull of $S \setminus D$ and radially project D onto a subdomain Ω of S^n as described earlier. Then Ω does not contain any hemisphere and so Theorem 4.6 applies. □

COROLLARY 4.8 (Polyhedral version of Example 4.4). *Let P be a convex polyhedron in $\mathbb{R}^{n+1}$ and let $\Gamma = (\Gamma_1, \ldots, \Gamma_m)$ be a collection of strictly convex closed hypersurfaces, each one contained interior to a face of P (with at most one in any face). Then, for $K_0 > 0$ sufficiently small, Γ bounds a smooth embedded K_0 hypersurface which can be represented as a radial graph.*

Rob Kusner observed that Theorem 4.6 holds in greater generality and the proof is the same. Let U be a domain and let $x : U \to S^n$ be an immersion. Let ρ be a positive function on U and define $X(p) = \rho(p)x(p)$, $p \in U$. Consider the problem of finding a strictly locally convex immersed K-hypersurface:

(4.5)
$$X : U \to \mathbb{R}^{n+1}, \qquad X(p) = \rho(p)x(p) \quad \text{for } p \in U, \qquad \rho(p) = \phi(p) \quad \text{for } p \in \partial U.$$

THEOREM 4.9. [**18**] *Suppose that no hemisphere can be isometrically embedded in U and assume there is a smooth immersed locally strictly convex subsolution $\underline{X} : U \to \mathbb{R}^{n+1}$ satisfying*

$$\underline{X}(p) = \underline{\rho}(p)x(p) \quad \textit{and} \quad K(\underline{X}(p)) \geq K \quad \textit{for } p \in U, \qquad \underline{\rho} = \phi \qquad \textit{on } \partial U.$$

Then there exists a smooth immersed K-hypersurface X satisfying (4.5) *and $\rho \leq \underline{\rho}$.*

Here is a more concrete form of Theorem 4.9:

COROLLARY 4.10. [**18**] *Let $\Gamma = (\Gamma_1, \ldots, \Gamma_m)$ and suppose there exists an immersed strictly locally convex hypersurface $\underline{X}$ satisfying $\underline{X} \cdot \nu < 0$ (for suitable choice of origin) everywhere and such that no subdomain of $\underline{X}$ is radially projected injectively onto a hemisphere of S^n. Then for K_0 small enough there is an immersed K_0-immersed hypersurface spanning Γ.*

Using Corollary 4.10 we can construct K-hypersurfaces of higher genus following a construction suggested in [**22**].

COROLLARY 4.11. [**18**] *For each positive integer k, there exists an embedded K-hypersurface of genus k.*

Before giving some idea of the proof of Theorem 4.5 (the proof of Theorem 4.6 is similar in spirit) we sketch its application [**34**] to the existence of complete embedded K-hypersurfaces in H^{n+1} with prescribed asymptotic boundary.

4.1. Complete K-hypersurfaces in hyperbolic space. We use the half-space model

$$H^{n+1} = \{(x, x_{n+1}) = (x_1, \ldots, x_{n+1}) : x_{n+1} \geq 0\} \quad \text{with metric } ds^2 = \frac{dx^2}{x_{n+1}^2}.$$

Suppose that we want to find a complete K-hypersurface with asymptotic boundary $\Gamma = \partial\Omega \subset \{x_{n+1} = 0\}$. In order to utilize Theorem 4.5 we vertically translate the domain Ω to the unit horosphere $P_1 = \{x_{n+1} = 1\}$ and look for a vertical graph $y = \log x_{n+1} = f(x), x \in \Omega$. We then write down the first and second fundamental forms of the graph in order to express the extrinsic Gauss curvature $K + 1$ as an expression in the second derivatives of f of Monge–Ampère type. The computation is straightforward but tedious and may be found in [**34**]. It turns out (just as in the proof of Theorem 4.6) that the equations simplify enormously if we use a new

variable $u = e^{2f} = x_{n+1}^2$. We arrive at the following (degenerate) boundary value problem:

$$\begin{aligned} \det(u_{ij} + 2\delta_{ij}) &= 2^n(K+1)\Big(1 + \frac{|\nabla u|^2}{4u}\Big)^{(n+2)/2} && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{4.6}$$

Although it is not essential, observe that the first of these equations is the classical Monge–Ampère equation for the dependent variable $u + |x|^2$ and so Theorem 4.5 would apply except that the boundary $u = 0$ makes the right-hand side possibly unbounded. Thus it is natural to approximate (4.6) the desired K-hypersurface by a K-hypersurface with boundary Γ translated up to a horosphere $P_c = \{x_{n+1} = c\}$; that is, we want to solve

$$\begin{aligned} \det(u_{ij} + 2\delta_{ij}) &= 2^n(K+1)\Big(1 + \frac{|\nabla u|^2}{4u}\Big)^{(n+2)/2} && \text{in } \Omega, \\ u &= c^2 && \text{on } \partial\Omega. \end{aligned} \tag{4.7}$$

To apply Theorem 4.5 we need to find a subsolution to (4.7). Since the horospheres are flat, $\underline{u} \equiv c^2$ is a strictly convex subsolution for $K \in (-1, 0)$ (since $K + 1 > 0$). Thus we have shown:

PROPOSITION 4.12. *For $K \in (-1, 0)$ there exists a smooth admissible solution $u_c \in C^\infty(\bar{\Omega})$ of (4.7).*

One must now analyze the behavior of the family u_c as c tends to zero. The first step is to show $|\nabla u|^2/4u$ is uniformly bounded independently of c. The final result is:

THEOREM 4.13. **[34]** *For $K \in (-1, 0)$ there is a smooth admissible solution $u \in C^\infty(\Omega) \cap C^{1,1}(\bar{\Omega})$ to (4.6). Hence there is a complete K-hypersurface S with asymptotic boundary $\Gamma = \partial\Omega$ which is a graph $y = \ln x_{n+1} = \frac{1}{2}\ln u$. Moreover, if $n = 3$ and Γ is a Jordan curve, then S is unique among all possible immersed locally convex solutions.*

The uniqueness statement for $n = 3$ arises because of the curious fact that the Jacobi operator $\tilde{L} = L + HK$ just as in $\mathbb{R}^{n+1}$ (see Theorem 3.4) which is elliptic since $H > 0$ and $K < 0$. This allows us to deform Γ from a small circle inside to a large circle outside and obtain a smooth family of solutions. The solution for a circle is unique (an equidistant sphere) and our family foliates the region of H^3 between the large sphere and the small sphere. Now suppose there was another K-hypersurface M with asymptotic boundary Γ. Because Γ is Jordan, it is not difficult to see that in fact M is embedded. Therefore the small equidistant sphere may be chosen inside M and the large equidistant sphere may be chosen to lie outside M. A simple maximum principle argument shows that S must be both inside M and outside M and so must equal M.

Theorem 4.13 has a more compelling interpretation in the ball model of H^3. Take Γ on the sphere at infinity and let C be the hyperbolic convex hull of Γ. Then for $K \in (-1, 0)$ there is a unique embedded K-surface with asymptotic boundary Γ in each component of the complement of C. As K varies between zero and negative one, these solutions foliate each component going from the sphere to ∂C. Since ∂C is a hyperbolic surface in the sense of Thurston, our solutions provide smooth approximations with constant negative curvature.

4.2. A priori estimates. We now sketch the main ideas in the proof of Theorem 4.5. It is convenient to rewrite our equation (4.1) as

$$\log\det u_{ij} = \log\psi(x,u,\nabla u) := f(x,u,\nabla u), \tag{4.8}$$

noting that the operator on the left-hand side is concave. We also define

$$L = u^{ij}\partial_i\partial_j - f_{p_i}\partial_i,$$

which is the essential part of the full linearization of (4.8); here (u^{ij}) is the inverse matrix to the strictly positive matrix (u_{ij}).

Now recall that we assume the existence of a locally strictly convex subsolution $\underline{u}$ satisfying (4.2).

DEFINITION 4.14.

$$A = \{w \in C^\infty(\bar\Omega) : (w_{ij}) > 0,\ w \geq \underline{u},\ w = \underline{u} = \phi \text{ on } \partial\Omega\}.$$

Hence, by the maximum principle, $\underline{u} \leq w \leq h$ in $\bar\Omega$, where h is the harmonic extension of ϕ. It follows that we have the C^1 estimate

$$|w| + |\nabla w| \leq C \quad \text{in } \bar\Omega$$

for a controlled constant C.

THEOREM 4.15. *Let $u \in A$ be a solution to* (4.1). *Then $|D^2u| \leq C$ in $\bar\Omega$ for a controlled constant C.*

REMARK 4.16. Using the Evans–Krylov interior regularity theory [**14**] and the boundary regularity results of [**3, 4, 25**], one can deduce from Theorem 4.15 a $C^{2+\alpha}(\bar\Omega)$ estimate for u.

The essential step in the proof of Theorem 4.15 is the estimate on $\partial\Omega$:

PROPOSITION 4.17. *Let $u \in A$ be a solution to* (4.1). *Then $|D^2u| \leq C$ on $\partial\Omega$ for a controlled constant C.*

PROOF. *Step 1.* Choose the origin of coordinates to be a point on $\partial\Omega$ (at which we will derive our estimates) with the x_n axis the interior normal direction and $x' = (x_1, \ldots, x_{n-1})$ tangential. We locally write $\partial\Omega$ as a graph:

$$x_n = \rho(x') = \frac{1}{2}\sum_{\alpha,\beta<n} B_{\alpha\beta}x_\alpha x_\beta + O(|x'|^3).$$

Since $u - \underline{u}(x', \rho(x')) \equiv 0$, we have

$$(u-\underline{u})_{\alpha\beta}(0) = -(u-\underline{u})_n(0)B_{\alpha\beta};$$

hence $|u_{\alpha\beta}(0)| \leq C$.

Step 2. We next show that $|u_{\alpha n}(0)| \leq C$. To this end, introduce the approximate tangential derivative

$$T_\alpha = \partial_\alpha + \sum_{\beta<n} B_{\alpha\beta}(x_\beta\partial_n - x_n\partial_\beta),$$

which differs from the exact tangential derivative $\partial_\alpha + \rho_\alpha\partial_n$ by $O(|x|^2)$. It is convenient to use $T = T_\alpha$ because it satisfies

$$L(Tu) = O(1),$$

as can be easily checked using the invariance of the left-hand side of (4.8) under rotations of coordinates. From this relation it follows that

$$\big|L(T(u-\underline{u}))\big| \le C\Big(1+\sum u^{ii}\Big), \qquad \big|T(u-\underline{u})\big| = O(|x|^2).$$

The following lemma from [**16**] constructs a comparison function v in terms of $u, \underline{u}, h$ and the distance $d(x,\partial\Omega)$ in a small neighborhood $\Omega_\delta = \{x \in \Omega : d(x,\partial\Omega) < \delta\}$ of $\partial\Omega$. The construction would be much simplified if we were to assume that $\det \underline{u}_{ij} \ge \psi(x,\underline{u},\nabla\underline{u}) + \epsilon$ as in [**18**].

LEMMA 4.18. *Let $v = (u-\underline{u}) + t(h-\underline{u}) - Nd(x,\partial\Omega)^2$ in Ω_δ. Then for t,δ sufficiently small and N sufficiently large,*

$$Lv \le -\frac{\epsilon}{4}\Big(1+\sum u^{ii}\Big)$$

and

$$v \ge 0 \quad \text{on } \partial\Omega_\delta, \qquad v = 0 \quad \text{on } \partial\Omega.$$

Now we can choose $A \gg B \gg 1$ such that

$$L\big(Av + B|x|^2 \pm T(u-\underline{u})\big) \le 0$$

and

$$Av + B|x|^2 \pm T(u-\underline{u}) \ge rhs\ 0 \qquad \text{on } \partial\Omega_\delta.$$

It follows from the maximum principle that

$$|u_{\alpha n}(0)| \le C.$$

Step 3. It remains to estimate $u_{nn}(0)$. Here we use the equation

$$A^{ij}u_{ij} = n \det u_{ij} = O(1)$$

(where A^{ij} is the cofactor matrix of (u_{ij})) to solve for u_{nn} in terms of the derivatives we have already estimated. We can do this if we know that A^{nn} is strictly positive, i.e., we need to know the strict tangential convexity of u. This strict tangential convexity in fact holds for all $w \in A$:

PROPOSITION 4.19. [**18**] *For some controlled constant c_0, we have*

$$\sum_{\alpha,\beta<n} w_{\alpha\beta}(0)\xi_\alpha\xi_\beta \ge c_0 > 0 \qquad \textit{for all } w \in A.$$

Of course, the point here is that no assumption is made on the geometry of $\partial\Omega$. For a proof, see [**18**].

REMARK 4.20. The case $\partial\Omega$ concave is actually the easiest, because

$$(u-\underline{u})_{\alpha\alpha} = -(u-\underline{u})_n B_{\alpha\alpha} \ge 0$$

(since $(u-\underline{u})_n > 0$ and $B_{\alpha\alpha} \le 0$).

This completes the sketch of the proof of Proposition 4.17. □

4.3. Degenerate Monge–Ampère equations and convex hulls. To get a better understanding of the general existence problem for K-hypersurfaces posed at the beginning of this article, it is of great interest to study totally degenerate Monge–Ampère equations. For a single smooth closed embedded submanifold Γ of $\mathbb{R}^{n+1}$ of codimension two, this corresponds to the geometric question of the existence of convex hypersurfaces $S^{\pm}$ with Gauss curvature $K(S^{\pm}) \equiv 0$. The hypersurfaces $S^{\pm}$ correspond to the boundaries of the convex hull $C(\Gamma)$, which is the convex region bounded by the two "convex caps" $S^{\pm}$ meeting along Γ. It is well-known that $S^{\pm}$ cannot be smooth, in general.

For Γ a graph over the boundary of a strictly convex domain Ω the lower cap S can be represented as a graph $x_{n+1} = u(x)$ where

$$u(x) = \max\{v(x) : v(x) \in C^{\circ}(\bar{\Omega}),\ v \text{ convex}, v \leq \phi \text{ on } \partial\Omega\}$$

(here Γ = graph ϕ over $\partial\Omega$).

The convex function u is a weak solution (in the Alexandrov sense or the viscosity sense) of the degenerate Monge–Ampère boundary value problem:

$$\begin{aligned} \det u_{ij} &\equiv 0 \qquad \text{in } \Omega, \\ u &= \phi \qquad \text{on } \partial\Omega. \end{aligned} \tag{4.9}$$

How regular is u? Many people have studied this question, including Rauch and Taylor [**30**], Trudinger and Urbas [**37**], and Caffarelli, Nirenberg and Spruck [**4**]. The optimal regularity was obtained in [**8**].

THEOREM 4.21. [**8**] *Assume Ω strictly convex with $\partial\Omega \in C^{3,1}$ and let $\phi \in C^{3,1}$. Then the unique admissible weak solution u of* (4.9) *is in $C^{1,1}(\bar{\Omega})$.*

The simple example $u = (1+y)^{2-\varepsilon}$ with $\varepsilon \in (0,1)$ on $B_1(0) \subset \mathbb{R}^2$ (taken from [**8**]) shows that this result is optimal.

An important generalization of this result was obtained by Guan [**16**]. Here, as in [**18**], we drop the assumption of strict convexity of Ω but assume that there is an admissible subsolution $\underline{u}$ in Ω for the given boundary data ϕ. More precisely:

THEOREM 4.22. [**16**] *Let Ω be a $C^{3,1}$ domain and $\phi \in C^{3,1}(\partial\Omega)$. Suppose there exists a locally strictly convex function $\underline{u} \in C^2(\bar{\Omega})$ with $u = \phi$ on $\partial\Omega$. Then there exists a unique locally convex weak solution of* (4.9) *in $C^{1,1}(\bar{\Omega})$.*

The extension of Guan's result to space curves (or submanifolds of codimension two) was obtained by Ghomi [**12, 13**].

DEFINITION 4.23. Suppose Γ is a smooth Jordan curve lying on an ovaloid O. Then:

(i) Γ is strictly convex, that is, through every point $x_0 \in \Gamma$, there exists a supporting plane H_{x_0} with $H_{x_0} \cap \Gamma = \{x_0\}$.
(ii) The curvature $k(\Gamma)$ never vanishes, that is, Γ has no inflection points.

We will call a Jordan curve Γ satisfying (i) and (ii) strictly convex. More generally, an m-dimensional closed embedded submanifold Γ of $\mathbb{R}^{n+1}$ is called strictly convex if

(i′) each point $x_0 \in \Gamma$ has a strict support plane H_{x_0,n_0} with $\Gamma\backslash\{x_0\}$ contained in one of the open half-spaces determined by H, say $\ell(x) = \langle x - x_0, n_0\rangle < 0$ for all $x \in \Gamma$ distinct from x_0 (where n_0 is the outer normal), and

(ii′) H_{x_0,n_0} is nonsingular, that is, x_0 is a nondegenerate critical point of ℓ. This last condition is equivalent to the condition that $\langle A_{n_0}X, X\rangle < 0$ for all nonzero $X \in T_{x_0}\Gamma$, where A_{n_0} is the second fundamental form of Γ at x_0 with respect to the normal direction n_0.

In his thesis, Ghomi proves the following:

THEOREM 4.24. [**12, 13**] *Every $C^{k,\alpha}$ strictly convex submanifold Γ lies on a $C^{k,\alpha}$ strictly convex ovaloid O, $\alpha \in [0,1]$.*

Now consider Γ, a strictly convex codimension-two submanifold in the sense of Ghomi, with $\Gamma \in C^{3,1}$. Then Γ lies on a $C^{2,1}$ ovaloid O and we let $D^{\pm}$ denote the two components of the complement of Γ on O. Choose one of these components, say D^- and choose an origin inside the convex hull of $O - D^-$ so that D^- radially projects onto a domain Ω contained in a hemisphere (say the upper hemisphere) of the unit sphere $S^n \subset \mathbb{R}^{n+1}$.

We look for S^-, the component of $\partial C(\Gamma)$ facing D^- as a radial graph over Ω:

$$\begin{aligned} X &= \rho(x)x \qquad && \text{for } x \in \Omega, \\ \rho &= \phi \qquad && \text{on } \partial\Omega, \end{aligned}$$

where $\Gamma = \phi(x)x$ for $x \in \partial\Omega$. By assumption, $\Omega, \phi \in C^{3,1}$. As explained earlier, it is simpler to work with $u = 1/\rho$; then the Gauss curvature $K(X)$ is related to u by

$$\begin{aligned} \frac{\det(\nabla_{ij}u + u\sigma_{ij})}{\sigma} &= \frac{K(u^2 + |\nabla u|^2)^{(n+2)/2}}{u^{n+2}} \qquad && \text{in } \Omega, \\ u &= \varphi = 1/\phi \qquad && \text{on } \partial\Omega. \end{aligned}$$

Here σ_{ij} denotes the metric on S^n, $\sigma = \det \sigma_{ij}$ and $\nabla_{ij}u$ is the Hessian of u with respect to the metric σ_{ij}.

Thus we are looking for an admissible ($\nabla_{ij}u + u\sigma_{ij} \geq 0$) weak solution of

$$\begin{aligned} \det(\nabla_{ij}u + u\sigma_{ij}) &\equiv 0 \quad \text{in } \Omega, \\ u &= \varphi \quad \text{on } \partial\Omega, \end{aligned} \tag{4.10}$$

where $\partial\Omega, \varphi \in C^{3,1}$.

It is convenient to approximate (4.10) with the nondegenerate problems

$$\begin{aligned} \frac{\det(\nabla_{ij}u + u\sigma_{ij})}{\sigma} &= \varepsilon\mu^{n+2} \quad \text{in } \Omega, \\ u &= \varphi \qquad \text{on } \partial\Omega, \end{aligned} \tag{4.11}$$

where μ is a fixed positive smooth function we will define in a moment. Since D^- is strictly convex there exists a strictly convex subsolution $u \in C^{3,1}(\bar{\Omega})$ with $u = \varphi$ on $\partial\Omega$ for $0 < \varepsilon \ll 1$. Applying the results of [**18**], we see that there exists an admissible solution of (4.11), $u^\varepsilon \in C^{3,\alpha}(\bar{\Omega})$ (we will see presently that u^ε is unique) with $|\nabla u^\varepsilon| \leq C$ independent of ε. Therefore u^ε converges uniformly to an admissible weak solution u of (4.10), also unique.

THEOREM 4.25. $u \in C^{1,1}(\bar{\Omega})$.

It seems difficult (but probably possible) to extend Guan's result to domains Ω of arbitrary geometry in S^n. However, we can proceed as follows. We parametrize

the upper hemisphere S^n_+ by choosing x to lie in the tangent plane to S^n at the north pole and setting

$$y = \frac{(x,1)}{\mu(x)} \in S^n, \qquad \mu(x) = \sqrt{1+x^2}.$$

Then

$$\begin{aligned}
\sigma_{ij} &= \langle y_{x_i}, y_{x_j} \rangle = \frac{1}{\mu^2}\Big(\delta_{ij} - \frac{x_i x_j}{\mu_2}\Big),\\
\sigma^{ij} &= (\sigma_{ij})^{-1} = \mu^2(\delta_{ij} + x_i x_j)\\
\sigma &= \det \sigma_{ij} = \mu^{-(2n+2)},\\
\Gamma^k_{ij} &= -\frac{1}{\mu^2}(x_i\delta_{kj} + x_j\delta_{ki}).
\end{aligned}$$

Set $\tilde{u}^\varepsilon(x) = \mu(x)u^\varepsilon(y(x))$, the degree-1 homogeneous extension of u restricted to the tangent plane. A calculation gives

$$\tilde{u}^\varepsilon_{x_i x_j} = \mu(\nabla_{ij}u^\varepsilon + u^\varepsilon\sigma_{ij}),$$

so

$$\det \tilde{u}^\varepsilon_{x_i x_j} = \sigma\mu^n \frac{\det(\nabla_{ij}u^\varepsilon + u^\varepsilon\sigma_{ij})}{\sigma} = \mu^{-(n+2)}\varepsilon\mu^{n+2} \equiv \varepsilon \qquad \text{in } \tilde{\Omega}$$

$$\tilde{u}^\varepsilon = \tilde{\varphi} \equiv \mu\varphi \quad \text{on } \partial\tilde{\Omega},$$

where $\tilde{\Omega}$ is the central projection of Ω from the origin onto the tangent plane at the north pole. Moreover, there exists a strictly convex subsolution $v = \tilde{U} = \mu U$ in $\tilde{\Omega}$ with the given boundary data $v = \tilde{\varphi} \equiv \mu\varphi$ on $\partial\tilde{\Omega}$ and the $\tilde{u}^\varepsilon$ converge uniformly to $\tilde{u} = \mu u$, a convex weak solution of

$$\begin{aligned}
\det \tilde{u}_{ij} &\equiv 0 \quad \text{in } \tilde{\Omega},\\
\tilde{u} &= \tilde{\varphi} \quad \text{on } \partial\tilde{\Omega},
\end{aligned}$$

where $\tilde{\Omega}, \tilde{\varphi} \in C^{3,1}$.

Applying Guan's results [**16**, Theorem 1.1] gives $\tilde{u} \in C^{1,1}(\bar{\tilde{\Omega}})$. Thus $u \in C^{1,1}(\bar{\Omega})$ as required.

COROLLARY 4.26. *S^- is $C^{1,1}$ up to Γ.*

References

[1] Andrews, B., Gauss curvature flow: the fate of the rolling stones. Invent. Math. 138 (1999), no. 1, 151–161.

[2] Andrews, B., Motion of hypersurfaces by Gauss curvature. Pacific J. Math. 195 (2000), no. 1, 1–34.

[3] Caffarelli, L., Nirenberg, L. and Spruck, J., The Dirichlet problem for nonlinear second-order elliptic equations. I. Monge–Ampère equation, Comm. Pure Appl. Math. 37 (1984), no. 3, 369–402.

[4] Caffarelli, L., Kohn, J. J., Nirenberg, L. and Spruck,J., The Dirichlet problem for nonlinear second-order elliptic equations. II. Complex Monge–Ampère, and uniformly elliptic, equations. Comm. Pure Appl. Math. 38 (1985), no. 2, 209–252.

[5] Caffarelli, L., Nirenberg, L. and Spruck, J., The Dirichlet problem for nonlinear second order elliptic equations III. Functions of the eigenvalues of the Hessian, Acta Math. 155 (1985), 261–301.

[6] Caffarelli, L., Nirenberg, L. and Spruck, J., Nonlinear second order elliptic equations IV. Starshaped compact Weingarten hypersurfaces, Current topics in partial differential equations, ed. by Y. Ohya, K. Kasahara, N. Shimakura, 1986, pp. 1–26, Kinokunize Co., Tokyo.

[7] Caffarelli, L., Nirenberg, L. and Spruck, J., Nonlinear second order elliptic equations V. The Dirichlet problem for Weingarten hypersurfaces, Comm. Pure Appl. Math. 41 (1988), 47–70.
[8] Caffarelli, L., Nirenberg, L. and Spruck, J., The Dirichlet problem for the degenerate Monge–Ampère equation, Rev. Mat. Iberoamericana, 2 (1986), no. 1–2, 19–27.
[9] Chang, S.-Y A., Gursky, M. and Yang, P., An equation of Monge–Ampère type in conformal geometry and four manifolds of positive Ricci curvature, Ann. of Math. (2) 155 (2002), no. 3, 709–787.
[10] Danielli, D., Garofalo, N. and Spruck, J., The method of moving spheres for elliptic Weingarten surfaces, preprint.
[11] Gårding, L., An inequality for hyperbolic polynomials, J. Math. and Mechanics 8 (1959), 957–965.
[12] Ghomi, M., Strictly convex submanifolds and hypersurfaces of positive curvature, Ph.D thesis (1998), Johns Hopkins University.
[13] Ghomi, M., Strictly convex submanifolds and hypersurfaces of positive curvature, J. Diff. Geometry 57 (2001), 273–299.
[14] Gilbarg, D. and Trudinger, N., Elliptic Partial Differential Equations of Second Order, 2nd ed., Springer, New York, 1983.
[15] Golubitsky, M. and Guillemin, V., Stable Mappings and Their Singularities, Springer, New York, 1973.
[16] Guan, B., The Dirichlet problem for Monge–Ampère equations in non-convex domains and spacelike hypersurfaces of constant Gauss curvature. Trans. Amer. Math. Soc. 350 (1998), no. 12, 4955–4971.
[17] Guan, B. and Spruck, J., Interior gradient estimates for solutions of prescribed curvature equations of parabolic type. Indiana Univ. Math. J. 40 (1991), no. 4, 1471–1481.
[18] Guan, B. and Spruck, J., Boundary-value problems on S^n for surfaces of constant Gauss curvature. Ann. of Math. (2) 138 (1993), no. 3, 601–624.
[19] Guan, B. and Spruck, J., The existence of hypersurfaces of constant Gauss Curvature with prescribed Boundary. J. Diff. Geom. 63 (2003), 1–29.
[20] Hardy, G. H., Littlewood, J. E. and Polya, G., Inequalities. Cambridge University Press 1952.
[21] Harvey, R. and Lawson, B., Calibrated geometries. Acta Math. 148, (1982), 47–157.
[22] Hoffman, D., Rosenberg, H. and Spruck, J., Boundary value problems for surfaces of constant Gauss curvature. Comm. Pure Appl. Math 45 (1992), 1051–1062.
[23] Hörmander, L., Notions of convexity. Progress in Mathematics 127, Birkhäuser, Boston, MA 1994.
[24] Huisken, G. and Ilmanen, T., The Riemannian Penrose inequality. Internat. Math. Res. Notices 1997 no. 20, 1045–1058.
[25] Krylov, N. V., Boundedly inhomogeneous elliptic and parabolic equations in a domain. Izvestia math. 47 (1983), 95–108.
[26] Li, A. and Li, Y. Y., On some conformally invariant fully nonlinear equations. Preprint.
[27] Massari, U. and Miranda, M., Minimal Surfaces of Codimension One. North-Holland, Amsterdam, 1984.
[28] McCuan, J., Symmetry via spherical reflection, J. Geom. Anal. 10 (2000), 545–564.
[29] McCuan, J., Symmetry via spherical reflection and spanning drops in a wedge. Pacific. J. Math. 180 (1997), 291–323.
[30] Rauch, J. and Taylor, B. A., The Dirichlet problem for the multi-dimensional Monge–Ampère equation. *Rocky Mountain J. Math.* **7** (1997), 345–364.
[31] Reilly, R., Applications of the Hessian operator in a Riemannian manifold. Indiana Univ. Math. J. 26 (1977), no. 3, 459–472.
[32] Reilly, R., Variational properties of functions of the mean curvatures for hypersurfaces in space forms. J. Differential Geometry 8 (1973), 465–477.
[33] Rosenberg, H., Hypersurfaces of constant curvature in space forms. Bull. Sci. Math. 117 (1993), no. 2, 211–239.
[34] Rosenberg, H. and Spruck, On the existence of convex hypersurfaces of constant Gauss curvature in hyperbolic space. J. Differential Geom. 40 (1994), 379–409.
[35] Serrin, J., A symmetry problem in potential theory. Arch. Rational Mech. Anal. 43 (1971), 304–318.
[36] Trudinger, N., The Dirichlet problem for the prescribed curvature equations. Arch. Rational Mech. Anal. 111 (1990), 153–179.

[37] N. Trudinger and J. Urbas, On second derivative estimates for equations of Monge-Ampère type. Bull. Austr. Math. Soc. 30 (1984), 321–334.
[38] Trudinger, N. and Wang, X.-J., On locally convex hypersurfaces with boundary. J. Reine Angew. Math. 551 (2002), 11–32.
[39] Viaclovsky, J., Conformally invariant Monge–Ampère equations: global solutions. Trans. AMS 352 (2000), 4371–4379.
[40] Yuan, Y., A priori estimates for solutions of fully nonlinear special Lagrangian equations. Ann. Inst. H. Poincaré Anal. Non Linéaire 18 (2001), 261–270.

Department of Mathematics, Johns Hopkins University, Baltimore, MD 21218
E-mail address: js@math.jhu.edu

Clay Mathematics Proceedings
Volume 2, 2004

Hyperbolic Surfaces of Constant Mean Curvature One with Compact Fundamental Domains

Hermann Karcher

Abstract. We construct surfaces of constant mean curvature one in hyperbolic 3-space $\mathbb{H}^3$. They have the symmetry group of a Platonic tessellation of $\mathbb{H}^3$ and therefore compact fundamental domains.

1. Introduction

In 1987 Bryant [Br] gave a "Weierstrass" representation of cmc1 surfaces (surfaces of constant mean curvature 1) in $\mathbb{H}^3$: A holomorphic null immersion $F : M^2 \to \mathrm{SL}(2,\mathbb{C})$ (i.e., $\langle F', F'\rangle_{\mathbb{C}} = 0$) gives a cmc1 immersion $f := FF^* : M^2 \to \mathbb{H}^3$, where $\mathbb{H}^3$ is the hypersurface $\{\det = 1\}$ in the four-dimensional space of positive definite hermitian symmetric 2×2-matrices. (The determinant is a quadratic function on this space.) Conversely, every cmc1 immersion f lifts to such a holomorphic null immersion into $\mathrm{SL}(2,\mathbb{C})$, the double cover of the isometry group.

The connection of cmc1 surfaces with holomorphic data is older; I learned from Lawson [La] how simply connected minimal surfaces in $\mathbb{R}^3$ give isometric cmc1 surfaces in $\mathbb{H}^3$: from the surface data of any minimal surface, namely Riemannian metric $g(\,,\,)$ and shape operator S, one gets surface data of a cmc1 surface by taking the same Riemannian metric and a new shape operator $S_\pm := S \pm \mathrm{id}$. This change does not affect the Codazzi equation, and the Gauss equation for $g(\,,\,)$, S in $\mathbb{R}^3$ becomes the Gauss equation for $g(\,,\,)$, $S_\pm$ in $\mathbb{H}^3$.

Both Bryant's Weierstrass representation and the described Lawson correspondence, can be used to construct cmc1 surfaces in $\mathbb{H}^3$. The Weierstrass representation point of view was developed by Umehara and Yamada [UY1–5]. They emphasize the following: while Bryant's Weierstrass data, namely a so-called secondary Gauss map g and the Hopf holomorphic quadratic differential Q, come from the minimal surface in $\mathbb{R}^3$ and correspond to a left invariant ODE for the immersion F into the group $\mathrm{SL}(2,\mathbb{C})$, they prefer a right invariant ODE for F in terms of the hyperbolic Gauss map G and the Hopf differential Q, where G is defined as the map from the surface to the conformal sphere at infinity of $\mathbb{H}^3$ obtained by following the geodesics from the surface in the direction of the mean curvature vector to infinity. This map is conformal for cmc1 surfaces (the normal map in the other direction to infinity is not conformal).

©2004 Clay Mathematics Institute

Umehara and Yamada give a connection between the two representations by noting that the map $F \to F^{-1}$ in the group leads to a pair of cmc1 immersions, FF^* and $F^{-1}F^{-1*}$, where the role of the hyperbolic and the secondary Gauss maps are interchanged. This method leads to a large collection of examples: together with Rossman [RUY] they construct from many of the *finite total curvature* minimal surfaces in $\mathbb{R}^3$ corresponding cmc1 cousin surfaces in $\mathbb{H}^3$ which have, up to some parameter adjustment, the same meromorphic (hyperbolic) Gauss map and Hopf differential as the minimal surfaces they started from.

More recently F. Pacard constructed cmc1 surfaces by functional analysis techniques; he connected horospheres by tiny necks.

In this paper we use the Lawson correspondence to construct, from solutions of polygonal Plateau problems in $\mathbb{R}^3$, cmc1 surfaces in $\mathbb{H}^3$ that are bounded by planar arcs of symmetry. Reflections in the symmetry planes, repeatedly applied to such *compact* fundamental domains, yield the complete surfaces. This is similar to the construction [La, Ka, Gb] of cmc1 surfaces in $\mathbb{R}^3$ from solutions of Plateau problems in $\mathbb{S}^3$, except that in the older situation we could choose the Plateau contours explicitly so that all the angles between the symmetry planes of the cmc1 surfaces come out correctly. In the present situation we have to consider two-parameter families of Plateau contours and show that the contour parameters can be chosen such that the angles between the symmetry planes of the cmc1 piece in $\mathbb{H}^3$ are correct. One of the two Plateau parameters simply scales the size of the polygonal contour and we have inequalities which control the effect of this change on the hyperbolic figure. Therefore we can find the correct parameter values with an iterated intermediate value argument (instead of a more complicated degree argument). In the [RUY] construction this scaling size appears as a real factor in front of the Hopf differential Q and leads to one-parameter families of solutions; in our case, such one-parameter families would exist if we were content with the solution of those free boundary value problems in which a cmc1 surface piece meets all faces of *some* Platonic polyhedron in convex curves. Since we look for surfaces in *tessellating* polyhedra, that is, with dihedral angles π/m, only one, two or three parameter values in these families give the desired surfaces. While the hyperbolic Gauss map is a known meromorphic map at the beginning of the constructions of [RUY] we do not even consider it for our examples. Note that every orbit in the sphere at infinity of the isometry group of a Platonic tessellation is dense so that the orbit of any Gauss value under the isometries of the surface is dense.

The first section of this paper deals with the Platonic tessellations of hyperbolic space. Their isometry groups have (as in the Euclidean cubical tessellation) tetrahedral fundamental domains in $\mathbb{H}^3$ from which one can build (by reflections in the tetrahedron faces) Platonic polyhedra, polyhedra, that is, on whose directed edges their symmetry group acts transitively. Further repeated reflections in the faces of the hyperbolic Platonic polyhedra give a tessellation of $\mathbb{H}^3$. The symmetry planes of the tessellation are also the symmetry planes of the looked-for cmc1 surfaces. We want to construct these surfaces from the assumption that they are cut in a similar way as Polthier's minimal surfaces in the following pictures. Namely, we expect them to be cut into congruent tiles which are bounded by only *four* arcs of symmetry. The corresponding Plateau contours in $\mathbb{R}^3$ would then be polygons with only four edges. We therefore start our construction by considering the Plateau solutions of all such polygons in $\mathbb{R}^3$.

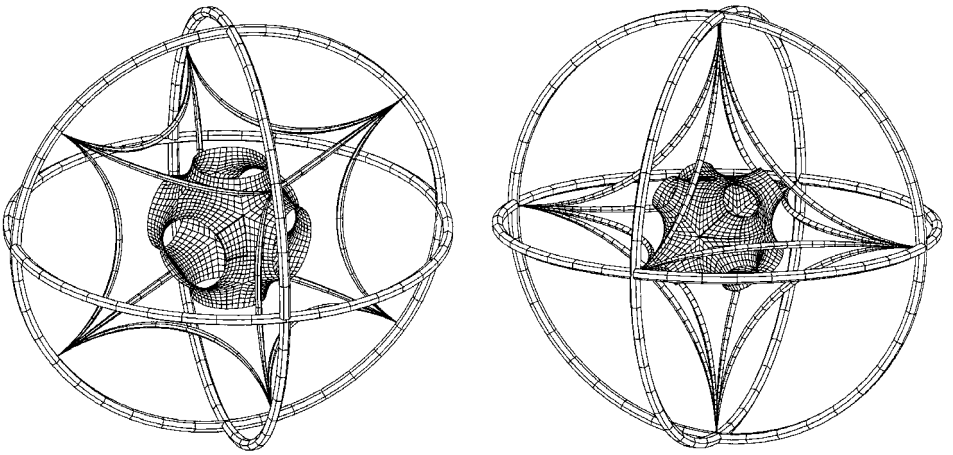


FIGURE 1. Hyperbolic cube and octahedron with vertices at ∞. Inside these solids are minimal surfaces that meet their faces orthogonally in symmetry lines that are close to circles. These minimal surfaces were constructed by K. Polthier; he also computed the numerical approximations and made these pictures. Note that the symmetry planes of the Platonic solids cut the minimal surfaces into fundamental domains which are bounded by four planar arcs of reflectional symmetry. The constant mean curvature surfaces that will be constructed in this paper have the same symmetries and meet the faces of the solids orthogonally, also in fairly circular curves.

In the second section we explain the geometric relations between the minimal patches in $\mathbb{R}^3$ and their cmc1 cousins in $\mathbb{H}^3$. Then we prove two comparison lemmas which rely on the simplicity of the Plateau contours: the projection of a (nonplanar) quadrilateral in the direction of an edge is a triangle and the Plateau solution is a *graph* over this triangle; this implies that the tangent planes along the vertical edge rotates monotonically (and not back and forth since that contradicts the graph property) so that the principal curvature function of the corresponding symmetry arc of the conjugate patch *does not change sign*. The two lemmas control the angles between the symmetry planes of the cmc1 patches well enough so that we can prove existence of a quadrilateral in $\mathbb{R}^3$ which spans a minimal surface patch whose cousin in $\mathbb{H}^3$ is the fundamental piece of the cmc1 surface with Platonic symmetry.

2. Platonic Tessellations of $\mathbb{H}^3$

We describe the compact and noncompact hyperbolic Platonic solids that we shall use. View a Euclidean Platonic polyhedron from its center, assume that the vertices are on a sphere of radius r and let $\mathbb{R}^3$ be the tangent space at $p \in \mathbb{H}^3$. Then consider the hyperbolic geodesics of length r from p in the direction towards the vertices; the hyperbolic convex hull of the endpoints is a hyperbolic Platonic solid with circumsphere of radius r. The dihedral angles of this polyhedron are smaller than the Euclidean dihedral angles; they decrease with growing r and converge to the Euclidean limit as $r \to 0$. In the limit $r \to \infty$ we obtain a Platonic solid with

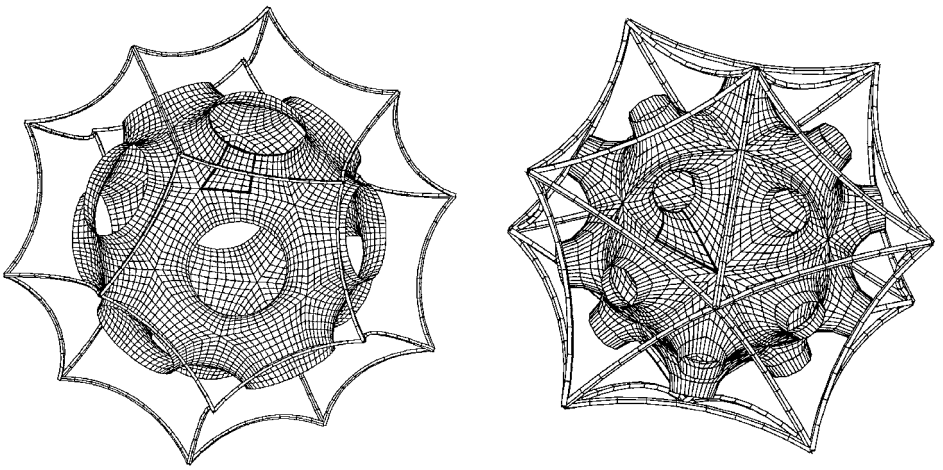


FIGURE 2. Hyperbolic 72°-dodecahedron and 120°-icosahedron. The surfaces inside the solids are the minimal surfaces constructed and computed by K. Polthier. Quadrilateral fundamental domains for the symmetry groups are highlighted.

vertices on the sphere at infinity; its dihedral angles can be seen on the intersection with a horosphere around a vertex. The intrinsic geometry of a horosphere is Euclidean, therefore this intersection is an equilateral triangle for the solids with trivalent vertices (tetrahedron, cube, dodecahedron), is a square for the octahedron and is a 108° regular pentagon for the icosahedron. Therefore we have the following tessellating (and for comparison in brackets [ ] some non-tessellating) Platonic solids in Euclidean (left column) and hyperbolic (right column) spaces. (The degrees refer to the dihedral angles; the last entry of each row has vertices at infinity in $\mathbb{H}^3$.)

[70.53°-tetrahedron]	60°-tetrahedron
90°-cube	72°-cube, 60°-cube
[116.57°-dodecahedron]	90°-dodecahedron, 72°-dodecahedron, 60°-dodecahedron
[109.47°-octahedron]	90°-octahedron
[138.19°-icosahedron]	120°-icosahedron, [108°-icosahedron not tessellating $\mathbb{H}^3$].

A Platonic solid is cut by its planes of symmetry into tetrahedra; all of them have one vertex at the center (C), the other three vertices are at the midpoint of a face (F), at the midpoint of an edge (E) and at a vertex of the solid (V). The dihedral angles of these tetrahedra are important for our construction: three of them, at the edges CE, FE, FV, are $\pi/2$; at the edge EV we have half the dihedral angle of the solid; at the edge CF the angle is π/k if the faces are k-gons; and at the edge CV the angle is π/n if n edges of the solid join to that vertex. (Compare the Euclidean cube.)

In Euclidean space the tetrahedron is called *self-dual* because the convex hull of the midpoints of its *faces* is again a tetrahedron; similarly, the Euclidean cube-octahedron or dodecahedron-icosahedron are called *dual pairs*. In hyperbolic space one also has self-dual Platonic solids and dual pairs. The convex hull of the midpoints of all those solids of a tessellation which have one vertex in common is again

a tessellating Platonic solid. It gives the *dual* tessellation; if these two tessellations are congruent we also call them self-dual. In Euclidean space the tessellation by cubes is self-dual. In hyperbolic space the tessellations by 72°-dodecahedra or by 120°-icosahedra are self-dual, while the tessellations by 72°-cubes and by 90°-dodecahedra are a dual pair. (The dual partners of the solids with infinite vertices are Platonic solids with infinitely many vertices. They have no Euclidean analogue and we will not use them here.)

If one projects the edges of a Platonic solid from the *center* C onto a sphere around the center, one obtains a Platonic tessellation of $\mathbb{S}^2$, and dual polyhedra give dual tessellations of $\mathbb{S}^2$. Similarly, one obtains from a *tessellation* by Platonic solids another Platonic tessellation of $\mathbb{S}^2$, if one intersects a small sphere around a *vertex* V with the faces of the solids meeting at that vertex. These spherical tessellations are very helpful to imagine the neighbors of a solid in a tessellation.

3. Comparison Between Euclidean Minimal and Their Cousin Hyperbolic Surfaces

We repeat some known facts (because we have to observe signs carefully) about *conjugate minimal surfaces in constant curvature spaces.* The notion refers to minimal surfaces that are simply connected (i.e., pieces or coverings). Let J be the (parallel) almost complex structure (i.e., the oriented 90°-rotation in each tangent space). Take the Riemannian metric $g(\,,\,)$ and change the shape operator S of a given minimal surface to $S_* := J \cdot S$; then S_* is trace free, symmetric and together with $g(\,,\,)$ satisfies the Gauss and Codazzi equations. Hence these data define an isometric but usually not congruent minimal surface, that is referred to as the conjugate surface. For geodesics c we have an interesting relation between (normal) curvature and torsion on the original and conjugate immersion:

$$\kappa := g(c', S \cdot c') = g(J \cdot c', J \cdot S \cdot c') = g(J \cdot c', S_* \cdot c') = -\tau_*,$$
$$\tau := g(c', J \cdot S \cdot c') = g(c', S_* \cdot c') = \kappa_*.$$

This is used to relate symmetry lines on a minimal surface and its conjugate: geodesic curvature lines lie in a plane orthogonal to the surface and reflection in this plane is a congruence of the minimal surface. On the conjugate immersion this curve has $\kappa_* = \tau = 0$, it is a straight line (i.e., a geodesic in the three-dimensional constant curvature space), and 180°-rotation around it is a congruence of the conjugate surface. The rotation speed of the tangent plane along such a straight line (segment) is $\tau_* = -\kappa$, and vice versa.

We use this as follows: consider a minimal surface which is the (simply connected) Plateau solution in a quadrilateral (non-planar) contour. Each projection in the direction of an edge is a (convex) triangle so that the Plateau solution is a *graph* over this triangle. The graph property implies that the tangent plane along the vertical edge cannot rotate back and forth, it can only rotate monotonically, i.e., the function $\tau = \kappa_*$ does *not change sign* along the edge. The total rotation of the tangent plane is the angle of the triangular projection below the vertical edge; of course, this angle is also the total rotation of the normal on the conjugate immersion. This means that the arc conjugate to the edge bounds, together with the normals at its end points, a convex domain, provided the normals intersect with an interior angle $\leq \pi$. Finally, the curvature functions $\kappa_* = \tau$ of arcs conjugate to adjacent edges have opposite sign: place these adjacent edges in a horizontal

plane; the two other edges then go either both up or both down. Since one is at an initial point the other at an end point of its horizontal neighbor, this means that the tangent planes along each pair of adjacent edges rotate in opposite directions.

4. Constant Mean Curvature Cousins of Minimal Surfaces

Since the sign of the second fundamental form can be changed by reversing the normal, we assume that the constant mean curvature is positive. Given a simply connected minimal surface in a space $M^3(K)$ of constant curvature K, take its Riemannian metric $g(\,,)$ and change its shape operator S to $S_+ := S + c \cdot \mathrm{id}$. Note that S_+ is symmetric with the same eigenvectors (principal curvature directions) as S; together with $g(\,,)$ it satisfies the Gauss and Codazzi equations for the space $M^3(K - c^2)$ and therefore defines a surface with constant mean curvature c in this space; it is referred to as a constant mean curvature cousin of the given minimal surface.

We will use this with $K = 0$, $c = 1$ and apply the construction to the conjugate of the Plateau solution of a quadrilateral contour; that is, we use the surface data $\{g, \hat{S} = J \cdot S + c \cdot \mathrm{id}\}$. This gives us cmc1 surfaces in $\mathbb{H}^3$ which are bounded by four planar symmetry arcs. The angles between adjacent symmetry arcs are given by the Platonic tessellation into which we plan to fit the cmc1 surface to be constructed. This determines the angles between adjacent edges of the quadrilateral Plateau contour. We also require that the angles between each pair of opposite symmetry planes (of the cmc1 surface piece) have the value needed for the chosen Platonic tessellation. To achieve this we need the remaining work (to find the correct quadrilateral contour).

5. The Effect of Scaling the Minimal Surface in $\mathbb{R}^3$

Consider the conjugate piece of the Plateau solution of a quadrilateral contour in $\mathbb{R}^3$ and the angle between the normals at the endpoints of each of the four planar symmetry arcs that bound the conjugate minimal surface piece. (This angle is of course the angle between a pair of opposite symmetry planes.) We know that the curvature function κ of each of the four boundary arcs of the conjugate Plateau surface does not change sign (recall that the tangent planes of the Plateau solution rotate monotonically along each of the four edges). We also observed above that adjacent arcs have curvature functions of opposite sign. If we scale the boundary arcs, $c \to \lambda c$, then their curvatures change as $\kappa \to (1/\lambda)\kappa$, and the corresponding curvatures of the hyperbolic cmc1 cousin boundary arcs are

$$\kappa_h \;=\; \frac{1}{\lambda}\kappa + 1, \text{ where } \kappa \text{ equals the torsion function } \tau \text{ of the Plateau patch.}$$

This means that the arcs with $\kappa \geq 0$ have $\kappa_h \geq 1$, i.e., they have a focal point along each normal, and independent of λ always on the same side (the limit infinity can be included). We call these the $+$arcs. The others, the $-$arcs, may still have focal points on the same side as the corresponding Euclidean arc for sufficiently small λ, but as λ increases, κ_h eventually changes sign along the whole arc. Curvature as a function of arc length determines a planar curve, and, in our case, the dependence on λ is continuous; moreover, for $\lambda \to 0$, the behavior of the hyperbolic arc converges to that of the Euclidean arc from which we started (this is best seen if we scale $\mathbb{H}^3$ to $M^3(-\lambda^2)$).

As long as the normals of the hyperbolic arc intersect on the same side as for the Euclidean arc (which they do for sufficiently small λ), we can compute the angle ω between the normals at the endpoints with the Gauss–Bonnet theorem. Let l be the length of the hyperbolic arc and A the area bounded by the arc and the normals at its endpoints.

For $+$arcs we have

$$\omega_+ = 2\pi - (\pi/2 + \pi/2 + (\pi - \omega_+)) = \int \kappa_h \, ds - A = \int \kappa \, ds + l - A.$$

For $-$arcs the normals intersect for small λ on the negatively curved side, that is, $-\kappa_h = |\kappa| - 1$; hence

$$\omega_- = \int -\kappa_h \, ds - A = \int |\kappa| \, ds - l - A.$$

Now we show that the angle ω_- decreases from its Euclidean value $\int \kappa \, ds$ to 0 as the scaling increases the hyperbolic length l from ~ 0 to a maximal length smaller than the obvious bound $\int |\kappa| \, ds$. And we show that the angle ω_+ increases from its Euclidean value $\int \kappa \, ds$ to $\pi/2$ as l is scaled up, but we do not obtain a *bound* on l that guarantees that $\omega_+ = \pi/2$ is reached before ℓ is increased to that bound. In both cases we require that the normals at the endpoints of the arc *do not meet the arc* before they intersect each other.

Consider a quarter arc of a very elongated ellipse and extend it a bit beyond the vertex of maximal curvature; the normal line at this endpoint meets the arc again before it intersects the normal line at the other endpoint. This explains why we require the *additional hypothesis* that we start with a Euclidean angle $\int |\kappa| \, ds \leq \pi/2$.

We deal with the $-$arcs first. For them, ω_- is at least by $-l$ smaller than the turning angle of the Euclidean arc. By making λ, hence l, larger we can indeed decrease ω_- to 0, even with the obvious a priori bound: $l < \int |\kappa| \, ds \leq \pi/2$. Note that the assumption $\int |\kappa| \, ds \leq \pi/2$ for the Euclidean turning angle is also sufficient to exclude (use Gauss–Bonnet) that the normals at the endpoints of the hyperbolic arc meet the arc before they intersect each other.

Next we deal with the $+$arcs. We will prove the inequality $A < l$ and we want to increase ω_+ to values well above the Euclidean turning angle $\int \kappa \, ds$ just by increasing λ. Unfortunately this means that we cannot reach $\omega_+ = \pi/2$ with an intermediate value argument; our following proof will reach $\pi/2$, but without an a priori bound on l. Such a bound would be convenient in view of such a bound for the $-$arcs. Therefore we cannot extend our construction from the above mentioned Platonic tessellations to tessellations by Platonic prisms (given by infinite orthogonal prisms over a Platonic tessellation of a hyperbolic plane).

LEMMA 1 (On the area A between a convex arc of hyperbolic length l and its end point normals.). *Consider a hyperbolic arc with curvature $\kappa_h \geq 1$ and such that the normals at the endpoints intersect each other with interior angle $\leq \pi$ (and do not meet the arc except at their foot point). Then*

$$A \leq \int_0^l (\kappa_h - \sqrt{\kappa_h^2 - 1}) \, ds = \int_0^l \frac{ds}{\kappa_h + \sqrt{\kappa_h^2 - 1}} \leq l.$$

PROOF. By assumption the arc and its end point normals bound a convex domain. Connect any interior point p to the endpoints of the arc by shortest

geodesics, both must meet the arc with an angle less than $\pi/2$. The endpoints are therefore *not* the nearest points on the arc from p and a nearest point q to p must exist as an *interior* point of the arc. Then p is not beyond a focal point on the normal of the arc in q. This means that we can get an upper bound for the area if we integrate in parallel coordinates of the arc along each normal up to the focal point. The focal distance r_f at a point of curvature $\kappa_h \geq 1$ is given by $\kappa_h = \coth r_f$. The line element in parallel coordinates (t, r) is

$$ds^2 = dr^2 + \big(\cosh r - \kappa_h(t) \sinh r\big)^2 \, dt^2.$$

Therefore we get for the area the bound

$$A \leq \int_0^l \int_0^{r_f(t)} \big(\cosh r - \kappa_h(t) \sinh r\big) \, dr \, dt = \int_0^l \big(\sinh\, r_f(t) - \kappa_h(t)(\cosh\, r_f(t) - 1)\big) \, dt.$$

Now use the relation $\kappa_h = \coth r_f$ to simplify the integrand:

$$\begin{aligned} \sinh\, r_f(t) - \kappa_h(t)(\cosh\, r_f(t) - 1) &= \sinh\, r_f(t) - \cosh\, r_f(t)^2 / \sinh\, r_f(t) + \kappa_h(t) \\ &= \kappa_h(t) - 1/\sinh r_f(t) = \kappa_h - \sqrt{\kappa_h^2 - 1}. \end{aligned}$$

The integrand is strictly less than 1 unless $\kappa_h(t) = 1$ and the focal distance is infinite; this limit is included in the proof. □

LEMMA 2 (On the change of the normal angle of +arcs under scaling). *Consider a fixed +arc on a minimal surface in $\mathbb{R}^3$ with curvature function $\kappa \geq 0$ and total turning angle $\int \kappa \, ds < \pi/2$. Take it so small that the normal angle ω_+ of its hyperbolic cousin is still $\leq \pi/2$. The +arc and its scalings define hyperbolic symmetry arcs on cmc1 surfaces with curvature $\kappa_h(s) = (1/\lambda)\kappa(s/\lambda) + 1$. As long as the normal angles satisfy $\omega_+(\lambda) \leq \pi/2$ we have*

$$\omega_+(\lambda) - \int \kappa \;=\; l(\lambda) - A(\lambda) \geq \int \frac{2\kappa(t)\, dt}{1 + 2\kappa/\lambda}, \tag{5–1}$$

$$\omega_+(\lambda) - \int \kappa \;=\; l(\lambda) - A(\lambda) \geq \sqrt{2\lambda} \int \frac{\sqrt{\kappa(t)}\, dt}{1 + \sqrt{2\kappa/\lambda}}. \tag{5–2}$$

PROOF. Insert the expression $\kappa_h(s) = (1/\lambda)\kappa(s/\lambda) + 1$ in the upper bound for A of the preceding lemma to get the lower bound

$$l(\lambda) - A(\lambda) \geq \int_0^{\lambda l} \big(1 - (\kappa_h(s) - \sqrt{\kappa_h(s)^2 - 1})\big) \, ds = \int_0^l \big(\sqrt{1 + 2\lambda/\kappa(t)} - 1\big)\kappa(t) \, dt.$$

Then use the equality $\sqrt{a} - 1 = (a-1)/(\sqrt{a} + 1)$ with $a = 1 + 2\lambda/\kappa$ and simplify in the denominator with $\sqrt{1 + 2x} \leq 1 + x$, respectively with $\sqrt{1 + x} \leq 1 + \sqrt{x}$ to get the two lower bounds. □

The lower bound (5–1) increases with λ to $\int 2\kappa \, dt$. If this makes ω_+ as large as we want (at most π), then we have an a priori upper bound for λ in terms of $\max \kappa$, i.e., a bound for the required length of the hyperbolic arc. The lower bound (5–2) increases with λ to infinity, but less explicitly.

In the first examples below one would get a better understanding if one had an upper bound better than $l(\lambda) - A(\lambda) \leq l(\lambda)$; focal point arguments do not give that.

6. Illustrative Examples: The Symmetric n-Noids

The symmetric n-noids were constructed in [UY3, p. 217, example 4.1]. With the conjugate cousin method they are simpler than the following *compact* fundamental domain examples. Therefore they are constructed here to explain this method. It will be sufficient to explain the cmc1 three-noid. For a qualitative picture place a small Euclidean minimal three-noid at the center of a ball model of $\mathbb{H}^3$ and let small horospheres around the limit points of the three axes of the half-catenoids grow until they touch the minimal three-noid; imagine that the three horospheres are connected by the central piece of the three-noid. Notice that the underlying Riemann surface is a 3-punctured sphere and the hyperbolic Gauss map is of degree two; the Weierstrass representation construction starts from here. Such a surface has the same planar symmetries as the minimal three-noid, one equator symmetry plane and three planes orthogonal to it which intersect in the normal line through the two umbilic points of the three-noid. These symmetry planes cut the surface into six simply connected (congruent) pieces, each of them bounded by three planar symmetry arcs. The first arc (a) comes from infinity to the umbilic point, the next (b) starts from there making a 60° angle with the first arc and meets the last one, the arc (c), in the equator plane orthogonally. Arc (c) then goes back to infinity. The length of the finite arc (b) we call $|b|$. The minimal surface in $\mathbb{R}^3$ of which the cmc1 surface patch is the conjugate cousin, is bounded by two half lines (a), (c) which are connected by a segment (b) which meets (a) under 60° and (c) under 90°. If we project this contour in the direction of (b), then we obtain an infinite sector, the angle ω of which parametrizes a one-parameter family of solutions. We need to assume (because our lemmas are not optimal) $\omega < \pi/2$. The Plateau solution is a graph, which implies that the tangent plane along (b) rotates only in one direction. We orient the normal so that the conjugate arc of (b) has curvature $\kappa \geq 0$, because we want to consider it as the +arc. (On the cmc1 surface we denote the corresponding boundary arcs also by (a), (b) and (c).)

Now we apply our lemmas. As long as $|b| < \pi/2 - \omega$ we have for the angle between the end point normals of the corresponding hyperbolic arc $\omega_+ \leq \omega + |b| < \pi/2$. On the other hand, the lower bound (5–2) for $l - A$ (which grows with λ to ∞) shows that we can scale the Plateau contour enough to make $\omega_+ = \pi/2$. This says that the normal of the cmc1 surface rotates along the finite symmetry arc (b) by $\pi/2$, or in other words, the symmetry plane of (a) meets the symmetry plane of (c) orthogonally. This completes the existence proof since the intersection line (in $\mathbb{H}^3$) of the symmetry planes of (a) and (b) meets the symmetry plane of (c) with the angle ω_+. Consequently, the reflections in these symmetry planes make a smooth surface out of six copies of the fundamental piece.

If the Euclidean turning angle ω is very close to $\pi/2$, then the lower bound (5–1) for $l(\lambda) - A(\lambda)$ shows that κ/λ has to be large, i.e., $|b|$ is small. In the other direction, one could suspect that $\omega \to 0$ implies $|b| \to \infty$, but, as mentioned in the lemma, we miss a sufficiently good upper bound for $l - A$ and cannot prove that.

7. The Quadrilateral Plateau Contours Needed for the Platonic Examples

The surfaces we want to construct are already vaguely suggested by the Schwarz P-surface, a triply periodic minimal surface in $\mathbb{R}^3$ which meets all the boundary squares of a cubical tessellation of $\mathbb{R}^3$ orthogonally in convex, almost circular curves.

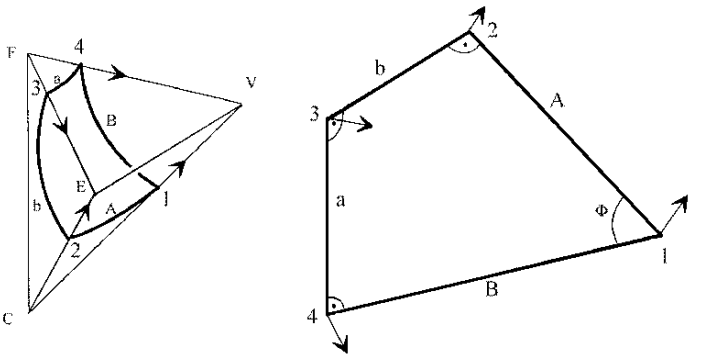


FIGURE 3. Patch in a fundamental tetrahedron with angles $\pi/2$ at edges 2, 3, 4, and quadrilateral polygon contour of conjugate patch. The arrows are normal vectors.

The pictures of Polthier's minimal surfaces in each hyperbolic Platonic solid give a more precise idea. Analogous surfaces of constant mean curvature greater than 1 can be described more easily: take a sphere inscribed in a Platonic solid and puncture it where it touches the faces in their midpoints; shrink the sphere a bit and replace the punctures by small necks which meet the faces orthogonally in convex curves around the face midpoints. This limit of surfaces with extremely small necks does not exist for cmc1 surfaces in $\mathbb{H}^3$ since the horospheres are too big for our Platonic solids. Therefore it is not a priori plausible that this picture will lead to an existence proof, and we will see at the end that it only succeeds because of explicit angle properties of our Euclidean Plateau contours.

The symmetry planes of any Platonic solid cut this solid into tetrahedral fundamental domains, described above; they also cut a Schwarz type surface inside the solid (see Polthier's pictures above) into fundamental pieces which lie in these tetrahedra and meet their faces orthogonally. I have indicated such quadrilateral fundamental domains on all of the preceding pictures. The simplest such situation has been studied by Smyth [Sm] for minimal surfaces in Euclidean tetrahedra: these surface patches meet each face of the given tetrahedron once; they miss two edges and they meet the other four edges orthogonally, so that the angle between the neighboring boundary arcs of the patch is the dihedral angle at that tetrahedron edge. This description implies a two-parameter family of quadrilateral contours for the conjugate minimal surface patch in $\mathbb{R}^3$ and the existence proof starts with these: one takes the Plateau solution to any such contour and proves that the symmetry planes of *its* conjugate patch indeed bound (up to scaling) the tetrahedron from which Smyth started.

Our initial situation is more special, since our tetrahedra are fundamental domains for the symmetry groups of (hyperbolic) Platonic solids. But the required control is more difficult since it cannot be done explicitly in terms of the conjugate contour. The fundamental surface patches do not meet two edges of the fundamental tetrahedron, neither the edge CF from the center to the face nor the edge EV from an edge midpoint of the Platonic solid to a vertex. The surface patch meets the other edges orthogonally so that the dihedral angles of the tetrahedron are also

the angles between the bounding arcs of the patch. The boundary arc (a) of the patch in the face FEV meets the boundary arc (b) in the face EFC with the dihedral angle $\pi/2$ of the edge FE. The boundary arc A of the patch in the face EVC and the boundary arc B in the face FVC meet with the dihedral angle $\phi = \pi/n$ of the edge VC (n is the number of edges of the solid that meet at the vertex V). The boundary arcs b and A meet with the dihedral angle $\pi/2$ of the edge CE, and the boundary arcs a and B meet with the dihedral angle $\pi/2$ of the edge FV. (To see the three dihedral angles $\pi/2$, note that the edge CF is orthogonal to the face FEV, and the edge VE is orthogonal to the face EFC.) The four angles at the vertices of the patch are also the angles of the quadrilateral conjugate contour, they are therefore easily achieved. However, the four planes of the boundary arcs of the patch also intersect in two lines which are *not met* by the patch: along the edge CF with a dihedral angle π/k (where k is the number of vertices of a face of the Platonic solid) and along the edge EV with the dihedral angle π/m, where $2\pi/m$ is the dihedral angle of the Platonic solid. The special form of this last angle is not needed for the conjugate construction, but it is needed for the solids to tessellate $\mathbb{H}^3$. This leaves us with the following two conditions: the normals at the endpoints of the arc a must intersect with the angle π/k, and the normals at the endpoints of the arc b must intersect with the angle π/m. Indeed, if we can construct such a patch, then completion by repeated reflection in the planes of its planar bounding arcs gives the desired cmc1 surface.

Therefore we can now describe the families of conjugate contours: start with two half lines A, B meeting under the angle ϕ, take a half line a meeting B under the angle $\pi/2$ and take the common perpendicular b of a and A. For each given ϕ this gives a two-parameter family of quadrilaterals. If we start with the Plateau solutions for these contours and take their conjugate cousin cmc1 patches then, indeed, these are bounded by planar symmetry arcs which meet under the correct angles $\pi/2$, $\pi/2$, $\pi/2$, ϕ; their planes therefore intersect in a hyperbolic tetrahedron which has already four correct dihedral angles. It remains to show with Lemmas 1 and 2 that we can choose the quadrilaterals to get also the other two dihedral angles correct. If the Platonic solid (see the above list) has k-gon faces, then the correct angle α_{cor} between the end point normals of the arc a has to be π/k, with $k = 3, 4, 5$. If the Platonic solid has dihedral angles $2\pi/m$, then the correct angle β_{cor} between the end point normals of the arc (b) has to be π/m, with $m = 3, 4, 5, 6$.

8. Explicit Determination of the Euclidean Data of the Contour Quadrilaterals

Let α be the angle through which the tangent plane of the Plateau solution rotates along the a-edge (this is also the angle between the direction vectors of b and $-B$); let β be the corresponding rotation angle for the b-edge. If we position b vertically, we see a right triangle: a, A are horizontal with angle β between them and the projection of B is orthogonal to a. The situation is similar if we position a vertically. This gives two relations:

$$a = A\cos\beta, \qquad b = B\cos\alpha.$$

Because of the right angles between a, B and also between b, A, we can express the diagonal between the other two vertices as $D^2 = a^2 + B^2 = b^2 + A^2$. We eliminate

a, b (or instead A, B) with the previous equations and obtain

$$B \sin\alpha = A \sin\beta, \qquad b \tan\alpha = a \tan\beta.$$

The length of the other diagonal gives $d^2 = a^2 + b^2 = A^2 + B^2 - 2AB\cos\phi$; eliminate first a, b then A/B to obtain

$$\cos\phi = \sin\alpha \cdot \sin\beta.$$

9. Control of the Angle Between the Endpoint Normals

We orient the normals to make the curvature of a positive and that of b negative. The strategy then is: For all a/b in a certain range we can, by simply scaling the quadrilateral from a very small size up, decrease β_h from the Euclidean value $\leq \pi/2$ to the correct value without increasing α_h beyond $\pi/2$. Moreover, for $a/b \ll 1$ the value of α_h stays below the correct value and for $a/b \approx 1$ the value of α_h stays above the correct value. The intermediate value argument with respect to the parameter a/b achieves the correct normal angles for a and b, so that the Schwarz type cmc1 surfaces exist for all the tessellating Platonic solids of $\mathbb{H}^3$.

Now the details. Recall that $\phi = \pi/n$, for n the number of edges at a polyhedron vertex; that $\alpha_{\text{cor}} = \pi/k$ for k the number of edges of a polyhedron face, and that $\beta_{\text{cor}} = \pi/m$ is half the dihedral angle of the polyhedron.

	tetrahedron	cube	dodecahedron	octahedron	icosahedron
$\pi/2 - \phi$	30	30	30	45	54
α_{cor}	60	45	36	60	60
β_{cor}	30	36, 30	45, 36, 30	45	60

First, if $a/b \ll 1$, we show how to achieve $\beta_h = \beta_{\text{cor}}$ with α_h staying below α_{cor}. In the limit $a/b \to 0$ we have the rotation angles $\alpha_e = \pi/2 - \phi$, $\beta_e = \pi/2$ of the Euclidean quadrilateral. As we scale the quadrilateral down, the hyperbolic angles α_h, β_h converge to the Euclidean ones. We check in the above list of correct angles that β_e is well above all desired values for β_h and α_e is smaller by at least $6^\circ = 36^\circ - 30^\circ = 60^\circ - 54^\circ$ than all the correct values α_{cor}. Therefore we start with a very small quadrilateral and scale it up at most to a length $l_b \leq \pi/2 - \beta_{\text{cor}}$ to achieve $\beta_h = \beta_{\text{cor}}$. If we start with $a/b < 6/90$, then the trivial bound $\alpha_h \leq \alpha_e + l_a$ shows that α_h does not increase above α_{cor}.

Secondly, we show how to achieve $\beta_h = \beta_{\text{cor}}$ while holding α_h between α_{cor} and $\pi/2$:

Consider the first three cases, which have $\phi = 60^\circ$. We start with quadrilaterals with $\alpha_e = \alpha_{\text{cor}}$ and check with $\sin\beta_e := \cos\phi / \sin\alpha_e$ that in all cases $\beta_e > \beta_{\text{cor}}$. Again, by scaling up from a very small size we can decrease β_h to β_{cor}, while α_h increases further and we only have to hold it below $\pi/2$ in (i) to (iii).

(i) The three dodecahedra: $\alpha_e = 36^\circ$, so $\sin\beta_e := \cos\phi / \sin\alpha_e = \sin 58.283^\circ$, and the desired correct values are $\beta_{\text{cor}} = 45^\circ, 36^\circ, 30^\circ$. Since $a < b$ we can decrease β_h to the smallest value 30° without α_h growing above $(36 + 28.283)^\circ$. This proves existence for the dodecahedra.

(ii) The two cubes: $\alpha_e = 45^\circ$ hence $\beta_e = 45^\circ$ and $a = b$. Therefore we can decrease, by scaling, β_h from the Euclidean limit down to the interesting values $\beta_{\text{cor}} = 36^\circ, 30^\circ$ without α_h growing by more than the needed decrease, that is, $\alpha_h \leq 60^\circ$. This proves existence for the cubes.

(iii) The 60°-tetrahedron: $\alpha_e = 60°$, so $\beta_e = 35.264°$ and $a/b = \tan\alpha/\tan\beta \leq 2.45$. Again we can decrease, by scaling, β_h from the Euclidean limit β_e down to the desired value $\beta_{\text{cor}} = 30°$ without increasing α_h to more than $(60 + 2.5 \cdot 5.3)° < 75°$. This proves existence for the infinite tetrahedron.

The 120°-icosahedron: Choose $\phi = 36°$ and $a = b$. Then

$$\cos 36° = (\sin 64.086°)^2,$$

showing that we need to scale only to a length $l_b \leq 4.086/180 \cdot \pi$ to decrease β_h from the Euclidean value 64.086° for very small quadrilaterals to the desired $\beta_{\text{cor}} = 60°$. Here α_h increases at most by the same amount because of $l_a = l_b$, in particular stays below $\pi/2$.

The 90°-octahedron: $\phi = 45°$;

$$\cos\phi = \sin\alpha \cdot \sin\beta = \sin 60° \cdot \sin 54.735°$$

shows that we have to decrease β_h by less than 10° from 54.735° to 45°, and from $a/b = \tan\alpha/\tan\beta = 1.225$ it follows that α_h does not increase more than 12.25°. □

References

[Br] Bryant, R.: Surfaces of mean curvature one in hyperbolic space. Astérisque 154–155 (1987), 321–347.

[Gb] Große-Brauckmann, K.: New surfaces of constant mean curvature. Math. Z. 214 (1992), 527–565.

[Ka] Karcher, H.: The triply periodic minimal surfaces of Alan Schoen and their constant mean curvature companions. Manuscripta Math. 64 (1989), 291–357.

[La] Lawson, B. H.: Complete minimal surfaces in S^3. Annals of Math. 92 (1970), 335–374.

[RUY] Rossman, W., Umehara, M., Yamada, K.: Irreducible constant mean curvature 1 surfaces in hyperbolic space with positive genus. Pages 561–584 in this volume.

[Sm] Smyth, B.: Stationary minimal surfaces with boundary on a simplex. Invent. Math. 76 (1984), 411–420.

[UY1] Umehara, M., Yamada, K.: A parametrization of the Weierstrass formulae and perturbation of some complete minimal surfaces of $\mathbb{R}^3$ into the hyperbolic 3-space. Journal für reine und angewandte Mathematik 432 (1992), 93–116.

[UY2] Umehara, M., Yamada, K.: Complete surfaces of constant mean curvature 1 in the hyperbolic 3-space, Annals of Math. 137 (1993), 611–638.

[UY3] Umehara, M., Yamada, K.: Surfaces of constant mean curvature c in $\mathbb{H}^3(-c^2)$ with prescribed hyperbolic Gauss map, Mathematische Annalen 304 (1996), 203–224.

[UY4] Umehara, M., Yamada, K.: Another construction of a CMC-1 surface in $\mathbb{H}^3$, Kyungpook Mathematical Journal 35 (1996), 831–849.

[UY5] Umehara, M., Yamada, K.: A duality on CMC-1 surfaces in the hyperbolic 3-space and a hyperbolic analogue of the Osserman inequality, Tsukuba J. Math. 21 (1997), 229–237.

Mathematisches Institut, der Universität Bonn, Beringstrasse 1, D-53115 Bonn, Germany

E-mail address: unm416@uni-bonn.de

Clay Mathematics Proceedings
Volume 2, 2004

Isoperimetric Inequalities of Minimal Submanifolds

Jaigyoung Choe

Abstract. In this paper we introduce various types of isoperimetric inequalities for minimal submanifolds in Euclidean space, the sphere, hyperbolic space, and Riemannian manifolds. Some inequalities for compound soap films, domains, nonpositively curved surfaces and harmonic maps are also discussed.

Contents

1. Introduction 325
2. Osserman and Schiffer's Proof 327
3. Li, Schoen and Yau's proof 331
4. The Cone Method 332
5. Hyperbolic Case 334
6. Variable Curvature Case 336
7. Varifolds and Flat Chains 341
8. Higher-Dimensional Minimal Submanifolds 343
9. The Calibration Method 346
10. Weak Isoperimetric Inequalities 347
11. Modified Volume 351
12. Relative Isoperimetric Inequality 353
13. Negatively Curved Surfaces 359
14. Isoenergy Inequalities 361
References 367

1. Introduction

The history of the isoperimetric problem begins with its legendary origins in the Problem of Queen Dido, told by Virgil in the *Aeneid*. Dido was a Phoenician princess from the city of Tyre. She fled by ship from Tyre when King Pygmalion, her tyrannical brother, murdered her husband to usurp her possessions. When Dido arrived in Africa at the site that was to become Carthage, she sought to purchase land from the natives. They told her they would sell only as much land as she

Supported in part by KOSEF 2000-2-10200-002-5.

©2004 Clay Mathematics Institute

could surround with a bull's hide. She accepted the terms and made the most of the situation. First she had her people cut a bull's hide into thin strips and tie them together to form a single, very long, closed cord. Then, by sheer intuition, she reasoned that she could encompass the most area by shaping the cord into the circumference of a circle. In this way she acquired a larger piece of land than she had coveted.

Here let us add our own fable to history: Dido was thus able to lead a comfortable life in the big land now called Byrsa. But her peaceful life did not last long; King Pygmalion, wanting ever more power and land, invaded Carthage, and Queen Dido was forced to flee again. This time she decided to move to Wonderland, where people inhabit a big soap film (minimal surface). There she purchased land surrounded by the same cord that she had used in Carthage. And Queen Dido asked herself whether her land in Wonderland was bigger than that in Carthage...

The first proof that Dido made the optimum choice in Carthage appears in the commentary of Theon to Ptolemy's *Almagest* and in the collected works of Pappus. The author of the proof is Zenodoros. However, Zenodoros' proof still contained a gap that was not filled in until the second half of the nineteenth century. A rigorous mathematical proof was given by Weierstrass in his lectures at the University of Berlin [**Sp**]. But Dido's Wonderland question has not been completely settled yet.

This paper concerns Queen Dido's new problem. We will summarize the results so far obtained by the author and others, and will present some new results. Many kinds of minimal submanifolds will be dealt with and various types of isoperimetric inequalities for them will be introduced.

Queen Dido's characterization of the circle is most succinctly expressed in the isoperimetric inequality

$$4\pi A \leq L^2,$$

where A is the area enclosed by a curve C of length L, and where equality holds if and only if C is a circle. The first result in regard to Dido's new problem is due to Carleman [**Ca**] who showed in 1921 that if S is a disk type minimal surface in $\mathbb{R}^n$ with area A and perimeter L, then

$$4\pi A \leq L^2 \tag{1–1}$$

with equality if and only if S is a flat disk. In other words, Queen Dido's land in Wonderland is smaller than that in Carthage. Then in 1933 Beckenbach and Radó [**BR**] generalized Carleman's method and showed that $4\pi A \leq L^2$ holds also for a disk type surface of nonpositive Gaussian curvature. Note that minimal surfaces have nonpositive Gaussian curvature.

It would seem at first glance that Carleman, Beckenbach and Radó's results provide a complete solution to Queen Dido's new problem, but that is not the case for some reason. The restriction to disk type surfaces is not a natural one. For example, a simple closed curve in $\mathbb{R}^3$ may bound a minimal surface of the type of a Möbius strip, or a surface with higher genus. One would like to know if $4\pi A \leq L^2$ also in those cases. A point worth noting when one drops simple connectivity is that the inequality $4\pi A \leq L^2$ does not hold for general surfaces of nonpositive Gaussian curvature. For example, on a long cylinder the perimeter remains fixed while the area can be made as big as one wishes. Also one can construct an example with a single boundary curve using a flat torus: the complement of a small disk is a domain of genus one on the torus whose perimeter can be made arbitrarily small.

In view of these examples it is something of a surprise that similar examples cannot be constructed within the class of minimal surfaces. This leads us to the following.

OPEN PROBLEM 1.1. *Prove that* $4\pi A \leq L^2$ *for any minimal surface in* $\mathbb{R}^n$.

The difficulty in this open problem is that the minimal surface $S \subset \mathbb{R}^n$ may have arbitrary topology, ∂S may have more than one component, and S may not be area minimizing. Here are two arguments which give some insight into why one might expect the classical isoperimetric inequality (1–1) to hold also on minimal surfaces. Consider a surface $S \subset \mathbb{R}^n$ that has least area among all surfaces with the same boundary curve ∂S. Let its area be A and its perimeter L. First, suppose that ∂S is connected. Let S' be the cone over ∂S with vertex at some point of ∂S. Then S' has the same boundary as S, and hence its area A' is not less than A. But since it is a cone, S' can be developed in $\mathbb{R}^2$ onto a domain $D \subset \mathbb{R}^2$, preserving its area and perimeter. By the classical isoperimetric inequality in the plane, $4\pi A \leq 4\pi A' \leq L^2$. Second, suppose ∂S is not connected, and let $C_1, \ldots, C_n$ be the distinct components of ∂S, $S_1, \ldots, S_n$ the Douglas–Radó solutions with $\partial S_i = C_i$, A_i the area of S_i, and L_i the length of C_i. Then $S_1 \cup \cdots \cup S_n$ has the same boundary as S, and hence its area $A_1 + \cdots + A_n$ is not less than A. But by the first argument for each S_i, $4\pi A_i \leq L_i^2$, and so $4\pi A \leq 4\pi(A_1 + \cdots + A_n) \leq L_1^2 + \cdots + L_n^2 \leq L^2$. These arguments, unfortunately, are valid only for area minimizing surfaces, while in Open Problem 1.1 S is an arbitrary minimal surface.

2. Osserman and Schiffer's Proof

First we take a look at the history of the isoperimetric inequality of a minimal surface in space. Carleman [**Ca**] used complex function theory in 1921 for the proof of (1–1). In 1959 Reid [**Re**] proved (1–1) for a minimal surface with connected boundary in $\mathbb{R}^3$. His proof is based on Wirtinger's inequality:

$$\int_0^{2\pi} \left(\frac{dy}{dt}\right)^2 dt \geq \int_0^{2\pi} y^2 dt, \tag{2–1}$$

where $y(t)$ is a smooth function with period 2π and $\int_0^{2\pi} y(t)\,dt = 0$. This proof was extended by Hsiung [**Hs**] to $\mathbb{R}^n$ in 1961. Then in 1975 a proof of (1–1) was obtained by Osserman–Schiffer [**OS**] for a doubly connected minimal surface in $\mathbb{R}^3$, and in 1977 by Feinberg [**Fe**] for a doubly connected minimal surface in $\mathbb{R}^n$. In 1983 Li–Schoen–Yau proved (1–1) for a minimal surface with two boundary components in $\mathbb{R}^3$. And in 1990 the author [**C1**] gave a proof for a minimal surface with two boundary components in $\mathbb{R}^n$.

In this section we shall give a proof for a minimal surface with connected boundary in $\mathbb{R}^n$. Also Osserman and Schiffer's result will be outlined.

Let S be an arbitrary surface in $\mathbb{R}^n$. If $\vec{H}$ is the mean curvature vector of S, and if x is the position vector, then a general formula for the area A of S is

$$2A = -\int_S \langle x-p,\, \vec{H}\rangle + \int_{\partial S} \langle x-p,\, \nu\rangle \tag{2–2}$$

where $p \in \mathbb{R}^n$ is arbitrary, and ν is the outward unit conormal to ∂S on S, i.e., the normal to ∂S which is tangent to S. This formula is a special case of the first variation formula of area, using the variation field of the 1-parameter family of

homothetic expansions about p. Therefore

$$4\pi A \leq L^2 - 2\pi \int_D \langle x-p,\, \vec{H}\rangle,$$

which is the desired isoperimetric inequality for a minimal surface S, if we can show

$$2\pi \int_{\partial S} \langle x-p,\, \nu\rangle \leq L^2. \tag{2–3}$$

Osserman gave a proof of this inequality in [**O1**], p. 209. As a matter of fact, we can prove that this inequality is a consequence of an isoperimetric inequality provided p is a point of ∂S and ∂S is connected. To see why, let η be the unit normal to ∂S which makes the smallest angle with $x-p$ among the normals to ∂S. Then $\langle x-p,\, \nu\rangle \leq \langle x-p,\, \eta\rangle$ and hence

$$2\pi \int_{\partial S} \langle x-p,\, \nu\rangle \leq 2\pi \int_{\partial S} \langle x-p,\, \eta\rangle.$$

From the definition of η it is not difficult to see that

$$\tfrac{1}{2} \int_{\partial S} \langle x-p,\, \eta\rangle = \mathrm{Area}(p \ast \partial S),$$

where $p \ast \partial S$ is the cone which is the union of the line segments from p to the points of ∂S. Therefore (2–3) follows from the isoperimetric inequality of the planar domain obtained by developing $p \ast \partial S$ on $\mathbb{R}^2$. Note here that in general $p \ast \partial S$ cannot be developed onto a planar domain if p is not from ∂S or ∂S is not connected. This kind of estimation of an integral by the area or angle of a cone will play a key role in Section 4 (see (4–1) and (4–2)).

We have thus proved the following.

THEOREM 2.1. *If S is a minimal surface in $\mathbb{R}^n$ with connected boundary, then $4\pi A \leq L^2$.*

Suppose now that S is minimal and has at least two boundary components. Let $\partial S = \bigcup C_i$, $L_i = \mathrm{Length}(C_i)$ and choose points $p_i \in C_i$. Then by (2–3)

$$2\pi \int_{C_i} \langle x-p_i, \nu\rangle \leq L_i^2. \tag{2–4}$$

But

$$\int_{C_i} \langle x-p_1, \nu\rangle = \int_{C_i} \langle x-p_i, \nu\rangle - \int_{C_i} \langle p_1-p_i, \nu\rangle, \tag{2–5}$$

so combining (2–4) and (2–5) with (2–2) gives

$$4\pi A \leq \sum_{i=1}^{k} L_i^2 + 2\pi \sum_{i=2}^{k} \int_{C_i} \langle p_i-p_1, \nu\rangle. \tag{2–6}$$

If the integral term in (2–6) could be made to vanish, one would get

$$4\pi A \leq \sum_{i=1}^{k} L_i^2 \tag{2–7}$$

which is strictly stronger than $4\pi A \leq L^2$. In fact (2–7) does hold for area minimizing surfaces as we have seen at the end of the previous section. On the other hand, (2–7) is not true in general. Consider a catenoidal waist S bounded by parallel circles which is the *horizon* when viewed from a point p in the axis of the catenoid.

In other words, the catenoid is tangent to the position vector from p along the boundary circles of S. By [**C6**, Proposition 1], $\mathrm{Area}(S) = \mathrm{Area}(p \ast \partial S)$ and so $\mathrm{Area}(S)$ will be greater than the sum of the areas A_i of the disks bounded by the separate circles. But for each of those disks we have $4\pi A_i = L_i^2$. Hence (2–7) fails to hold.

Returning to (2–6), let us think of geometric conditions under which the integral term of (2–6) can be made to vanish. A useful observation is the following. Applying (2–2) to a minimal surface S with $\partial S = \bigcup C_i$ yields

$$\int_{\partial S} \langle p, \nu \rangle = \int_{\partial S} \langle x, \nu \rangle - 2A,$$

for any p. Since the right-hand side is independent of p, it follows that

$$\sum \int_{C_i} \nu = \int_{\partial S} \nu = 0.$$

Actually this integral is called the *flux* of S along ∂S (see [**Fa**] p. 81).

THEOREM 2.2. (a) *Let S be a minimal surface in $\mathbb{R}^n$ with $\partial S = \bigcup_{i=1}^k C_i$. If the flux $\int_{C_i} \nu$ of S along C_i vanishes for each i, then*

$$4\pi A \leq \sum_{i=1}^k L_i^2.$$

(b) [**O1**] *Let S be a minimal surface in $\mathbb{R}^n$ with two boundary components, i.e., $\partial S = C_1 \cup C_2$. If no hyperplane separates C_1 from C_2, then*

$$4\pi A \leq L_1^2 + L_2^2.$$

PROOF. (a) follows from (2–6) since $p_i - p_1$ is a constant vector on each C_i. The flux $\int_{C_1} \nu = -\int_{C_2} \nu$ may or may not be a zero vector. In any case, let Π be a hyperplane which is orthogonal to the flux and intersects both C_1 and C_2. Choose $p_1 \in C_1 \cap \Pi$ and $p_2 \in C_2 \cap \Pi$. Then $\langle p_2 - p_1, \int_{C_2} \nu \rangle = 0$, so (2–6) proves (b). □

Suppose that a hyperplane separates C_1 from $\bigcup_{i=2}^k C_i$. Then it is easy to see that the flux $\int_{C_1} \nu$ of S along C_1 is nonzero. This observation and Theorem 2.2 lead to the following.

OPEN PROBLEM 2.3. *If S is a minimal surface in $\mathbb{R}^n$ with $\partial S = \bigcup_{i=1}^k C_i$ such that none of $C_1, \ldots, C_k$ can be separated from the others by a hyperplane, then*

$$4\pi A \leq \sum_{i=1}^k L_i^2.$$

The inequality of Theorem 2.2 (b) is equivalent to

$$L^2 - 4\pi A \geq 2L_1 L_2.$$

Likewise, for an arbitrary doubly connected minimal surface in $\mathbb{R}^3$, Osserman and Schiffer [**OS**] (see p. 297) proved the following.

THEOREM 2.4. [**OS**] *For any doubly connected minimal surface in $\mathbb{R}^3$,*

$$L^2 - 4\pi A \geq 2L_1 L_2 (1 - \log 2). \tag{2–8}$$

Let $S \subset \mathbb{R}^n$ be a doubly connected minimal surface. S is parametrized by a conformal harmonic map ψ of an annulus $r_1 < |z| < r_2$ into $\mathbb{R}^n$, where ψ is assumed to extend continuously to the boundary circles and to map them onto Jordan curves C_1, C_2. Let $L(r)$ be the length of the image of the circle $|z| = r$, for $r_1 < r < r_2$. The only case of interest is that where C_1 and C_2 are rectifiable, of length say, L_1 and L_2. In that case, $\lim_{r \to r_i} L(r) = L_i$ for $i = 1, 2$. The key lemma for the case $n = 3$ is this:

PROPOSITION 2.5. [**OS**] *The function $L(r)$ satisfies*

$$\frac{d^2 L}{d(\log r)^2} \geq L \tag{2–9}$$

with equality possible in only two cases: if ψ is a conformal map onto a planar annulus, or if the image of ψ is a catenoid bounded by a pair of coaxial circles in parallel planes.

Using this equality together with (2–2) and the specific expressions for $L(r)$ on a catenoid, they prove Theorem 2.4.

The proof of Proposition 2.5 depends on the Weierstrass representation formula for minimal surfaces in $\mathbb{R}^3$, and does not go through for $n > 3$. But Feinberg [**Fe**] noted that a weaker form of (2–9) will yield a weaker form of (2–8). Specifically,

$$\frac{d^2 L}{d(\log r)^2} > \frac{2}{\pi^2} L. \tag{2–10}$$

He proved this by deriving an analog of the Wirtinger inequality (2–1) but "without the squares". Namely,

$$\int_0^{2\pi} y(t)dt = 0 \;\Rightarrow\; \int_0^{2\pi} |y'(t)|\, dt > \frac{2}{\pi} \int_0^{2\pi} |y(t)|\, dt.$$

Note that unlike (2–1), the inequality here is a strict one. However, the constant $2/\pi$ is best possible. The inequality $4\pi A \leq L^2$ for all doubly connected minimal surfaces in $\mathbb{R}^n$ follows from (2–10).

A final remark for doubly connected surfaces. We have learned from Carleman, Feinberg and Beckenbach–Radó that disk type minimal surfaces, doubly connected minimal surfaces and disk type nonpositively curved surfaces all satisfy the inequality $4\pi A \leq L^2$. However, as mentioned in the Introduction, doubly connected nonpositively curved surfaces do not satisfy the inequality because there are long cylinders with fixed perimeter whose area can be arbitrarily large. But is there any way of controlling the area of a cylinder with fixed perimeter? One way would be to fix the conformal structure of the cylinder. Consider all flat doubly connected surfaces which are conformally equivalent. It is not difficult to show that among them the flat annulus A with boundary curves of equal length satisfies the best isoperimetric inequality. Motivated by this, we propose the following.

OPEN PROBLEM 2.6. *Let S be a nonpositively curved annulus which is conformally equivalent to a flat annulus A with boundary curves of equal length. Show that*

$$\frac{\text{Area}(S)}{\text{Length}(\partial S)^2} \leq \frac{\text{Area}(\mathsf{A})}{\text{Length}(\partial \mathsf{A})^2}.$$

3. Li, Schoen and Yau's proof

In this section we briefly introduce Li–Schoen–Yau's proof for a minimal surface with two boundary components [**LSY**]. Their proof works also for a minimal surface with connected boundary.

First we rederive (2–2) for minimal S. Fix a point $p \in \mathbb{R}^n$ and define $r(x) = \operatorname{dist}(p,x), x = (x_1,\dots,x_n)$. Given an m-dimensional submanifold $N \subset \mathbb{R}^n$, it is well known that

$$\triangle x = (\triangle x_1,\dots,\triangle x_n) = \vec{H},$$

where $\triangle$ is the intrinsic Laplacian on N and $\vec{H}$ is the mean curvature vector of N. Hence the rectangular coordinate functions $x_1,\dots,x_n$ are harmonic on a minimal submanifold of $\mathbb{R}^n$. If we take p as the origin, then

$$\triangle r^2 = \sum \triangle x_i^2 = 2\sum x_i \triangle x_i + 2\sum |\nabla x_i|^2 = 2m \text{ on } N. \tag{3–1}$$

Integrating this over $N = S$ for $m = 2$ yields

$$4\,\mathrm{Area}(S) = \int_S \triangle r^2 = \int_{\partial S} 2r\frac{\partial r}{\partial \nu}, \tag{3–2}$$

where ν is the outward unit conormal to ∂S. Translating S suitably, we may assume $\int_{\partial S} x_i = 0$. Then

$$\begin{aligned} 4\,\mathrm{Area}(S) &= \int_{\partial S} 2r\frac{\partial r}{\partial \nu} \le 2\int_{\partial S} r \le 2\,\mathrm{Length}(\partial S)^{1/2}\left(\int_{\partial S}\sum x_i^2\right)^{1/2} \\ &\le \frac{1}{\pi}\,\mathrm{Length}(\partial S)^{3/2}\left(\int_{\partial S}\sum\left(\frac{dx_i}{ds}\right)^2\right)^{1/2} = \frac{1}{\pi}\,\mathrm{Length}(\partial S)^2 \end{aligned}$$

(on the last line, the inequality follows by (2–1) or the Poincaré inequality, and the equality because $\sum (dx_i/ds)^2 = 1$). This gives $4\pi A \le L^2$ when ∂S is connected.

In case ∂S is not connected, we cannot use the Poincaré inequality as above. However, note that $\int_{\partial S}\sum (dx_i/ds)^2$ is invariant under translations in $\mathbb{R}^n$ and that $\int_{\partial S}\sum x_i^2$ may or may not be invariant depending on the choice of translations. Note also that the disconnected boundary ∂S may be made into a connected curve provided a suitable translation is applied to each component of ∂S. With these observations Li, Schoen and Yau introduced the following.

DEFINITION 3.1. *A set $\Gamma \subset \mathbb{R}^n$ is* weakly connected *if there exists a rectangular coordinate system $\{x_i\}_{i=1}^n$ of $\mathbb{R}^n$ such that for every affine hypersurface $H^{n-1} = \{x_i = \text{constant}\}$ in $\mathbb{R}^n$, H does not separate Γ.*

Therefore a connected set is obviously weakly connected. The key idea of [**LSY**] is that if ∂S is weakly connected, then one can find a family of suitable translations which leave $\int_{\partial S}\sum x_i^2$ unchanged and make ∂S into a connected curve C. Consequently, the Poincaré inequality can be applied to C, and thereby yielding $4\pi A \le L^2$ for S as above. In this way they proved

THEOREM 3.2. [**LSY**] *If S is a minimal surface in $\mathbb{R}^n$ with weakly connected boundary ∂S, then S satisfies $4\pi A \le L^2$.*

And then applying the maximum principle to the family of homothetic contractions of a catenoid, they showed that if the boundary of a minimal surface S in $\mathbb{R}^3$ has two components and is not weakly connected, then S cannot be connected.

In this way they proved that a minimal surface in $\mathbb{R}^3$ with one or two boundary components satisfies $4\pi A \le L^2$.

4. The Cone Method

The author extended Li–Schoen–Yau's theorem by taking a more geometric point of view [**C1**]. In their proof they used the inequality $\frac{\partial r}{\partial \nu} \le 1$. But if η is the unit normal to ∂S which makes the smallest angle with ∇r, then

$$\frac{\partial r}{\partial \nu} \le \frac{\partial r}{\partial \eta} \le 1.$$

In fact η is the outward unit conormal to ∂S on the cone $p \ast \partial S$, the union of the line segments from p to the points of ∂S. Although $p \ast \partial S$ is not minimal, the identity $\triangle r^2 = 4$ holds there too because

$$\triangle r^2 = 4 + 2\langle x, \triangle x\rangle = 4 + 2\langle x, \vec{H}\rangle = 4,$$

where $\vec{H}$ is the mean curvature vector of $p \ast \partial S$, which is perpendicular to x. Therefore

(4–1)
$$4\,\mathrm{Area}(S) = \int_S \triangle r^2 = \int_{\partial S} 2r\frac{\partial r}{\partial \nu} \le \int_{\partial S} 2r\frac{\partial r}{\partial \eta} = \int_{p \ast \partial S} \triangle r^2 = 4\,\mathrm{Area}(p \ast \partial S),$$

which gives an area comparison between S and $p \ast \partial S$.

A nice thing about the cone $p \ast \partial S$ is that $p \ast \partial S$ is flat and hence is locally developable. If ∂S is connected, cut along a line segment l from p to a point of ∂S and then one can develop $p \ast \partial S$ into a cone $O \ast C$ on $\mathbb{R}^2$. $O \ast C$ has the same area as $p \ast \partial S$ and C has the same length as ∂S. C may not be a closed curve, but we can show that it is a curve with self-intersection if p is an interior point of S as follows. Choose a point $p \in S$ and show that (Lemma 4.5 below)

$$\triangle \log r \ge 2\pi\delta_p.$$

Then

(4–2)
$$2\pi \le \int_S \triangle \log r = \int_{\partial S} \frac{1}{r}\frac{\partial r}{\partial \nu} \le \int_{\partial S} \frac{1}{r}\frac{\partial r}{\partial \eta} = \mathrm{Angle}(\partial S, p),$$

where $\mathrm{Angle}(\partial S, p)$ is the *angle* of ∂S viewed from p. In other words, $\mathrm{Angle}(\partial S, p)$ equals 2π times the density of $p \ast \partial S$ at p. This angle estimate implies that ∂S rotates around p by at least $360°$ and consequently C should intersect itself. Then cutting $O \ast C$ into two pieces and pasting them appropriately gives rise to a domain $D \subset \mathbb{R}^2$ with

$$\mathrm{Area}(D) \ge \mathrm{Area}(O \ast C) \ge \mathrm{Area}(S), \quad \mathrm{Length}(\partial D) = \mathrm{Length}(C) = \mathrm{Length}(\partial S)$$

(see [**C1**, Lemma 1] for the construction of D). Therefore the classical isoperimetric inequality for D gives rise to (1–1) for S.

So far ∂S has been assumed to be connected. However, even if ∂S is not connected, C as defined above may behave like a connected curve. This motivates the following.

DEFINITION 4.1. *A set $\Gamma \subset \mathbb{R}^n$ is said to be* radially connected *from $p \in \mathbb{R}^n$ if $\{r : r = \mathrm{dist}(p, q),\ q \in \Gamma\}$ is a connected interval.*

If ∂S is radially connected from p, then we can apply to $p \ast\!\!\ast \partial S$ the argument of "cutting and inserting and pasting" to obtain a cone $O \ast\!\!\ast C \subset \mathbb{R}^2$ with C connected. Moreover, if p is in S, then C has a self-intersection and so we can obtain the domain $D \subset \mathbb{R}^2$ as above and hence the desired isoperimetric inequality for S. See [**C1**, Theorem 1] for more details. Thus we have outlined the proof of the following.

THEOREM 4.2. [**C1**] *If S is a minimal surface whose boundary is radially connected from a point of the surface, then it satisfies $4\pi A \leq L^2$.*

Although there is no relationship between radial connectivity and weak connectivity, we have a stronger corollary than Li–Schoen–Yau's: If S is in $\mathbb{R}^n$ such that ∂S has two components, then ∂S is radially connected from a point in S which is a midpoint between the two boundary components, and hence (1–1) holds for such S.

The volume comparison as in (4–1) holds for higher-dimensional minimal submanifolds as well with the same proof:

PROPOSITION 4.3. [**C1**]. *If $N \subset \mathbb{R}^n$ is an m-dimensional minimal submanifold and p is a point in $\mathbb{R}^n$, then*

$$\mathrm{Vol}(N) \leq \mathrm{Vol}(p \ast\!\!\ast \partial N).$$

Similarly, the angle estimate (4–2) holds in higher dimension also. To show this we need to define the geometric quantity *angle*:

DEFINITION 4.4. [**C1**]. *Let $M \subset \mathbb{R}^n$ be a k-dimensional rectifiable set and p a point in $\mathbb{R}^n$. We define the* k-dimensional angle *of M from p,* $\mathrm{Angle}(M,p)$ *to be the k-dimensional mass of $(p \ast\!\!\ast M)^\infty \cap \partial B_p^n(1)$ counting multiplicity, where $(p \ast\!\!\ast M)^\infty$ is the infinite cone obtained by indefinitely extending $p \ast\!\!\ast M$ across M and $\partial B_p^n(1)$ is the unit sphere with center at p in $\mathbb{R}^n$.*

Note that

$$\mathrm{Angle}(M,p) = (k+1)\omega_{k+1}\Theta^{k+1}(p \ast\!\!\ast M, p),$$

where ω_{k+1} is the volume of a unit ball in $\mathbb{R}^{k+1}$ and $\Theta^{k+1}(p \ast\!\!\ast M, p)$ denotes the $(k+1)$-dimensional density of $p \ast\!\!\ast M$ at p. Using this, we can also define the angle of a set in a Riemannian manifold.

LEMMA 4.5. [**C1**]. (a) $\log r$ *is subharmonic on a minimal surface $N^2 \subset \mathbb{R}^n$. More precisely, $\frac{1}{2\pi}\triangle \log r \geq \delta_p$, the Dirac delta function with singularity at p.*
(b) *On an m-dimensional minimal submanifold $N \subset \mathbb{R}^n$, $m \geq 3$, $\triangle r^{2-m} \leq m\,\omega_m \delta_p$.*

PROOF. Use $\triangle r^2 = 2n$, $|\nabla r| \leq 1$ on N. And recall that in $\mathbb{R}^m$ $\triangle r^{2-m} = m\,\omega_m \delta_O$ and that near p N approximates to $\mathbb{R}^m$. □

Integrating these Laplacians and using the facts $\frac{\partial r}{\partial \nu} \leq \frac{\partial r}{\partial \eta}$ and $\int_{\partial N} r^{1-m}\frac{\partial r}{\partial \eta} = \mathrm{Angle}(\partial N, p)$, we get the following angle estimate.

PROPOSITION 4.6. [**C1**]. *If p is an interior point of an m-dimensional minimal submanifold $N \subset \mathbb{R}^n$, then*

$$\mathrm{Angle}(\partial N, p) \geq m\,\omega_m$$

and equality holds if and only if N is totally geodesic and star-shaped with respect to p.

5. Hyperbolic Case

The author and Gulliver [**CG1**], [**CG2**] investigated the possibility of extending the cone method to a minimal surface S in S^n and H^n. We showed that the area comparison $\text{Area}(S) \leq \text{Area}(p \ast \partial S)$ does hold for $S \subset H^n$ but not for $S \subset S^n$ (see [**CG1**, Proposition 2, Remark 1]), whereas the angle estimate $\text{Angle}(\partial S, p) \geq 2\pi$ holds for $S \subset S^n$ as well as for $S \subset H^n$ ([**CG2**, Proposition 2]). As a result we proved the isoperimetric inequality

$$4\pi A \leq L^2 - A^2$$

for a minimal surface $S \subset H^n$ whose boundary is radially connected from a point of S.

Just as $p \ast \partial S \subset \mathbb{R}^n$ is flat away from p, $p \ast \partial S \subset H^n$ has constant Gaussian curvature -1 away from p. Therefore we can also develop $p \ast \partial S \subset H^n$ on H^2 as we did $p \ast \partial S \subset \mathbb{R}^n$ on $\mathbb{R}^2$. In H^n, however, we shall take a different approach: apply Bol's isoperimetric inequality. Bol showed that a simply connected domain D on a surface satisfies the inequality

$$4\pi A \leq L^2 + (\sup_D K)A^2,$$

where K is the Gaussian curvature of the surface [**Bo**]. In this section Bol's isoperimetric inequality will be used on a smooth surface which is very close to $p \ast \partial S$.

Given a k-dimensional submanifold N in a Riemannian manifold M, one can define the *extrinsic Laplacian* $\overline{\triangle}$ and *intrinsic Laplacian* $\triangle$ on N as follows. Let $e_1, \ldots, e_k$ be orthonormal vector fields on a domain of N. Define

$$\overline{\triangle} f = \sum_i \overline{\nabla}^2 f(e_i, e_i) \qquad \text{and} \qquad \triangle f = \sum_i \nabla^2 f(e_i, e_i),$$

where $\overline{\nabla}$ and ∇ are the Riemannian connections of M and N, respectively. Then one can easily prove the following.

LEMMA 5.1. *$\overline{\triangle} f = \triangle f - \vec{H} f$, where $\vec{H}$ is the mean curvature vector of N.*

By this lemma the intrinsic Laplacian on a minimal submanifold or a cone can be replaced with the extrinsic Laplacian which is easier to compute.

Let $G(r)$ be a Green's function of S^k (or H^k), whose gradient is $\sin^{1-k} r \nabla r$ (or $\sinh r^{1-k} r \nabla r$, respectively). For instance,

$$G(r) = \log \frac{\sin r}{1 + \cos r} \text{ on } S^2 \quad \text{and} \quad G(r) = \log \frac{\sinh r}{1 + \cosh r} \text{ on } H^2.$$

Note that $\triangle G(r) = k\omega_k \delta_p$. Assuming $S^k \subset S^n$ and $H^k \subset H^n$, one can extend G to a rotationally symmetric function $G(r)$ on S^n and H^n. Then by Lemma 5.1 we have

LEMMA 5.2. [**CG2**] *$\triangle G(r) \geq k\omega_k \delta_p$ on minimal $N \subset S^n$ or H^n, and $\triangle G = k\omega_k \Theta \delta_p$ on $p \ast \partial N$, where Θ is the density of $p \ast \partial N$ at p.*

DEFINITION 5.3. *Let $C \subset S^n$ be a k-dimensional rectifiable set and p a point in S^n. The k-dimensional* angle $\text{Angle}(C, p)$ *of C viewed from p is defined by*

$$\text{Angle}(C, p) = \frac{\text{Vol}[(p \ast C) \cap \partial B_p^n(t)]}{\sin^k t},$$

where $\partial B_p^n(t)$ is the geodesic sphere of radius $t < \operatorname{dist}(p, C)$ centered at p, and the volume is measured counting multiplicity. Clearly the angle does not depend on t. There is obviously an analogous definition for the angle of $C \subset H^n$ viewed from $p \in H^n$.

Note that

$$\operatorname{Angle}(C, p) = (k+1)\omega_{k+1}\Theta^{k+1}(p \ast C, p),$$

where $\Theta^{k+1}(p \ast C, p)$ is the $(k+1)$-dimensional density of $p \ast C$ at p. Now we integrate the Laplacians of Lemma 5.2 on N and on $p \ast \partial N$. Using $\partial r/\partial \nu \leq \partial r/\partial \eta$ as in the previous section, we can prove the following.

PROPOSITION 5.4. [**CG2**]. *Let N be a k-dimensional minimal submanifold in S^n or H^n and p an interior point of N. In case of $N \subset S^n$, we assume that $\operatorname{dist}(p, q) \leq \pi/2$ for all $q \in N$. Then*

$$\operatorname{Angle}(\partial N, p) \geq k\omega_k.$$

Equality holds if and only if N is totally geodesic and star-shaped with respect to p.

Now let $\alpha(r)$ be the volume of the geodesic ball of radius r in H^k and let $f(r)$ be a function on H^k such that $\nabla f = \frac{\alpha}{\alpha'}\nabla r$. Then $f(r) = \log(1 + \cosh r)$ for $k = 2$. If we extend f to a rotationally symmetric function $f(r)$ on $H^n \supset H^k$, then by Lemma 5.1 we have

LEMMA 5.5. [**CG1**] *$\triangle f \geq 1$ on minimal $N^k \subset H^n$ and $\triangle f = 1$ on H^k and a k-dimensional cone in H^n.*

Here the vertex of the cone should be the point p from which the distance r is measured. Therefore integrating these Laplacians, we get the following.

PROPOSITION 5.6. [**CG1**] *Let N be a minimal submanifold of H^n and p any point in H^n. Then*

$$\operatorname{Vol}(N) \leq \operatorname{Vol}(p \ast \partial N);$$

equality holds if and only if $p \in N$, and N must be totally geodesic and star-shaped with respect to p.

If ∂S is radially connected from p, then after applying the cutting and pasting arguments one can think of $p \ast \partial S$ as a cone with connected boundary. Now Proposition 5.4 implies that for $k = 2$ and $N = S$, the Gaussian curvature of $p \ast \partial S$ at p is either $-\infty$ if $\operatorname{Angle}(\partial S, p) > 2\pi$, or -1 if $\operatorname{Angle}(\partial S, p) = 2\pi$. In the case of $-\infty$ one can slightly perturb $p \ast \partial S$ near p to a smooth surface Σ of Gaussian curvature ≤ -1; see [**CG1**, Lemma 3]. Then Bol's isoperimetric inequality for Σ implies

$$4\pi \operatorname{Area}(p \ast \partial S) \leq \operatorname{Length}(\partial S)^2 - \operatorname{Area}(p \ast \partial S)^2$$

(see [**CG1**, Lemma 4] for details). Hence using Proposition 5.6 and the monotonicity of the quadratic function $4\pi A + A^2$ for positive area A, we can prove the following.

THEOREM 5.7. [**CG1**] *If S is a minimal surface in hyperbolic space H^n whose boundary is radially connected from a point of the surface, then S satisfies $4\pi A \leq L^2 - A^2$, with equality if and only if S is a geodesic disk in a totally geodesic $H^2 \subset H^n$.*

6. Variable Curvature Case

So far we have considered minimal surfaces in a Riemannian manifold $\overline{M}$ of constant sectional curvature K. In this section we extend the results of the preceding sections to minimal surfaces in a manifold M of variable curvature $\leq K$. The main obstacle to this extension is that one cannot prove the area comparison

$$\text{Area}(S) \leq \text{Area}(p \ast\!\ast \partial S)$$

for a minimal surface $S \subset M$. But we will get around this difficulty by comparing the area of S with that of a cone in $\overline{M}$.

When we study a Riemannian manifold of variable curvature the comparison theorems are very useful tools. Among them, the one that we need most for our purposes is the Hessian comparison theorem for the distances in M and in $\overline{M}$. Let $r(x)$ be the distance from a fixed point p to x in M and denote the Hessian of r by $\nabla^2 r$. Assume that γ is a geodesic from p to q and v is a vector at q perpendicular to γ. Then $\nabla^2 r(v,v)$ is the second variation of the length of γ associated with the Jacobi field X along γ satisfying $X(p) = 0$ and $X(q) = v$. The Jacobi field minimizes the second variation among all vector fields along γ with the same boundary conditions. Therefore if the sectional curvature of M^n is bounded from above by that of a space form $\overline{M}^n$ which has a distance function $\bar{r}$ with $\bar{r}(\cdot) = \text{dist}(\bar{p}, \cdot)$ and the Riemannian connection denoted by $\overline{\nabla}$, then one gets the Hessian comparison

$$\nabla^2 r(v,v) \geq \overline{\nabla}^2 \bar{r}(u,u), \tag{6–1}$$

where u is a vector at $\bar{q} \in \overline{M}$ with $|u| = |v|$ and $\bar{r}(\bar{q}) = r(q)$, which is perpendicular to the geodesic $\bar{\gamma} \subset \overline{M}$ from $\bar{p}$ to $\bar{q}$ (see [**SY1**] p. 4).

From this comparison one can obtain the following lemmas on the Laplacian of some functions of distance.

LEMMA 6.1. [**C4**] *Let S be a minimal surface in a simply connected Riemannian manifold M of sectional curvature bounded above by a constant K. Define $r(x) = \text{dist}(p,x)$ for fixed $p \in M$.*

(I) *If $K = 0$, we have on S:*

(a) $\triangle r^2 \geq 4$;

(b) $\triangle r \geq \dfrac{1}{r}(2 - |\nabla r|^2)$;

(c) $\triangle \log r \geq 2\pi\delta_p$ *if* $p \in S$.

(II) *If $K = -k^2 < 0$, then*

(d) $\triangle r \geq k(2 - |\nabla r|^2) \coth kr$;

(e) $\triangle \log(1 + \cosh kr) \geq -K$;

(f) $\triangle \log \dfrac{\sinh kr}{1 + \cosh kr} \geq 2\pi\delta_p$ *if* $p \in S$;

(g) $\triangle \log \sinh kr \geq 2\pi\delta_p - K$ *if* $p \in S$.

(III) *If $K = k^2 > 0$, then*

(h) $\triangle r \geq k(2 - |\nabla r|^2) \cot kr$;

(i) $\triangle \log \sin kr \geq 2\pi\delta_p - K$ *if* $p \in S$ *and* $r \leq \dfrac{\pi}{2k}$;

(j) $\triangle \log \dfrac{\sin kr}{1+\cos kr} \geq 2\pi\delta_p$ *if* $p \in S$ *and* $r \leq \dfrac{\pi}{2k}$.

PROOF. Denote the metrics of $\overline{M}$ and M by $\bar{g}$ and g, respectively. Assume first that $K = 0$ and $\overline{M} = \mathbb{R}^n$. A straightforward computation in orthonormal coordinates of $\mathbb{R}^n$ gives

$$\overline{\nabla}^2 \bar{r}^2 = 2\bar{g}. \tag{6–2}$$

Since

$$\nabla^2 r^2 = 2r\nabla^2 r + 2\nabla r \otimes \nabla r \quad \text{and} \quad \overline{\nabla}^2 \bar{r}^2 = 2\bar{r}\overline{\nabla}^2 \bar{r} + 2\overline{\nabla}\bar{r} \otimes \overline{\nabla}\bar{r},$$

(6–1) and (6–2) imply

$$\nabla^2 r^2 \geq 2g.$$

This and Lemma 5.1 imply (a). For (b) we compute

$$\triangle r = \operatorname{div}\nabla(r^2)^{1/2} = \operatorname{div}\frac{1}{2r}\nabla r^2 = \frac{1}{2r}\triangle r^2 - \frac{1}{2r^2}\langle \nabla r,\ 2r\nabla r\rangle \geq \frac{1}{r}(2 - |\nabla r|^2).$$

Similarly for (c)

$$\triangle \log r = \operatorname{div}\frac{1}{r}\nabla r = \frac{1}{r}\triangle r - \frac{1}{r^2}|\nabla r|^2 \geq \frac{2}{r^2}(1 - |\nabla r|^2) \geq 0.$$

Near p, however, S can be identified with T_pS, the tangent plane of S at p, on which $\triangle \log r = 2\pi\delta_p$ with respect to the Euclidean metric. Therefore on S, $\triangle \log r \geq 2\pi\delta_p$.

Assume now that $K = -k^2 < 0$. For any circle $C \subset \overline{M}$ of radius a with center at $\bar{p}$, the length of C equals $l(a) = (2\pi/k)\sinh ka$. So the geodesic curvature of C is $l'(a)/l(a) = k\coth ka$. Hence the principal curvature of the geodesic sphere $\Sigma \subset \overline{M}$ of radius a with center at $\bar{p}$ is $k\coth ka$ everywhere on Σ in any direction. Note here that both the tangent space and the normal line to Σ are the eigenspaces of the Hessian of $\bar{r}$, $\overline{\nabla}^2 \bar{r}$. Therefore one can easily see that

$$\overline{\nabla}^2 \cosh k\bar{r} = (k^2 \cosh k\bar{r})\bar{g}, \tag{6–3}$$

and hence

$$\overline{\nabla}^2 \bar{r} = k\coth k\bar{r}(\bar{g} - \overline{\nabla}\bar{r} \otimes \overline{\nabla}\bar{r}). \tag{6–4}$$

Thus Lemma 5.1, (6–1), (6–4) and the fact that $\nabla r (= \operatorname{grad} r)$ is an eigenvector of $\nabla^2 r$ with eigenvalue zero prove (d). Then

$$\begin{aligned}\triangle \log(1+\cosh kr) &= \operatorname{div}\frac{k\sinh kr}{1+\cosh kr}\nabla r = \frac{k^2}{1+\cosh kr}|\nabla r|^2 + \frac{k\sinh kr}{1+\cosh kr}\triangle r \\ &\geq \frac{2k^2\cosh kr + k^2|\nabla r|^2(1-\cosh kr)}{1+\cosh kr} \geq k^2,\end{aligned}$$

which gives (e). Now we have

$$\begin{aligned}\triangle \log \frac{\sinh kr}{1+\cosh kr} &= \operatorname{div}\frac{k}{\sinh kr}\nabla r = -\frac{k^2\cosh kr}{\sinh^2 kr}|\nabla r|^2 + \frac{k}{\sinh kr}\triangle r \\ &\geq \frac{2k^2\cosh kr(1-|\nabla r|^2)}{\sinh^2 kr} \geq 0.\end{aligned}$$

However,

$$f(r) = \frac{1}{2\pi} \log \frac{\sinh kr}{1 + \cosh kr}$$

is a fundamental solution of $\triangle$ on $H^2(K)$ since

$$\frac{1}{2\pi f'(r)} = \frac{1}{k} \sinh kr$$

is the length of a Jacobi field. So (f) follows. Adding (e) to (f) gives (g).

Assume finally that $K = k^2 > 0$. As in the case of $K < 0$ above, one can show that in $\overline{M}^n$ of constant sectional curvature K,

$$\overline{\nabla}^2 \cos k\bar{r} = -(k^2 \cos k\bar{r})\bar{g}$$

and

$$\overline{\nabla}^2 \bar{r} = k \cot k\bar{r}(\bar{g} - \overline{\nabla}\bar{r} \otimes \overline{\nabla}\bar{r}),$$

from which (h) follows. And then

$$\begin{aligned} \triangle \log \frac{\sin kr}{1 + \cos kr} = \operatorname{div} \frac{k}{\sin kr} \nabla r &= -\frac{k^2 \cos kr}{\sin^2 kr} |\nabla r|^2 + \frac{k}{\sin kr} \triangle r \\ &\geq \frac{2k^2 \cos kr}{\sin^2 kr} (1 - |\nabla r|^2) \geq 0. \end{aligned}$$

As in (f),

$$f(r) = \frac{1}{2\pi} \log \frac{\sin kr}{1 + \cos kr}$$

is a fundamental solution of $\triangle$ on $S^2(K)$ since

$$\frac{1}{2\pi f'(r)} = \frac{1}{k} \sin kr$$

is the length of a Jacobi field. Thus (j) follows. For (i) we compute

$$\begin{aligned} \triangle \log \sin kr = \operatorname{div} \left(\frac{k \cos kr}{\sin kr} \nabla r \right) &= -k^2 \csc^2 kr |\nabla r|^2 + k \cot kr \triangle r \\ &\geq k^2 \csc^2 kr [2 \cos^2 kr - (1 + \cos^2 kr) |\nabla r|^2] \geq -k^2. \end{aligned}$$

Note that

$$\lim_{r \to 0} \frac{\dfrac{d}{dr} \log \sin kr}{\dfrac{d}{dr} \log \dfrac{\sin kr}{1 + \cos kr}} = 1,$$

which proves (i). □

LEMMA 6.2. [**C4**] *Let $\Gamma = \bar{p} \times \bar{C}$ be the cone from $\bar{p}$ over a curve $\bar{C}$ in a Riemannian manifold $\overline{M}$ of nonpositive constant sectional curvature $K = -k^2$ and define $\bar{r}(x) = \operatorname{dist}(\bar{p}, x), \bar{p} \in \overline{M}$. Then, on Γ,*

(a) $\triangle \bar{r}^2 = 4$ *if* $K = 0$, *while* $\triangle \log(1 + \cosh k\bar{r}) = -K$ *if* $K < 0$.

(b) $\triangle \log \bar{r} = \alpha \delta_{\bar{p}}$ *if* $K = 0$, *while* $\triangle \log \dfrac{\sinh k\bar{r}}{1 + \cosh k\bar{r}} = \alpha \delta_{\bar{p}}$ *if* $K < 0$, *where* $\alpha = \operatorname{Angle}(\bar{C}, \bar{p})$.

PROOF. On Γ, $\nabla \bar{r}$ is perpendicular to $\vec{H}$, the mean curvature vector of Γ; hence Lemma 5.1 implies that for any function f of distance $\bar{r}$, $\overline{\triangle} f = \triangle f$. Moreover $|\nabla \bar{r}| \equiv 1$ on Γ. It follows from (6–2) and (6–3) that all the inequalities in Lemma 6.1(a), (c), (e) and (f) become equalities. This proves the lemma except for the constant α. The constant 2π that appears in the Laplacian of the fundamental solution on $\mathbb{R}^2$ and H^2 comes from the limit as $a \to 0$ of the circumference of the circle of radius a with center at $\bar{p}$ divided by a. Similarly, if $\Sigma_{\bar{p}}(a)$ denotes the geodesic sphere of radius a with center at $\bar{p}$, α equals

$$\lim_{a \to 0} \frac{1}{a} \operatorname{Length}(\Gamma \cap \Sigma_{\bar{p}}(a)),$$

which is called the *angle* of $\bar{C}$ viewed from $\bar{p}$ and denoted $\operatorname{Angle}(\bar{C}, \bar{p})$. □

Now we have the main theorem as follows.

THEOREM 6.3. [**C4**] *Let S be a minimal surface in a complete simply connected Riemannian manifold M with sectional curvature bounded above by a nonpositive constant K. If ∂S is radially connected from a point of S, then S satisfies the isoperimetric inequality*

$$4\pi A \le L^2 + KA^2, \tag{6–5}$$

where equality holds if and only if S is a geodesic disk in a surface of constant Gaussian curvature K.

PROOF. First suppose $K < 0$. Integrate Lemma 6.1(e) to get

$$\begin{aligned} -K \operatorname{Area}(S) &\le \int_S \triangle \log(1 + \cosh kr) = \int_{\partial S} \frac{k \sinh kr}{1 + \cosh kr} \frac{\partial r}{\partial \nu} \\ &\le \int_{\partial S} \frac{k \sinh kr}{1 + \cosh kr} \frac{\partial r}{\partial \eta} = \int_{\partial S} \frac{k \sinh kr}{1 + \cosh kr} \sqrt{1 - \langle \nabla r, \tau \rangle^2}, \end{aligned}$$

where ν, η are as in the preceding sections and τ is a unit tangent to ∂S.

Now the key step in the extension to the variable curvature case is to carry the last integral above over to the simply connected space form $\overline{M}$ of sectional curvature K. Let $C_1, \ldots, C_l$ be the components of ∂S. Choose $q_i \in C_i$ for each $i = 1, \ldots, l$, and take $\bar{q}_1, \ldots, \bar{q}_l \in \overline{M}$ in such a way that $r(q_i) = \bar{r}(\bar{q}_i)$. Suppose that each curve C_i is parametrized by $c_i(s)$ with arclength parameter s such that $q_i = c_i(0) = c_i(\lambda_i)$, $\lambda_i = \operatorname{Length}(C_i)$. Then we construct a curve $\bar{C}_i$ in $\overline{M}$ starting from $\bar{q}_i$ and parametrized by $\bar{c}_i(s)$ with arclength parameter $s \in [0, \lambda_i]$ and $\bar{c}_i(0) = \bar{q}_i$ such that the unit tangent vector $\bar{c}_i'(s)$ makes an angle of $\cos^{-1}\langle \nabla r, c_i'(s) \rangle$ with $\nabla \bar{r}$. Of course the curve $\bar{C}_i$ is not unique; but given a two-dimensional infinite cone $\bar{p} \ast \bar{C}$ containing $\bar{q}_i$, one can uniquely determine a curve $\bar{C}_i$ on $\bar{p} \ast \bar{C}$ with the prescribed properties. Since $\bar{p} \ast \bar{C}$ is developable, one can also assume without loss of generality that $\bar{c}_i(0) = \bar{c}_i(\lambda_i)$, or equivalently, that $\bar{C}_i$ is closed. Anyhow, r on C_i coincides with $\bar{r}$ on $\bar{C}_i$ in the sense that

$$r(c_i(s)) = \bar{r}(\bar{c}_i(s)) \quad \text{and} \quad \langle \nabla r, c_i'(s) \rangle = \langle \nabla \bar{r}, \bar{c}_i'(s) \rangle.$$

Hence

$$-K\,\mathrm{Area}(S) \le \sum_{i=1}^{l} \int_{C_i} \frac{k\sinh kr}{1+\cosh kr}\sqrt{1-\langle \nabla r, c_i'(s)\rangle^2}$$
$$= \sum_{i=1}^{l} \int_{\bar{C}_i} \frac{k\sinh k\bar{r}}{1+\cosh k\bar{r}}\sqrt{1-\langle \nabla \bar{r}, \bar{c}_i'(s)\rangle^2}.$$

If $\bar{\eta}$ is the outward unit conormal to $\bar{C}_i$ on $\bar{p}\ast\bar{C}_i$, then

$$\mathrm{Area}(S) \le -\frac{1}{K}\sum_{i=1}^{l}\int_{\bar{C}_i}\frac{k\sinh k\bar{r}}{1+\cosh k\bar{r}}\frac{\partial \bar{r}}{\partial \bar{\eta}} = \sum_{i=1}^{l}\int_{\bar{p}\ast\bar{C}_i}\frac{1}{-K}\triangle\log(1+\cosh k\bar{r})$$
$$= \sum_{i=1}^{l}\mathrm{Area}(\bar{p}\ast\bar{C}_i) \ \ (\text{by Lemma 6.2(a)}) = \mathrm{Area}(\bar{p}\ast\bar{C}),$$

with $\bar{C} = \bigcup_{i=1}^{l}\bar{C}_i$. Also it follows from the definition of $\bar{C}_i$ that

$$\mathrm{Length}(\partial S) = \mathrm{Length}(\bar{C}).$$

On the other hand, integrating Lemma 6.1(f) over S and Lemma 6.2(b) over $\bar{p}\ast\bar{C}$ as above, we get

$$2\pi \le \int_S \triangle\log\frac{\sinh kr}{1+\cosh kr} = \int_{\partial S}\frac{k}{\sinh kr}\frac{\partial r}{\partial \nu} \le \int_{\partial S}\frac{k}{\sinh kr}\frac{\partial r}{\partial \eta}$$
$$= \int_{\bar{C}}\frac{k}{\sinh k\bar{r}}\frac{\partial \bar{r}}{\partial \bar{\eta}} = \int_{\bar{p}\ast\bar{C}}\triangle\log\frac{\sinh k\bar{r}}{1+\cosh k\bar{r}} = \mathrm{Angle}(\bar{C},\bar{p}).$$

Moreover, since $r|_{\partial S}$ coincides with $\bar{r}|_{\bar{C}}$, $\bar{C}$ is also radially connected from $\bar{p}$. Hence from the cutting and pasting arguments and the approximation argument as in [**CG1**, Lemma 4] it follows that

$$4\pi\mathrm{Area}(\bar{p}\ast\bar{C}) \le \mathrm{Length}(\bar{C})^2 + K\mathrm{Area}(\bar{p}\ast\bar{C})^2.$$

Therefore using the area comparison obtained above and the monotonicity of the quadratic function $4\pi A - KA^2$ of $A > 0$, we obtain the desired isoperimetric inequality for S in case $K < 0$.

If equality holds in the isoperimetric inequality, then

$$\mathrm{Area}(S) = \mathrm{Area}(\bar{p}\ast\bar{C})$$

and therefore equality should hold in Lemma 6.1(e). Consequently equality holds in (6–1) and $|\nabla r| \equiv 1$ on S as we easily see in the proof of Lemma 6.1(e). It follows that $S = p\ast\partial S$ and, by the Index Lemma, S is constantly curved and hence totally geodesic. Thus Schmidt's theorem [**Sm**] completes the proof in case $K < 0$.

Second, suppose $K = 0$. Lemma 6.1(a) and Lemma 6.2(a) imply

$$\mathrm{Area}(S) \le \mathrm{Area}(\bar{p}\ast\bar{C}),$$

and Lemma 6.1(c) and Lemma 6.2(b) imply

$$2\pi \le \mathrm{Angle}(\bar{C},\bar{p}).$$

Thus the theorem follows from the arguments of Section 4. □

In showing $\text{Area}(S) \leq \text{Area}(\bar{p} \ast \bar{C})$ in the proof of Theorem 6.3, the main idea is to construct the cone $\bar{p} \ast \bar{C}$ in $\overline{M}$ such that

$$\int_{\partial S} \frac{k \sinh kr}{1 + \cosh kr} \frac{\partial r}{\partial \eta} = \int_{\bar{C}} \frac{k \sinh k\bar{r}}{1 + \cosh k\bar{r}} \frac{\partial \bar{r}}{\partial \bar{\eta}},$$

in other words,

$$r|_{\partial S} = \bar{r}|_{\bar{C}} \text{ and } (\partial r/\partial \eta)|_{\partial S} = (\partial \bar{r}/\partial \bar{\eta})|_{\bar{C}}.$$

It is interesting to remark that this idea can be interpreted as *giving a constant curvature metric* $\hat{g}$ to $p \ast \partial S$ that preserves r, the length of ∂S and the angle between ∇r and ∂S. This new metric $\hat{g}$ plays a key role in the proof of the embeddedness of a minimal surface in a Riemannian manifold in [**CG3**].

7. Varifolds and Flat Chains

Some minimal surfaces in $\mathbb{R}^3$, like compound soap films, contain singular curves. They are not smooth but smooth almost everywhere. In some literature they are called stationary varifolds or area minimizing currents. Here it is interesting to ask whether the isoperimetric inequality $4\pi A \leq L^2$ holds also for these surfaces with singularities. In [**C2**] the author gave an affirmative answer; moreover he derived a new type of optimal isoperimetric inequality for certain types of soap films with singularities.

First we state sharp isoperimetric inequalities for domains in the plane where only a specific part of the boundary counts toward the length.

LEMMA 7.1. [**C2**] *Let l_1 and l_2 be the rays emanating from a point O with an angle of $\theta \leq \pi$. Let C be a curve from a point of l_1 to a point of l_2.*

(a) *Suppose that C lies in the smaller sector of the two formed by the rays (C may lie in either sector if $\theta = \pi$). Define D as the domain bounded by l_1, l_2, and C. Then*

$$2\theta \, \text{Area}(D) \leq \text{Length}(C)^2,$$

and equality holds if and only if C is a circular arc perpendicular to the rays.

(b) *If C lies in the larger sector, then*

$$2\pi \, \text{Area}(D) \leq \text{Length}(C)^2,$$

where equality holds if and only if C is a semicircle perpendicular to one of the two rays.

DEFINITION 7.2. *A* compound Jordan curve *is a one-dimensional rectifiable connected set in $\mathbb{R}^n$ which is the union of finitely many Jordan curves (= homeomorphic images of a circle).*

LEMMA 7.3. [**C2**] *If C is a compound Jordan curve in $\mathbb{R}^n$ and p a point of C, then*

$$4\pi \, \text{Area}(p \ast C) \leq \text{Length}(C)^2.$$

Equality holds if and only if $p \ast C$ can be developed, by cutting and inserting, one-to-one onto a disk.

DEFINITION 7.4. (a) *Suppose V is an m-dimensional varifold of locally bounded first variation in $\mathbb{R}^n$, and that Z is the generalized boundary of V with the generalized boundary measure σ. Assume Z is $(m-1)$-rectifiable. Let*

$$\psi(x) = \lim_{\rho \to 0} \frac{\sigma(B_x(\rho))}{\mathsf{H}^{m-1}(Z \cap B_x(\rho))}, \quad x \in Z.$$

Then define the varifold boundary *∂V of V to be the varifold $\underline{\underline{v}}(Z, \psi)$. In other words, ∂V is the $(m-1)$-dimensional rectifiable varifold with support Z and multiplicity ψ. Clearly $\mu_{\partial V} = \sigma$.*

(b) *For an m-varifold $V = \underline{\underline{v}}(M, \theta)$, the* varifold cone *$p \divideontimes V$ from p over V is the $(m+1)$-varifold $\underline{\underline{v}}(p \divideontimes M, \bar{\theta})$, where $\bar{\theta}(y) = \theta(x)$ whenever y lies on the line segment from p to $x \in M$.*

PROPOSITION 7.5. [**C2**] *Let V be an m-varifold of locally bounded first variation in $\mathbb{R}^n$. If the generalized boundary Z of V is rectifiable and V is stationary in $\mathbb{R}^n \sim Z$, then for any $p \in \mathbb{R}^n$,*

$$\mathbf{M}(V) \leq \mathbf{M}(p \divideontimes \partial V).$$

PROOF. By the first variation formula for the mass of V with the variation field Y,

$$\delta V(Y) = \int_Z \langle \nu, Y \rangle \, d\sigma.$$

Take Y to be the radial vector field defined by $Y(x) = x - p$. Then Y is the initial velocity vector field of the 1-parameter family of homothetic expansions $\{\phi_t\}$ given by $\phi_t(x) = (1+t)(x-p) + p$. Hence

$$\mathbf{M}(\phi_{t\sharp} V) = (1+t)^m \mathbf{M}(V),$$

and so

$$\delta V(Y) = \frac{d}{dt}(1+t)^m \mathbf{M}(V)|_{t=0} = m\mathbf{M}(V).$$

On the other hand, since Z is rectifiable, Z has tangent spaces almost everywhere and ν is normal to Z. Let $\eta(x)$ be a unit vector which is perpendicular to Z at $x \in Z$ and lies in the subspace of $\mathbb{R}^n$ spanned by $Y(x) = x - p$ and the tangent space to Z at x. Taking the negative of η if necessary, one may assume $\langle \eta, Y \rangle \geq 0$. It is not difficult to see that

$$\langle \nu, Y \rangle \leq \langle \eta, Y \rangle.$$

Let $r(x) = |Y(x)|$. Then dr is the 1-form dual to the unit radial vector field $Y/|Y|$. Hence

$$m\,\mathbf{M}(V) = \int_Z \langle \nu, Y \rangle \, d\sigma \leq \int_Z \langle \eta, Y \rangle \, d\sigma = \int_Z Y \lrcorner (dr \wedge d\sigma) = m\,\mathbf{M}(p \divideontimes \partial V),$$

where $\lrcorner$ denotes the interior multiplication. □

LEMMA 7.6. [**C2**] *Let $W = \underline{\underline{v}}(Z, \psi)$ be a rectifiable 1-varifold in $\mathbb{R}^n$ with $\psi \geq 1$ and let p be a point in Z. If Z is a compound Jordan curve, then*

$$4\pi \mathbf{M}(p \divideontimes W) \leq \mathbf{M}(W)^2.$$

THEOREM 7.7. [**C2**] *Suppose that V is a 2-varifold of locally bounded first variation in $\mathbb{R}^n$, the generalized boundary Z of V is rectifiable, and V is stationary in $\mathbb{R}^n \sim Z$. If the multiplicity of ∂V is ≥ 1 and Z is a compound Jordan curve, then*

$$4\pi\, \mathbf{M}(V) \leq \mathbf{M}(\partial V)^2.$$

PROOF. Use Proposition 7.5 and Lemma 7.6. □

The inequality of Theorem 7.7 resembles the classical isoperimetric inequality. But here we shall see a new optimal isoperimetric inequality for some specific soap films as follows.

DEFINITION 7.8. *Let $Y^k \subset \mathbb{R}^3$ be the union of k great semicircles on a sphere meeting at the north and south poles at equal angles of $2\pi/k$. Define $\mathbf{Y}_2^k \subset \mathbf{I}_2^k$ to be the set of 2-dimensional flat chains T mod k in $\mathbb{R}^n$ with multiplicity 1 almost everywhere such that* $\mathrm{spt}\partial T$ *is homeomorphic to Y^k and the associated varifold $V = \underline{\underline{v}}(\mathrm{spt}T, \theta)$ is locally of bounded first variation in $\mathbb{R}^n$.*

THEOREM 7.9. [**C2**] *Suppose that T is a 2-dimensional area minimizing flat chain mod k in $\mathbf{Y}_2^k$. If $C_1, C_2, \ldots, C_k$ are the curves that constitute* $\mathrm{spt}\partial T$ *and have common end points, then*

$$2\pi\, \mathbf{M}(T) \leq \sum_{i=1}^{k} \mathrm{Length}(C_i)^2.$$

And equality holds if and only if $\mathrm{spt}\, T$ *is the union of k flat half-disks meeting each other along the common diameter.*

Let Y be the union of three half-disks meeting each other along their common diameter at equal angles of 120 degrees. Let T be the intersection with the unit ball $B_p(1)$ of an infinite cone from p through the 1-skeleton of a regular tetrahedron whose center of mass is p. In [**Ta**] J. Taylor proved that the disk, Y and T are the only three cones that are area minimizing under Lipschitz maps leaving the boundary fixed. In view of this fact and Theorem 7.9, we would like to propose the following problem.

OPEN PROBLEM 7.10. *Suppose that V is a 2-varifold with multiplicity 1 almost everywhere and is locally of bounded first variation in $\mathbb{R}^n$ such that V is stationary outside the rectifiable boundary* $\mathrm{spt}\,\partial V$. *Suppose also that* $\mathrm{spt}V$ *is homeomorphic to T. Let $C_1, C_2, \ldots, C_6 \subset \mathrm{spt}\partial V$ be the curves that constitute* $\mathrm{spt}\,\partial V$ *and lie between four junctions of* $\mathrm{spt}\,\partial V$. *Show that*

$$2\cos^{-1}(-\tfrac{1}{3})\, \mathbf{M}(V) \leq \sum_{i=1}^{6} \mathrm{Length}(C_i)^2,$$

where equality holds if and only if $\mathrm{spt}V$ *is a homothetic expansion (or contraction) of T.*

8. Higher-Dimensional Minimal Submanifolds

So far we have considered two-dimensional minimal surfaces only. In this section we study the isoperimetric inequality of higher-dimensional minimal submanifolds.

Given a domain D in $\mathbb{R}^m$ it is well known that if ω_m is the volume of a unit ball in $\mathbb{R}^m$, then

$$m^m\omega_m \mathrm{Vol}(D)^{m-1} \leq \mathrm{Vol}(\partial D)^m \tag{8–1}$$

and equality holds if and only if D is a ball. In view of Open Problem 1.1 it is tempting to conjecture

OPEN PROBLEM 8.1. *Any m-dimensional minimal submanifold N of $\mathbb{R}^n$ satisfies the classical isoperimetric inequality*

$$m^m\omega_m \mathrm{Vol}(N)^{m-1} \leq \mathrm{Vol}(\partial N)^m,$$

where equality holds if and only if N is a ball in an m-plane of $\mathbb{R}^n$.

This problem is far less settled than Open Problem 1.1. The only two cases that are known to hold are i) when ∂N lies on the $(n-1)$-dimensional sphere centered at a point of N (by monotonicity) and (ii) when N is area minimizing (by Almgren [**Al**]).

8.1. Monotonicity. Closely related to the isoperimetric inequality of a minimal submanifold N is the monotonicity property: the volume of $N \cap B_p(r)$ divided by the volume of the geodesic ball of radius r is a nondecreasing function of r. This property has been proved for $\mathbb{R}^n$ and $\mathbf{H}^n$ [**An**], but not for $\mathbf{S}^n_+$ (but see [**GS**], p. 353).

LEMMA 8.2. (Monotonicity) *Let N be an m-dimensional minimal submanifold in $\mathbb{R}^n$ and r the distance in $\mathbb{R}^n$ from $p \in \mathbb{R}^n$. Then $\mathrm{Vol}(N \cap B_p(r))/r^m$ is a monotonically nondecreasing function of r for $0 < r < \mathrm{dist}(p, \partial N)$.*

PROOF. Write $N_r = N \cap B_p(r)$. Integrate (3–1) ($\triangle r^2 = 2m$) to obtain

$$m\mathrm{Vol}(N_r) = \frac{1}{2}\int_{N_r} \triangle r^2 = \int_{\partial N_r} r|\nabla r|. \tag{8–2}$$

Denote the volume forms on N and ∂N by dv and dS_r, respectively. Then we have

$$dv = \frac{1}{|\nabla r|} dS_r dr.$$

Then

$$\frac{d}{dr}\int_{N_r} |\nabla r|^2 dv = \int_{\partial N_r} |\nabla r| dS_r.$$

Hence

$$m\mathrm{Vol}(N_r) = r\frac{d}{dr}\int_{N_r} |\nabla r|^2\, dv = r\frac{d}{dr}\mathrm{Vol}(N_r) - r\frac{d}{dr}\int_{N_r}\left(1 - |\nabla r|^2\right) dv \leq r\frac{d}{dr}\mathrm{Vol}(N_r).$$

In the inequality above we used the fact that $|\nabla r| \leq 1$ on N. Hence

$$\frac{d}{dr}\,\frac{\mathrm{Vol}(N_r)}{r^m} \geq 0. \qquad \square$$

THEOREM 8.3. *Let N be an m-dimensional minimal submanifold of $\mathbb{R}^n$. If ∂N lies in a sphere centered at a point p of N, then*

$$m^m\omega_m \mathrm{Vol}(N)^{m-1} \leq \mathrm{Vol}(\partial N)^m.$$

Equality holds if and only if N is a ball.

PROOF. Let R be the radius of the sphere. Applying $|\nabla r| \leq 1$ to (8–2) gives

$$m\mathrm{Vol}(N) = \int_{\partial N} r|\nabla r| \leq R \cdot \mathrm{Vol}(\partial N).$$

Since $\lim_{r\to 0} \mathrm{Vol}(N_r)/r^m = \omega_m$, from Lemma 8.2 we get

$$\omega_m \leq \mathrm{Vol}(N)/R^m.$$

Hence

$$m\mathrm{Vol}(N) \leq \omega_m^{-1/m}\mathrm{Vol}(N)^{1/m}\mathrm{Vol}(\partial N),$$

which gives the desired inequality. Equality holds if and only if N is a cone with density at the center equal to 1 if and only if N is a ball. □

8.2. Almgren's proof. We have seen at the end of Section 1 that Open Problem 1.1 is true for area minimizing surface. Likewise it is shown by Almgren [**Al**] that Open Problem 8.1 is true for area minimizing submanifolds in $\mathbb{R}^n$. A submanifold is said to be *area minimizing* if its volume is less than or equal to the volume of every other submanifold having the same boundary.

THEOREM 8.4. [**Al**] *If N is an m-dimensional area minimizing submanifold of $\mathbb{R}^n$, then*

$$m^m\omega_m\mathrm{Vol}(N)^{m-1} \leq \mathrm{Vol}(\partial N)^m$$

with equality if and only if N is a ball.

Although the proof involves some complicated technicalities, the basic idea is very elegant and not hard to understand. So we shall introduce the argument sketched by B. White in [**Wh**] for the special case of 2-dimensional surfaces in $\mathbb{R}^3$ which generalizes to higher dimension.

PROPOSITION 8.5. [**Al**], [**Wh**] *If S is an area minimizing surface in $\mathbb{R}^3$ with area π, then its perimeter is greater than or equal to 2π, with equality if and only if S is a disk.*

PROOF. Among all area minimizing surfaces with area π, let S be the one whose perimeter is as short as possible. It follows that S minimizes the ratio

$$\frac{\mathrm{Length}(\partial S)^2}{\mathrm{Area}(S)}$$

among all area minimizing surfaces (because the ratio is invariant with respect to dilations). Of course since a disk of area π and circumference 2π is area minimizing, ∂S must have length $\leq 2\pi$; our goal is to show that it has length exactly 2π.

Let C_t be a 1-parameter family of curves with $C_0 = \partial S$, and for each C_t let S_t be a surface of least area with boundary $\partial S_t = C_t$. Because $\mathrm{Length}(C_t)^2/\mathrm{Area}(S_t)$ attains its minimum at $t = 0$ we have

$$0 = \Big(\frac{d}{dt}\Big)_{t=0} \frac{\mathrm{Length}(C_t)^2}{\mathrm{Area}(S_t)}$$

and therefore

$$(8\text{–}3)\quad 0 = \frac{2}{\pi}\,\mathrm{Length}(C_0)\Big(\frac{d}{dt}\Big)_{t=0}\mathrm{Length}(C_t) - \frac{1}{\pi^2}\,\mathrm{Length}(C_0)^2\Big(\frac{d}{dt}\Big)_{t=0}\mathrm{Area}(S_t).$$

Now if in the 1-parameter family of curves C_t, the initial velocity of each point $x \in C_0$ is $v(x)$, then

$$\left(\frac{d}{dt}\right)_{t=0} \text{Length}(C_t) = -\int_{C_0} \langle v, \kappa \rangle \quad \text{and} \quad \left(\frac{d}{dt}\right)_{t=0} \text{Area}(S_t) = \int_{C_0} \langle v, \nu \rangle,$$

where $\kappa(x)$ is the curvature vector of C_0 at x and ν is the outward unit conormal to ∂S. The first of these formulas is the first variation formula for the length of the curves C_t and the second is the first variation formula for the area of the minimal surfaces S_t. Combining these formulas with (8–3) gives

$$0 = \int_{C_0} \big\langle v,\, 2\pi\kappa + \text{Length}(C_0)\nu \big\rangle.$$

Because this holds for every vector field v, it follows that $2\pi\kappa + \text{Length}(C_0)\nu$ must be identically 0. That is, the curve C_0 has constant curvature $\frac{1}{2\pi}\text{Length}(C_0)$. Recall that $\text{Length}(C_0) \leq 2\pi$ (we are trying to prove equality), so C_0 has curvature everywhere less than or equal to 1. The proposition then follows from the following result.

PROPOSITION 8.6. [**Al**], [**Wh**] *If C is a closed k-dimensional submanifold in $\mathbb{R}^m$ with mean curvature everywhere less than or equal to k, then the volume of C is greater than or equal to the volume of the unit k-sphere, with equality if and only if C is congruent to the unit k-sphere.*

PROOF FOR $k = 1$, $m = 3$. Let K be the convex hull of C. Let $n : \partial K \to S^2$ be the Gauss map, which assigns to each $x \in \partial K$ the outward unit normal $n(x)$ to K at x. Note that at a corner of K there are many outward unit normals so the map n is multivalued.

The first observation is that n maps $\partial K \sim C$ to a set of zero area in S^2. To see this, consider for example the case where C consists of two congruent circles with one above the other so that K is a cylinder. Then n maps the sides of the cylinder to a great circle and the top and bottom to a pair of points.

On the other hand, the image of ∂K under n is all of S^2, so the image of $C \cap \partial K$ under n must have area 4π. It is not too hard to see that an infinitesimal piece of C at x of length ds is mapped to a set of area at most

$$2\,|\kappa(x)|\,ds,$$

which is less than or equal to $2\,ds$ because $|\kappa(x)| \leq 1$. Thus

$$\frac{\text{Area}(n(C))}{\text{Length}(C)} \leq 2.$$

But the area of $n(C)$ is 4π, therefore the length of C is at least 2π. □

The key ingredients in the proof of Proposition 8.5 are the existence of the area minimizing surface S_t and the continuity of $\text{Area}(S_t)$ to $\text{Area}(S)$ as $t \to 0$, both of which would not be valid if S were a general (not necessarily area minimizing) minimal surface.

9. The Calibration Method

There are numerous proofs for the original problem of Dido; among these we will introduce the most recent one given by Hélein [**He**] in 1994. His proof, in fact, holds even for curves on a sphere and on a hyperbolic plane as follows.

THEOREM 9.1. *Let S be a surface of Gaussian curvature K ($=1$ or -1). If a closed curve C on S has length L and encloses a domain of area A, then*

$$4\pi A \leq L^2 + KA^2,$$

where equality holds if and only if C is a geodesic circle.

PROOF [**He**]. Let D be the domain enclosed by the smooth curve C on S. (x, y) denotes a point in $D \times \partial D$. For fixed $y \in \partial D$, cover D with the set of all circular arcs emanating from y and perpendicular to ∂D at y. Let $V(x, y)$ be the unit tangent vector to the arc pointing away from y. Then $V(x, y)$ is a unit vector field on $D \times \partial D$. One can easily compute

$$\mathrm{div} V = \frac{1 + f'(r)}{f(r)} \langle \nu, \nabla r \rangle$$

where ν is the unit inward normal to ∂D at y, $r = \mathrm{dist}(x, y)$ and $f(r) = r$, $\sin r$ or $\sinh r$ depending on whether $K = 0$, $K = +1$ or $K = -1$. Moreover one can show that if $r = \mathrm{dist}(x, z), x, z \in D$, then

$$\mathrm{div} \frac{1 + f'(r)}{f(r)} \nabla r = 4\pi \delta_x - K.$$

Let ω, dl be the volume forms of D, ∂D, respectively. Then we have a two-form $\alpha = V \lrcorner\, \omega \wedge dl$ where $\lrcorner$ denotes the interior multiplication such that

$$d\alpha = \mathrm{div} V \omega \wedge dl = \frac{1 + f'(r)}{f(r)} \langle \nu, \nabla r \rangle \, \omega \wedge dl.$$

Therefore

$$\begin{aligned} \int_{D \times \partial D} d\alpha &= \int_D \Big(\int_{\partial D} \frac{1 + f'(r)}{f(r)} \langle \nu, \nabla r \rangle \, dl \Big) \omega = \int_D \left(\int_D d\Big(\frac{1 + f'(r)}{f(r)} \nabla r \lrcorner\, \omega \Big) \right) \omega \\ &= \int_D \left(\int_D \mathrm{div} \Big(\frac{1 + f'(r)}{f(r)} \nabla r \Big) \omega \right) \omega = \int_D \Big(\int_D (4\pi \delta_x - K)\, \omega \Big) \omega \\ &= \int_D (4\pi - KA)\, \omega = 4\pi A - KA^2. \end{aligned}$$

On the other hand

$$\int_{D \times \partial D} d\alpha = \int_{\partial D \times \partial D} \alpha \;\leq\; \int_{\partial D \times \partial D} dl \wedge dl = L^2. \tag{9–1}$$

Thus we get $4\pi A \leq L^2 + KA^2$. In (9–1) inequality becomes equality if and only if V is perpendicular to ∂D at the end $\neq y$ of the circular arc, which occurs if and only if ∂D is a circle. □

10. Weak Isoperimetric Inequalities

It would be beautiful if Hélein's argument generalized to minimal surfaces as well. But various attempts made by the author ended up with no results. In this section, however, we will exploit Simon's argument which resembles Hélein's (see [**CG2**], p. 181) and obtain an isoperimetric inequality which is not sharp but which holds for *all* minimal surfaces. We will also present Ros's argument which improves Simon's inequality, and derive Sobolev-type inequalities related with the weak isoperimetric inequalities.

10.1. Weak inequalities.

THEOREM 10.1. [**C4**], [**CG2**] *Let S be a minimal surface in a complete simply connected Riemannian manifold M with sectional curvature bounded above by a constant K. If $K \leq 0$, then*

$$2\pi A \leq L^2 + KA^2. \tag{10–1}$$

In case $K > 0$, (10–1) *holds under the additional assumption* $\operatorname{diam}(S) \leq \frac{\pi}{2\sqrt{K}}$.

PROOF. (i) $K = -k^2 < 0$. Integrating Lemma 6.1(g) for fixed $p \in S$, we get

$$2\pi - KA \leq \int_S \triangle \log \sinh kr \leq \int_{\partial S} k \coth kr. \tag{10–2}$$

Recall that $r(x) = \operatorname{dist}(p, x)$ for fixed $p \in M$. Since (10–2) holds for any $p \in S$ we can let p vary on S and integrate (10–2) over S and apply Fubini's theorem to obtain

$$\begin{aligned} 2\pi A - KA^2 &\leq \int_S \int_{\partial S} k \coth kr = \int_{\partial S} \int_S k \coth kr \\ &\leq \int_{\partial S} \int_S \triangle r \text{ (by Lemma 6.1(d))} \\ &= \int_{\partial S} \int_{\partial S} \frac{\partial r}{\partial \nu} \leq L^2. \end{aligned}$$

(ii) $K = 0$. Integrate Lemma 6.1(c) twice and apply Lemma 6.1(b) as above.
(iii) $K > 0$. Integrate Lemma 6.1(i) twice and apply Lemma 6.1(h). □

THEOREM 10.2. [**CG2**] *Let N be an m-dimensional minimal submanifold of a complete simply connected Riemannian manifold M^n with sectional curvature bounded above by a negative constant $-k^2$. Then*

$$k(m-1)\operatorname{Vol}(N) \leq \operatorname{Vol}(\partial N).$$

PROOF. On a space form $\overline{M}^n$ of sectional curvature $-k^2$, we have $\overline{\nabla}^2 r = k \coth kr(g - dr \otimes dr)$, where g is the metric tensor of $\overline{M}$. Hence the Hessian comparison (6–1) and Lemma 5.1 imply that on N,

$$\triangle r \geq k(m - |\nabla r|^2) \coth kr. \tag{10–3}$$

Therefore

$$\begin{aligned} \triangle \log \cosh kr &= \operatorname{div}\left(k \frac{\sinh kr}{\cosh kr} \nabla r\right) = \frac{k^2}{\cosh^2 kr} |\nabla r|^2 + k \frac{\sinh kr}{\cosh kr} \triangle r \\ &\geq k^2(m - |\nabla r|^2) \geq k^2(m-1). \end{aligned}$$

Hence

$$k(m-1)\operatorname{Vol}(N) \leq \frac{1}{k} \int_N \triangle \log \cosh kr \leq \int_{\partial N} \frac{\sinh kr}{\cosh kr} \leq \operatorname{Vol}(\partial N). \qquad \square$$

10.2. Simon's and Ros's methods. The weak isoperimetric inequality

$$2\pi A \le L^2$$

for all minimal surfaces in $\mathbb{R}^n$ was originally proved by L. Simon (see [**Bm**, p. 318] or [**O2**, p. 1210]). We review Simon's argument briefly. As in (3–1), we have $\triangle r^2 = 4$ on a minimal surface $S \subset \mathbb{R}^n$. Hence

$$\triangle \log r = \frac{2}{r^2}(1 - |\nabla r|^2) \ge 2\pi\delta_p \tag{10–4}$$

and

$$\triangle r = \frac{1}{r}(2 - |\nabla r|^2) \ge \frac{1}{r}. \tag{10–5}$$

Integrating (10–4) over S yields

$$2\pi \le \int_{y\in S} \triangle \log r_x(y) \le \int_{y\in\partial S} \frac{1}{r_x(y)} \frac{\partial r_x(y)}{\partial \nu} \le \int_{y\in\partial S} \frac{1}{r_x(y)}, \tag{10–6}$$

where $r_x(y) = \operatorname{dist}(x, y)$ for some fixed $x \in \mathbb{R}^n$ and ν is the outward unit conormal to ∂S. Integrating (10–5) over S gives

$$\int_{x\in S} \frac{1}{r_y(x)} \le \int_{x\in S} \triangle r_y(x) = \int_{x\in\partial S} \frac{\partial r_y(x)}{\partial \nu}. \tag{10–7}$$

Let x vary over S, integrate (10–6) over S and use Fubini's theorem and (10–7) to get

$$\begin{aligned} 2\pi A &\le \int_{x\in S}\int_{y\in\partial S} \frac{1}{r_x(y)} = \int_{y\in\partial S}\int_{x\in S} \frac{1}{r_x(y)} = \int_{y\in\partial S}\int_{x\in S} \frac{1}{r_y(x)} \\ &\le \int_{y\in\partial S}\int_{x\in\partial S} \frac{\partial r_y(x)}{\partial \nu} \le \int_{y\in\partial S}\int_{x\in\partial S} 1 = L^2. \end{aligned} \tag{10–8}$$

Recently A. Ros improved this inequality by the factor of $\sqrt{2}$. His idea goes as follows. First, note that one can write

$$\frac{\partial r_y(x)}{\partial \nu} = \frac{\langle x-y, \nu(x)\rangle}{|x-y|}.$$

Note also that the roles of x and y can be interchanged in (10–8). Hence by adding up each expression (10–8) for x and y, we get

$$4\pi A \le \int_{y\in\partial S}\int_{x\in\partial S} \frac{\langle x-y, \nu(x)-\nu(y)\rangle}{|x-y|}.$$

Therefore

$$\begin{aligned} 4\pi A &\le \int_{y\in\partial S}\int_{x\in\partial S} |\nu(x) - \nu(y)| \\ &\le L\left(\int_{y\in\partial S}\int_{x\in\partial S} |\nu(x) - \nu(y)|^2\right)^{1/2} \qquad \text{(by the Hölder inequality)} \\ &= L\left(\int_{y\in\partial S}\int_{x\in\partial S} (2 - 2\langle \nu(x), \nu(y)\rangle)\right)^{1/2} \\ &= L\left(\int_{y\in\partial S}\int_{x\in\partial S} 2\right)^{1/2} \qquad \text{(since } \textstyle\int_{x\in\partial S} \nu(x) = 0 \text{ on minimal } S) \\ &= \sqrt{2}L^2. \end{aligned}$$

Thus:

THEOREM 10.3. (ROS) *For any minimal surface in* $\mathbb{R}^n$,

$$2\sqrt{2}\,\pi A \le L^2.$$

The author has recently heard that A. Stone also obtained this result [**St**].

10.3. Sobolev-type inequalities. Due to the analytic nature of the proofs of Theorem 10.1 and 10.2 we can derive, applying the same argument, the Sobolev-type inequalities corresponding to the above isoperimetric inequality. As is well known, one can recover the isoperimetric inequality from the Sobolev-type inequality using characteristic functions as test functions. For more Sobolev-type inequalities, see [**CG2**].

PROPOSITION 10.4. [**CG2**] *Let* f *be a compactly supported smooth nonnegative function on a minimal surface* S *in a complete simply connected Riemannian manifold* M^n *with sectional curvature bounded above by a constant* $K = \pm k^2$. *If* $K = k^2$, *assume also that* $\operatorname{diam}(S) \le \pi/(2k)$. *Then*

$$2\pi \int_S f^2 \le \left(\int_S |\nabla f|\right)^2 + K\left(\int_S f\right)^2.$$

PROOF. For $K = k^2$ we have from Lemma 6.1(i)

$$\operatorname{div}(f\nabla \log \sin kr) \ge \langle \nabla f\,, k\cot kr \nabla r\rangle - Kf + 2\pi f\,\delta_p.$$

Integrating both sides, we see that for a fixed $p = y \in S$

$$2\pi f(y) \le k\int_{x\in S} |\nabla f(x)| \cot kr_y(x) + K\int_{x\in S} f(x),$$

where $r_y(x) = \operatorname{dist}(y,x)$. Moreover by Lemma 6.1(h)

$$\operatorname{div}(f\nabla r) \ge \langle \nabla f, \nabla r\rangle + kf\cot kr.$$

So

$$k\int_{y\in S} f(y)\cot kr_x(y) \le \int_{y\in S} |\nabla f(y)|.$$

Therefore

$$\begin{aligned} 2\pi\int_S f^2 &\le \int_{y\in S} f(y)\left(k\int_{x\in S}|\nabla f(x)|\cot kr_y(x) + K\int_{x\in S} f(x)\right)\\ &= \int_{x\in S}|\nabla f(x)|\left(k\int_{y\in S} f(y)\cot kr_x(y)\right) + K\left(\int_S f\right)^2\\ &\le \left(\int_S |\nabla f|\right)^2 + K\left(\int_S f\right)^2. \end{aligned}$$

A similar proof is valid for $K = -k^2$. □

PROPOSITION 10.5. [**CG2**] *Let* f *be a nonnegative smooth function with compact support on an* m*-dimensional minimal submanifold* N *in a complete simply connected Riemannian manifold of sectional curvature bounded above by a negative constant* $-k^2$. *Then*

$$k(m-1)\int_N f \le \int_N |\nabla f|.$$

PROOF. From (10–3) we have

$$\operatorname{div}(f\nabla r) \geq \langle \nabla f, \nabla r\rangle + k(m-1)f.$$

Integrate both sides over N. □

11. Modified Volume

Unlike the isoperimetric inequalities $4\pi A \leq L^2$ and $k^k\omega_k \operatorname{Vol}(D)^{k-1} \leq \operatorname{Vol}(\partial D)^k$ in space (see (1–1) and (8–1)), the inequality (6–5) $4\pi A \leq L^2 + KA^2$ for minimal surfaces in space forms of curvature K has a correction term. In this section, however, we introduce a *modified volume* $M_p(N)$ of a k-dimensional minimal submanifold N of a space form and obtain an isoperimetric inequality like (8–1) with no correction term:

$$k^k\omega_k M_p(N)^{k-1} \leq \operatorname{Vol}(\partial N)^k.$$

DEFINITION 11.1. *Let p be a point in S^n and let $r(x)$ be the distance from p to x in S^n. Given a k-dimensional submanifold N of S^n, the* modified volume $M_p(N)$ of N with center at p *is defined to be*

$$M_p(N) = \int_N \cos r.$$

Similarly for $N \subset H^n$, define

$$M_p(N) = \int_N \cosh r.$$

Obviously $M_p(N) \leq \operatorname{Vol}(N)$ for $N \subset S^n$, and $M_p(N) \geq \operatorname{Vol}(N)$ for $N \subset H^n$. Suppose that S^n is embedded in $\mathbb{R}^{n+1}$ with p the north pole $(0,\dots,0,1)$ and that H^n is embedded as the hypersurface $\Sigma, x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1, x_{n+1} > 0$, of $\mathbb{R}^{n+1}$ with the Minkowski metric $ds^2 = dx_1^2 + \cdots + dx_n^2 - dx_{n+1}^2$ such that p becomes the point $(0,\dots,0,1) \in \Sigma$. Note that $\cos r$ is the Jacobian of the projection of S^n into $x_{n+1} = 0$ and that in the Minkowski space

$$dx_i = \cosh r\, dr,\ i = 1,\dots,n.$$

Therefore we have the following.

LEMMA 11.2. [**CG2**] *$M_p(U), U \subset S^n$ or H^n, is the Euclidean volume of the orthogonal projection of U into the horizontal hyperplane $x_{n+1} = 0$, counting orientation.*

It is well known that $\sin r$ and $\sinh r$ are the lengths of Jacobi fields in S^2 and H^2, respectively. Hence it is easy to show that

$$\overline{\nabla}^2 \cos r = -(\cos r)g \text{ on } S^n,\ \ \overline{\nabla}^2 \cosh r = (\cosh r)g \text{ on } H^n$$

(see [**CG2**, Lemma 3]). Hence by Lemma 5.1, if $N \subset S^n$ or H^n is a k-dimensional minimal submanifold or a cone, then

$$\triangle \cos r = -k\cos r,\ \ \triangle \cosh r = k\cosh r, \tag{11–1}$$

where the vertex of the cone should be the point p from which the distance function r is measured. Integrating (11–1) and following the proof of Proposition 4.3 or Proposition 5.6, we get the following.

PROPOSITION 11.3. [**CG2**] *For any minimal submanifold N in S^n or H^n and any point p in S^n or H^n,*

$$M_p(N) \le M_p(p \divideontimes \partial N). \tag{11–2}$$

We now have the comparison formulas that we need: Proposition 5.4 and Proposition 11.3. With these in our hands we can use the arguments of developing and cutting and pasting as in Section 4 and prove

THEOREM 11.4. [**CG2**] *Let S be a minimal surface in S^n and p a point of S. Assume that $r \le \pi/2$ on S. If ∂S is radially connected from p, that is, $\{s : s = \operatorname{dist}(p,q), q \in \partial S\}$ is a connected interval, then $4\pi M_p(S) \le \operatorname{Length}(\partial S)^2$. Equality holds if and only if S is a totally geodesic disk with center at p.*

The same inequality is false for minimal surfaces in H^n. In fact, among domains in H^2 with prescribed boundary length, the modified area has no upper bound. Moreover, our proof fails in H^n because the projection in Minkowski space $\mathbb{R}^{n+1}$ from H^n onto the hyperplane $x_{n+1} = 0$ is a length-increasing map. In Theorem 11.6 below, however, it will be shown that the same inequality holds, even in higher dimension, in case ∂S lies in a sphere of H^n centered at p.

As we have seen in Section 8, the monotonicity of the volume of a minimal submanifold is closely related to the isoperimetric inequalities. The monotonicity has been proved for $\mathbb{R}^n$ and for H^n [**An**], but not for S^n (see however [**GS**], p. 353). The next proposition shows that modified volume enjoys the monotonicity in all three cases.

PROPOSITION 11.5. [**CG2**] *Let N be a k-dimensional minimal submanifold in S^n (H^n, respectively) and r the distance in S^n (H^n, respectively) from p. Then $M_p(N \cap B_p(r))/\sin^k r$ ($M_p(N \cap B_p(r))/\sinh^k r$, respectively) is a monotonically nondecreasing function of r for $0 < r < \min(\pi/2, \operatorname{dist}(p, \partial N))$ ($0 < r < \operatorname{dist}(p, \partial N)$, respectively).*

PROOF. Define $N_r = N \cap B_p(r) \subset S^n$. Then

$$M_p(N_r) = -\frac{1}{k}\int_{N_r} \triangle \cos r = \frac{1}{k}\int_{\partial N_r} \sin r \frac{\partial r}{\partial \nu} = \frac{1}{k}\sin r \int_{\partial N_r} |\nabla r|.$$

Denote the volume forms of N and ∂N_r by dv and dS_r, respectively. Then $dv = \frac{1}{|\nabla r|} dS_r dr$. Hence

$$\frac{d}{dr}\int_{N_r} \cos r |\nabla r|^2 dv = \frac{d}{dr}\int_0^r \int_{\partial N_r} \cos r |\nabla r| dS_r dr = \cos r \int_{\partial N_r} |\nabla r|.$$

Therefore

$$\begin{aligned} M_p(N_r) &= \frac{\sin r}{k \cos r}\cos r \int_{\partial N_r} |\nabla r| = \frac{\sin r}{k\cos r}\frac{d}{dr}\int_{N_r} \cos r |\nabla r|^2 \\ &\le \frac{\sin r}{k\cos r}\frac{d}{dr}\int_{N_r} \cos r = \frac{\sin r}{k\cos r}\frac{d}{dr} M_p(N_r). \end{aligned}$$

Hence

$$\frac{d}{dr}\log[M_p(N_r)/\sin^k r] \ge 0.$$

Thus $M_p(N_r)/\sin^k r$ is nondecreasing; similarly for $N \subset H^n$. □

A very special case of the radially connected boundary occurs when ∂N lies in a geodesic sphere. In this case the conclusion of Theorem 11.4 may be extended to hyperbolic space, and the minimal submanifold may have any dimension.

THEOREM 11.6. [**CG2**] *Let N be a k-dimensional minimal submanifold in S^n or H^n. Assume that ∂N lies in a geodesic sphere centered at a point p of N and that r is the distance in S^n or H^n from p. Furthermore, in case of $N \subset S^n$, assume $r \leq \pi/2$ on N. Then*

$$k^k \omega_k M_p(N)^{k-1} \leq \mathrm{Vol}(\partial N)^k.$$

Equality holds if and only if N is a totally geodesic ball centered at p.

PROOF. Assume $N \subset H^n$ and let R be the radius of the geodesic sphere in which ∂N lies. Then

$$\begin{aligned} M_p(N) &= \frac{1}{k}\int_N \triangle \cosh r = \frac{1}{k}\int_{\partial N} \sinh r \frac{\partial r}{\partial \nu} \\ &= \frac{1}{k}\sinh R \int_{\partial N} \frac{\partial r}{\partial \nu} \leq \frac{1}{k}\sinh R \cdot \mathrm{Vol}(\partial N). \end{aligned}$$

Since $\lim_{r\to 0} M_p(N \cap B_p(r))/\sinh^k r = \omega_k$, we obtain from Proposition 11.5

$$M_p(N)/\sinh^k R \geq \omega_k.$$

Hence

$$M_p(N) \leq \frac{1}{k}\omega_k^{-1/k} M_p(N)^{1/k} \mathrm{Vol}(\partial N)$$

and so the desired inequality follows. Obviously equality holds if and only if N is a cone with density at the center equal to 1, or equivalently, N is a totally geodesic ball. A similar proof holds for $N \subset S^n$. □

12. Relative Isoperimetric Inequality

By the classical isoperimetric inequality (8–1) for $D \subset \mathbb{R}^n$ we have

$$n^n \omega_n \mathrm{Vol}(D)^{n-1} \leq \mathrm{Vol}(\partial D)^n. \tag{12–1}$$

An immediate consequence of this inequality is that if H is a closed half space of $\mathbb{R}^n$ and D is a subset of H, then

$$\tfrac{1}{2}\, n^n \omega_n \mathrm{Vol}(D)^{n-1} \leq \mathrm{Vol}(\partial D \sim \partial H)^n$$

and equality holds if and only if D is a half ball with the flat part of its boundary contained in ∂H. This follows if one applies (12–1) to the union of D and its mirror image across ∂H. Then a natural question to ask is the following.

OPEN PROBLEM 12.1. *If $C \subset \mathbb{R}^n$ is a convex domain and D is a subset of $\mathbb{R}^n \sim C$, does D satisfy the isoperimetric inequality*

$$\tfrac{1}{2}\, n^n \omega_n \mathrm{Vol}(D)^{n-1} \leq \mathrm{Vol}(\partial D \sim \partial C)^n? \tag{12–2}$$

Does equality hold if and only if $C = H$ and D is a half ball with the flat part of its boundary lying in ∂H?

Inequality (12–2) is called the *relative* isoperimetric inequality, C is called the supporting set of D, and $\mathrm{Vol}(\partial D \sim \partial C)$ is called the relative volume of ∂D. For $n = 2$ it is easy to prove (12–2): just reflect the convex hull of D about its linear boundary.

A partial answer for $n \geq 3$ was obtained by I. Kim [**K2**]; he showed that if $U = \{(x, y) \in \mathbb{R}^2 : y \geq f(x), f'' \geq 0\}$, then (12–2) holds for $C = U \times \mathbb{R}^{n-2}$. In this section we shall first see that the relative isoperimetric inequality holds if C is a graph which is symmetric about $n-1$ hyperplanes of $\mathbb{R}^n$ [**C5**]. In particular, (12–2) holds when C is a ball. The tools of [**C5**] are Gromov's method of using the divergence theorem and Steiner's method of symmetrization. Then we shall give an outline of the proof of the relative isoperimetric inequality which has been obtained recently by the author and Ritoré [**CR**].

12.1. Gromov's method. In [**Gr**] Gromov gave a new proof of the classical isoperimetric inequality. As F. Morgan pointed out to us, Knothe [**Kn**] and Berger [**Bg**] also used the same method as Gromov. His proof is based on a volume-preserving map whose divergence is bigger than or equal to the dimension of space. Here we shall see how Gromov's method can be adapted for our purpose and why the convexity of the supporting set is necessary.

THEOREM 12.2. [**C5**] *Let C be a convex domain in $\mathbb{R}^n$ and D a subset of $\mathbb{R}^n \sim C$ with piecewise C^1 boundary. Suppose that every normal vector η to $\partial D \cap \partial C$ toward the exterior of D does not point upward, that is, $\langle \eta, \partial/\partial x^n \rangle \leq 0$ for the unit vertical vector $\partial/\partial x^n$. Suppose also that there exist vertical hyperplanes $\Pi_1, \ldots, \Pi_{n-1}$ which are mutually perpendicular such that C and D are symmetric about each of them. Then*

$$\tfrac{1}{2}\, n^n \omega_n \mathrm{Vol}(D)^{n-1} \leq \mathrm{Vol}(\partial D \sim \partial C)^n,$$

where equality holds if and only if D is a half ball.

PARTIAL PROOF. First define a C^1 map $\phi_D : D \to [0,1]^n$ by

$$\phi_D(x_1, \ldots, x_n) = (\phi_1, \ldots, \phi_n), \ \phi_i = \frac{\overline{v}_i}{v_i},$$

$$v_i = L^{n-i+1}\{(a_1, \ldots, a_n) \in D : a_j = x_j,\ 1 \leq j \leq i-1,\ -\infty \leq a_k \leq \infty,\ i \leq k \leq n\},$$
$$\overline{v}_i = L^{n-i+1}\{(a_1, \ldots, a_n) \in D : a_j = x_j,\ 1 \leq j \leq i-1,\ -\infty \leq a_i \leq x_i,$$
$$-\infty \leq a_k \leq \infty,\ i+1 \leq k\},$$

where L^k is k-dimensional Lebesgue measure. Then $\phi_i = \phi_i(x_1, \ldots, x_i)$ and the Jacobian matrix of ϕ_D, $(\partial\phi_i/\partial x_j)$, is lower triangular with diagonal entries $\partial\phi_i/\partial x_i = v_{i+1}/v_i$ and $\partial\phi_n/\partial x_n = 1/v_n$. Therefore

$$\det\left(\frac{\partial \phi_i}{\partial x_j}\right) = \frac{1}{v_1}.$$

Similarly, define $\phi_B : B \to [0,1]^n$ where B is the half ball

$$\text{(12–3)} \qquad \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_n \geq 0,\ \sum x_i^2 \leq (2\omega_n^{-1}\mathrm{Vol}(D))^{2/n}\right\}.$$

Note that $\mathrm{Vol}(B) = \mathrm{Vol}(D) = v_1$. Like ϕ_D the Jacobian determinant of ϕ_B equals $1/v_1$. Let $\psi : D \to B$ be defined by $\psi = \phi_B^{-1} \circ \phi_D$. Then the Jacobian determinant of ψ equals 1. In other words, ψ is a volume-preserving map.

Now consider a vector field V on D defined by $V(x) =$ *the position vector of* $\psi(x), x \in D$. Since the Jacobian matrix of ψ is also lower triangular, it follows from the arithmetic-geometric mean inequality that

$$\text{(12–4)} \qquad n = n(\det D\psi)^{1/n} \leq \mathrm{div} V.$$

Let Π_n be the horizontal hyperplane $\{x_n = 0\}$ and let $U_1, \dots, U_{2^{n-1}}$ be the congruent subsets of Π_n separated by the vertical hyperplanes $\Pi_1, \dots, \Pi_{n-1}$. Translating C and D in a suitable way we may assume that each Π_i contains $(0, \dots, 0)$. Define the projection $p : \mathbb{R}^n \to \Pi_n$ by $p(x_1, \dots, x_n) = (x_1, \dots, x_{n-1}, 0)$. By the divergence theorem applied to (12–4)

$$n\mathrm{Vol}(D) \le \int_{\partial D \sim \partial C} \langle V, \eta \rangle + \int_{\partial D \cap \partial C} \langle V, \eta \rangle, \tag{12–5}$$

where η is the outward unit normal to ∂D. By (12–3) we have

$$|V| \le (2\,\omega_n^{-1}\mathrm{Vol}(D))^{1/n} \text{ on } \partial D \sim \partial C. \tag{12–6}$$

By the symmetry of C and D about $\Pi_1, \dots, \Pi_{n-1}$ and by the convexity of C, we get

$$\langle V, \eta \rangle \le 0 \text{ on } \partial D \cap \partial C. \tag{12–7}$$

This is because if $x \in \partial D \cap \partial C$ and $p(x) \in U_k, 1 \le k \le 2^{n-1}$, then both $\psi(x)$ and $-p(q_\eta)$ lie in U_k, where $q_\eta \in \mathbb{R}^n$ is the point whose position vector is η. Therefore it follows from (12–5), (12–6), and (12–7) that

$$n\mathrm{Vol}(D) \le (2\,\omega_n^{-1}\mathrm{Vol}(D))^{1/n}\mathrm{Vol}(\partial D \sim \partial C),$$

which implies (12–2).

See [**C5**] for the case of equality. □

12.2. Steiner's symmetrization. One of the oldest and most powerful methods in isoperimetric inequalities is Steiner's symmetrization [**S2**]. The key idea of this method is that given k functions

$$x_n = f_1(x_1, \dots, x_{n-1}), \quad \dots, \quad x_n = f_k(x_1, \dots, x_{n-1}),$$

the volume of the graph of the average function of $f_1, \dots, f_k$ is not bigger than the average of the volumes of the graphs of $f_1, \dots, f_k$. This volume estimate is based on the simple inequality for k vectors in $\mathbb{R}^n$: $|v_1 + \cdots + v_k| \le |v_1| + \cdots + |v_k|$. Here, using the symmetrization method, we shall improve Theorem 12.2.

THEOREM 12.3. [**C5**] *Let C be a convex domain in $\mathbb{R}^n$, D a subset of $\mathbb{R}^n \sim C$ with piecewise C^1 boundary, and Π_n a horizontal hyperplane $\{x_n = 0\}$. Suppose that both $\partial D \sim \partial C$ and $\partial D \cap \partial C$ are graphs over a closed set $A \subset \Pi_n$. If A is symmetric about $n-1$ vertical hyperplanes $\Pi_1, \dots, \Pi_{n-1}$ which are mutually perpendicular, then*

$$\tfrac{1}{2}\, n^n \omega_n \mathrm{Vol}(D)^{n-1} \le \mathrm{Vol}(\partial D \sim \partial C)^n,$$

where equality holds if and only if D is a half ball.

PARTIAL PROOF. Let $f_0, g_0 : A \to \mathbb{R}$ be the functions defined by

$$x_n = f_0(x_1, \dots, x_{n-1}), x_n = g_0(x_1, \dots, x_{n-1})$$

such that $\partial D \sim \partial C, \partial D \cap \partial C$ are the graphs of f_0, g_0, respectively. Let G be the group of isometries of $\mathbb{R}^n$ generated by $n-1$ horizontal reflections which leave $\Pi_1, \dots, \Pi_{n-1}$ fixed, respectively. G consists of 2^{n-1} elements, say, $r_1, \dots, r_{2^{n-1}}$. Define $f_i = f_0 \circ r_i$ and $g_i = g_0 \circ r_i, i = 1, \dots, 2^{n-1}$. Also define $f = 2^{1-n} \sum_{i=1}^{2^{n-1}} f_i$, $g = 2^{1-n} \sum_{i=1}^{2^{n-1}} g_i$. Since $f_0 \ge g_0$ on A and $f_0 = g_0$ on ∂A, we have $f \ge g$ on A

and $f = g$ on ∂A. Hence graph(f) and graph(g) enclose a domain $\hat{D}$, and it is easy to see that

$$\text{Vol}(D) = \text{Vol}(\hat{D}). \tag{12–8}$$

Note also that $\hat{D}$ is symmetric about $\Pi_1, \ldots, \Pi_{n-1}$ and graph$(g) \subset \partial\hat{D}$ is a subset of $\partial\hat{C}$ for some convex domain $\hat{C}$. Moreover

$$\begin{aligned}
\text{Vol}(\partial\hat{D} \sim \partial\hat{C}) = \text{Vol}(\text{graph}(f)) &= \int_A \left| \left(2^{1-n} \sum_i \frac{\partial f_i}{\partial x_1}, \ldots, 2^{1-n} \sum_i \frac{\partial f_i}{\partial x_{n-1}}, 1 \right) \right| \\
&= 2^{1-n} \int_A \left| \left(\sum_i \frac{\partial f_i}{\partial x_1}, \ldots, \sum_i \frac{\partial f_i}{\partial x_{n-1}}, 2^{n-1} \right) \right| \\
&\leq 2^{1-n} \int_A \sum_{i=1}^{2^{n-1}} \left| \left(\frac{\partial f_i}{\partial x_1}, \ldots, \frac{\partial f_i}{\partial x_{n-1}}, 1 \right) \right| \\
&= 2^{1-n} \sum_{i=1}^{2^{n-1}} \text{Vol}(\text{graph}(f_i)) = \text{Vol}(\text{graph}(f_0)) = \text{Vol}(\partial D \sim \partial C).
\end{aligned}$$

This plus (12–8), together with Theorem 12.2 applied to $\hat{C}, \hat{D}$, gives the desired inequality.

See [**C5**] for the proof of the equality case. □

Although the symmetry assumption is required in Theorems 12.2 and 12.3, it is not necessary in case the convex set C is a ball:

THEOREM 12.4. [**C5**] *If C is a ball in $\mathbb{R}^n$ and D is a subset of $\mathbb{R}^n \sim C$ with rectifiable boundary, then*

$$\tfrac{1}{2}\, n^n \omega_n \text{Vol}(D)^{n-1} \leq \text{Vol}(\partial D \sim \partial C)^n$$

with equality if and only if D is a half ball.

It is easy to prove this theorem once we know that the isoperimetric region of the complement of a ball is rotationally symmetric about a line through the center of the ball.

LEMMA 12.5. [**C5**] *Outside a ball $C \subset \mathbb{R}^n$ there exists a set $\tilde{D}$ whose boundary has the least relative volume* $\text{Vol}(\partial\tilde{D} \sim \partial C)$ *among all sets outside C with the same volume as $\tilde{D}$. In fact, $\partial\tilde{D} \sim \partial C$ is a spherical cap perpendicular to ∂C and $\partial\tilde{D} \cap \partial C$ lies in an open hemisphere of ∂C.*

PROOF. The existence of $\tilde{D}$ can be obtained by following the compactness argument in [**Sp**], pp. 441-444. Obviously $\partial\tilde{D} \sim \partial C$ has constant mean curvature and makes 90° with ∂C. We claim that $\tilde{D}$ is rotationally symmetric about a line. Suppose not. Then there exists an $(n-3)$-dimensional great sphere S in ∂C such that $\tilde{D}$ is not symmetric about any hyperplane containing S. Choose a hyperplane Π containing S that divides $\tilde{D}$ into $\tilde{D}_1$ and $\tilde{D}_2$ of equal volume. Suppose without loss of generality that $\text{Vol}(\partial\tilde{D}_1 \sim (\partial C \cup \Pi)) \leq \text{Vol}(\partial\tilde{D}_2 \sim (\partial C \cup \Pi))$. Let $\tilde{D}_3$ be the mirror image of $\tilde{D}_1$ across Π and define $\tilde{D}_{13}$ to be the union of the closures of $\tilde{D}_1$ and $\tilde{D}_3$. If $\partial\tilde{D} \sim \partial C$ intersects Π at 90°, then the unique continuation property of the constant mean curvature hypersurfaces implies that $\tilde{D}$ is symmetric about Π, contradicting our hypothesis. Therefore some part of $\partial\tilde{D}_{13} \sim \partial C$ should be not

C^1 along Π. Then we can slightly perturb $\tilde{D}_{13}$ along this singular part to get a set $D' \subset \mathbb{R}^n \sim C$ such that

$$\mathrm{Vol}(D') = \mathrm{Vol}(\tilde{D}_{13}) = \mathrm{Vol}(\tilde{D}),$$

and

$$\mathrm{Vol}(\partial D' \sim \partial C) < \mathrm{Vol}(\partial \tilde{D}_{13} \sim \partial C) \leq \mathrm{Vol}(\partial \tilde{D} \sim \partial C).$$

But this contradicts the least relative volume property of $\partial \tilde{D}$. Hence $\tilde{D}$ must be rotationally symmetric about a line l. Now let $\{q\} = (\partial \tilde{D} \sim \partial C) \cap l$ and take a spherical cap A through q which is rotationally symmetric about l and has the same mean curvature as $\partial \tilde{D} \sim \partial C$. Since $\partial \tilde{D} \sim \partial C$ is tangent to A at q, we can apply the maximum principle and conclude that $\partial \tilde{D} \sim \partial C$ itself is a spherical cap. Then $\partial \tilde{D} \cap \partial C$ is a subset of an open hemisphere of ∂C. □

12.3. Dimensions three and four. The author and M. Ritoré have recently proved (12–2) in case $n = 3$ [**CR**]. Here we give an idea of the proof. Given a convex set $C \subset \mathbb{R}^n$, define the *relative isoperimetric profile* of $\mathbb{R}^n \sim \overline{C}$, $I_C : \mathbb{R}^+ \to \mathbb{R}^+$, by

$$I_C(v) = \inf_D \{\mathrm{Area}(\partial D \sim \partial C) : D \subset \mathbb{R}^n \sim \overline{C},\ \mathrm{Vol}(D) = V\}.$$

Let $\mathbb{H}^+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_n \geq 0\}$ be the upper half space of $\mathbb{R}^n$. Then the relative profile of $\mathbb{R}^n \sim \mathbb{H}^+$ is given by

$$I_{\mathbb{H}^+}(V) = n \left(\frac{\omega_n}{2}\right)^{1/n} V^{(n-1)/n},$$

and the relative isoperimetric inequality (12–2) is equivalent to

$$I_C(\mathrm{Vol}(D)) \geq I_{\mathbb{H}^+}(\mathrm{Vol}(D)), \tag{12–9}$$

with equality if and only if D is a half ball. How can one prove (12–9)? The idea is the following. First we shall take the first variation of (12–9) to get

(12–10)

$$\sup_{p \in \partial D \sim \partial C} \{H(p) : H(p) \text{ is the mean curvature of } \partial D \text{ at } p\} \geq H_0(\mathrm{Area}(\partial D \sim \partial C))$$

where $H_0(\mathrm{Area}(\partial D \sim \partial C))$ is the mean curvature of the hemisphere of area $\mathrm{Area}(\partial D \sim \partial C)$ in $\mathbb{R}^n$.

Then (12–10) follows from

$$\int_{\partial D \sim \partial C} H^2 \geq 2\pi. \tag{12–11}$$

We prove this inequality by using a method of conformal geometry: If k_1 and k_2 are the principal curvatures of $\partial D \sim \partial C$, then $\int_{\partial D \sim \partial C} (k_1 - k_2)^2$ is invariant under the conformal change of metric in $\mathbb{R}^3$. Note that

$$\frac{1}{4} \int_{\partial D \sim \partial C} (k_1 - k_2)^2 = \int_{\partial D \sim \partial C} (H^2 - K),$$

where K is the Gaussian curvature. Then (12–11) can be obtained by conformally blowing up $\partial D \sim \partial C$ around its boundary point and by using the convexity of ∂C.

In [**CR**] a new type of relative isoperimetric inequality is also proved:

Let C_1, C_2 be convex domains in $\mathbb{R}^3$ and D a subset of $\mathbb{R}^3 \sim (C_1 \cup C_2)$. If ∂C_1 and ∂C_2 make an angle of at least θ, then

$$18\theta \mathrm{Vol}(D)^2 \leq \mathrm{Area}(\partial D \sim \partial(C_1 \cup C_2))^3$$

with equality if and only if C_1 and C_2 are half spaces with an angle of θ and $\partial D \sim \partial(C_1 \cup C_2)$ is part of a sphere perpendicular to the planes ∂C_1 and ∂C_2.

The author also proved the relative isoperimetric inequality (12–2) in $\mathbb{R}^4$ [**C7**]. Croke's arguments in [**Cr**] work nicely in the relative setting as well due to the following observation: The distance between two rays in $\mathbb{R}^n$ emanating from a point grows linearly, while the distance between two rays in D grows faster after than before the rays hit and bounce off ∂C.

To be precise, the relative isoperimetric inequalities of [**CR**] and [**C7**] hold in a Riemannian manifold M^n $(n = 3, 4)$ of nonpositive curvature. These results partially answer the Open Problem 12.6 below.

It would be interesting if one could derive a version of the relative isoperimetric inequality for minimal surfaces. Therefore, combining Open Problems 8.1 and 12.1, one can propose the following.

OPEN PROBLEM 12.6. *Given a convex domain C in $\mathbb{R}^n$ and an m-dimensional minimal submanifold N outside C such that N is orthogonal to ∂C along $\partial C \cap \partial N$, prove that*

$$\tfrac{1}{2}\, m^m \omega_m \mathrm{Vol}(N)^{m-1} \leq \mathrm{Vol}(\partial N \sim \partial C)^m,$$

where equality holds if and only if N is a half ball.

I. Kim [**K1**] obtained a partial result for this open problem when N is two-dimensional. He proved that if S is a minimal surface in a Riemannian manifold M of constant sectional curvature $K \leq 0$, S lies outside a convex set C in M and is orthogonal to ∂C, and $\partial S \sim \partial C$ is connected or radially connected from a point of $\partial S \cap \partial C$, then

$$2\pi\, \mathrm{Area}(S) \leq \mathrm{Length}(\partial S \sim \partial C)^2 + K\, \mathrm{Area}(S)^2$$

and equality holds if and only if S is a totally geodesic half disk. For the proof of this, he first showed that

$$\mathrm{Area}(S) \leq \mathrm{Area}(p \ast (\partial S \sim \partial C)) \text{ for any } p \in C$$

and

$$\mathrm{Angle}(\partial S \sim \partial C, p) \geq \pi \text{ for any } p \in \partial S \cap \partial C.$$

Then he used the method of developing and cutting and pasting as in Section 4. Kim also extended Küster's linear isoperimetric inequality; Küster [**Ku**] proved

$$\mathrm{Vol}(N) \leq \frac{R}{m}\, \mathrm{Vol}(\partial N)$$

for an m-dimensional minimal submanifold N of $\mathbb{R}^n$ contained in a closed ball of radius R. Kim obtained a linear isoperimetric inequality for a minimal submanifold N^m in a complete simply connected Riemannian manifold M^n with sectional curvature bounded above by a nonpositive constant K:

$$\mathrm{Vol}(N) \leq \frac{\alpha_{m,K}(R)}{\alpha'_{m,K}(R)}\, \mathrm{Vol}(\partial N),$$

where N is contained in a geodesic ball of radius R in M and $\alpha_{m,K}(R)$ denotes the volume of the geodesic ball of radius R in the m-dimensional space form of sectional curvature K.

13. Negatively Curved Surfaces

As we have seen in Section 1, it was Carleman [**Ca**] who first showed that the classical isoperimetric inequality $4\pi A \leq L^2$ still holds for some curved surfaces: disk type minimal surfaces in space. Then in 1926 Weil [**We**] obtained the same result for disk type surfaces of negative Gaussian curvature. Thereafter a variety of different methods were employed by a dozen mathematicians to prove the same or more general inequalities; Bol [**Bo**] used parallel curves and Alexandrov [**Ax**] used the method of polyhedral approximation. Huber [**Hu**] improved the inequality of Carleman and its generalization to subharmonic functions by Beckenbach and Radó [**BR**]. In this section we give a new simple proof of $4\pi A \leq L^2$ for nonpositively curved surfaces using the maximum principle: Given a disk type nonpositively curved surface S, we construct a domain S in $\mathbb{R}^2$ with area larger than that of S and perimeter equal to that of S. Then the inequality for S follows immediately from the classical isoperimetric inequality for S.

THEOREM 13.1. *If S is a simply connected nonpositively curved surface, then $4\pi A \leq L^2$.*

Here we will give two proofs of this theorem. One is a geometric proof given in [BZ], and the other is more analytic in nature.

GEOMETRIC PROOF (OUTLINE). Consider the following special case. Suppose that S has connected smooth boundary all of whose parallel curves $l_t = \{x \in S : \operatorname{dist}(x, \partial S) = t\}$, except the furthest one l_r, are smooth simple closed curves. Denote by $A(t)$ the area of the set $S_t = \{x \in S : \operatorname{dist}(x, \partial S) < t\}$. Under these assumptions $A'(t) = l(t)$, where $l(t)$ is the length of l_t. But the first variation formula for $l(t)$ says

$$\frac{dl}{dt} = -\int_{l_t} k ds$$

where k is the geodesic curvature of l_t with respect to the inward normal to l_t. Assume that $A''(t) = l'(t)$ also exists and is continuous in [0,r). By the Gauss–Bonnet formula for S_t, we have

$$A''(t) = -\int_{l_t} k ds = -\int_{S_t} K dA - \int_{\partial S} k ds.$$

Since $K \leq 0$, the Gauss–Bonnet formula for S implies

$$A''(t) \leq -\int_S K dA - \int_{\partial S} k ds = -2\pi. \tag{13–1}$$

Assume here that $r = \sup\{t : l_t \neq \phi\}$. Multiplying (13–1) by $2A'(t) \geq 0$ and integrating from 0 to r, we get

$$A'(r)^2 - A'(0)^2 \leq -4\pi[A(r) - A(0)].$$

Since $A'(0) = L$, $A(0) = 0$, $A(r) = A$, this yields

$$L^2 - 4\pi A \geq A'(r)^2 \geq 0,$$

which is the desired inequality.

In general, the assumptions on the structure of the parallel curves l_t and the differentiability of $A(t)$ do not hold. Nevertheless it is possible to obtain a rigorous proof along these lines. Such a proof is presented in [**BZ**], pp. 20–27. In order to

overcome the technical difficulties, the argument is carried out for polyhedra. The general case follows by passing to the limit. □

ANALYTIC PROOF. Let x and y be isothermal coordinates on S. Then the metric and the Gaussian curvature K of S can be written as

$$ds^2 = e^{2\lambda}(dx^2 + dy^2), \qquad K = -e^{-2\lambda}\triangle\lambda. \tag{13–2}$$

By the curvature assumption we have

$$\triangle\lambda \geq 0 \text{ on } S. \tag{13–3}$$

Let h be the solution to the Dirichlet problem

$$\triangle h = 0 \text{ on } S, \qquad h = \lambda \text{ on } \partial S, \tag{13–4}$$

and introduce a surface $\tilde{S}$ which is S with the flat metric $\tilde{g} = e^{2h}(dx^2 + dy^2)$. Actually $\tilde{S}$ is the image of S in the complex plane under the holomorphic map $\phi(z)$ such that $\log|\phi'(z)| = h(x,y), z = x + iy$. Define $\tilde{A} = \text{Area}(\tilde{S})$, $\tilde{L} = \text{Length}(\partial\tilde{S})$. Note that the boundary condition (13–4) implies $\tilde{L} = L$. From (13–3), (13–4) and the maximum principle we get $\tilde{A} \geq A$. But we have $4\pi\tilde{A} \leq \tilde{L}^2$ for $\tilde{S} \subset \mathbb{R}^2$. Hence

$$4\pi A \leq 4\pi\tilde{A} \leq \tilde{L}^2 = L^2. \qquad \square$$

In the analytic proof we used the Dirichlet boundary value problem. By solving the mixed boundary value problem, instead, we can get a relative isoperimetric inequality:

THEOREM 13.2. **[C5]** *Let S be a disk type surface of nonpositive Gaussian curvature. Suppose that ∂S is the disjoint union of Γ_1 and Γ_2 such that Γ_1 is connected and concave, i.e., if $c(s)$ is an arclength parametrization of Γ_1, then $c''(s)$ vanishes or points outward from S. Then*

$$2\pi\text{Area}(S) \leq \text{Length}(\Gamma_2)^2 \tag{13–5}$$

and equality holds if and only if S is a flat half-disk.

PROOF. The proof is similar to the analytic proof of the preceeding theorem. The difference is that here we solve the mixed boundary value problem

$$\triangle h = 0 \text{ on } S, \qquad h = \lambda \text{ on } \Gamma_2, \qquad \frac{\partial h}{\partial \nu} = 0 \text{ on } \Gamma_1$$

and that the concavity of the free part Γ_1 implies

$$\frac{\partial\lambda}{\partial\nu} \leq 0,$$

where ν is the outward unit normal to Γ_1. This is because

$$\begin{aligned} 0 &\geq \left\langle \nabla e^{-\lambda}\partial/\partial y e^{-\lambda}\frac{\partial}{\partial y}, e^{-\lambda}\frac{\partial}{\partial x}\right\rangle = -\left\langle e^{-\lambda}\frac{\partial}{\partial y}, \nabla_{e^{-\lambda}\frac{\partial}{\partial y}} e^{-\lambda}\frac{\partial}{\partial x}\right\rangle \\ &= -e^{-3\lambda}\left\langle \frac{\partial}{\partial y}, \nabla_{\partial/\partial y}\frac{\partial}{\partial x}\right\rangle = -e^{-3\lambda}\left\langle \frac{\partial}{\partial y}, \nabla_{\partial/\partial x}\frac{\partial}{\partial y}\right\rangle \\ &= -\tfrac{1}{2}e^{-3\lambda}\frac{\partial}{\partial x}\left|\frac{\partial}{\partial y}\right|^2 = -e^{-\lambda}\frac{\partial\lambda}{\partial x} = \frac{\partial\lambda}{\partial\nu}. \end{aligned}$$

Then, by the maximum principle, $h \geq \lambda$ on S and hence $\tilde{A} \geq A$. Clearly $\text{Length}(\Gamma_2)$ remains the same under the new metric $\tilde{g}$ and Γ_1 is a line segment in $\tilde{S}$. Therefore

(13–5) follows from the relative isoperimetric inequality for $\tilde{S} \subset \mathbb{R}^2$. See [**C5**] for more details. □

In this section we have seen that nonpositively curved two-dimensional surfaces satisfy the same isoperimetric inequality as $\mathbb{R}^2$. In regard to higher-dimensional nonpositively curved Riemannian manifolds, Aubin conjectured that in the sense of the isoperimetric inequality, $\mathbb{R}^n$ is more efficient than any complete simply connected Riemannian manifold M^n of nonpositive sectional curvature. More precisely, he conjectured that for any domain D in M^n,

$$n^n \omega_n \mathrm{Vol}(D)^{n-1} \leq \mathrm{Vol}(\partial D)^n$$

and equality holds if and only if D is a Euclidean ball. Recently Kleiner [**Kl**] and Croke [**Cr**] proved this inequality in M^3 and M^4, respectively; but this conjecture is still open for $n \geq 5$. Extending Aubin's conjecture to the case of relative isoperimetric inequality, we would like to propose the following:

OPEN PROBLEM 13.3. *Let C be a convex domain in a complete simply connected Riemannian manifold M^n of nonpositive sectional curvature and D a subset of $M \sim C$. Prove that*

$$\tfrac{1}{2}\, n^n \omega_n \mathrm{Vol}(D)^{n-1} \leq \mathrm{Vol}(\partial D \sim \partial C)^n,$$

where equality holds if and only if D is a Euclidean half ball.

As mentioned in the preceding section, this open problem has been solved for $n = 3$ in [**CR**] and for $n = 4$ in [**C7**].

14. Isoenergy Inequalities

So far we have studied many isoperimetric inequalities that relate the volume of a domain with that of its boundary. In this last section, however, we shall consider a *map* from a domain into a manifold and derive an inequality that relates the interior energy of the map with its boundary energy.

Consider a C^2 harmonic map u from a closed unit ball $\bar{B} \subset \mathbb{R}^n$ to $\mathbb{R}^k, n \geq 2$. Define $E(u)$ and $E(u|_{\partial B})$ to be the energy of the map u and of the restriction of u to ∂B, respectively. Then is there any relationship between $E(u)$ and $E(u|_{\partial B})$ that resembles the isoperimetric inequality? Here we answer this question affirmatively; we obtain a relationship in a sharp form, called the *isoenergy* inequality, for a general target manifold N^k as well as $\mathbb{R}^k$. First, if N is a nonpositively curved k-dimensional Riemannian manifold, then we show

$$(n-1)E(u) \leq E(u|_{\partial B}),$$

where equality holds when $N = \mathbb{R}^k$ and u is a homothety or an orthogonal projection composed with a homothety. Second, when N^k is any Riemannian manifold, we prove

$$(n-2)E(u) \leq E(u|_{\partial B}),$$

where equality holds if $N = S^{n-1} \subset \mathbb{R}^n$ and $u(x) = x/|x|$.

In Subsection 14.2, the method of the isoenergy inequality will enable us to estimate an upper bound for the first eigenvalue of the Laplacian on a minimal submanifold in the sphere.

14.1. Isoenergy inequality via monotonicity. Assume that M^n, N^k are Riemannian manifolds with N^k isometrically embedded in $\mathbb{R}^m$. We look at a bounded map $u : M \to N$ whose first derivatives are in L^2; such a map is thought of as a map $u = (u_1, \dots, u_m) : M \to \mathbb{R}^m$ having image almost everywhere in N. Then the *energy* $E(u)$ of u is defined by

$$E(u) = \int_M |\nabla u|^2,$$

where $|\nabla u|^2 = \sum_{i=1}^m |\nabla u_i|^2$, ∇u_i being the gradient of u_i on M. $|\nabla u|^2$ is called the *energy density* of u. The critical points of $E(u)$ on the space of maps are referred to as *harmonic maps*. Thus $u \in C^2$ is harmonic if and only if

$$\Delta_M\, u \;\perp\; T_u N. \tag{14–1}$$

A harmonic map u is *stationary* if its energy is critical with respect to variations of the type $u \circ F_t$, where $F_t : M \to M$ is a smooth path of diffeomorphisms of M fixing the boundary. It can be shown that stationary harmonic maps satisfy the monotonicity property for the scale invariant energy in balls. We state an equivalent form of the monotonicity in the following lemma.

LEMMA 14.1. *Let $B_\rho = \{x \in \mathbb{R}^n : |x| < \rho\}$ and $B = B_1$. If $u : B_{1+\epsilon} \to N^k$, $\epsilon > 0$, is a stationary harmonic map, we have*

$$(n-2)\int_B |\nabla u|^2 = \int_{\partial B} \left(|\nabla u|^2 - 2\left|\frac{\partial u}{\partial r}\right|^2 \right), \qquad r = |x|. \tag{14–2}$$

PROOF. The monotonicity formula [**Pr**], [**Sc**] says

$$\rho^{2-n}\int_{B_\rho} |\nabla u|^2 - \sigma^{2-n}\int_{B_\sigma} |\nabla u|^2 \;=\; 2\int_{B_\rho - B_\sigma} r^{2-n}\left|\frac{\partial u}{\partial r}\right|^2,$$

for $0 < \sigma < \rho < 1+\epsilon$. Noting that $\int_{\partial B_\rho} f = \frac{d}{d\rho}\int_{B_\rho} f$ for almost all ρ, differentiate the formula with respect to ρ and set $\rho = 1$. □

The first isoenergy inequality of this section holds for a stationary harmonic map of $B_{1+\epsilon}$ into an arbitrary target manifold.

THEOREM 14.2. [**C3**] *Let $n \geq 3$ and suppose that $u : B_{1+\epsilon} \to N^k$, $\epsilon > 0$, is a stationary harmonic map into a Riemannian manifold N. Then*

$$(n-2)E(u|_B) \leq E(u|_{\partial B}),$$

where equality can be attained if $N = S^{n-1} \subset \mathbb{R}^n$ and $u(x) = x/|x|$.

PROOF. Let $\overline{\nabla} u_i$ denote the gradient of u_i on ∂B. Observe that

$$E(u|_{\partial B}) = \int_{\partial B} \sum_i |\overline{\nabla} u_i|^2 = \int_{\partial B} \left(|\nabla u|^2 - \left|\frac{\partial u}{\partial r}\right|^2 \right).$$

It follows from (14–2) that

$$(n-2)E(u|_B) = E(u|_{\partial B}) - \int_{\partial B}\left|\frac{\partial u}{\partial r}\right|^2, \tag{14–3}$$

which gives the desired inequality. If $u(x) = x/|x|$, then $|\partial u/\partial r| = 0$ and hence equality holds. □

Remark 1. (i) Lemma 14.1 and Theorem 14.2 fail to hold for nonstationary harmonic maps. See [**Po**][**Ri**] for such maps.

(ii) We should remark, in relation to Theorem 14.2, J.C.Wood's theorem that any smooth harmonic map u ($n \geq 2$) which is constant on ∂B is constant [**Wo**] (see also [**KW**]); the case for weakly harmonic maps is still open (see [**Sc**] for the definition of weakly harmonic maps).

(iii) When $n = 3, 4, 5, 6$, there is a sequence $\{\phi_i\}$ of C^2 harmonic maps $\phi_i : \bar{B} \to S^n \subset \mathbb{R}^{n+1}$ (see [**SY2**]) such that $\phi_i(x) = (x, 0)$ for $x \in \partial B$, $E(\phi_i) < E(\phi_{i+1})$, and

$$n - 2 = \inf_i \frac{E(\phi_i|_{\partial B})}{E(\phi_i)}.$$

Now we prove the isoenergy inequality for a harmonic map from $\bar{B}$ into $\mathbb{R}^k$. Although it is a special case of the isoenergy inequality for harmonic maps into a nonpositively curved space (Theorem 14.4), we state it independently because the proof of the Euclidean case is different and interesting in its own right.

THEOREM 14.3. [**C3**] *Suppose that u is a smooth harmonic map from $\bar{B} \subset \mathbb{R}^n, n \geq 2$, into $\mathbb{R}^k$. Then we have the isoenergy inequality*

$$(n-1)E(u) \leq E(u|_{\partial B}),$$

where equality holds if and only if u is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^k$.

PROOF. (14–1) implies

$$\Delta u_i = 0, \qquad i = 1, \ldots, k.$$

Hence

$$E(u) = \frac{1}{2}\int_B \Delta \sum_i u_i^2 = \int_{\partial B} \sum_i u_i \frac{\partial u_i}{\partial r} \leq \left(\int_{\partial B} \sum_i u_i^2\right)^{1/2} \left(\int_{\partial B} \left|\frac{\partial u}{\partial r}\right|^2\right)^{1/2},$$

where, without loss of generality, we assume $\int_{\partial B} u_i = 0$ for $i = 1, \ldots, k$. Using (14–3) and the fact that $n - 1$ is the first eigenvalue of the Laplacian on ∂B, one sees that the right-hand side of the preceding display is at most

$$\left(\frac{1}{n-1}\int_{\partial B} \sum_i |\overline{\nabla} u_i|^2\right)^{1/2} \left(E(u|_{\partial B}) - (n-2)E(u)\right)^{1/2}.$$

Hence by combining the inequalities above one gets

$$E(u)^2 \leq \frac{1}{n-1} E(u|_{\partial B})\big(E(u|_{\partial B}) - (n-2)E(u)\big),$$

which gives the desired isoenergy inequality. Moreover equality holds if and only if u_i is a constant multiple of $\partial u_i/\partial r$ and

$$\Delta_{\partial B}\, u_i + (n-1)u_i = 0, \qquad i = 1, \ldots, k,$$

which holds if and only if u is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. □

THEOREM 14.4. [**C3**] *If u is a smooth harmonic map from $\bar{B} \subset \mathbb{R}^n, n \geq 2$, to a k-dimensional Riemannian manifold N of nonpositive sectional curvature, then*

$$(n-1)E(u) \leq E(u|_{\partial B}).$$

PROOF. The Bochner formula [**EL**] says that if $u : M^n \to N^k$ is harmonic, then

(14–4)
$$\tfrac{1}{2}\Delta|\nabla u|^2 = ||\nabla' du||^2 - \sum_{\alpha,\beta} R_N(u_*e_\alpha, u_*e_\beta, u_*e_\alpha, u_*e_\beta) + \sum_i \mathrm{Ric}_M(u^*\theta_i, u^*\theta_i),$$

where ∇' is the pullback connection from TN, $e_1, \ldots, e_n$ is an orthonormal basis for TM and $\theta_1, \ldots, \theta_k$ is orthonormal for T^*N. Hence for $M = \bar{B}$ and N nonpositively curved, $|\nabla u|^2$ is subharmonic. Since the mean value of a subharmonic function on a sphere of radius r centered at the origin is monotonically nondecreasing as a function of r, one can deduce that

$$\frac{E(u)}{\omega_n} \le \frac{1}{n\omega_n}\int_{\partial B} |\nabla u|^2 = \frac{1}{n\omega_n}\left((n-2)E(u) + 2\int_{\partial B}\left|\frac{\partial u}{\partial r}\right|^2\right),$$

where equality follows from (14–2). So

$$E(u) \le \int_{\partial B}\left|\frac{\partial u}{\partial r}\right|^2. \tag{14–5}$$

Then adding (14–3) to (14–5) gives the isoenergy inequality. □

Remark 2. In case u is a harmonic map from a ball B_ρ of radius ρ into N, one obviously has

$$(n-1)E(u) \le \rho E(u|_{\partial B_\rho}).$$

When the target manifold N is nonpositively curved we have an extension theorem by Eells–Sampson [**ES**] and Hamilton [**Ha**]: Given $\phi \in C^3(\bar{B}, N)$, there is a harmonic map $u \in C^2(\bar{B}, N)$ such that $u = \phi$ on ∂B, and u is homotopic to ϕ. Since this theorem allows us to impose a condition on $u|_{\partial B}$, e.g., conformality, one can obtain a mixture of the isoenergy inequality and the isoperimetric inequality as follows.

COROLLARY 14.5. [**C3**] *Suppose N^k is nonpositively curved and let $B^n = \{x \in \mathbb{R}^n : |x| < 1\}$, $n = 2, 3$.*

(a) *If $u : \bar{B}^2 \to N$ is harmonic and $u|_{\partial B}$ is a constant speed map, then*
$$4\pi\mathrm{Area}(u(B)) \le 2\pi E(u) \le \mathrm{Length}(u(\partial B))^2.$$

(b) *If $u : \bar{B}^3 \to N$ is harmonic and $u|_{\partial B}$ is conformal, then*
$$E(u) \le \mathrm{Area}(u(\partial B)).$$

(c) *If $u : \bar{B}^n \to N$ is harmonic and $u|_{\partial B}$ is conformal, then*
$$(n\omega_n)^{3-n}E(u)^{n-1} \le \mathrm{Vol}(u(\partial B))^2.$$

PROOF. The first inequality in (a) is well known. For the second, use the constant speed condition and Theorem 14.4. Part (b) is a special case of (c). For (c), let k^2 be the conformal factor of $u|_{\partial B}$. Then

$$\begin{aligned} E(u) &\le \frac{1}{n-1}E(u|_{\partial B}) = \int_{\partial B} k^2 \le \left(\int_{\partial B} 1\right)^{(n-3)/(n-1)}\left(\int_{\partial B} k^{n-1}\right)^{2/(n-1)} \\ &= (n\omega_n)^{(n-3)/(n-1)}\,\mathrm{Vol}(u(\partial B))^{2/(n-1)}. \end{aligned}$$

□

14.2. Eigenvalue estimate. Given an $(n-1)$-dimensional minimal submanifold Σ in $S^l \subset \mathbb{R}^{l+1}$, $O*\Sigma$ is the cone from the origin O of $\mathbb{R}^{l+1}$ over Σ, that is, the union of the unit line segments from O to the points of Σ. It is well known that $O*\Sigma$ is an n-dimensional minimal submanifold of $\mathbb{R}^{l+1}$. In this section we want to consider the isoenergy inequality of a harmonic map u from $O*\Sigma$ into $\mathbb{R}^k$. In the proof of the isoenergy inequality of Theorem 14.3 we used the fact that $n-1$ is the first eigenvalue of the Laplacian on S^{n-1}. However, we do not know the exact value of the first eigenvalue $\lambda_1(\Sigma)$ of the minimal submanifold $\Sigma \subset S^l$. Therefore, instead of deriving an isoenergy inequality, we obtain, by reversing the argument, an upper bound of the first eigenvalue in terms of the energy of the harmonic map u and its boundary energy. To do this, we need the following monotonicity on $O*\Sigma$.

LEMMA 14.6. *Let Σ be an $(n-1)$-dimensional submanifold of $S^l \subset \mathbb{R}^{l+1}$. If u is a harmonic map from $O*\Sigma$ into $\mathbb{R}^k$ which is C^2 up to and including the boundary Σ, then*

$$(n-2)\int_{O*\Sigma} |\nabla u|^2 = \int_\Sigma \left(|\nabla u|^2 - 2\left|\frac{\partial u}{\partial r}\right|^2 \right).$$

PROOF. Let $\Sigma_\rho = \{x \in O*\Sigma : |x| < \rho\}$. Note that the quantity

$$\Theta(\rho) = \rho^{2-n}\int_{\Sigma_\rho} |\nabla u|^2 \tag{14–6}$$

is invariant under scaling. More precisely, if we denote $u^\rho(x) = u(\rho x)$, then

$$\Theta(\rho) = \int_{\Sigma_1} |\nabla u^\rho|^2.$$

So at $\rho = 1$ we see that

$$\frac{d}{d\rho}\Theta(\rho) = 2\sum_{i=1}^k \int_{\Sigma_1} \left\langle \nabla u_i, \nabla \frac{du_i^\rho}{d\rho} \right\rangle = 2\sum_{i=1}^k \int_\Sigma \langle x, \nabla u_i\rangle \frac{du_i^\rho}{d\rho} - 2\sum_{i=1}^k \int_{\Sigma_1} \triangle u_i \frac{du_i^\rho}{d\rho}.$$

Since $\langle x, \nabla u_i\rangle = du_i^\rho/d\rho$, $du^\rho/d\rho = \partial u/\partial r$ on Σ and $\triangle u_i = 0$ on Σ_1, we get

$$\left(\frac{d}{d\rho}\Theta(\rho)\right)_{\rho=1} = 2\int_\Sigma \left|\frac{\partial u}{\partial r}\right|^2. \tag{14–7}$$

Thus (14–6) and (14–7) complete the proof. □

THEOREM 14.7. *Let Σ be an $(n-1)$-dimensional submanifold of $S^l \subset \mathbb{R}^{l+1}$ and let $u : O*\Sigma \to \mathbb{R}^k$ be a harmonic map which is C^2 up to and including the boundary Σ. Then*

$$\lambda_1(\Sigma) \le \frac{E(u|_\Sigma)}{E(u)}\left(\frac{E(u|_\Sigma)}{E(u)} - n + 2\right). \tag{14–8}$$

PROOF. We follow the proof of Theorem 14.3 and use Lemma 14.6. So

$$E(u) = \frac{1}{2}\int_{O*\Sigma} \Delta \sum_i u_i^2 = \int_\Sigma \sum_i u_i \frac{\partial u_i}{\partial r} \le \left(\int_\Sigma \sum_i u_i^2\right)^{1/2} \left(\int_\Sigma \left|\frac{\partial u}{\partial r}\right|^2\right)^{1/2}$$
$$\le \left(\frac{1}{\lambda_1(\Sigma)}\int_\Sigma \sum_i |\overline{\nabla} u_i|^2\right)^{1/2} \left(E(u|_\Sigma) - (n-2)E(u)\right)^{1/2}.$$

Hence

$$\lambda_1(\Sigma)E(u)^2 + (n-2)E(u)E(u|_\Sigma) - E(u|_\Sigma)^2 \le 0,$$

which gives (14–8). □

COROLLARY 14.8. *Let Σ be an embedded minimal hypersurface in S^n. If $u : O*\Sigma \to \mathbb{R}^k$ is harmonic and C^2 up to and including the boundary, then*

$$(n-1)E(u) \le (-n+2+\sqrt{n^2-2n+2})E(u|_\Sigma).$$

PROOF. Combine (14–8) with Choi–Wang's estimate [**CW**]:

$$\frac{n-1}{2} \le \lambda_1(\Sigma).$$ □

Remark 3. In a sense (14–8) is similar to Chavel's estimate [**Ch**]:

$$\lambda_1(S) \le \frac{(n-1)}{n^2}\frac{A^2}{V^2},$$

where V is the volume of an n-dimensional minimal submanifold S of an m-dimensional complete simply connected nonpositively curved Riemannian manifold M and A is the volume of ∂S. Chavel's estimate, when applied to $S = O*\Sigma \subset M = \mathbb{R}^{n+1}$ with Σ^{n-1} minimal in $S^n \subset \mathbb{R}^{n+1}$, implies $\lambda_1(\Sigma) \le n-1$, which is nothing new. Also our estimate (14–8), when applied to $u =$ identity, draws the same conclusion because $(n-1)E(\mathrm{id}) = E(\mathrm{id}|_\Sigma)$. But should there exist a harmonic map $u : O*\Sigma \to \mathbb{R}^k$ satisfying

$$(n-1)E(u) \;>\; E(u|_\Sigma), \tag{14–9}$$

then one would be able to conclude from (14–8) that

$$\lambda_1(\Sigma) \;<\; n-1, \tag{14–10}$$

which would disprove Yau's conjecture [**Ya**, Problem 100]. In fact, since $\mathrm{Ric}_{O*\Sigma}$ is nonpositive, one could deduce from the Bochner formula (14–4) that $|\nabla u|^2$ is strictly superharmonic provided

$$||\nabla' du||^2 + \sum_i \mathrm{Ric}_{O*\Sigma}(u^*\theta_i, u^*\theta_i) \;<\; 0.$$

Then the argument of the proof of Theorem 14.4 would imply (14–9).

OPEN PROBLEM 14.9. *Does there exist a harmonic map $u : O*\Sigma \to \mathbb{R}^k$ satisfying* (14–9)*?*

OPEN PROBLEM 14.10. *Let Σ be an $(n-1)$-dimensional embedded minimal hypersurface of S^n. A map $u : \Sigma \to S^n \subset \mathbb{R}^{n+1}$ is said to be* balanced *if $\int_\Sigma u$ equals the zero vector in $\mathbb{R}^{n+1}$. Does there exist a balanced energy minimizing map of Σ into S^n which is different from the identity? If it exists, is its energy smaller than that of the identity?*

If the answer to Open Problem 14.10 is affirmative, then one gets (14–10) since

$$\lambda_1(\Sigma)\mathrm{Vol}(\Sigma) = \lambda_1(\Sigma)\int_\Sigma |u|^2 \le \int_\Sigma |\nabla u|^2 \;<\; E(\mathrm{id}) = (n-1)\mathrm{Vol}(\Sigma).$$

For a minimal surface of codimension ≥ 2, that case really occurs. Let $\psi : S^2(\sqrt{3}) \to S^4$ be the two-to-one locally isometric minimal immersion whose image is the Veronese surface V diffeomorphic to the projective plane. Let $\widetilde{V}$ be the

double covering of V and define $u : \widetilde{V} \to S^4$ by $u = \frac{1}{\sqrt{3}}\psi^{-1}$. Then u is a balanced harmonic map satisfying

$$E(u) = \tfrac{1}{3}E(\mathrm{id}).$$

Indeed $\lambda_1(\widetilde{V}) = \frac{2}{3}$.

References

[Ax] A. D. Alexandrov, *Isoperimetric inequalities for curved surfaces*, Dokl. Akad. Nauk USSR **47** (1945), 235-238.

[Al] F. J. Almgren, Jr., *Optimal isoperimetric inequalities*, Indiana University Math. J. **35** (1986), 451-547.

[An] M. Anderson, *Complete minimal varieties in hyperbolic space*, Invent. Math. **69** (1982), 477-494.

[BR] E. F. Beckenbach and T. Radó, *Subharmonic functions and surfaces of negative curvature*, Trans. Amer. Math. Soc, **35** (1933), 662-674.

[Bg] M. Berger, *Geometry* II, Springer, New York, 1977.

[Be] F. Bernstein, *Über die isoperimetrische Eigenschaft des Kreises auf der Kugeloberfläche und in der Ebene*, Math. Ann. **60** (1905), 117-136.

[Bo] G. Bol, *Isoperimetrische Ungleichung für Bereiche auf Flächen*, Jber. Deutsch. Math.-Verein. **51** (1941), 219-257.

[Bm] E. Bombieri, *An introduction to minimal currents and parametric variational problems*, Mathematical Reports **2** (1985) Harwood Academic Publishers.

[BZ] Y. D. Burago and V. A. Zalgaller, *Geometric inequalities*, Grundlehren der mathematischen Wissenschaften 285, Springer, New York, 1988.

[Ca] T. Carleman, *Zur Theorie der Minimalflächen*, Math. Z. **9** (1921), 154-160.

[Ch] I. Chavel, *On A. Hurwitz' method in isoperimetric inequalities*, Proc. Amer. Math. Soc. **71** (1978), 275-279.

[C1] J. Choe, *The isoperimetric inequality for a minimal surface with radially connected boundary*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), **17** (1990), 583-593.

[C2] J. Choe, *Sharp isoperimetric inequalities for stationary varifolds and area minimizing flat chains mod k*, Kodai Math. J. **19** (1996), 177-190.

[C3] J. Choe, *The isoenergy inequality for a harmonic map*, Houston J. Math. **24** (1998), 649-654.

[C4] J. Choe, *The isoperimetric inequality for minimal surfaces in a Riemannian manifold*, J. reine angewandte Mathematik **506** (1999), 205-214.

[C5] J. Choe, *Relative isoperimetric inequality for domains outside a convex set*, to appear in Archives Inequalities Appl.

[C6] J. Choe, *Index, vision number, and stability of complete minimal surfaces*, Arch. Rat. Mech. Anal. **109** (1990), 195-212.

[C7] J. Choe, *The double cover relative to a convex set and the relative isoperimetric inequality*, preprint.

[CG1] J. Choe and R. Gulliver, *The sharp isoperimetric inequality for minimal surfaces with radially connected boundary in hyperbolic space*, Invent. Math. **109** (1992), 495-503.

[CG2] J. Choe and R. Gulliver, *Isoperimetric inequalities on minimal submanifolds of space forms*, manuscripta math. **77** (1992), 169-189.

[CG3] J. Choe and R. Gulliver, *Embedded minimal surfaces and total curvature of curves in a manifold*, Math. Research Letters **10** (2003), 343-362.

[CR] J. Choe and M. Ritoré, The relative isoperimetric inequality outside a convex set, preprint.

[CW] H.I. Choi and A.-N. Wang, *A first eigenvalue estimate for minimal hypersurfaces*, J. Diff. Geom. **18** (1983), 559-562.

[Cr] C. Croke, *A sharp four dimensional isoperimetric inequality*, Comment. Math. Helv. **59** (1984), 187-192.

[EL] J. Eells and L. Lemaire, *A report on harmonic maps*, Bull. London Math. Soc. **10** (1978), 1-68.

[ES] J. Eells and J. Sampson, *Harmonic mappings of Riemannian manifolds*, Amer. J. Math. **86** (1964), 109-160.

[Fa] Y. Fang, *Lectures on minimal surfaces in* $\mathbb{R}^3$, Proc. Centre for Mathematics and its Applications, Australian Nat. Univ. **35** (1996).

[Fe] J. Feinberg, *The isoperimetric inequality for doubly connected minimal surfaces in* $\mathbb{R}^n$, J.d'Anal. Math. **32** (1977), 249-278.

[Fi] F. Fiala, *Le problème des isopérimètres sur les surfaces ouvertes à courbure positive*, Comment. Math. Helv. **13** (1940/41), 293-346.

[Gr] M. Gromov, *Isoperimetric inequalities in Riemannian manifolds*, In Asymptotic Theory of Finite Dimensional Normed Spaces, Lecture Notes in Math. **1200**, Appendix I, 114-129. Berlin: Springer Verlag, 1986.

[GS] R. Gulliver and P. Scott, *Least area surfaces can have excess triple points*, Topology **26** (1987), 345-359.

[Ha] R. Hamilton, *Harmonic maps of manifolds with boundary*, Lecture Notes in Math. **471**, Springer, 1975.

[He] F. Hélein, *Inégalité isoperimetrique et calibrations*, Annales de l'Institut Fourier **44** (1994), 1211-1218. *Isoperimetric inequalities and calibrations*, Progress in Partial Differential Equations: the Metz surveys, M. Chipot and I. Shafrir ed., Pitman Research Notes in Mathematics, Series 345, Longman (1996). Prepublication 1996 numero 9602.

[Hs] C. C. Hsiung, *Isoperimetric inequalities for two-dimensional Riemannian manifolds with boundary*, Ann. of Math. (2) **73** (1961), 213-220.

[Hu] A. Huber, *On the isoperimetric inequality on surfaces of variable Gaussian curvature*, Ann. Math. **60** (1954), 237-247.

[KW] H. Karcher and J.C. Wood, *Non-existence results and growth properties for harmonic maps and forms*, J. Reine Angew. Math. **353** (1984), 165-180.

[K1] I. Kim, *Relative isoperimetric inequality and linear isoperimetric inequality for minimal submanifolds*, manuscripta math. **97** (1998), 343-352.

[K2] I. Kim, *An optimal relative isoperimetric inequality in concave cylindrical domains in* $\mathbb{R}^n$, J. Inequalities Appl. **1** (2000), 97-102.

[Kl] B. Kleiner, *An isoperimetric comparison theorem*, Invent. math. **108** (1992), 37-47.

[Kn] H. Knothe, *Contributions to the theory of convex bodies*, Michigan Math. J. **4** (1957), 39-52.

[Ku] A. Küster, *On the linear isoperimetric inequality*, manuscripta math. **53** (1985), 255-259.

[LSY] P. Li, R. Schoen and S.-T. Yau, *On the isoperimetric inequality for minimal surfaces*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), **11** (1984), 237-244.

[O1] R. Osserman, *Isoperimetric and related inequalities*, Proc. Symposia Pure Math. vol. **27**, Amer. Math. Soc. Providence, 1975, 207-215.

[O2] R. Osserman, *The isoperimetric inequality* Bull. Amer. Math. Soc. **84** (1978), 1182-1238.

[OS] R. Osserman and M. Schiffer, *Doubly connected minimal surfaces*, Arch. Rational Mech. Anal. **58** (1975), 285-307.

[Po] C.-C. Poon, *Some new harmonic maps from* B^3 *to* S^2, J. Diff. Geom., **34** (1991), 165-168.

[Pr] P. Price, *A monotonicity formula for Yang–Mills fields*, manuscripta math. **43** (1983), 131-166.

[Re] W. T. Reid, *The isoperimetric inequality and associated boundary problems* J. Math. Mech. **8** (1959), 897-906.

[Ri] T. Rivière, *Everywhere discontinuous harmonic maps into spheres*, Acta Math., **175** (1995), 197-226.

[Sm] E. Schmidt, *Über die isoperimetrische Aufgabe im m-dimensionalen Raum konstanter negativer Krümmung. I. Die isoperimetrischen Ungleichungen in der hyperbolischen Ebene und für Rotationskörper im n-dimensionalen hyperbolischen Raum*, Math. Z. **46** (1940), 204-230.

[Sc] R. Schoen, *Analytic aspects of the harmonic map problem*, Math. Sci. Res. Inst. Publ. Vol. **2**, Springer, Berlin, 1984, 321-358.

[SY1] R. Schoen and S.-T. Yau. *Lectures on Differential Geometry*, International Press, 1994.

[SY2] R. Schoen and S.-T. Yau. *Lectures on Harmonic Maps*, International Press, 1995.

[Sp] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, **4**, Publish or Perish, Berkeley, 1979.

[S1] J. Steiner, *Sur le maximum et le minimum des figures dans le plan sur la sphère et dans l'espace en général*, J. Reine Angew. Math. **24** (1842), 93-152.

[S2] J. Steiner, *Einfach Beweise der isoperimetrische Hauptsätze*, J. Reine Angew. Math. **18** (1838), 281-296. Reprinted: Gesammelte Werke. Bronx, NY: Chelsea Publ. Co., 1971 (reprint of 1881-1882 ed.).

[St] A. Stone, *On the isoperimetric inequality on a minimal surface*, preprint.

[Ta] J. Taylor, *The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces*, Ann. of Math. (2) **103** (1976), 489-539.

[We] A. Weil, *Sur les surfaces à courbure négative*, C. R. Acad. Sci., Paris **182** (1926), 1069-1071.

[Wh] B. White, *Some recent developments in differential geometry*, Math. Intelligencer **11** (1989), 41-47.

[Wo] J.C. Wood, *Non-existence of solutions to certain Dirichlet problems for harmonic maps, I*, preprint.

[Ya] S.-T. Yau, *Problem Section*, Seminar on Differential Geometry, Annals Math. Studies, Princeton Univ. Press, **102** (1982), 669-706.

Department of Mathematics, Seoul National University, Seoul, 151-742, Korea
E-mail address: choe@math.snu.ac.kr

Clay Mathematics Proceedings
Volume 2, 2004

Complete Nonorientable Minimal Surfaces in $\mathbb{R}^3$

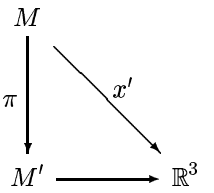
Francisco Martín

Abstract. An introduction to complete nonorientable minimal surfaces is given: known examples, methods of construction, geometrical and topological properties. Some of the more interesting unsolved questions in this area are presented: existence of examples with the lowest total curvature, problems related with the generalized Gauss map and stability.

A nonorientable surface, M', immersed in $\mathbb{R}^3$ does not have a global Gauss map, therefore it does not have a well defined mean curvature. So what is a nonorientable minimal surface? We say that a nonorientable surface M' immersed in $\mathbb{R}^3$, $x' : M' \to \mathbb{R}^3$ is minimal if, and only if, the mean curvature of any orientable piece of M' is zero.

We will be concentrating on complete nonorientable minimal surfaces. It is well known that a complete embedded surface in $\mathbb{R}^3$ must be orientable, so we cannot expect embeddedness in our family. In other words all the surfaces that we can construct have self-intersections. Nevertheless, this is a very interesting class of surfaces, for several reasons. Throughout these notes, we will try to explain some of them.

But first we need a method of construction for such surfaces. The idea is quite simple. We consider the orientable double covering of M', $\pi : M \to M'$ and the composition:



In M we consider the pull-back of the Riemannian metric of M', then x is a minimal immersion of an orientable surface in $\mathbb{R}^3$, so we can represent it by using standard Weierstrass data

$$\left(\overbrace{G}^{\text{Gauss map}} , \underbrace{dh}_{\text{height differential}} \right)$$

©2004 Clay Mathematics Institute

or (Φ_1, Φ_2, Φ_3). If you consider $x'(M')$ and $x(M)$ as sets of points in $\mathbb{R}^3$, they are equal. The only difference is that x passes twice through each point of this set, the first time with a normal vector and the second time with the opposite vector.

However, in this case the Weierstrass representation has to satisfy a compatibility condition with the oriented double covering. In order to introduce this condition, let us consider the antiholomorphic order two deck transformation associated to this covering, $I : M \to M$. We can identify M' with the quotient of M by the group $\langle I \rangle$ generated by I:

$$M' \equiv \frac{M}{\langle I \rangle}.$$

Since $x \circ I = x$, it is not hard to see that the antiholomorphic involution has the following effect on the Weierstrass representation:

$$I^*(\Phi_j) = \overline{\Phi}_j, \quad j = 1, 2, 3, \tag{0–1}$$

or, equivalently,

$$G \circ I = -\frac{1}{\overline{G}} \quad \text{and} \quad I^*(dh) = \overline{dh}.$$

Conversely, if (M, G, dh) are the Weierstrass data of an orientable minimal immersion $x : M \to \mathbb{R}^3$ and I is an antiholomorphic involution without fixed points in M satisfying the last equations, we can define a minimal immersion of the nonorientable surface $M' = M/\langle I \rangle$ in $\mathbb{R}^3$ satisfying $x' = x \circ \pi$. From now on we label (M, I, G, dh) as the Weierstrass representation of M'.

As in the orientable case, complete nonorientable minimal surfaces of finite total curvature have some very special properties that are not shared by general minimal surfaces. Let $X' : M' \to \mathbb{R}^3$ be a complete nonorientable minimal immersion with finite total curvature and (M, I, G, dh) the Weierstrass data of X'.

Under these assumptions, using Huber's and Osserman's theorems [**2, 11**], then M is conformally diffeomorphic to a compact Riemann surface $\overline{M}$ punctured in a finite number of points $\{P_1, \ldots, P_r\}$ and (G, dh) extends meromorphically to $\overline{M}$.

The geometry of the ends $\{P_1, \ldots, P_r\}$ is rigidly controlled by the singularities of G and dh at these points.

W. H. Meeks and L. P. Jorge [**3**] showed how these singularities determine the asymptotic behavior of the orientable surface M around each end, giving geometric meaning to the numbers

$$\nu_{P_j} = \max\{\mathrm{Ord}(\Phi_i, P_j)\,;\, i = 1, 2, 3\} - 1, \qquad j = 1, 2, \ldots, r.$$

In fact, ν_{P_j} is the winding number of the closed curve obtained by intersecting M with a sufficiently large sphere (always $\nu_{P_j} \geq 1$).

An interesting formula arising from the classical theory of compact Riemann surfaces, involving the topology of M, the degree of G, and the numbers ν_{P_j}, is the following:

$$2\ \deg(G) = -\chi(\overline{M}) + \sum_{j=1}^{r} (\nu_{P_j} + 1).$$

This formula, which is related to the Cohn-Vossen inequality, was given by Gackstatter, and later Meeks and Jorge gave it geometrical meaning. We will call it the Jorge–Meeks formula.

Now, going back to the nonorientable setting, it is not difficult to see that the anticonformal involution I also extends to $\overline{M}$ and we have:

$$r = 2s, \quad \{P_1, \dots, P_r\} = \{Q_1, \dots, Q_s, I(Q_1), \dots, I(Q_s)\}.$$

Therefore, G has a well defined degree and $2\mathcal{C}(M') = \mathcal{C}(M) = -4\pi \deg(G)$.

The fact that $G \circ I = -1/\overline{G}$, implies the existence of a map $g : M' \to \mathbb{RP}^2$ in such a way that the following diagram is commutative:

$$\begin{array}{ccc} M & \xrightarrow{\quad G \quad} & \overline{\mathbb{C}} \\ \Big\downarrow{\scriptstyle \pi} & & \Big\downarrow{\scriptstyle \pi_0} \\ M' & \xrightarrow{\quad g \quad} & \mathbb{RP}^2 \end{array}$$

where $\pi_0 : \overline{\mathbb{C}} \to \mathbb{RP}^2 \equiv (\overline{\mathbb{C}}/\langle I_0 \rangle)$ is the natural projection, and $I_0(z) = -1/\bar{z}$ is the antipodal map.

Since $\deg(\pi) = \deg(\pi_0) = 2$, the degree of g is also well defined and $\deg(g) = \deg(G)$. In particular $\mathcal{C}(M') = -2\pi \deg(G)$. Note that

$$\overline{M}' = \frac{\overline{M}}{\langle I \rangle}$$

is a compact nonorientable conformal surface, and

$$M' = \overline{M}' - \{\pi(Q_1), \dots, \pi(Q_s)\}.$$

In this setting, the Jorge–Meeks formula can be reformulated as

$$\deg(g) = -\chi(\overline{M}') + \sum_{i=1}^{s} (n_i + 1), \tag{0–2}$$

where $n_i = \nu_{P_i} = \nu_{I(P_i)}$.

Moreover, in the nonorientable case, we have stronger restrictions on the topology of M' (or M). Meeks showed that:

THEOREM 1 (Meeks [**10**]). *Let $\overline{M}'$ be a compact nonorientable conformal surface, and $M' = \overline{M}' - \{P_1, \dots, P_s\}$. If $X' : M' \to \mathbb{R}^3$ is a complete minimal immersion with finite total curvature, then the Euler characteristic $\chi(\overline{M}')$ of $\overline{M}'$ and $\mathcal{C}(M')/2\pi$ are congruent modulo 2.*

This theorem is the consequence of the following topological lemma.

LEMMA 1 (Meeks [**10**]). *Let M_1 and M_2 be two compact surfaces without boundary, and consider $p : M_1 \to M_2$ a branched covering map. Then:*

(1) $\chi(M_2)$ odd implies that $\chi(M_1)$ and $\deg(p)$ are both either even or odd.

(2) $\chi(M_2)$ even yields that $\chi(M_1)$ is even too.

The proof of this result can be found in the article of Meeks mentioned above, and we omit it.

PROOF OF THEOREM 1. Consider $G : \overline{M}' \to \mathbb{RP}^2$ the generalized Gauss map of X'. Then G is a branched covering map. Taking into account that $\chi(\mathbb{RP}^2) = 1$, we apply Lemma 1 and obtain $\deg(G) = -\mathcal{C}(M')/(2\pi) \equiv \chi(\overline{M}')(\mathrm{mod}\ 2)$. □

As a consequence of the monotonicity formula, Kusner proved the following theorem:

THEOREM 2 (Kusner [**5**]). *Let $X' : M' \to \mathbb{R}^3$ be a connected complete nonorientable minimal immersion with finite total curvature. Following (0–2), define*

$$n(M') = \sum_{i=1}^{n} n_i.$$

Then, for any $p \in \mathbb{R}^3$, the cardinal number of ${X'}^{-1}(p)$ is at most $n(M') - 1$.

Roughly speaking, $n(M')$ represents the number of times that the surface M' passes through infinity. Hence, this theorem basically asserts that the number of times that x' achieves a finite value is less than the number of times that x' passes through infinity.

As we mentioned before, no properly embedded surface in $\mathbb{R}^3$ is nonorientable. Hence, and as a consequence of Theorem 2, we have

COROLLARY 1 (Kusner [**5**]). *There are no complete nonorientable minimal surfaces with finite total curvature and two embedded ends.*

Up until now we have only stated theoretical results, but we also want to present some examples. Among all the nonorientable minimal surfaces, those which have critical total curvature (from the point of view of the Jorge–Meeks formula) are especially interesting.

1. Nonorientable Minimal Surfaces of Least Total Curvature

Let $X' : M' \to \mathbb{R}^3$ be a complete nonorientable minimal surface with finite total curvature. As in the orientable case, we say that X' has critical total curvature if and only if $|\mathcal{C}(M')| \leq |\mathcal{C}(M'')|$, where $X'' : M'' \to \mathbb{R}^3$ is any complete nonorientable minimal surface with the same genus as M'. Looking at the nonorientable analogue of the Jorge–Meeks formula,

$$\deg(g) = -\chi(\overline{M}') + \sum_{i=1}^{s}(n_i + 1);$$

this means that the degree of the generalized Gauss map is the least possible among the surfaces with the same genus.

We know that $n_i \geq 1$ for $i = 1, \ldots, n$, and that the two numbers $\deg g$ and $\chi(\overline{M}')$ have the same parity. So $\sum_{i=1}^{s}(n_i + 1)$ has to be an even integer greater than or equal to 2.

Taking into account that a minimal surface with only one embedded end and finite total curvature must be a plane, the degree of g is the least possible if and only if $\sum_{i=1}^{s}(n_i + 1) = 4$; in other words, $|\mathcal{C}(M')|$ is critical if and only if $\mathcal{C}(M') = -2\pi(\text{genus}(M') + 2)$ (i.e., $\text{degree}(g) = \text{genus}(M') + 2$), so either $s = 1$ and $n_1 = 3$ or $s = 2$ and $n_1 = n_2 = 1$. Kusner's Corollary states that the second case cannot occur, so only the first one holds.

Are there any examples of this kind? The answer is affirmative at least in the case of genus one (a once punctured projective plane) and in the case of genus two (a once punctured Klein bottle). We are going to briefly describe these examples.

Meeks' minimal Möbius strip. In 1981, William Meeks constructed the first example of a nonorientable minimal surface with critical total curvature. The Weierstrass data are quite simple. We consider $M = \mathbb{C} - \{0\}$ and $I(z) = -1/\bar{z}$. Define

$$G(z) = z^2 \frac{z+1}{z-1}, \qquad dh = i\frac{z^2-1}{z^2}\, dz.$$

So the Weierstrass 1-forms are

$$(\Phi_1, \Phi_2, \Phi_3) = \left(\frac{i}{2}\left(\frac{(z-1)^2}{z^4} - (z+1)^2\right) dz,\ -\frac{1}{2}\left(\frac{(z-1)^2}{z^4} + (z+1)^2\right) dz,\ i\frac{z^2-1}{z^2} dz \right),$$

which obviously satisfy the conditions to be the Weierstrass representation of a complete nonorientable minimal surface. Furthermore, it is clear that

$$\text{Residue}(\Phi, 0) = \text{Residue}(\Phi, \infty) = 0.$$

Hence, Φ has no real periods, so the minimal immersion $X : M \to \mathbb{R}^3$, $X = \text{Re}\left(\int \Phi\right)$ is well defined.

Since $\deg G = 3$, we have $\mathcal{C}(M) = -12\pi$.

Taking $M' = M/\langle I \rangle$, X induces a complete nonorientable minimal immersion with finite total curvature, $X' : M' \to \mathbb{R}^3$, satisfying $\mathcal{C}(M') = -6\pi$. Observe that M' is homeomorphic to $\mathbb{RP}^2 - \{p(0)\}$, which has the topological type of a Möbius strip. The surface $X'(M')$ has two symmetries. The nontrivial one is induced by $T(z) = \bar{z}$, and corresponds to a reflection about the x_2-axis, which is contained in the surface.

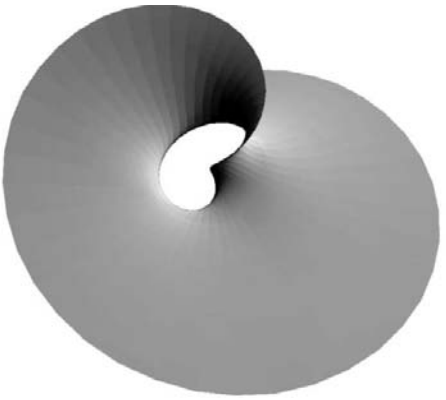


FIGURE 1. Meeks' minimal Möbius strip.

Moreover, this surface was characterized by Meeks [**10**] as the unique[1] minimal surface with total curvature -6π.

A minimal Klein bottle with one end. The second (and last) known example with critical curvature was discovered by López and it has the topology of a Klein Bottle minus one point. The total curvature in this case is -8π.

The Weierstrass representation is more complicated than in the previous example. Let $\overline{M}_r$ be the algebraic curve

$$\overline{M}_r = \left\{ (z,u) \in (\mathbb{C} \cup \{\infty\})^2 \ \Big/\ z^2 = \frac{u(u-r)}{ru+1} \right\}, \quad r \in \mathbb{R}.$$

Elementary arguments from algebraic geometry tell us that $\overline{M}_r$ is a conformal torus. Let I denote the antiholomorphic involution without fixed points defined by

$$I : \overline{M}_r \longrightarrow \overline{M}_r, \qquad I(z,u) = \left(\frac{1}{\bar{z}}, \frac{-1}{\bar{u}}\right).$$

[1]As usual, we speak of uniqueness up to homotheties and rigid motions.

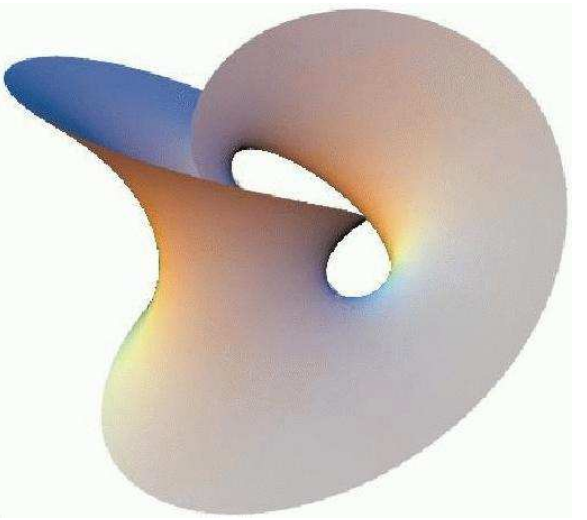


FIGURE 2. The minimal Klein bottle of total curvature -8π.

Label $M_r = \overline{M}_r - \{(0,0),(\infty,\infty)\}$. M_r is a Riemann surface of genus 1, and I leaves M_r invariant. We want to define a proper conformal minimal immersion of $M_r' \equiv M_r/\langle I\rangle$ in $\mathbb{R}^3$, for a suitable r.

First, define the following meromorphic data:

$$G = z\frac{u-1}{u+1} \quad dh = i\frac{u^2-1}{u^2}du \tag{1–1}$$

on M_r.

Then $\Phi_1 = \frac{1}{2}(1/G - G)dh$, $\Phi_2 = \frac{i}{2}(1/G + G)dh$, $\Phi_3 = dh$ satisfy the conditions to be the Weierstrass representation of a minimal immersion on M_r, and the compatibility conditions with respect to I. So, as we have said, if $X : M_r \to \mathbb{R}^3$, $X = \operatorname{Re}\int(\Phi_1, \Phi_2, \Phi_3)$ is well defined, then it induces a minimal immersion $X' : M_r' \to \mathbb{R}^3$ satisfying $X = \pi \circ X'$, where $\pi : M_r \to M_r'$ is the natural projection.

However, the period problem associated to the immersion is a bit more complicated, because we now have to deal with the homology of the torus. I'll give just an idea of how to solve the period problem, because it is representative of the way we solve period problems in nonorientable settings.

X is well defined if and only if, for $j = 1,2,3$, Φ_j have no real periods, that is, $\operatorname{Re}\int_\gamma \Phi_j = 0$ for every closed curve γ in M_r. It is easy to check that $\operatorname{Residue}(\Phi_j, 0) = \operatorname{Residue}(\Phi_j, \infty) = 0$, for $j = 1,2,3$. So it suffices to prove $\operatorname{Re}\int_\gamma \Phi_j = 0$ for any closed curve γ lying in $\overline{M}_r$ (not containing the ends).

If γ is a closed curve in (the torus) $\overline{M}_r$,

$$\int_\gamma \Phi_j = \int_{I_*(\gamma)} I^*(\Phi_j) = \int_{I_*(\gamma)} \overline{\Phi}_j,$$

and so:

$$\operatorname{Re}\left(\int_\gamma \Phi_j\right) = \frac{1}{2}\int_{\gamma + I_*(\gamma)} \Phi_j.$$

What remains to be proved is that

$$\int_{\gamma + I_*(\gamma)} \Phi_j = 0, \quad \gamma \in \Gamma,$$

where Γ is a homology basis of $\overline{M}_r$. Taking into account that Φ_3 is exact and using the symmetries of the surface, one can prove that the period problem reduces to

$$\int_{-1/r}^{0} G\, dh = 0.$$

If we define $f : \mathbb{R} \to \mathbb{R}$ as $f(r) = -i\int_{-1/r}^{0} G\, dh$, it is not hard to see that $f < 0$ on $]-\infty, -\frac{1}{2}[$ and $f > 0$ on $]-\frac{1}{3}, +\infty[$. Furthermore, a simple computation gives that $f'(r) > 0$ for all $r \in [-\frac{1}{2}, -\frac{1}{3}]$. Therefore, there is only one $r_0 \in]-\frac{1}{2}, -\frac{1}{3}[$ that solves the period problem.

What are the symmetries of the López surface? The symmetry group is generated by the reflections around two orthogonal straight lines contained in the Klein bottle. These isometries are induced by the anticonformal mappings

$$S_1, S_2 : \overline{M}_r \longrightarrow \overline{M}_r, \quad S_1(z,u) = (-\bar{z}, \bar{u}), \quad S_2(z,u) = (\bar{z}, \bar{u}).$$

It is straightforward to check that $S_j \circ I = I \circ S_j$, $j = 1, 2$. Then both maps can be induced in the quotient $M_r' = M_r\langle I \rangle$ and they represent a 180^o rotation about the x_1-axis and x_2-axis, respectively.

Why this interest in the symmetries? Because they represent an important difference between the orientable and nonorientable cases.

In the orientable case, all the examples with critical total curvature [**15**] have the same symmetry group as Enneper's surface (see Figure 3), that is to say, eight symmetries: two reflections around orthogonal straight lines contained in the surface, two reflections on vertical planes and their compositions.

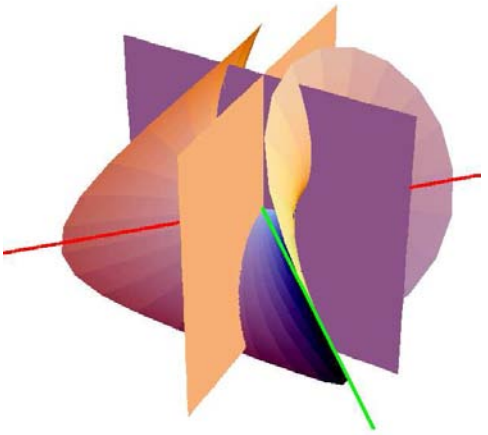


FIGURE 3. The symmetries of Enneper's surface.

In the nonorientable setting, the number of symmetries is smaller and it seems to depend on the genus. It has been conjectured that the symmetry group alternates depending on the genus parity; if the genus is odd, the group is generated by a

straight line contained in the surface, and if the genus is even it is generated by two orthogonal straight lines contained in the surface.

This lack of symmetries increases the difficulty of finding new examples of complete nonorientable minimal surfaces with critical total curvature. It becomes more difficult to determine the complex structure associated to the oriented double covering and find the corresponding Weierstrass representation.

As mentioned, for arbitrary nonorientable genus greater than 2, the following conjecture is still open:

CONJECTURE 1. *For any $g > 2$, there is a (unique) complete, nonorientable, minimal surface of genus g with least total curvature.*

2. Highly Symmetric Nonorientable Examples

By increasing the number of symmetries we also increase the total curvature and the degree of the Gauss map. However, if the group of symmetries is large enough, elementary topological arguments determine, up to conformal transformations, the underlying complex structure of such a surface. Then it is not hard to describe the Weierstrass data arising out of these kinds of examples and obtain uniqueness theorems.

Basically, two ways exist to construct new examples of highly symmetric minimal surfaces:

(1) In the first one, the genus of $\overline{M}$ is fixed and the number of ends increases. Among these surfaces we emphasize a family of immersed projective planes with p ($p \geq 3$, p odd) embedded flat ends and total curvature $-2\pi(2p-1)$, by Kusner [**5**].

To be more precise, given p odd, $p \geq 3$, define

$$M_p = \overline{\mathbb{C}} - \left\{ z \in \mathbb{C} \Big/ z^{2p} + 2\frac{\sqrt{2p-1}}{p-1} z^p - 1 = 0 \right\}, \quad I(z) = -\frac{1}{\bar{z}},$$

$$G = \frac{z^{p-1}\left(z^p - \sqrt{2p-1}\right)}{\sqrt{2p-1}\, z^p + 1}, \quad \frac{1}{G}\, dh = \frac{i\left(\sqrt{2p-1}\, z^p + 1\right)^2}{\left(z^{2p} + 2\dfrac{\sqrt{2p-1}}{p-1} z^p - 1\right)^2}\, dz.$$

It is straightforward to check that (Φ_1, Φ_2, Φ_3) has no residues at the ends, and satisfies (0–1). So the minimal immersion $X_p = \mathrm{Re}(\int \Phi)$ is well defined, and induces a minimal immersion X_p' of the nonorientable surface $M_p' = M_p/\langle I \rangle$ in $\mathbb{R}^3$. The surface $X_p'(M_p')$ contains p straight lines which lie in a plane and meet at equal angles. The dihedral group of order $2p$ acts on $X_p'(M_p')$ by reflections about these lines.

(2) In the second one, the number of ends is fixed and the genus of $\overline{M}$ increases. Inside these kinds of minimal surfaces we emphasize a family of complete nonorientable highly symmetric minimal surfaces with arbitrary topology and one end, constructed by López and myself [**8**]. For each topology we constructed the most symmetric example. Furthermore, if the Euler characteristic of the closed associated surface is even, the examples minimize the energy (or the degree of the Gauss map) among the surfaces with their symmetry.

The Weierstrass data are as follows ($\overline{M}_{k\,m\,r}$ is a Riemann surface of genus $m(k-1)$):

$$\overline{M}_{k\,m\,r} = \left\{ (z,w) \in \mathbb{C}^2 : z^k = \frac{w(w^m - r)}{rw^m + 1} \right\}, \quad M_{k\,m\,r} = \overline{M}_{k\,m\,r} - \{(0,0), (\infty,\infty)\},$$

$$I_1 : \overline{M}_{k\,m\,r} \longrightarrow \overline{M}_{k\,m\,r}, \qquad I_1(z,w) = \left(\frac{1}{\bar{z}}, -\frac{1}{\bar{w}}\right),$$

$$G = z^{k-1}\frac{w^m - 1}{w^m + 1}, \qquad dh = i\frac{w^{2m} - 1}{w^{m+1}}\,dw,$$

where $k \geq 2$, $m \geq 1$ and m is odd, r is suitable and $r \in \mathbb{R} - \{0, -1\}$.

When $k \geq 2$ is even, we also have

$$\overline{M}_{k\,m\,r} = \left\{(z,w) \in \mathbb{C}^2 : z^k = \frac{w(w^m - r)}{rw^m + 1}\right\}, \quad M_{k\,m\,r} = \overline{M}_{k\,m\,r} - \{(0,0), (\infty,\infty)\},$$

$$I_2 : \overline{M}_{k\,m\,r} \longrightarrow \overline{M}_{k\,m\,r}, \quad I_2(z,w) = \left(-\frac{1}{\bar{z}}, -\frac{1}{\bar{w}}\right),$$

$$G = z^{k-1}\frac{w^m - 1}{w(w^m + 1)} \qquad dh = i\frac{w^{2m} - 1}{w^{m+1}}\,dw,$$

where, as above, $m \geq 1$ is odd, r is suitable and $r \in \mathbb{R} - \{0, -1\}$.

For each k, m the surfaces $M_{k\,m\,r}$ intersect the $x_1\,x_2$-plane in km straight lines which meet at equal angles at the origin, and the dihedral group $\mathcal{D}(k\,m)$ acts on $M_{k\,m\,r}$ by reflections about these lines.

If $k = 2$ and $m = 1$, we obtain the once punctured Klein bottle of total curvature -8π described above.

Figure 4 corresponds to the surface $M_{3\,1}$. This surface is the result of gluing three Meeks Möbius strips symmetrically.

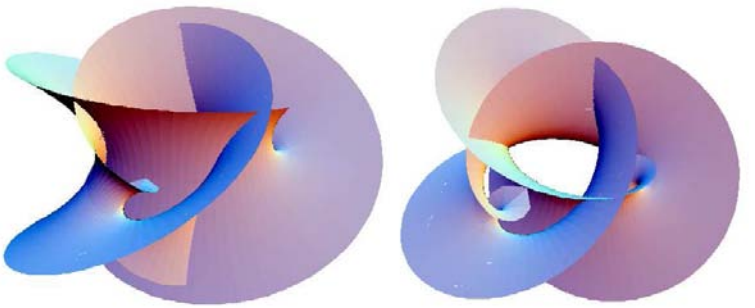


FIGURE 4. The surface $M_{3\,1}$.

We end these notes by discussing some other unsolved problems in the theory of complete nonorientable minimal surfaces.

The first one is related to the generalized Gauss map $g : M' \to \mathbb{RP}^2$. In all the known nonorientable examples with finite total curvature, the Gauss map is surjective (onto). Bearing in mind Osserman's theorem [**11**] about the Gauss map of a complete minimal surface with finite total curvature, the generalized Gauss map cannot omit more than one point in the projective plane (two points in the sphere). So the question is:

> Are there any complete nonorientable minimal surfaces with finite total curvature whose generalized Gauss map omits one point in $\mathbb{RP}^2$?

The second problem that we would like to mention is related to the stability. Marty Ross [**13, 14**] has proved that any complete nonorientable minimal surface with finite total curvature is unstable. It is still unknown whether the same theorem is valid if we suppress the hypothesis of finite total curvature, or if there is a complete, *stable*, nonorientable minimal surface.

References

[1] D. Hoffman, W. H. Meeks. *Embedded minimal surfaces of finite topology.* Annals of Mathematics **131** (1990), 1–34.

[2] A. Huber, *On subharmonic functions and differential geometry in the large.* Comment. Math. Helv. **32** (1957), 13–72.

[3] L. Jorge, W. H. Meeks. *The topology of complete minimal surfaces of finite total gaussian curvature.* Topology **22** (1983), 203–221.

[4] H. Karcher. *Construction of minimal surfaces.* Surveys in Geometry, 1–96, University of Tokyo, 1989. (Also: Lecture Notes No. 12, SFB256, Bonn, 1989).

[5] R. Kusner. *Conformal Geometry and complete minimal surfaces* . Bull. Amer. Math. Soc. (N.S.) **17**:2 (1987), 291–295.

[6] F. J. López. *A complete minimal Klein Bottle in* $\mathbb{R}^3$. Duke Mathematical Journal **71**:1 (1993), 23–30.

[7] F. J. López. *On complete nonorientable minimal surfaces with low total curvature.* Trans. Amer. Math. Soc. **348**:7 (1996), 2737–2758.

[8] F. J. López, F. Martín. *Complete nonorientable minimal surfaces and symmetries.* Duke Math. J. **79**:3 (1995), 667–686.

[9] F. J. López, F. Martín. *Complete nonorientable minimal surfaces with the highest symmetry group.* Amer. J. Math. **119**:1 (1997), 55–81.

[10] W. H. Meeks. *The classification of complete minimal surfaces in* $\mathbb{R}^3$ *with total curvature greater than* -8π. Duke Mathematical Journal **48** (1981), 523–535.

[11] R. Osserman. *Global properties of minimal surfaces in* E^3 *and* E^n. Ann. Math. (2) **80** (1963), 392–404.

[12] R. Osserman. *A survey of minimal surfaces.* Dover Publications, New York, second edition, 1986.

[13] M. Ross. *Complete nonorientable minimal surfaces in* $\mathbb{R}^3$. Comment. Math. Helv. **67**:1 (1992), 64–76.

[14] M. Ross. *The second variation of nonorientable minimal submanifolds.* Trans. Amer. Math. Soc. **349**:8 (1997), 3093–3104.

[15] M. Weber, M. Wolf. *Minimal surfaces of least total curvature and moduli spaces of plane polygonal arcs.* Geom. Funct. Anal. **8**:6 (1998), 1129–1170.

Departamento de Geometría y Topología, Universidad de Granada, 18071 Granada, Spain

E-mail address: fmartin@goliat.ugr.es

Clay Mathematics Proceedings
Volume 2, 2004

Some Picard-Type Results for Properly Immersed Minimal Surfaces in $\mathbb{R}^3$

Francisco J. López

Abstract. These notes are devoted to some Picard-type theorems for properly immersed minimal surfaces with finite topology and compact boundary in $\mathbb{R}^3$, and summarize the main results obtained in the recent paper [**9**].

The classical Picard Theorem asserts that a meromorphic function defined in a once punctured disc which omits three points of the Riemann sphere $\overline{\mathbb{C}}$ extends meromorphically to the puncture.

Likewise, we can ask for the geometry of complete minimal surfaces in $\mathbb{R}^3$ whose Gauss map or the immersion itself omits some subsets or $\overline{\mathbb{C}}$ or $\mathbb{R}^3$, respectively.

The case of the Gauss map is classic, and was extensively treated by Osserman and Fijimoto. However, in these notes we deal with Picard type results for proper minimal immersions in $\mathbb{R}^3$. To be more precise, we are mainly interested in studying those closed subsets $F \subset \mathbb{R}^3$ for which the following statement holds:

> **Main Statement:** *If S is a properly immersed minimal surface in $\mathbb{R}^3$ of finite topology, compact boundary and eventually disjoint from F, then S has finite total curvature.*

We also deal with the same question when the conclusion is that the surface has finite type or is parabolic.

By definition, two closed sets are eventually disjoint if they meet in a compact subset of $\mathbb{R}^3$. A surface has finite topology if it is homeomorphic to a compact surface (with possibly nonempty boundary) minus a finite set of interior points. Moreover, a properly immersed minimal surface in $\mathbb{R}^3$ of finite topology is said to be of finite type if it is parabolic (that is to say, conformally equivalent to a compact Riemann surface minus a finite set of interior points) and the Weierstrass representation (g, η) of the surface can be parameterized by meromorphic data on the compactification of the surface: $\phi_3 = g\eta$ and dg/g extend meromorphically to the compactification. This notion was introduced by Rosenberg [**17**].

It is well known that

$$\text{finite total curvature} \Rightarrow \text{finite type} \Rightarrow \text{parabolicity},$$

but

$$\text{parabolicity} \not\Rightarrow \text{finite type} \not\Rightarrow \text{finite total curvature}.$$

We now review the already established results in this area:

©2004 Clay Mathematics Institute

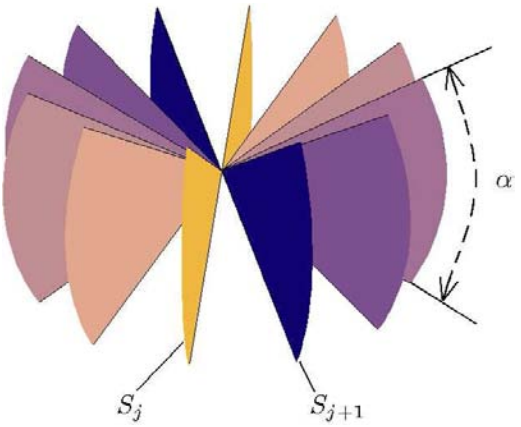


FIGURE 1. A finite collection $\mathcal{F}$ of sectors of angle α.

- The *strong half-space Theorem*, by Hoffman and Meeks [**7**], asserts that the only properly immersed minimal surfaces in $\mathbb{R}^3$ without boundary and contained in a half-space are planes.
- Collin, Kusner, Meeks and Rosenberg [**4**] have proved that any properly immersed minimal surface in $\mathbb{R}^3$ contained in a half-space is *parabolic* in the sense that bounded harmonic functions are determined by their boundary values. In particular, if the surface has finite topology and nonempty boundary, it is conformally equivalent to a compact Riemann surface with boundary and a finite number of interior points removed (see also [**13**]).
- There exist properly immersed minimal surfaces without boundary in $\mathbb{R}^3$ with the conformal type of a disc, and so, not parabolic. We refer to the recent paper by S. Morales [**15**].
- Another interesting result is *the cone Lemma*, by Hoffman and Meeks [**6**]. It asserts that any properly immersed minimal surface in $\mathbb{R}^3$ of finite topology and compact boundary which is eventually disjoint from a shallow enough cone has finite total curvature.
- We have quite a lot of information for properly embedded minimal surfaces. So, very recently, Colding and Minicozzi [**2**] have proved that any properly embedded minimal surface of finite topology and lying above a shallow enough downward facing cone has finite total curvature.
- Finally, we should mention that any properly embedded minimal surface with more than one end has finite total curvature, by Meeks and Rosenberg [**12**] and Collin [**3**], and if the surface has only one end, then it is of finite type, by Meeks and Rosenberg [**14**].

1. The Main Theorems: New Picard Type Closed Subsets.

We are going to show some new closed subsets F in $\mathbb{R}^3$ for which the Main Statement holds. The nature of these closed subsets is very simple: they are defined as the union of a finite collection of convex noncompact domains with polygonal boundary in different planes.

1.1. A finite collection of planar sectors. The first closed subset we consider is the union of a finite number of noncompact planar sectors $\mathcal{F} = \{S_1, \ldots, S_r\}$, like the one shown in Figure 1. The collection of sectors must satisfy some basic properties, that roughly speaking we have summarized as follows:

(1) They lie in pairwise nonparallel planes and are pairwise eventually disjoint. For the sake of simplicity, we will also assume that the planes containing the sectors are vertical and all the sectors have the same horizontal bisector plane, although these conditions are not necessary.
(2) All the sectors in the collection have the same angle $\alpha \in]0, \pi]$, that in the following will be called the angle of $\mathcal{F}$.
(3) The convex hull of $|\mathcal{F}| \stackrel{\text{def}}{=} \bigcup_{j=1}^{r} S_j$ is $\mathbb{R}^3$.
(4) The sectors are homogeneously distributed in $\mathbb{R}^3$, that is to say, the angle made by the planes containing two contiguous sectors is always the same.

Then we have obtained the following theorem:

THEOREM 1.1 (The Sectors Theorem [**9**]). *For any $\alpha \in]0, \pi]$, there exists a finite collection of sectors $\mathcal{F}_\alpha$ satisfying the above properties and of angle α for which the Main Statement is valid, that is to say:*

> *Any properly immersed minimal surface in $\mathbb{R}^3$ of finite topology and compact boundary eventually disjoint from $|\mathcal{F}_\alpha|$ has finite total curvature.*

The number of sectors of the collection $\mathcal{F}_\alpha$ in the theorem is not arbitrary. In fact, $\sharp(\mathcal{F}_\alpha)$ must be greater than or equal to an integer number r_α depending on the angle α. Moreover, the function $\alpha \to r_\alpha$ is decreasing, and

$$\lim_{\alpha \to 0} r_\alpha = +\infty.$$

Furthermore, as we are going to see in the next section, it is natural to conjecture that $\lim_{\alpha \to \pi} r_\alpha = 3$.

The proof of the existence of $\mathcal{F}_\alpha$ is constructive, and so it can be applied to numerical algorithms. Therefore, and using a computer, it is possible to obtain numerical estimates of the function r_α, for $\alpha \in]0, \pi]$.

1.2. Three vertical half-planes. The second closed subset F we are going to consider is the union of three disjoint and vertical half-planes Π_1, Π_2 and Π_3 not contained in a wedge of $\mathbb{R}^3$. We will distinguish two cases:

1.2.1. *The convex hull of F is $\mathbb{R}^3$.* Assume that the three half-planes do not lie in a half-space (see Figure 2, (a)). In this case, we have obtained that:

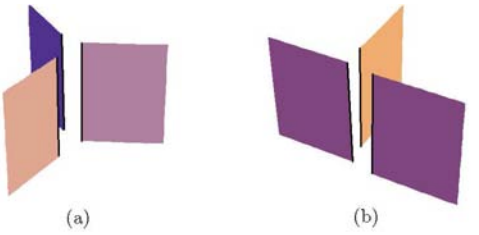


FIGURE 2. The half-planes Π_1, Π_2 and Π_3.

THEOREM 1.2 (First theorem of the three half-planes [**9**]). *If a properly immersed minimal surface in $\mathbb{R}^3$ of finite topology and compact boundary, namely S, is disjoint from $\bigcup_{j=1}^{3} \Pi_j$, then it has finite total curvature. Furthermore, all the ends of S are planar and with horizontal limit normal vector.*

1.2.2. *The convex hull of F is a half-space.* Assume now that the convex hull of the three half-planes is a half-space (see Figure 2, (*b*)). In this case, we have proved:

THEOREM 1.3 (Second theorem of the three half-planes [**9**]). *If a properly immersed minimal surface in $\mathbb{R}^3$ of finite topology and compact boundary, namely S, is disjoint from $\bigcup_{j=1}^{3} \Pi_j$, then it is parabolic. Furthermore, if S has bounded curvature, then S has finite total curvature.*

1.3. The planar complement of three parallel city maps. Finally, consider the closed subset F shown in Figure 3.

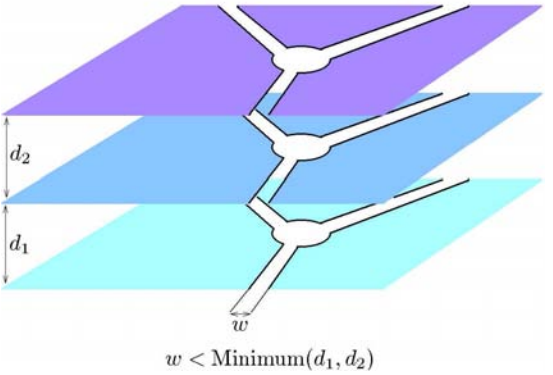


FIGURE 3. The complement of three parallel city maps.

It consists of the union of the complements of three city maps in three parallel planes. By definition, a city map in a plane is a tubular neighbourhood of a disc and a finite set of half-lines intersecting the disc. We also assume that the city maps satisfy the following geometrical conditions:

(i) They are parallel, in the sense that they have been constructed by translating the same city map orthogonally to the plane containing the city map.
(ii) The width of the streets is less than the distance between the planes.

Under these conditions, we have proved that:

THEOREM 1.4 (The City Maps Theorem [**9**]). *If a properly immersed minimal surface in $\mathbb{R}^3$ of finite topology and compact boundary, namely S, is disjoint from F, (that is to say, it meets the three parallel planes in the city maps), then it is parabolic. Furthermore, if S has bounded curvature, then S is of finite type.*

2. The key step in the proof of the theorems: the existence of finite planes.

The key step in the proof of the previously mentioned results is the existence of planes having the property of splitting the surface into a finite number of components. Given $S = X(M)$, where $X : M \to \mathbb{R}^3$ is a proper minimal immersion, and

a plane Σ, this condition means that the set $X^{-1}(\mathbb{R}^3 - \Sigma)$ has a finite collection of connected components. In the following, planes of this kind will be called *finite planes.* If the surface S is of finite topology, this condition is equivalent to saying that S intersects Σ transversally except at a finite set of points, and the intersection of S and Σ has a finite number of components. By definition, a component of $S - \Sigma$ is the image under X of a component of $X^{-1}(\mathbb{R}^3 - \Sigma)$.

Why is the existence of finite planes important?

Basically for two reasons:

Firstly, and as we have mentioned before, because any properly immersed minimal surface in $\mathbb{R}^3$ of finite topology and compact boundary admitting a finite plane is parabolic [**13**]. Furthermore, if in addition the surface has bounded curvature, then it is of finite type (see [**18**] and [**16**]).

Secondly, because the existence of finite planes yields quite a lot information about the Gauss map of the surface. As a matter of fact, it is possible to prove that:

THEOREM 2.1 (The Exceptional Values Theorem [**9**]). *If a properly immersed minimal surface in $\mathbb{R}^3$ of finite topology and compact boundary has a finite plane Σ, then its Gauss map takes on the two normal vectors of Σ a finite number of times.*

This theorem was inspired by some nice ideas of Fang and Meeks [**5**].

As a consequence of the classical Picard theorem,

COROLLARY 2.1 (The two finite planes theorem [**9**]). *A properly immersed minimal surface in $\mathbb{R}^3$ of finite topology has finite total curvature if and only if it has two nonparallel finite planes.*

PROOF. Let S be a a properly immersed minimal surface of finite topology admitting two nonparallel finite planes. Then, from the above theorem, its Gauss map has four exceptional values. Since the surface is parabolic, Picard's theorem implies that the Gauss map extends analytically to the ends, and so S has finite total curvature. The converse is trivial. □

With these ingredients in mind, we can now sketch the proof of the above results:

Parabolicity will be deduced from the existence of a finite plane, *finite type* will be deduced from the existence of a finite plane and the additional hypothesis of having bounded curvature, and *finite total curvature* will be deduced from the existence of at least two nonparallel finite planes.

2.1. Geometrical requirements to the existence of finite planes. The existence of finite planes is the natural geometrical consequence of some nonexistence theorems for properly immersed nonflat minimal surfaces in $\mathbb{R}^3$ with planar boundary in truncated noncompact tetrahedral domains of $\mathbb{R}^3$.

I will now go through these results:

2.1.1. *Minimal surfaces with planar boundary in truncated tetrahedral domains.* Let C be a truncated tetrahedral noncompact domain like the one illustrated in Figure 4.

Note that C has four noncompact faces and only one compact one. The compact face $F_0(C)$ corresponds to a rectangle. We label as $\mathfrak{h}(C)$ and $\mathfrak{o}(C)$ as the height and width of $F_0(C)$, respectively. We also label $\vartheta(C) \in]0, \pi[$ as the *vertical angle* of C, that is to say, the angle made by the two planes containing the top and

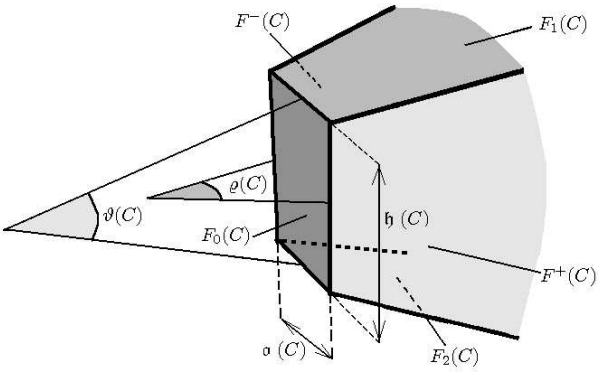


FIGURE 4. A truncated tetrahedral domain C.

bottom noncompact faces $F_1(C)$ and $F_2(C)$ of C, and likewise we label $\varrho(C)$ as the *horizontal angle*, that is to say, the angle made by the two planes containing the lateral noncompact faces $F^+(C)$ and $F^-(C)$. Then we get:

THEOREM 2.2 (The First Nonexistence Theorem [**11**]). *Let C be the truncated tetrahedral domain in Figure 4.*

If the vertical angle $\vartheta(C)$ is small enough in terms of the horizontal angle $\varrho(C)$, and the width $\mathfrak{o}(C)$ is large enough in terms of the height $\mathfrak{h}(C)$ and the angles $\vartheta(C)$ and $\varrho(C)$, then there are not properly immersed nonflat minimal surfaces in $\mathbb{R}^3$ contained in C and with boundary lying in $F^+(C) \cup F^-(C)$. In other words, a minimal surface like the one illustrated in Figure 5 does not exist.

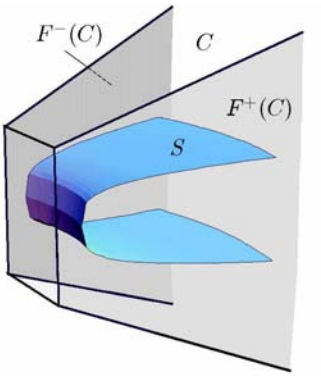


FIGURE 5. A surface S in C satisfying $\partial(S) \subset F^+(C) \cup F^-(C)$.

The version for $\theta(C) = 0$ of the above result is specially interesting, and was first studied in [**10**]. In this case, the nonexistence theorem works for any $\varrho(C) \in]0\pi]$.

2.1.2. *Minimal surfaces with planar boundary in truncated tetrahedral cylinders.* We can also obtain nonexistence theorems of minimal surfaces with planar boundary in solid right cylinders over quadrilaterals.

Let C be the solid right cylinder over the quadrilateral D illustrated in Figure 6.

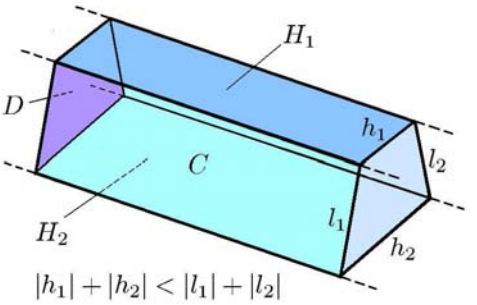


FIGURE 6. The solid right cylinder C over the quadrilateral D.

THEOREM 2.3 (The Second Nonexistence Theorem [**10**],[**11**]). *If the sum of the lengths of the horizontal edges h_1 and h_2 of D is less than the sum of the lengths of the nonhorizontal ones l_1 and l_2, then there are no properly immersed nonflat minimal surfaces in $\mathbb{R}^3$ contained in C, disjoint from D, and whose boundary lies in the horizontal faces H_1 and H_2 of C.*

PROOF. The proof of this theorem is elementary. It is based on a suitable use of some Jenkins–Serrin graphs (see [**8**]) over D as barriers for the maximum principle application.

Reasoning by contradiction, assume there is a surface S as indicated in the theorem. In Figure 7 we have illustrated the idea of the proof when D is a rectangle.

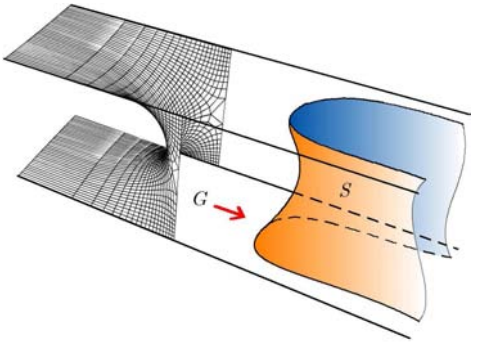


FIGURE 7. A surface S contained in C, disjoint from D, with boundary lying in $H_1 \cup H_2$, and a Jenkins–Serrin graph G disjoint from S.

Let G be the Jenkins–Serrin graph in Figure 7. The existence of this Jenkins–Serrin graph is derived from the condition regarding the lengths of the edges of D.

Without loss of generality, we can suppose that distance$(S, \partial(H_j)) > 0, j = 1, 2$. If not, substitute C for another solid right cylinder C' over a quadrilateral D' being as high as D and a little wider than D.

Up to translating G far enough toward the left of the figure, G does not touch the surface, but if we translate it toward the right, following the arrow in Figure 7,

both surfaces make contact for the first time. Since this contact is not at infinity, the surfaces meet at interior points, which contradicts the maximum principle. □

3. Proof of the Main Theorems

We are going to sketch the proof of the theorems in Section 1.

3.1. Proof of the Sectors Theorem.

PROOF. Let S be a properly immersed minimal surface in $\mathbb{R}^3$ with finite topology being eventually disjoint from a collection of sectors $\mathcal{F}$ of angle α satisfying the conditions (i), (ii), (iii) and (iv) stated in Subsection 1.1.

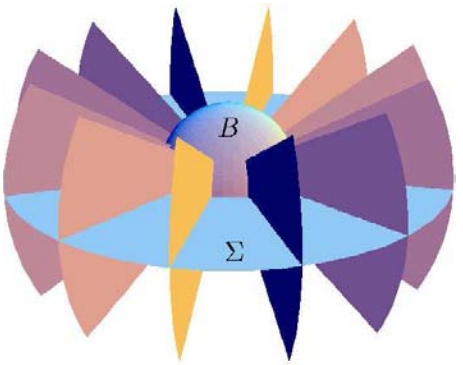


FIGURE 8. The collection of planar sectors $\mathcal{F}$, the plane Σ and the ball B.

Let Σ be a plane bisecting the sectors of the collection (see Figure 8), and let us see that Σ is a finite plane for a suitable choice of $\mathcal{F}$.

Let B be a ball containing $\partial(S)$ and $|\mathcal{F}| \cap S$. Since X is proper, it suffices to prove that any component of $S-\Sigma$ meets B. Otherwise, there would be a component of $S - \Sigma$ passing through a domain enclosed by two contiguous sectors, the plane Σ and B from top to bottom (see Figure 9). However, if the angle made by the planes containing two contiguous sectors is small enough in terms of the angle of the sectors and the radius of B is large enough, then the First Nonexistence Theorem applies for surfaces S' contained in any of these regions with boundary $\partial(S')$ lying in the top and bottom sides of this domain, which produces a contradiction.

This shows that Σ is a finite plane, and since the same argument works for planes close enough to Σ, the theorem holds by a direct application of the Two Finite Planes Theorem. □

3.2. Proofs of the first and second theorems of the three half-planes.

For the proofs of the First and Second Three Planes Theorems, some information about the geometry of properly immersed minimal surfaces in $\mathbb{R}^3$ contained in a wedge is needed.

THEOREM 3.1 (The Wedge Theorem [**9**]). *Let S be a properly immersed minimal surface homeomorphic to a closed half-plane and contained in a vertical wedge W of $\mathbb{R}^3$ (i.e., a wedge with vertical axis) of angle less than π. Assume that the*

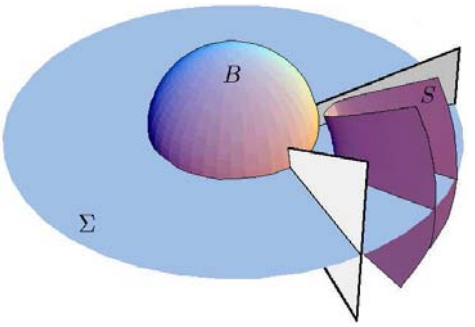


FIGURE 9. A piece of S passing through two consecutive planar sectors in $\mathcal{F}_\alpha$.

boundary of S lies in a vertical solid right circular cylinder, but S does not (see Figure 10). Then S lies in a slab, and outside a vertical solid right circular cylinder of large enough radius, S is a graph.

Moreover, the Gauss map of S uniformly converges, as the distance to the axis of W goes to infinity, to the normal vector of the slab.

The last means that the image under the Gauss map of any sequence of points in S whose distance to the axis of the wedge diverges to $+\infty$ is a sequence in the sphere converging to the normal vector of the slab.

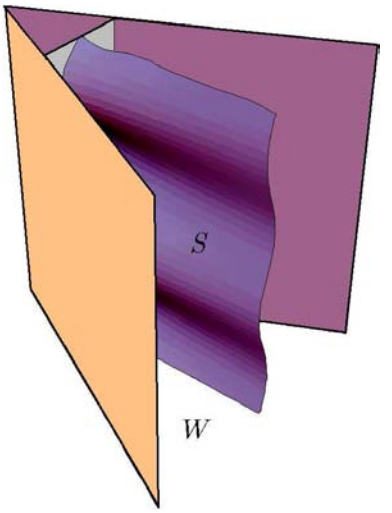


FIGURE 10. A surface S lying in the wedge W.

PROOF OF THE FIRST THEOREM OF THE THREE HALF-PLANES. Label Π_1, Π_2 and Π_3 as the three vertical pairwise disjoint half-planes in the statement of the theorem, and let $\mathcal{C}$ be a vertical right circular cylinder containing the boundary of S and the boundaries of the three half-planes. Then $\mathbb{R}^3 - \left(\mathcal{C} \cup \bigcup_{j=1}^{3} \Pi_j\right)$ consists of three truncated vertical wedges of angles less than π.

Let W be one of these truncated wedges. Let us see that there is a far enough plane Σ which is parallel to the axis of W, is disjoint from $\mathcal{C}$, meets W in a planar strip, and such that $S-(W\cap\Sigma)$ has a finite number of components. Indeed, if the distance from Σ to the axis of W is large enough, we can find two planes Σ_1 and Σ_2 parallel to Σ and disjoint from $\mathcal{C}$ satisfying: (i) Σ lies in the slab bounded by $\Sigma_1\cup\Sigma_2$; and (ii) if C_j is the solid tetrahedral cylinder enclosed by Σ_j and Σ in W, then the Second Nonexistence Theorem applies for surfaces with boundary in the faces $\Sigma\cap W$ and $\Sigma_j\cap W$ of C_j, $j=1,2$. Therefore, all the components of $S\cap C_j$ must pass through C_j from top to bottom, and so, they meet any compact section of C_j, $j=1,2$. Since the immersion is proper, we infer that $S\cap C_j$ has a finite number of components, $j=1,2$, and thus $S-(W\cap\Sigma)$ contains a finite number of components.

Since S has finite topology, we can take Σ_2 sufficiently far in such a way that $(W\cap S)-\Sigma_2$ consists of a finite set of discs for which the Wedge Theorem applies. Therefore, and outside of a vertical cylinder with large enough radius, S looks like the union of finitely many simply connected minimal graphs $G_1,\ldots,G_k$ contained in slabs $\mathcal{S}_1,\ldots,\mathcal{S}_k$, respectively. This setting can be found in Figure 11.

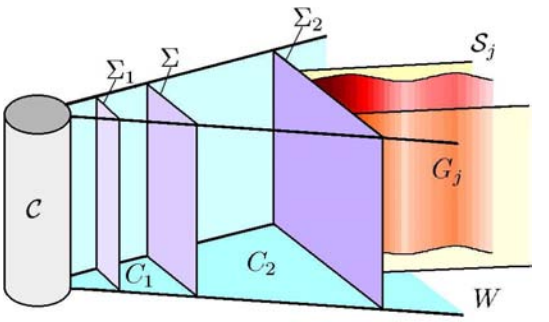


FIGURE 11. The wedge W, the cylinder $\mathcal{C}$, the planes Σ_1, Σ, Σ_2, the tetrahedral cylinders C_1 and C_2, a slab $\mathcal{S}_j$ and the graph G_j.

Finally, thanks to the Wedge Theorem, is it easy to check that any plane transverse to the cylinder $\mathcal{C}$ (that is to say, meeting $\mathcal{C}$ in a compact subset) is a finite plane for S, and so, S has finite total curvature. □

PROOF OF THE SECOND THEOREM OF THE THREE HALF-PLANES. Denote by Π_1, Π_2 and Π_3 the three vertical pairwise disjoint half-planes in the statement of the theorem, and let $\mathcal{C}$ be a vertical right circular cylinder containing the boundary of S and the boundaries of the three half-planes. Then $\mathbb{R}^3-\left(C\cup\bigcup_{j=1}^3\Pi_j\right)$ consists of two truncated vertical wedges W_1 and W_2 of angle less than π, and a domain Ω which contains a half-space and is contained in a half-space.

Following the proof of the first theorem of the three half-planes, there exists a vertical right circular cylinder $\mathcal{C}'$ such that $(S-\mathcal{C}')\cap(W_1\cup W_2)$ is the union of finitely many simply connected minimal graphs $G_1,\ldots,G_k$ contained in slabs $\mathcal{S}_1,\ldots,\mathcal{S}_k$, respectively, satisfying the properties stated in the Wedge Theorem. Therefore, it is not hard to see that any far enough plane Π disjoint from Ω meets $\bigcup_{j=1}^k G_j$ in a finite number of divergent Jordan arcs, and so, Π is a finite plane for S. Hence, S is parabolic.

If in addition S has bounded curvature, using Xavier's [**18**] and Rodriguez–Rosenberg's [**16**] ideas we get that S is of finite type. Let us see that S has finite total curvature.

In the following, and up to a rigid motion, we will suppose that Ω contains a horizontal half-space.

Reasoning by contradiction, assume that S contains an annular end A of infinite total curvature, and label (g, Φ_3) as its Weierstrass representation. Up to a biholomorphism, $A \equiv \overline{\mathbb{D}} - \{0\}$, and it is well known that $g = Pe^Q$, where P and Q are meromorphic in $\overline{\mathbb{D}}$ and Q has a pole at 0. Furthermore, $\Phi_3(z) = R(z)dz$, $z \in \overline{\mathbb{D}} - \{0\}$, where R is meromorphic in $\overline{\mathbb{D}}$ too. See [**17**] for details.

Since the only asymptotic values of g are 0 and ∞, then the Wedge Theorem implies that the slabs S_j, $j = 1, \ldots, k$ must be horizontal, that is to say, A (and in a similar way, the whole of S) lies in a horizontal half-space. No annular end of infinite total curvature lies in a slab (see the Cone Lemma [**6**] or the Sectors Theorem), and thus, the third coordinate function of A is proper. In other words, and up to scaling, $\Phi_3(z) = dz/z$. Therefore, the Gauss curvature is given by:

$$K(z) = -\left(\frac{4|z||P'/P + Q'||dz|}{(1/|g| + |g|)^2}\right)^2, \quad z \in \overline{\mathbb{D}} - \{0\},$$

and thus, K is not bounded on the noncompact set $|g|^{-1}(1)$, which is absurd and proves the theorem. □

Proof of the City Maps Theorem. For the sake of simplicity, we only consider the case of three parallel city maps having only three streets, as in Figure 3. Furthermore, we assume that S is nonflat.

Label Π_1, Π_2 and Π_3 as the three parallel planes containing the city maps, and assume that Π_2 lies in the slab $\mathcal{S}$ bounded by Π_1 and Π_3. Moreover, label $\mathcal{S}_1$ and $\mathcal{S}_3$ as the slabs bounded by $\Pi_1 \cup \Pi_2$ and $\Pi_2 \cup \Pi_3$, respectively.

The key step is to prove that Π_2 is a finite plane for S.

Without loss of generality, we will suppose that the solid right circular cylinder $\mathcal{C}$ over the city squares contains $\partial(S)$ in its interior. Let C_1^j, C_2^j and C_3^j denote the three pairwise disjoint truncated solid right cylinders over quadrilaterals in $\mathcal{S}_j - \mathcal{C}$ determined by the three couples of parallel streets in $\Pi_2 \cup \Pi_j$, $j = 1, 2$. Finally, call T_1^j, T_2^j and T_3^j the three truncated wedges of $\mathcal{S}_j - \left(\mathcal{C} \cup \bigcup_{h=1}^3 C_h^j\right)$, $j = 1, 2$. See Figure 12.

To see that Π_2 is a finite plane, it suffices to prove that $\mathcal{S}_1 \cap S$ and $\mathcal{S}_3 \cap S$ have a finite number of components. Suppose this is not true. In the following, and without loss of generality, we assume that $\mathcal{S}_1 \cap S$ has an infinite number of components. Since the immersion is proper, only a finite number of them can meet the compact set $\mathcal{C}_1 \stackrel{\text{def}}{=} \mathcal{C} \cap \mathcal{S}_1$, and so, there is a component $S_0 \subset \mathcal{S}_1 \cap S$ disjoint from $\mathcal{C}_1$ (in fact, an infinite number of components). Since S goes into and out of $\mathcal{S}_1$ through the streets, then either S_0 is contained in a cylinder C_h^1, $h = 1, 2, 3$, or S_0 passes through a truncated wedge T_h^1, $h = 1, 2, 3$, from left to right. See Figure 12.

If the radius of the city squares is large enough, the First Nonexistence Theorem applies for minimal surfaces in W_h^1 with boundary in the vertical sides of W_h^1. That is to say, there are no properly immersed minimal surfaces contained in W_h^1 with

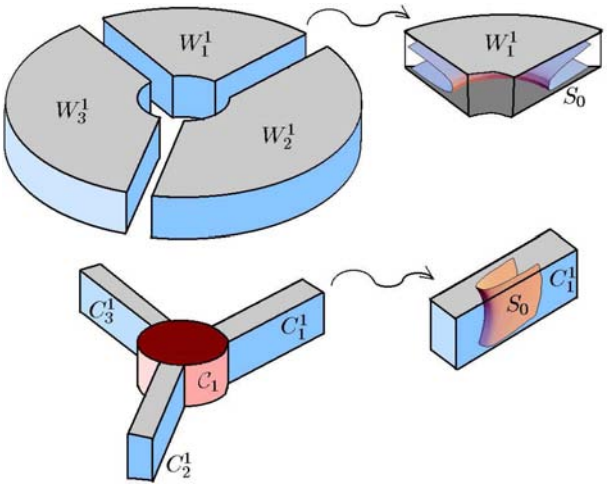


FIGURE 12. The different pieces of the slab $\mathcal{S}_1$ and the surface S_0.

boundary in its vertical faces, $h = 1, 2, 3$. Therefore, we deduce that $S_0 \cap T_h^1 = \varnothing$, $h = 1, 2, 3$, and so S_0 lies in one of the half-cylinders C_h^1, $j = 1, 2, 3$.

On the other hand, since the width of the streets is less than the distance between the planes containing the city maps, then the Second Nonexistence Theorem applies for surfaces with boundary in the horizontal faces of C_h^1 (which correspond to parallel streets), and so, $S_0 \cap C_h^1 = \varnothing$, which is obviously absurd.

We conclude that Π_2 is a finite plane, and from Meeks–Rosenberg [**13**], S is parabolic. If in addition S has bounded curvature, Rodriguez and Rosenberg's results [**16**] imply that S is of finite type. □

4. Further Remarks

(i) The hypothesis of having finite topology is fundamental in all the theorems in this lecture. Scherk singly periodic minimal surfaces are the counterexamples.

(ii) The same occurs with the hypothesis of having bounded curvature in the second part of the second theorem of the three half-planes and the City Maps Theorem. Indeed, there exists a simply connected properly immersed nonflat minimal surface in $\mathbb{R}^3$ without boundary which is not of finite type and is contained in the union of a half-space and a slab orthogonal to the half-space:

Consider on $\mathbb{C}$ the Weierstrass data $(g = e^{e^z}, \Phi_3 = dz)$. The induced immersion $X : \mathbb{C} \to \mathbb{R}^3$, $X(z) = \operatorname{Re} \int_1^z \left(\frac{1}{2}(1/g - g), \frac{I}{2}(1/g + g), 1\right) \Phi_3$, is proper and singly periodic. To be more precise, $X(\mathbb{C})$ is invariant under a horizontal translation, see Figure 13.

(iii) The Cone Lemma can be deduced as a corollary of the Sectors Theorem. Indeed, take a collection of sectors $\mathcal{F}_\alpha$, as in the theorem. Then insert a shallow enough cone $\mathcal{C}$ which is disjoint from all the sectors. See Figure 14. By the convex hull property (see [**7**] or [**11**]), any properly immersed noncompact minimal surface in $\mathbb{R}^3$ with compact boundary S and eventually disjoint from $\mathcal{C}$ is also eventually disjoint from the two convex regions determined by $\mathcal{C}$. Therefore, S lies in the nonconvex region bounded by $\mathcal{C}$, and so, it is eventually disjoint from $|\mathcal{F}_\alpha|$. If, in

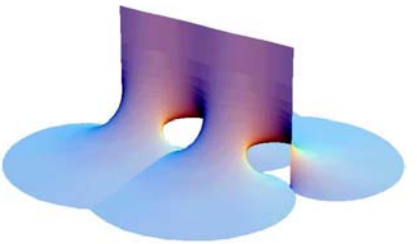


FIGURE 13. A view of $X(\mathbb{C})$.

addition, S has finite topology, then the Sectors Theorem implies that S has finite total curvature.

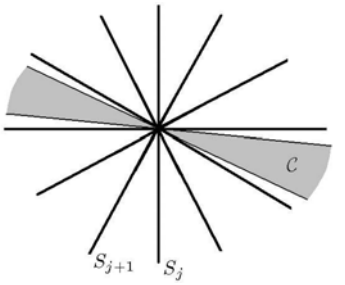


FIGURE 14. The projection on a horizontal plane of $|\mathcal{F}_\alpha|$ and $\mathcal{C}$.

(iv) Finally, using the ideas given throughout this lecture it is possible to obtain the following characterization of planes [**9**]:

The plane is the only properly immersed minimal surface in $\mathbb{R}^3$ *of finite topology and empty boundary satisfying either of the following conditions:*

- *It has only one end and is disjoint from three vertical half-planes with convex hull* $\mathbb{R}^3$.
- *It has only one end, bounded curvature and is disjoint from three vertical half-planes with convex hull a half-space.*
- *It admits two nonparallel planes of* $\mathbb{R}^3$ *meeting the surface in a straight line (which depends on the plane).*

Concerning the third result, Choe and Soret [**1**] have obtained the same conclusion for properly embedded minimal surfaces with finite total curvature. In their theorem, they assume that there is only one plane intersecting the surface in a straight line. However, there are counterexamples showing that the analogous statement fails for immersed surfaces.

References

[1] Choe, J., Soret, M. *Nonexistence of certain complete minimal surfaces with planar ends.* Comment. Math. Helv. **75** (2000), 189–199.

[2] Colding, Tobias H., Minicozzi, William P. *Complete properly embedded minimal surfaces in* $\mathbb{R}^3$. Preprint.

[3] Collin, P. *Topologie et courbure des surfaces minimales proprement plongées de* $\mathbb{R}^3$. Ann. Math. **145** (1997), 1–31.

[4] Collin, P., Kusner, R., Meeks, W. H., Rosenberg, H. *The topology, geometry and conformal structures of properly embedded minimal surfaces.* Preprint.
[5] Fang, Y., Meeks, W. H. *Some global properties of complete minimal surfaces of finite topology in* $\mathbb{R}^3$. Topology, **30**:1 (1991), 9–20.
[6] Hoffman, D., Meeks, W. H. *The asymptotic behavior of properly embedded minimal surfaces of finite topology.* J. Am. Math. Soc. **2**:4 (1989), 667–681.
[7] Hoffman, D., Meeks, W. H. *The strong half-space theorem for minimal surfaces.* Inventiones Math. **101** (1990), 373–377.
[8] Jenkins, H.,Serrin, J. *Variational problems of minimal surface type. II. Boundary value problems for the minimal surface equation.* Arch. Rat. Mech. Anal. **21** (1966), 321–342.
[9] López, Francisco J. *Some Picard theorems for minimal surfaces.* Trans. Amer. Math. Soc. 356 (2004), 703–733.
[10] López, Francisco J., Martín, F. *Minimal surfaces in a wedge of a slab.* Comm. in Ann. and Geom. **9**:4 (2001), 683-723.
[11] López, Francisco J. *Minimal surfaces in a cone.* Annals of Global Analysis and Geometry, **20**:3 (2001), 253-299.
[12] Meeks, W. H. and Rosenberg, H. *The geometry an conformal structure of properly embedded minimal surfaces of finite topology in* $\mathbb{R}^3$. Invent. Math. **114** (1993), 625–639.
[13] Meeks, W. H., Rosenberg, H. *Maximum principles at infinity with applications to minimal and constant mean curvature surfaces.* Preprint.
[14] Meeks, W. H., Rosenberg, H. *The uniqueness of the helicoid and the asymptotic geometry of properly embedded minimal surfaces with finite topology.* Preprint.
[15] Morales, S. *On the existence of a proper minimal surface in* $\mathbb{R}^3$ *with the conformal type of a disc.* To appear in GAFA.
[16] Rodriguez, L., Rosenberg, H. *Minimal surfaces in* $\mathbb{R}^3$ *with one end and bounded curvature.* Manuscripta Math. **96**:1 (1998), 3–7.
[17] Rosenberg, H. *Minimal surfaces of finite type.* Bull. Soc. Math. France **123** (1995), 351–359.
[18] Xavier, F. *Why no new complete simply-connected embedded minimal surfaces have been found since 1776.* Preprint.

Departamento de Geometría y Topología, Universidad de Granada, E–18071 Granada, España

E-mail address: fjlopez@goliat.ugr.es

Clay Mathematics Proceedings
Volume 2, 2004

Optimal Isoperimetric Inequalities for Three-Dimensional Cartan–Hadamard Manifolds

Manuel Ritoré

Abstract. We present a new proof of the following result: consider a complete, simply connected, three-dimensional manifold, whose sectional curvatures are bounded from above by some constant $c \leqslant 0$. Then its isoperimetric profile is bounded from below by the one of Euclidean space if $c = 0$, or by the one of hyperbolic space of constant curvature c if $c < 0$.

1. Introduction

The following conjecture appeared in Aubin [**2**], in Burago–Zalgaller [**3**] and in Gromov–Lafontaine–Pansu [**9**]. It can also be stated in terms of Sobolev inequalities in Riemannian manifolds [**10**]. Recall that the *isoperimetric profile* I_M of a manifold M is the function $I_M : (0, \operatorname{vol} M) \to \mathbb{R}^+$ given by

$$I_M(V) = \inf\{\operatorname{area} \partial\Omega : \Omega \Subset M \text{ has smooth boundary and } \operatorname{vol}\Omega = V\}.$$

Conjecture 1.1. *Let M^n be a complete, simply connected, n-dimensional manifold, whose sectional curvatures satisfy inequality $K_{sec} \leqslant c \leqslant 0$, for some constant $c \leqslant 0$. Then the isoperimetric profile I_M of M^n is bounded from below by the isoperimetric profile I_c of the complete and simply connected space M_c^n whose sectional curvatures are equal to c. This implies:*

$$\operatorname{area}(\partial\Omega) \geqslant I_c(\operatorname{vol}\Omega), \tag{1–1}$$

for any domain $\Omega \subset M$ with smooth boundary. Moreover, if equality holds in (1–1)*, then Ω is isometric to the geodesic ball of volume* $\operatorname{vol}(\Omega)$ *in M_c^n.*

A *Cartan–Hadamard* manifold is a complete, simply connected Riemannian manifold whose sectional curvatures are nonpositive. Such a manifold is diffeomorphic, via the exponential map at any point, to the Euclidean space of the same dimension. A set Ω that satisfies $\operatorname{area}\partial\Omega = I_M(\operatorname{vol}\Omega)$ is called an *isoperimetric domain.* Isoperimetric domains could not exist in Riemannian manifolds, as shown in [**17**, Thm. 2.16].

©2004 Clay Mathematics Institute

The above conjecture has been referred to in the literature as the Cartan–Hadamard conjecture [**10**, 8.2], or as the Aubin conjecture [**16**, 17.3].

This conjecture was already proved in the two-dimensional case by A. Weil [**21**]. In fact it follows from the classical isoperimetric inequality for discs in surfaces with Gauss curvature bounded from above. If $K \leqslant K_0$, then

$$L^2 \geqslant 4\pi A - K_0 A^2,$$

where A denotes the area of a set, and L its perimeter. If the surface is a plane, then a classical argument (filling the holes of a region) shows that this inequality is also valid for any domain of arbitrary topological type.

Conjecture 1.1 has been proved by C. Croke [**6**] when $n = 4$ and $c = 0$. Croke obtained generic inequalities of the form $\operatorname{area} \partial\Omega \geqslant \beta_n I_0(\operatorname{vol}\Omega)$, where $\beta_n \leqslant 1$ and equality hold only for $n = 4$.

Two more reasons to believe that the above conjecture should hold are: (i) inequality (1–1) is true for geodesic balls in Cartan–Hadamard manifolds by classical comparison theorems, and (ii) the inequality

$$\operatorname{area}(\partial\Omega) \geqslant \varepsilon_n I_c(\operatorname{vol}\Omega),$$

holds, where $\varepsilon_n < 1$ are constants depending only on the dimension n of the manifold, see Croke [**5**], Hoffman–Spruck [**11**] (see also Michael–Simon [**13**]), and Burago–Zalgaller [**3**].

Conjecture 1.1 has been proved by B. Kleiner [**12**] for any $c \leqslant 0$ in dimension 3. In his proof there is a scheme common to any dimension. He only uses dimension three to prove the following proposition, applying the Gauss–Bonnet formula over the two-dimensional boundary of an isoperimetric domain:

PROPOSITION 1.2. *Let M^3 be a complete, simply connected, 3-dimensional manifold, with sectional curvatures bounded from above by a constant $c \leqslant 0$. Let Ω be a compact set with $C^{1,1}$ boundary Σ. Then*

$$\max_\Sigma H_\Sigma \geqslant H_c(\operatorname{area}\Sigma),$$

where H_c is the mean curvature in the model space M_c^3 of the geodesic ball of area equal to $\operatorname{area}\Sigma$.

In these notes we shall use the terms area and volume to refer to $(n-1)$ and n-dimensional Hausdorff measures, respectively.

2. Proof of Conjecture 1.1 Using Proposition 1.2

We show that Conjecture 1.1 is true in any dimension if the analogue of Proposition 1.2 is valid. As we said before, isoperimetric domains may not exist in a noncompact manifold. To solve this problem we shall work in geodesic balls in M.

The following result summarizes what we can say about isoperimetric domains in a manifold with boundary (in a geodesic ball in our case):

THEOREM 2.1 (Existence and regularity of isoperimetric domains in manifolds with boundary). *Let B^n be a compact manifold with smooth boundary ∂B^n. Let $V \in (0, \operatorname{vol} B^n)$. Then there is a domain $\Omega \subset B^n$ with boundary $\Sigma = \partial\Omega$ such that*

(i) $\operatorname{vol}\Omega = V$, $\operatorname{area}\Sigma = I_B(V)$.
(ii) $\Sigma = \partial\Omega$ *is* $C^{1,1}$ *in a neighborhood of* ∂B.
(iii) *There is a singular set* $\Sigma_{sing} \subset \Sigma \cap \operatorname{int} B$ *of Hausdorff dimension less than or equal to* $n-8$ *such that* $(\Sigma \cap \operatorname{int} B) - \Sigma_{sing}$ *is a smooth hypersurface with constant mean curvature* H.
(iv) *The mean curvature* h *of* Σ *is defined almost everywhere (except in a set of* $\mathcal{H}^{n-1}$*-measure zero), and we have* $h \leqslant H$.

Moreover, if Ω_n *is a sequence of isoperimetric domains in* B *such that* $\operatorname{vol}(\Omega_i) \to V$, *then* $\operatorname{area}\partial\Omega_i \to I_B(V)$.

Existence of isoperimetric domains follows from classical theorems of Geometric Measure Theory for finite perimeter sets. Regularity of $\Sigma - \Sigma_{sing}$ in the interior of B is obtained from Gonzalez, Massari, Tamanini [**8**]. For $C^{1,1}$ regularity near ∂B, one must consult White [**22**] and Stredulinski–Ziemer [**20**]. The last line in the statement implies the continuity of the isoperimetric profile, which also follows from Gallot [**7**].

The proof of (iv) is obtained by taking p in the regular part of Σ and q in the intersection $\Sigma \cap \partial B$. Consider functions u, v, defined in neighborhoods of p, q, respectively, so that $v \geqslant 0$ and strictly positive near q and $\int_\Sigma u\, d\Sigma = \int_\Sigma v\, d\Sigma$. Then we get a variation $(-u+v)\,n$, where n is the inner normal to Σ, that fixes the volume of Ω, and pushes Σ towards Ω near q. The derivative of the area for this variation is nonnegative, since Ω is isoperimetric, and equals

$$\int_\Sigma nH\, u\, d\Sigma - \int_\Sigma nh\, v\, d\Sigma \geqslant 0,$$

and so

$$nH \geqslant \frac{\displaystyle\int_\Sigma nh\, v\, d\Sigma}{\displaystyle\int_\Sigma v\, d\Sigma},$$

from which the claim follows.

2.1. We now prove Conjecture 1.1. Choose a geodesic ball B that contains the domain $\Omega \subset M$. We recall that the isoperimetric profile I_B is continuous by Theorem 2.1.

Let Ω_v be an isoperimetric domain in B of volume $v = \operatorname{vol}\Omega$, and let H_v be the (constant) mean curvature of the regular part of $\Sigma_v = \partial\Omega_v$ in the interior of B.

If Proposition 1.2 is true for any dimension, then $H_v \geqslant H_c(\operatorname{area}\Sigma_v)$, and equality holds for a geodesic ball of area equal to $\operatorname{area}\Sigma_v$ in a space of constant curvature c.

Choose a deformation Ω_V with support in the regular part of Σ_v in the interior of B parametrized with respect to volume (it is enough to consider a normal deformation uN, where $u \geqslant 0$). Then we have, for $\Delta V < 0$,

$$\begin{aligned}\frac{I_B(v+\Delta V) - I_B(v)}{\Delta V} &\geqslant \frac{\operatorname{area}\partial\Omega_{v+\Delta V} - \operatorname{area}\Sigma_v}{\Delta V} \geqslant (nH_v + \varepsilon(\Delta V)) \\ &\geqslant (nH_c(I_B(v)) + \varepsilon(\Delta V)) > 0,\end{aligned}$$

which implies that I_B is strictly monotone and, so, smooth almost everywhere. Moreover, if I_B is smooth in v, then

$$I'_B(v) \geqslant nH_c(\text{area}\,\Sigma_v). \tag{2–1}$$

Now we are ready to finish the proof, since translating the profile M^3_c to left and right we obtain a foliation of the upper half-plane in $\mathbb{R}^2$, and inequality (2–1) follows since the function I_B meets this foliation transversally, so that the profile lies above M^3_c, since $I_B(0) = I_c(0) = 0$.

If equality holds in $I_B(v) \geqslant I_c(v)$, then we have equality of the profiles for any $V \in (0, v)$, so that I_B is smooth, equality holds in (2–1) for any value $V \in (0, v)$ and, so, Ω is isometric to a ball of volume v in space M^3_c by Proposition 1.2.

3. Kleiner's Proof of Proposition 1.2

Proposition 1.2 is trivial if $\Sigma \subset M^3$ is a sphere, since

$$4\pi = \int_\Sigma K\, dA = \int_\Sigma (K_{sec} + \kappa_1\kappa_2)\, dA \leqslant \int_\Sigma (c + \kappa_1\kappa_2)\, dA \leqslant (c + H^2)\,\text{area}\,\Sigma,$$

and equality holds if and only if $K_{sec} \equiv c$ over Σ and the surface is totally umbilical. In case Σ is a geodesic sphere in a space of constant curvature c, equality holds in the above inequality. This shows that

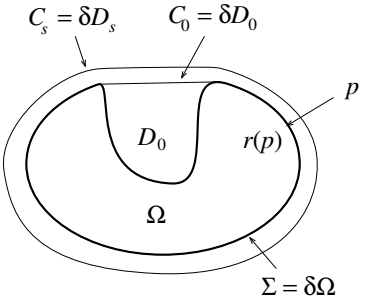
$$H \geqslant H_c(\text{area}\,\Sigma).$$

If equality holds in the above inequality, then Σ is a totally umbilical sphere so that the sectional curvature of the tangent plane equals c and Σ has the same second fundamental form as of the sphere of area Σ in M^3_c. It follows from Theorem 7 in [**19**] that the domain enclosed by Σ is a geodesic ball in M^3_ε.

Now assume that Σ is any $C^{1,1}$ surface which encloses a domain Ω. Consider the closed convex hull D_0 of Ω. The set D_0 is convex, and compact since it is contained in a (convex) ball of M^3. Of course nothing is known about the regularity of ∂D_0, so that by using ideas of Almgren [**1**], we consider the domains

$$D_s = \{x \in M^3 /\ d(x, D_0) \leqslant s\}.$$

We know that D_s is convex, that $r : M^3 - \text{int}\, D_0 \to \Sigma$ is well defined and nonincreasing, and that $C_s = \partial D_s$ is homeomorphic to $\mathbb{S}^2$.



Set $r_s = r\big|_{C_s}$. Since C_s is a $C^{1,1}$ surface, by Rademacher's Theorem (a Lipschitz function is smooth almost everywhere) its Gauss curvature and its Gauss–Kronecker curvature GK (product of principal curvatures) exist and the total curvature of C_s equals 4π. Then

$$\begin{aligned}
4\pi = \int_{C_s} K &= \int_{C_s} (K_{sec} + GK_{C_s}) \leqslant \int_{C_s} (c + GK_{C_s}) \\
&= \int_{r_s^{-1}(\Sigma)} (c + GK_{C_s}) + \int_{C_s - r_s^{-1}(\Sigma)} (c + GK_{C_s}) \\
&\leqslant \int_{r_s^{-1}(\Sigma)} (c + H_s^2) + c\,\mathrm{area}(C_s - r_s^{-1}(\Sigma)) + \int_{C_s - r_s^{-1}(\Sigma)} GK_{C_s} \\
&\leqslant \int_{r_s^{-1}(\Sigma)} (c + H_s^2) + \int_{C_s - r_s^{-1}(\Sigma)} GK_{C_s}.
\end{aligned}$$

In the first inequality we have bounded K_{sec} by c, and in the second one GK_{C_s} by the mean curvature of C_s. Equality holds in the above inequality if and only if $K_{sec} = c$ along C_s and $r_s^{-1}(\Sigma)$ is totally umbilical.

We now treat the integrals that appear in the last line.

We first check that

$$\lim_{s\to 0} \int_{r_s^{-1}(\Sigma)} (c + H_s^2) \leqslant (c + H_0^2)\,\mathrm{area}(\Sigma \cap C_0). \tag{3–1}$$

We only have to take into account that if C_s is C^2 in p (this happens for almost every $p \in C_s$) and $p \in C_s$ and $r_s(p) \in C_0 \cap \Sigma$, then we have

$$2H_s(p) \leqslant 2H_0 - s\,(\mathrm{Ric}_-),$$

where Ric_- is the infimum of the Ricci curvatures in the unit sphere at every point of C_s. Passing to the limit when $s \to 0$ we have (3–1). The way of getting the above inequality is to apply the formula

$$\frac{d(2H_t)(p)}{dt} = -(\mathrm{Ric}(N,N) + |\sigma|^2) \leqslant -\mathrm{Ric}(N,N),$$

and integrate with respect to t between 0 and s. If equality holds in (3–1), then $\mathrm{area}\, C_0 \cap \Sigma = \mathrm{area}\,\Sigma$, so that $\mathrm{vol}\, D_0 = \mathrm{vol}\,\Omega$, whence we conclude $D_0 = \Omega$. It follows that Σ is convex.

Now we check that $\lim_{s\to 0} \mathrm{area}(r_s^{-1}(\Sigma)) = \mathrm{area}(C_0 \cap \Sigma)$. We use the area formula for Lipschitz maps and get

$$\int_{r_s^{-1}(\Sigma)} \mathrm{Jac}(r_s)\, dC_s = \mathrm{area}(C_0 \cap \Sigma),$$

so that

$$\begin{aligned}
\mathrm{area}(r_s^{-1}(\Sigma)) &= \int_{r_s^{-1}(\Sigma)} \mathrm{Jac}(r_s)\, dC_s + \int_{r_s^{-1}(\Sigma)} (1 - \mathrm{Jac}(r_s))\, dC_s \\
&= \mathrm{area}(C_0 \cap \Sigma) + \int_{r_s^{-1}(\Sigma)} (1 - \mathrm{Jac}(r_s))\, dC_s \to \mathrm{area}(C_0 \cap \Sigma),
\end{aligned}$$

since $\mathrm{Jac}(r_s)$ converges uniformly to 1.

We finally see

$$\lim_{s\to 0}\int_{C_s - r_s^{-1}(\Sigma)} GK_{C_s} = 0.$$

Note first that if $p \in C_0 - \Sigma$ is a smooth point, then $GK_{C_0}(p) = 0$: otherwise we could push D_0 near p towards the interior of D_0 to contradict the convex hull property of D_0.

Fix now $s_0 > 0$ and write

$$\int_{C_s - r_s^{-1}(\Sigma)} GK_{C_s}\, dC_s = \int_{C_{s_0} - r_{s_0}^{-1}(\Sigma)} (GK_{C_s} \circ r_{s_0 s})\, \mathrm{Jac}(r_{s_0 s})\, dC_{s_0},$$

where $r_{s_0 s} = r_s^{-1} \circ r_{s_0}$. The right integral is uniformly bounded because the second fundamental form of C_s in $q = r_{s_0 s}(p)$ applied to a unit vector $e \in T_q C_s$ equals

$$\langle A_s(e), e\rangle = \frac{\langle J', J\rangle}{|J|^2},$$

where J is a Jacobi field along the geodesic $\gamma(s) = r_{s_0 s}(p)$ that has modulus 1 over C_{s_0}, orthogonal to C_{s_0}, and such that $J/|J| = e$. J is induced by a family of orthogonal geodesics leaving from C_{s_0}. The quantity $\langle A_s(e), e\rangle$ is bounded by the classical comparison theorems for geodesics starting from a submanifold with sectional curvature bounded from below [**4**, §7.3]. We remark that the second fundamental form of C_s is uniformly bounded from above, since every point in C_s is supported by a ball of radius s.

By the preceding discussion, $GK_{C_s} \circ r_{s_0 s}$ converges to 0 for almost every point of $C_{s_0} - r_{s_0}^{-1}(\Sigma)$, so that the integral of GK_{C_s} converges to 0 in $C_{s_0} - r_{s_0}^{-1}(\Sigma)$.

Then we conclude

$$4\pi \leqslant (c + H_0^2)\,\mathrm{area}\,\Sigma,$$

which implies

$$H_0 \geqslant H_c(\mathrm{area}\,\Sigma).$$

If equality holds in the above inequality then, an analysis of the possibilities following Kleiner [**12**], yields

- Σ has constant mean curvature equal to the one of the geodesic ball of the same area in M_c^3.
- The sectional curvature of the tangent plane to Σ equals c.
- Σ is totally umbilical.

Then a result by Schroeder–Ziller [**19**] implies that Σ is a geodesic ball in M_c^3.

4. A New Proof of Proposition 1.2

4.1. The Euclidean case. Consider first the case $K_{sec} \leqslant 0$. Let Σ be an embedded $C^{1,1}$ compact surface. This surface has principal curvatures defined almost everywhere. We shall prove that

$$\int_\Sigma H^2\, dA \geqslant 4\pi.$$

Since the integral of the squared mean curvature of any geodesic sphere in Euclidean space equals 4π, we get the inequality in the statement of Proposition 1.2. Later we shall study the case of equality.

Let $p \in \Sigma$ and let $d(q)$ measure the distance to p. We consider the conformal metric $g_\varepsilon = \rho_\varepsilon^2\, g = e^{2u_\varepsilon}\, g$, where

$$\rho_\varepsilon = \frac{2\varepsilon}{1+\varepsilon^2 d^2}, \qquad u_\varepsilon = \log\left(\frac{2\varepsilon}{1+\varepsilon^2 d^2}\right).$$

In case M is the Euclidean space this metric is the one obtained by applying a conformal transformation to the metric of the sphere and projecting this metric orthogonally to the Euclidean space by means of stereographic projection.

Now we verify that

$$\int_\Sigma H^2\, dA \geqslant 4\pi,$$

where equality holds if and only if Ω is flat (vanishing sectional curvatures).

By taking into account the well-known relation between conformal metrics we get

$$e^{2u_\varepsilon} K_\varepsilon \geqslant K + e^{2u_\varepsilon}, \tag{4–1}$$

where K_ε and K are the sectional curvatures of a given plane of M for the metrics g_ε and g, respectively. From now on we shall assume that they are the ones of the tangent plane to Σ. So

$$\begin{aligned}\int_\Sigma H^2\, dA &= \int_\Sigma (H^2+K)\, dA - \int_\Sigma K\, dA = \int_\Sigma ((H_\varepsilon^2)+K_\varepsilon)\, dA_\varepsilon - \int_\Sigma K\, dA \\ &\geqslant \int_\Sigma H_\varepsilon^2\, dA_\varepsilon + \int_\Sigma dA_\varepsilon \geqslant \int_\Sigma dA_\varepsilon,\end{aligned}$$

where in the first equality we have used the conformal invariance of $\int (H^2+K_{sec})\, dA$, and in the first inequality we have used inequality (4–1). The limit of the last integral can be computed by passing to polar (ambient) coordinates, or taking into account that it corresponds geometrically to blowing up the surface Σ at the point p with a spherical metric. Therefore

$$\lim_{\varepsilon\to\infty} \int_\Sigma dA_\varepsilon = 4\pi.$$

From the two last inequalities we obtain the desired estimate.

To analyze what happens when equality holds we need a more accurate estimate of the expression of the curvatures in the conformal metrics. So we write

$$e^{2u_\varepsilon} K_\varepsilon = K - \left(\frac{\varepsilon^2}{1+\varepsilon^2 d^2}\right)^2 4d^2 + \left(\frac{\varepsilon^2}{1+\varepsilon^2 d^2}\right)\left(\nabla^2 d^2(X,X) + \nabla^2 d^2(Y,Y)\right),$$

where X, Y is an orthonormal basis of the tangent plane to Σ. From this formula we get

$$4\pi = \int_\Sigma H^2\,dA = \int_\Sigma (H^2+K)\,dA - \int_\Sigma K\,dA = \int_\Sigma ((H_\varepsilon^2)+K_\varepsilon)\,dA_\varepsilon - \int_\Sigma K\,dA$$
$$= \int_\Sigma dA_\varepsilon + \int_\Sigma \left(\frac{\varepsilon^2}{1+\varepsilon^2 d^2}\right)\left(\nabla^2 d^2(X,X)+\nabla^2 d^2(Y,Y)-4\right)dA + \int_\Sigma H_\varepsilon^2\,dA_\varepsilon.$$

We already know that the first integral converges to 4π when $\varepsilon \to \infty$. So the limit of the remaining integrals is 0. Since $\nabla^2 d^2(X,X) \geqslant 2$ for any $|X|=1$ we obtain that both integrals are positive and, in particular,

$$\nabla^2 d^2(X,X) = \nabla^2 d^2(Y,Y) = 2.$$

Standard comparison theorems in Riemannian Geometry show that, if the geodesic starting from p leaves the enclosed domain Ω in a nontangential way, then $\nabla^2 d^2 = 2\,g$ at the hitting point. Standard comparison shows that $\nabla d^2 \equiv 2\,g$ along the geodesic. Moving slightly the geodesic we get a cone so that $\nabla d^2 \equiv 2\,g$ inside this cone. Since every point in the interior of Ω can be connected with Σ by a minimizing geodesic hitting Σ orthogonally, we conclude that every point inside Σ is flat and so Ω is flat.

4.2. We now see what happens if $(\max_\Sigma H^2)\,\mathrm{area}(\Sigma) = 4\pi$. In this case, in addition to Ω being flat, it follows that the mean curvature of the boundary is constant. For any domain of this type, Ros [**18**] and Montiel–Ros [**15**] have proved that

$$3\ \mathrm{vol}\,\Omega \leqslant \frac{1}{H}\ \mathrm{area}\,\Sigma,$$

and equality holds if and only if Ω is isometric to a geodesic ball in Euclidean space. But the classical Minkowski formula

$$3\ \mathrm{vol}\,\Omega = \frac{1}{H}\ \mathrm{area}\,\Sigma,$$

holds in Ω since the function $(1/2)\,d^2$ has Hessian on Ω proportional to twice the identity matrix. From this we conclude our proof of Proposition 1.2 in the flat case.

4.3. The hyperbolic case $K_{sec} \leqslant -1$. In the hyperbolic case one has to consider the family of conformal metrics

$$g_\varepsilon = \left(\frac{2\varepsilon}{(1-\varepsilon^2)+(1+\varepsilon^2)\,\cosh d}\right) g, \qquad \varepsilon > 1.$$

This family of metrics is obtained by writing the spherical metric in a disc D of $\mathbb{R}^n$ via stereographic projection in terms of the hyperbolic metric of constant curvature -1 in D. So we obtain $e^{2u_\varepsilon} K_\varepsilon \geqslant K + e^{2u_\varepsilon} + 1$ and

$$\int_\Sigma (-1+H^2)\,dA \geqslant \int_\Sigma dA_\varepsilon.$$

As in the previous case one proves that $\lim_{\varepsilon\to\infty} \int_\Sigma dA \to 4\pi$, which yields the desired estimate.

To analyze equality it is more convenient to write

$$e^{2u_\varepsilon} K_\varepsilon = K + 1 + e^{2u_\varepsilon} + \Big(\frac{1+\varepsilon^2}{(1-\varepsilon^2)+(1+\varepsilon^2)\cosh d}\Big)\big(\nabla^2 \cosh d(X,X) + \nabla^2 \cosh d(Y,Y) - 2\cosh d\big).$$

We recall that by classical comparison theorems, when $K_{sec} \leqslant -1$ we get

$$\nabla^2 \cosh d \geqslant \cosh d \, \langle\, , \rangle\,,$$

so that the factor in the previous displayed line is nonnegative. Hence

$$4\pi = \int_\Sigma (-1+H^2)\, dA = \int_\Sigma dA_\varepsilon + \int_\Sigma H_\varepsilon^2\, dA_\varepsilon + \int_\Sigma \Big(\frac{1+\varepsilon^2}{(1-\varepsilon^2)+(1+\varepsilon^2)\cosh d}\Big) \times \big(\nabla^2 \cosh d(X,X) + \nabla^2 \cosh d(Y,Y) - 2\cosh d\big)\, dA.$$

Letting $\varepsilon \to \infty$ and taking into account that $\lim_{\varepsilon\to\infty} \int_\Sigma dA_\varepsilon = 4\pi$ we deduce that the remaining positive integrals tend to 0 when $\varepsilon \to \infty$. In particular

$$\nabla^2 \cosh d(X,X) = \nabla^2 \cosh d(Y,Y) = \cosh d.$$

By standard comparison theorems, and arguing as in the Euclidean case, we conclude that the metric in Ω is hyperbolic.

4.4. If

$$\max_\Sigma (-1+H^2)\, \mathrm{area}\, \Sigma = 4\pi,$$

then H is constant. Moreover, from [**14**, Theorem 9] we conclude, by taking inner parallels

$$\int_\Sigma \big(\cosh d + H \sinh d \, \langle \partial/\partial d, N\rangle\big)\, dA \geqslant 0,$$

and equality holds only when Σ is a geodesic sphere. But since the metric in Ω is hyperbolic we have $\nabla^2 \cosh d = 2\,\langle\, , \rangle$, so that

$$\int_\Sigma \big(\cosh d + H \sinh d \, \langle \partial/\partial d, N\rangle\big)\, dA = 0,$$

and Proposition 1.2 also follows in the hyperbolic case.

References

1. F. Almgren, *Optimal isoperimetric inequalities*, Bull. Amer. Math. Soc. (N.S.) **13** (1985), no. 2, 123–126. MR **86j:**49084
2. Thierry Aubin, *Problèmes isopérimétriques et espaces de Sobolev*, J. Differential Geometry **11** (1976), no. 4, 573–598. MR 56 #6711
3. Yu. D. Burago and V. A. Zalgaller, *Geometric inequalities*, Springer-Verlag, Berlin, 1988, Translated from the Russian by A. B. Sosinskiĭ, Springer Series in Soviet Mathematics. MR **89b:**52020
4. Isaac Chavel, *Riemannian Geometry: a Modern Introduction*, Cambridge Univ. Press, Cambridge, 1993.
5. Christopher B. Croke, *Some isoperimetric inequalities and eigenvalue estimates*, Ann. Sci. École Norm. Sup. (4) **13** (1980), no. 4, 419–435. MR **83d:**58068
6. ———, *A sharp four-dimensional isoperimetric inequality*, Comment. Math. Helv. **59** (1984), no. 2, 187–192. MR **85f:**53060

7. Sylvestre Gallot, *Inégalités isopérimétriques et analytiques sur les variétés riemanniennes*, Astérisque (1988), no. 163-164, 5–6, 31–91, 281 (1989), On the geometry of differentiable manifolds (Rome, 1986). MR **90f:**58173
8. E. Gonzalez, U. Massari, and I. Tamanini, *On the regularity of boundaries of sets minimizing perimeter with a volume constraint*, Indiana Univ. Math. J. **32** (1983), no. 1, 25–37. MR **84d:**49043
9. Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Birkhäuser Boston Inc., Boston, MA, 1999, Based on the 1981 French original [MR 85e:53051], With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates. MR **2000d:**53065
10. Emmanuel Hebey, *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*, Courant Institute of Mathematical Sciences, New York, 1999. MR **2000e:**58011
11. David Hoffman and Joel Spruck, *Sobolev and isoperimetric inequalities for Riemannian submanifolds*, Comm. Pure Appl. Math. **27** (1974), 715–727. MR 51 #1676
12. Bruce Kleiner, *An isoperimetric comparison theorem*, Invent. Math. **108** (1992), no. 1, 37–47. MR **92m:**53056
13. J. H. Michael and L. M. Simon, *Sobolev and mean-value inequalities on generalized submanifolds of R^n*, Comm. Pure Appl. Math. **26** (1973), 361–379. MR 49 #9717
14. Sebastián Montiel, *Unicity of constant mean curvature hypersurfaces in some Riemannian manifolds*, Indiana Univ. Math. J. **48** (1999), no. 2, 711–748. MR **2001f:**53131
15. Sebastián Montiel and Antonio Ros, *Compact hypersurfaces: the Alexandrov theorem for higher order mean curvatures*, Differential geometry, Longman Sci. Tech., Harlow, 1991, pp. 279–296. MR **93h:**53062
16. Frank Morgan, *Geometric measure theory*, third ed., Academic Press Inc., San Diego, CA, 2000, A beginner's guide. MR 1 775 760
17. Manuel Ritoré, *Constant geodesic curvature curves and isoperimetric domains in rotationally symmetric surfaces*, Comm. Anal. Geom. **9**, (2001) no. 5, 1093–1138.
18. Antonio Ros, *Compact hypersurfaces with constant higher order mean curvatures*, Rev. Mat. Iberoamericana **3** (1987), no. 3-4, 447–453. MR **90c:**53160
19. V. Schroeder and W. Ziller, *Local rigidity of symmetric spaces*, Trans. Amer. Math. Soc. **320** (1990), no. 1, 145–160.
20. Edward Stredulinsky and William P. Ziemer, *Area minimizing sets subject to a volume constraint in a convex set*, J. Geom. Anal. **7** (1997), no. 4, 653–677. MR **99k:**49089
21. André Weil, *Sur les surfaces a courbure negative*, C. R. Acad. Sci. Paris **182** (1926), 1069–1071.
22. Brian White, *Existence of smooth embedded surfaces of prescribed genus that minimize parametric even elliptic functionals on 3-manifolds*, J. Differential Geom. **33** (1991), no. 2, 413–443. MR **92e:**58048

Departamento de Geometría y Topología, Universidad de Granada, E–18071 Granada, España

E-mail address: ritore@ugr.es

Clay Mathematics Proceedings
Volume 2, 2004

Embedded Minimal Disks

Tobias H. Colding and William P. Minicozzi II

Part I. Main Theorem: The Limit Foliation and the Singular Curve

This paper is a survey of our results about embedded minimal disks. Unlike our expository article [**CM15**], this paper is intended for readers with some background knowledge on minimal surfaces. However even an expert reader may find it worthwhile to look at [**CM15**] first.

We start with two key examples of embedded minimal disks in $\mathbf{R}^3$. The first are minimal graphs over simply connected domains and the second is the double spiral staircase (see [**CM15**]) known as the helicoid.

EXAMPLE 1. Graphs of functions $u : \Omega \to \mathbf{R}$ where $\Omega \subset \mathbf{R}^2$ is simply connected and u satisfies the second order nonlinear elliptic equation in divergence form (the so-called minimal surface equation)

$$\operatorname{div}\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right) = 0. \tag{I.0.1}$$

EXAMPLE 2 (Helicoid). See Figure 1. The minimal surface in $\mathbf{R}^3$ parametrized by

$$(s\cos t, s\sin t, -t) \text{ where } s,\, t \in \mathbf{R}. \tag{I.0.2}$$

One of our main theorems is that every embedded minimal disk can either be modeled by a minimal graph or by a piece of the helicoid depending on whether the curvature is small or not; see Theorem I.0.5 below. We will in this survey discuss some of the ingredients in the proof of this and how those ingredients fit together.

To be able to discuss the helicoid some more and give a precise meaning to its being a double spiral staircase, we will need the notion of a multivalued graph; see Figure 2. Let D_r be the disk in the plane centered at the origin and of radius r and let $\mathcal{P}$ be the universal cover of the punctured plane $\mathbf{C} \setminus \{0\}$ with global polar coordinates (ρ, θ) so $\rho > 0$ and $\theta \in \mathbf{R}$. An N-valued graph of a function u on the annulus $D_s \setminus D_r$ is a single valued graph over

$$\{(\rho, \theta) : r \leq \rho \leq s,\, |\theta| \leq N\,\pi\}. \tag{I.0.3}$$

Colding was partially supported by NSF Grant DMS 9803253 and an Alfred P. Sloan Research Fellowship, and Minicozzi by NSF Grant DMS 9803144 and an Alfred P. Sloan Research Fellowship.

©2004 Clay Mathematics Institute

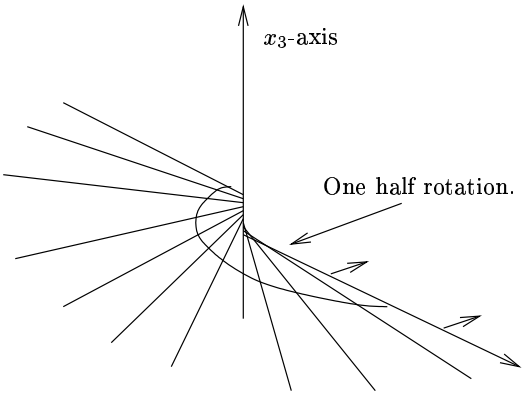


FIGURE 1. The helicoid is obtained by gluing together two ∞-valued graphs along a line. The two multivalued graphs are given in polar coordinates by $u_1(\rho,\theta) = -\theta$ and $u_2(\rho,\theta) = -\theta + \pi$. In either case $w(\rho,\theta) = -2\,\pi$.

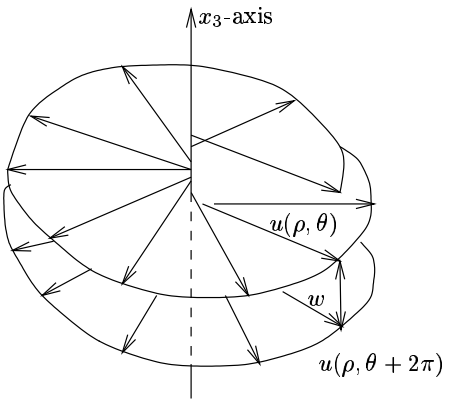


FIGURE 2. The separation of a multivalued graph. (Here the multivalued graph is shown with negative separation.)

The multivalued graphs that we will consider will never close up; in fact they will all be embedded. Note that embedded corresponds to the separation never vanishing. Here the separation (see Figure 2) is the function given by

$$w(\rho,\theta) = u(\rho,\theta+2\pi) - u(\rho,\theta). \tag{I.0.4}$$

If Σ is the helicoid, then $\Sigma \setminus x_3 - \text{axis} = \Sigma_1 \cup \Sigma_2$, where Σ_1, Σ_2 are ∞-valued graphs. Σ_1 is the graph of the function $u_1(\rho,\theta) = -\theta$ and Σ_2 is the graph of the function $u_2(\rho,\theta) = -\theta + \pi$. In either case the separation $w = -2\,\pi$. A multivalued minimal graph is a multivalued graph of a function u satisfying the minimal surface equation (I.0.1).

Here, as in [**CM6**] and [**CM8**], we have normalized so our embedded multivalued graphs have negative separation. This can be achieved after possibly reflecting in a plane.

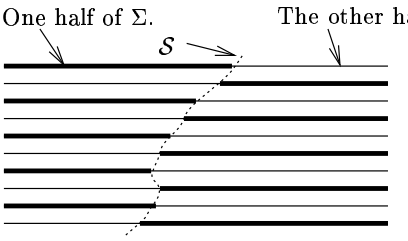


FIGURE 3. Theorem I.0.5: the singular set, $\mathcal{S}$, and the two multivalued graphs.

Now let $\Sigma_i \subset B_{2R}$ be a sequence of embedded minimal disks with $\partial\Sigma_i \subset \partial B_{2R}$. Clearly (after possibly going to a subsequence) either (1) or (2) occur:

(1) $\sup_{B_R \cap \Sigma_i} |A|^2 \leq C < \infty$ for some constant C.
(2) $\sup_{B_R \cap \Sigma_i} |A|^2 \to \infty$.

In (1) (by a standard argument) the intrinsic ball $\mathcal{B}_s(y_i)$ is a graph for all $y_i \in B_R \cap \Sigma_i$, where s depends only on C. Thus the main case is (2), which is the subject of the next theorem.

Using the notion of multivalued graphs, our main theorem can now be stated:

THEOREM I.0.5. [**CM6**, Theorem 0.1] (See Figure 3.) Let $\Sigma_i \subset B_{R_i} = B_{R_i}(0) \subset \mathbf{R}^3$ be a sequence of embedded minimal disks with $\partial\Sigma_i \subset \partial B_{R_i}$ where $R_i \to \infty$. If $\sup_{B_1 \cap \Sigma_i} |A|^2 \to \infty$, then there exists a subsequence, Σ_j, and a Lipschitz curve $\mathcal{S} : \mathbf{R} \to \mathbf{R}^3$ such that after a rotation of $\mathbf{R}^3$:

(1) $x_3(\mathcal{S}(t)) = t$. (That is, $\mathcal{S}$ is a graph over the x_3-axis.)
(2) Each Σ_j consists of exactly two multivalued graphs away from $\mathcal{S}$ (which spiral together).
(3) For each $1 > \alpha > 0$, $\Sigma_j \setminus \mathcal{S}$ converges in the C^α-topology to the foliation, $\mathcal{F} = \{x_3 = t\}_t$, of $\mathbf{R}^3$.
(4) $\sup_{B_r(\mathcal{S}(t)) \cap \Sigma_j} |A|^2 \to \infty$ for all $r > 0$, $t \in \mathbf{R}$. (The curvatures blow up along $\mathcal{S}$.)

In (2) and (3), that the $\Sigma_j \setminus \mathcal{S}$ are multivalued graphs and converge to $\mathcal{F}$ means that, for each compact subset $K \subset \mathbf{R}^3 \setminus \mathcal{S}$ and j sufficiently large, $K \cap \Sigma_j$ consists of multivalued graphs over (part of) $\{x_3 = 0\}$ and $K \cap \Sigma_j \to K \cap \mathcal{F}$ in the sense of graphs.

Theorem I.0.5 (like many of the other results discussed below) is modeled by the helicoid and its rescalings. Take a sequence $\Sigma_i = a_i\,\Sigma$ of rescaled helicoids where $a_i \to 0$. The curvatures of this sequence are blowing up along the vertical axis. The sequence converges (away from the vertical axis) to a foliation by flat parallel planes. The singular set $\mathcal{S}$ (the axis) then consists of removable singularities.

Of the many different paths that one could choose through our results about embedded minimal disks, we have chosen here one which discusses some of our more elementary results in greater detail and then only gives a very rough overview of some of our key results which are more difficult. Our hope is that by doing so this survey can serve as a reading guide for our papers where the reader is eased into the subject and then is shown the anatomy of the proof of our main theorem.

One of these more elementary themes is that of analysis of multivalued embedded minimal graphs—from estimates on the growth and decay of the separation between consecutive sheets of such graphs to the results about why such graphs are proper in a certain qualitative sense if they are contained in larger embedded minimal disks. These results about multivalued graphs are among what is discussed in some detail in the first part where we also discuss some of the other main results that go into the proof of our main theorem that any embedded minimal disk is either a graph of a function or can be approximated by a piece of a helicoid. In addition, we also discuss in some detail in the first part why the singular set, that is the set of points of large curvature in an embedded minimal disk, must all be contained in a curve which is a Lipschitz graph over a straight line.

In the second part we come to a less elementary result. Namely, that of why near a point of large curvature of an embedded minimal disk there must inside the disk be a small multivalued graph. Here small means on the scale of 1 over the square root of the maximum of the curvature. The much less elementary, but key, result of why such small multivalued graphs extend to large ones is only discussed very briefly in the second section of that second part.

In the third part we discuss the crucial one-sided curvature estimate. This is the estimate that gives a cone condition for all the points of large curvature. Namely, given a point of large curvature of an embedded minimal disk, then the cone condition is that all the other points of large curvature must lie within a double convex cone of the initial point of large curvature. Iterating this condition gives that the set of all the points of large curvature is the Lipschitz graph over a straight line. The one-sided curvature estimates uses all the results discussed in the first part.

In the final part we discuss what the differences are between the so-called local and global case. The local case is where we have a sequence of embedded minimal disks in a ball of fixed radius in $\mathbf{R}^3$—the global case is where the disks are in a sequence of expanding balls with radii tending to infinity. As we will see in the final part, then in the local case we can get limits with singularities. In the global case this does not happen because in fact any limit is always a foliation by flat parallel planes (this is assuming that the curvatures of the sequence are blowing up).

There are a number of important results that go into the proof of our main theorem about embedded minimal disks that are either not discussed here or are barely mentioned. One of these is why given a point of large curvature in an embedded minimal disk there are points of large curvature nearby above and below. This was one of the key results proven in [**CM5**] but is quite technical and thus has been omitted from this survey.

Let x_1, x_2, x_3 be the standard coordinates on $\mathbf{R}^3$. For $y \in \Sigma \subset \mathbf{R}^3$ and $s > 0$, the extrinsic and intrinsic balls are $B_s(y)$, $\mathcal{B}_s(y)$.

K_Σ is the sectional curvature of a smooth compact surface Σ and when Σ is immersed A_Σ will be its second fundamental form (so when Σ is minimal, then $|A|^2 = -2\,\mathrm{K}_\Sigma$). When Σ is oriented, $\mathbf{n}_\Sigma$ is the unit normal.

Using Theorems I.0.5 and I.3.3, W. Meeks and H. Rosenberg proved that the plane and helicoid are the only complete properly embedded simply-connected minimal surfaces in $\mathbf{R}^3$, [**MeRo**].

I.1. Estimates for Multivalued Minimal Graphs

We will later see that, just like the helicoid, general embedded minimal disks with large curvature at some interior point can be built out of multivalued graphs. This will be particularly useful once we have a good understanding of general embedded multivalued minimal graphs. In this section, we discuss three basic, but useful, estimates for such graphs:

- A gradient estimate for the separation which implies sublinear growth of w when there are enough sheets.
- A curvature estimate for 2-valued minimal graphs whose separation grows sublinearly.
- Sharp logarithmic upper and lower bounds for the separation when there is a growing number of sheets.

(See also [**CM7**] for further discussion of these estimates, their analogs for minimal annuli, and their implications.)

The first of these basic estimates was obtained in [**CM3**] where we showed (essentially by a gradient estimate) that the separation between the sheets grows sublinearly (in Theorem I.1.6 below we will discuss an improvement of this when the number of sheets grows sufficiently fast). Precisely, in [**CM3**] we showed that:

PROPOSITION I.1.1. [**CM3**, Proposition II.2.12] Given $\alpha > 0$, there exists N_α so if Σ is an embedded N_α-valued minimal graph over $D_{\mathrm{e}^{N_\alpha} R} \setminus D_{\mathrm{e}^{-N_\alpha}}$ of u and $1 \leq \rho \leq R$, then

$$\rho^{-\alpha} \leq \frac{w(\rho, 0)}{w(1, 0)} \leq \rho^{\alpha}. \tag{I.1.2}$$

Thus, choosing $\alpha < 1$, (I.1.2) gives the sublinear growth of the separation w. This sublinear growth of the separation is the main benefit of having at least N_α sheets. It comes from integrating the gradient bound proven in Proposition II.2.12 in [**CM3**]

$$\frac{|\nabla w|}{|w|} \leq \frac{\alpha}{\rho}. \tag{I.1.3}$$

Many of our results on embedded multivalued graphs apply as long as (I.1.2) holds and we have at least two sheets.

To get the better logarithmic bounds on the separation, one needs curvature estimates for embedded multivalued graphs. It follows from Heinz's curvature estimate for minimal graphs [**CM1**, Theorem 2.4] that (away from its boundary) a multivalued minimal graph has quadratic curvature decay

$$|A|^2 \leq C\, r^{-2}. \tag{I.1.4}$$

This scale-invariant estimate (I.1.4) does not require embeddedness and can easily be seen to be sharp without any further assumptions on the graph. However, using the embeddedness — and, in particular, the sublinear growth of the separation that embeddedness implies — we showed in Corollary 2.3 of [**CM8**] that the curvature of a multivalued embedded minimal graph decays faster than quadratically. That is, we showed that

$$|A|^2 \leq C\, r^{-2-5/6} \tag{I.1.5}$$

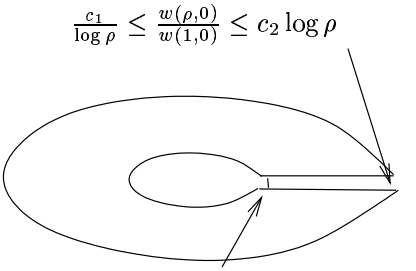


FIGURE 4. Theorem I.1.6: The sharp logarithmic upper and lower bounds for the separation.

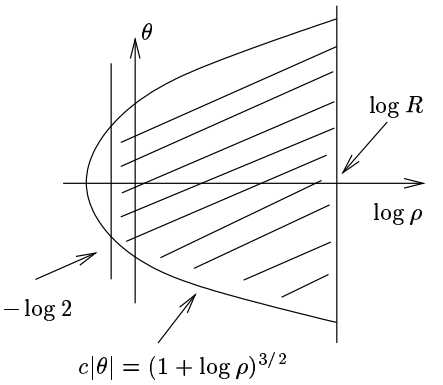


FIGURE 5. The domain of u in Theorem I.1.6 in $(\log \rho, \theta)$-coordinates.

for embedded 2-valued minimal graphs whose separation grows sublinearly; compare (I.1.2). (Note that by [**CM3**], there exists N_α so this applies to any N_α-valued embedded minimal graph.) The important point for the applications in [**CM8**] is that the faster than linear decay on the Hessian of u that (I.1.5) implies together with the nonlinear form of (I.0.1) gives that $|\Delta u|$ decays faster than quadratically; see equation (3.6) in [**CM8**].

The preceding faster than quadratic bound (I.1.5) used the nonlinearity of the minimal surface equation, much as nonlinearity is needed for proving the Bernstein theorem. In other situations the nonlinearity seems more to add difficulties than to be of any help. In those situations one tries to model the minimal surface equation by the Laplace equation and use that, from (I.1.5), if u is a multivalued solution of the minimal surface equation, then $|\Delta u|$ decays faster than quadratically so that u "becomes more and more like a harmonic function". We will now take advantage of this to discuss some analysis of such multivalued solutions that we will need later. This analysis is from [**CM8**].

The first such result is the following sharp upper and lower logarithmic bound:

THEOREM I.1.6. [**CM8**, Theorem 1.3] (See Figures 4 and 5.) Given $c > 0$, there exists c_1, c_2 such that if u satisfies (I.0.1) on $\{(\rho,\theta)\,|1/2 \leq \rho \leq R$ and $c|\theta| \leq (1+\log \rho)^{3/2}\}$ and $w < 0$ together with a slight condition on the growth of u and

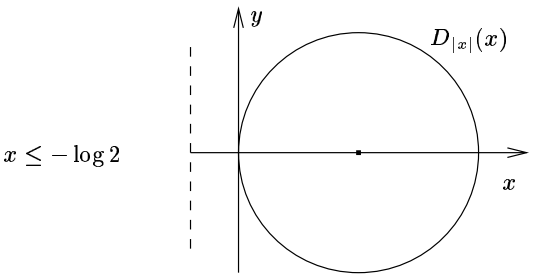


FIGURE 6. Proof of Theorem I.1.6: Applying the mean value equality to a negative harmonic function defined in a half-space.

w (see equation (1.5) in [**CM8**]), then for $2 \leq \rho \leq R^{1/2}$,

$$\frac{c_1}{\log \rho} \leq \frac{w(\rho,0)}{w(1,0)} \leq c_2 \log \rho. \tag{I.1.7}$$

The idea of the lower bound for the separation, i.e., $w(\rho,0)/w(1,0) \geq c_1/\log\rho$: If u is as in Theorem I.1.6, then $w < 0$ is almost harmonic and it can be seen to follow from (I.1.5) and the form of (I.0.1) that so is (the conformally transformed function) $\tilde{w}(x,y) = w(\mathrm{e}^x, y)$. Moreover, $\tilde{w} < 0$ is defined on a domain that is conformally close to a half-disk [**CM8**, Lemma 3.2]. Suppose for a moment that $\tilde{w}$ was actually harmonic and defined on the half-space $\{x > -\log 2\}$; then, by the mean value equality (see Figure 6) and the sign on $\tilde{w}$,

$$w(\mathrm{e}^x,0) = \tilde{w}(x,0) = \frac{1}{2\pi x}\int_{\partial D_x(x,0)} \tilde{w} \leq \frac{1}{2\pi x}\int_{D_1 \cap \partial D_x(x,0)} \tilde{w}. \tag{I.1.8}$$

By the Harnack inequality, it would then follow from (I.1.8) that

$$w(\mathrm{e}^x,0) = \tilde{w}(x,0) \leq c_1\, \tilde{w}(0,0)/x,$$

which is the desired lower bound. The upper bound follows similarly or by an inversion formula; see Section 3 of [**CM8**].

By Theorem I.1.6, the fastest possible decay for $w(\rho,0)/w(1,0)$ is $c_1/\log\rho$. Hence if $\tilde{w}(x+iy) = w(\mathrm{e}^x, y)$, then the fastest possible decay for $\tilde{w}(x,0)/\tilde{w}(0,0)$ is c_1/x and, as mentioned above, by using (I.1.5) it can be seen that $\tilde{w}$ is almost harmonic. This decay is achieved for the harmonic function $\tilde{v}(z) = -\mathrm{Re}\, z^{-1} = -x/(x^2+y^2) < 0$ (see Figure 7) and if

$$\begin{aligned} u(\rho,\theta) &= \int_0^\theta v(\rho,y)\,dy = \int_0^\theta \tilde{v}(\log\rho, y)\,dy \\ &= \int_0^\theta \frac{-\log\rho\, dy}{(\log\rho)^2 + y^2} = -\arctan\frac{\theta}{\log\rho}, \end{aligned} \tag{I.1.9}$$

then the graph of u is an embedded ∞-valued harmonic graph lying in a slab, i.e., $|u| \leq \pi/2$, and hence in particular is not proper. Note also that if u is given by (I.1.9), then $u_\theta = v$ and w/v is uniformly bounded above and below. We next want to rule out not only this as an example of one half of an embedded minimal disk, but more generally any ∞-valued minimal graph in a half-space.

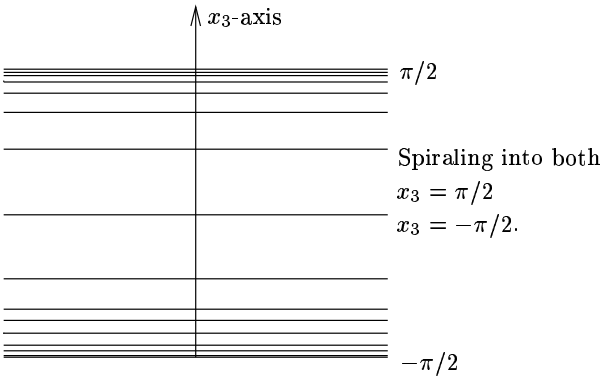


FIGURE 7. Properness; e.g., need to rule out that one of the multivalued graphs can contain a graph like $\arctan(\theta/\log\rho)$, where (ρ, θ) are polar coordinates.

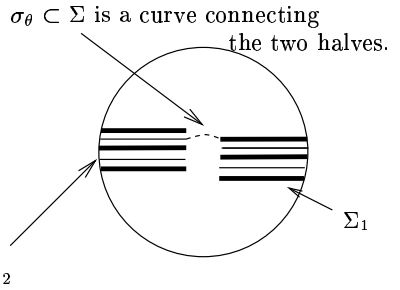


FIGURE 8. The short curves σ_θ in Theorem I.2.1 connecting the two multivalued graphs.

I.2. Towards Removability of Singularities

As mentioned above, we next want to rule out that any ∞-valued graph which lies in a half-space can be one half of an embedded minimal disk. This is done in Theorem I.2.1 below where the short curves σ_θ in the minimal disk will connect two multivalued sub-graphs and thus each sub-graph is essentially one half of the disk. These curves are needed to conclude properness since, as mentioned above, just having one ∞-valued minimal graph in a slab is not in itself a contradiction.

There is a second key assumption needed to prove this properness: The outer radii R_i must be going to infinity. We refer to this as the global case. The alternative, the local case, is when the R_i are bounded. The difference between these two cases is that in the local case it is possible to have non-proper limits which spiral infinitely into a plane. We will return to this in Part IV.

Let us illustrate in an example how these curves σ could be chosen. Here, to be consistent with [**CM8**], we use the helicoid which spirals downward so $w < 0$. If Σ is the helicoid, i.e., $\Sigma = (s\cos t, s\sin t, -t)$ where $s, t \in \mathbf{R}$, then $\Sigma \setminus \{s = 0\}$ consists of two ∞-valued graphs Σ_1, Σ_2 and the curves $\sigma_t = \Sigma \cap \{x_3 = t\}$ are short curves connecting the two halves; see Figure 8.

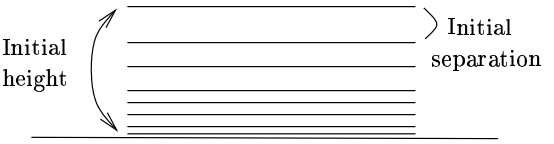


FIGURE 9. The flux argument in the proof of Theorem I.2.1: The initial height and separation.

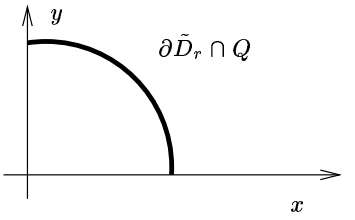


FIGURE 10. The flux argument in the proof of Theorem I.2.1: The curve where the average is calculated over.

THEOREM I.2.1. (See Corollary 1.2 in [**CM8**] for the precise statement.) Let Σ_i be a sequence as in Theorem I.0.5 and suppose that $\Sigma_{i,1}$, $\Sigma_{i,2}$ are multivalued graphs in Σ_i that spiral together (one inside the other). We claim that if $\Sigma_{i,1}$, $\Sigma_{i,2}$ can be connected by short curves in Σ_i (see Figure 8), then they cannot spiral into a plane. That is, they cannot accumulate in finite height.

IDEA OF THE PROOF OF THEOREM I.2.1. Let Σ_1, Σ_2 be two ∞-valued graphs of u_1, u_2 that spiral together, are part of an embedded minimal disk, and can be connected by short curves in the disk. We claim that the two graphs must grow out of any half-space. Suppose that they are contained in the half-space $u_i \geq 0$ and that they spiral downward, i.e., $w_i < 0$; we will get a contradiction.

Using that the curvature of a multivalued embedded minimal graph decays faster than quadratically will allow us (as in Theorem I.1.6) to model the graphs Σ_1, Σ_2 by graphs of harmonic functions. So suppose for a moment that both u_1 and u_2 are harmonic. We will show, using a flux argument, that if the separation is large compared with the initial height (see Figure 9), then the fact that the two graphs are part of an embedded minimal disk will eventually force the graphs to grow out of the half-space that they are assumed to lie in. Namely, set

$$\tilde{u}_i(x,y) = u_i(\mathrm{e}^x, y) \tag{I.2.2}$$

(i.e., make the conformal change $\rho = \mathrm{e}^x$ and $\theta = y$) and let Q be the quarter-space $\{(x,y) : x,y \geq 0\}$ and $\tilde{D}_r = \{(x,y) : x^2 + y^2 \leq r^2\}$. To see that each of the two graphs would grow out of the half-space, set (see Figure 10)

$$\mathrm{Av}_i(r) = \frac{1}{r} \int_{\partial \tilde{D}_r \cap Q} \tilde{u}_i. \tag{I.2.3}$$

The claim follows once we show that there are constants C_1 and C_2 depending on the ratio of the initial height with the initial separation so that for $r \geq 2$,

$$\mathrm{Av}_1(r) + \mathrm{Av}_2(r) - \mathrm{Av}_1(1) - \mathrm{Av}_2(1) \leq -C_1 \log^2 r + C_2 \log r. \tag{I.2.4}$$

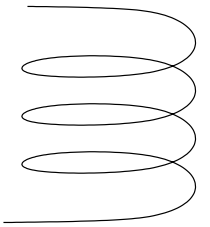


FIGURE 11. The flux argument in the proof of Theorem I.2.1: One of the two spiraling curves; $\theta \to (\theta, u_i(1,\theta))$.

Namely, note that if (I.2.4) holds, then the left-hand side of (I.2.4) goes to negative infinity as $r \to \infty$, contradicting that each $\tilde{u}_i > 0$. Notice that we can take $r \to \infty$ to get the contradiction precisely because the radii $R_i \to \infty$ in Theorem I.0.5. This is the key place where we use that $R_i \to \infty$ (see Part IV for an example which shows this is necessary).

To see (I.2.4), note that when u_i is harmonic, then so is $\tilde{u}_i$ and by Stokes' theorem

$$\text{Av}_i'(r) = \frac{1}{r}\int_{\partial \tilde{D}_r \cap Q} \frac{d\tilde{u}_i}{dn} = \frac{1}{r}\int_0^r \frac{\partial \tilde{u}_i}{\partial y}(s,0)\,ds + \frac{1}{r}\int_0^r \frac{\partial \tilde{u}_i}{\partial x}(0,s)\,ds. \tag{I.2.5}$$

Since the separation is roughly equal to $2\pi\,(u_i)_\theta$, by combining (I.2.5) with the lower bound for the separation from Theorem I.1.6 we get that (more or less)

$$\text{Av}_i'(r) \leq -\frac{C_1 \log r}{2\,r} + \frac{1}{r}\int_0^r \frac{\partial \tilde{u}_i}{\partial x}(0,s)\,ds. \tag{I.2.6}$$

Since the two spiraling curves (i.e., $\theta \to (\theta, u_i(1,\theta))$ for $i = 1,2$, see Figure 11) together with the two short curves that are assumed to exist bounds a disk (see Figure 12), we get by Stokes' theorem (see Lemma 5.1 of [**CM8**]) that

$$\begin{aligned} &\left| \int_0^r \frac{\partial \tilde{u}_1}{\partial x}(0,s)\,ds + \int_0^r \frac{\partial \tilde{u}_2}{\partial x}(0,s)\,ds \right| \\ &\qquad = \left| \int_0^r \frac{\partial u_1}{\partial \rho}(1,\theta)\,d\theta + \int_0^r \frac{\partial u_2}{\partial \rho}(1,\theta)\,d\theta \right| \leq C_2. \end{aligned} \tag{I.2.7}$$

Here, we bounded the flux along the short curves by the length of the curves since each $|\nabla u_i| \leq 1$ (as is the case for the restrictions of the coordinate functions). Adding the bounds on Av_1' and Av_2' in (I.2) and substituting (I.2.7) gives

$$\text{Av}_1'(r) + \text{Av}_2'(r) \leq -\frac{C_1 \log r}{r} + \frac{C_2}{r}. \tag{I.2.8}$$

Integrating (I.2.8) gives (I.2.4).

In the general case where u_i satisfies the minimal surface equation there are a number of difficulties that have to be dealt with; see Lemma 4.1 of [**CM8**]. □

Now this was a little analysis of multivalued solutions of the minimal surface equation and can all be found in [**CM8**]. It illustrates that once we show existence of multivalued minimal graphs in embedded minimal disks and know that such graphs extend as graphs with a sufficiently rapidly growing number of sheets, then we get a removable singularity theorem.

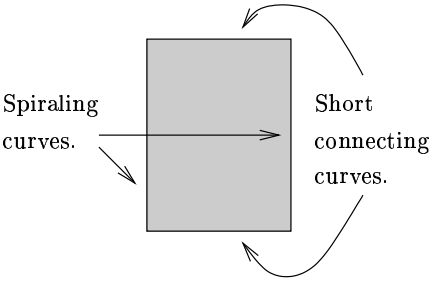


FIGURE 12. The flux argument in the proof of Theorem I.2.1: The two short curves together with the two spiraling curves bounds a disk in Σ.

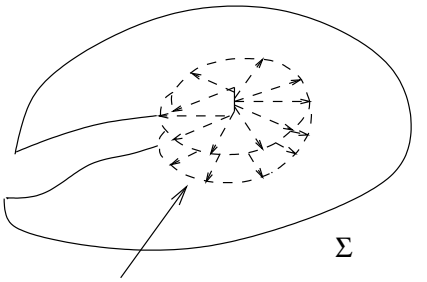


FIGURE 13. Part 1 of the proof of Theorem I.3.1; see Theorem II.0.4: finding a small multivalued graph in a disk near a point of large curvature.

I.3. Existence of Multivalued Graphs and the One-Sided Curvature Estimate

We now come to our key results for embedded minimal disks. These are some of the main ingredients in the proof of Theorem I.0.5. The first says that if the curvature of such a disk Σ is large at some point $x \in \Sigma$, then nearby x a multivalued graph forms (in Σ) and this extends (in Σ) almost all the way to the boundary. More precisely:

THEOREM I.3.1. [**CM4**, Theorem 0.2] See Figure 13 and Figure 14. Given $N \in \mathbf{Z}_+$, $\epsilon > 0$, there exist C_1, $C_2 > 0$. Let $0 \in \Sigma \subset B_R \subset \mathbf{R}^3$ be an embedded minimal disk, $\partial\Sigma \subset \partial B_R$. If $\max_{B_{r_0} \cap \Sigma} |A|^2 \geq 4\, C_1^2\, r_0^{-2}$ for some $R > r_0 > 0$, then there exists (after a rotation) an N-valued graph $\Sigma_g \subset \Sigma$ over $D_{R/C_2} \setminus D_{2r_0}$ with gradient $\leq \epsilon$ and $\Sigma_g \subset \{x_3^2 \leq \epsilon^2\,(x_1^2 + x_2^2)\}$.

As a consequence of Theorem I.3.1, one easily gets that if $|A|^2$ is blowing up near 0 for a sequence of embedded minimal disks Σ_i, then there is a sequence of 2-valued graphs $\Sigma_{i,d} \subset \Sigma_i$, where the 2-valued graphs start off on a smaller and smaller scale (namely, r_0 in Theorem I.3.1 can be taken to be smaller as the curvature gets larger). Consequently, by the sublinear separation growth, such 2-valued graphs collapse and, hence, a subsequence converges to a smooth minimal graph through 0. To be precise, given any fixed $\rho > 0$, (I.1.2) bounds the separation

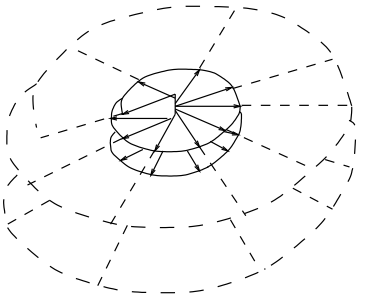


FIGURE 14. Part 2 of the proof of Theorem I.3.1; see Theorem II.0.6: extending a small multivalued graph in a disk.

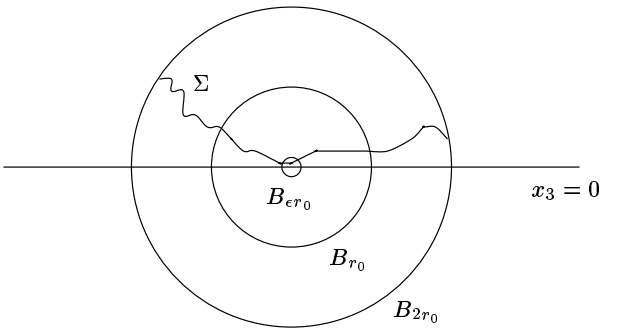


FIGURE 15. Theorem I.3.3: the one-sided curvature estimate for an embedded minimal disk Σ in a half-space with $\partial\Sigma \subset \partial B_{2r_0}$: The components of $B_{r_0} \cap \Sigma$ intersecting $B_{\epsilon r_0}$ are graphs.

w at $(\rho, 0)$ by

$$|w(\rho,0)| \leq \left(\frac{\rho}{r_0}\right)^{\alpha} |w(r_0,0)| \leq 2\pi\,\epsilon\,\rho^{\alpha} r_0^{1-\alpha}, \tag{I.3.2}$$

and this goes to 0 as $r_0 \to 0$ since $\alpha < 1$. The bound $|w(r_0,0)| \leq 2\pi\,\epsilon\,r_0$ in (I.3.2) came from integrating the gradient bound on the graph around the circle of radius r_0. (Here 0 is a removable singularity for the limit.) Moreover, if the sequence of such disks is as in Theorem I.0.5, i.e., if $R_i \to \infty$, then the minimal graph in the limit is entire and hence, by Bernstein's theorem [**CM1**, Theorem 1.16], is a plane.

The second key result is the curvature estimate for embedded minimal disks in a half-space. This theorem says roughly that if an embedded minimal disk lies in a half-space above a plane and comes close to the plane, then it is a graph over the plane. Precisely, this is the following theorem:

THEOREM I.3.3. [**CM6**, Theorem 0.2] (See Figure 15.) There exists $\epsilon > 0$, such that if $\Sigma \subset B_{2r_0} \cap \{x_3 > 0\} \subset \mathbf{R}^3$ is an embedded minimal disk with $\partial\Sigma \subset \partial B_{2r_0}$, then for all components Σ' of $B_{r_0} \cap \Sigma$ which intersect $B_{\epsilon r_0}$,

$$\sup_{\Sigma'} |A_\Sigma|^2 \leq r_0^{-2}. \tag{I.3.4}$$

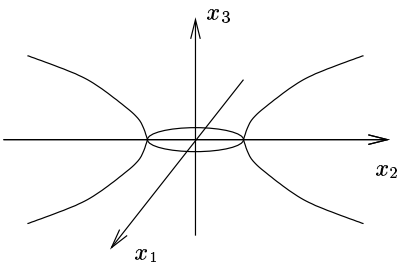


FIGURE 16. The catenoid given by revolving $x_1 = \cosh x_3$ around the x_3-axis.

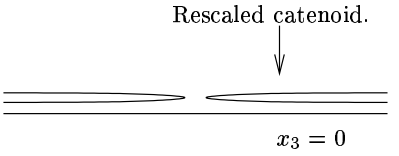


FIGURE 17. Rescaling the catenoid shows that simply connected (and embedded) is needed in the one-sided curvature estimate.

Using the minimal surface equation and that Σ' has points close to a plane, it is not hard to see that, for $\epsilon > 0$ sufficiently small, (I.3.4) is equivalent to the statement that Σ' is a graph over the plane $\{x_3 = 0\}$.

An embedded minimal surface Σ which is as in Theorem I.3.3 is said to satisfy the (ϵ, r_0)-*effective one-sided Reifenberg condition*; compare appendix A of [**CM6**] and the appendix of [**ChC**]. We will often refer to Theorem I.3.3 as *the one-sided curvature estimate.*

Note that the assumption in Theorem I.3.3 that Σ is simply connected is crucial as can be seen from the example of a rescaled catenoid. The catenoid is the minimal surface in $\mathbf{R}^3$ given by $(\cosh s \cos t, \cosh s \sin t, s)$ where $s, t \in \mathbf{R}$; see Figure 16. Under rescalings this converges (with multiplicity 2) to the flat plane; see Figure 17. Likewise, by considering the universal cover of the catenoid, one sees that embedded, and not just immersed, is needed in Theorem I.3.3.

As an almost immediate consequence of Theorem I.3.3 and a simple barrier argument we get that if in a ball two embedded minimal disks come close to each other near the center of the ball, each of the disks is a graph:

COROLLARY I.3.5. (Corollary 0.4 in [**CM6**]). See Figure 18. There exist $c > 1$, $\epsilon > 0$ as follows: Let Σ_1, $\Sigma_2 \subset B_{cr_0} \subset \mathbf{R}^3$ be disjoint embedded minimal surfaces with $\partial\Sigma_i \subset \partial B_{cr_0}$ and $B_{\epsilon r_0} \cap \Sigma_i \neq \varnothing$. If Σ_1 is a disk, then for all components Σ_1' of $B_{r_0} \cap \Sigma_1$ which intersect $B_{\epsilon r_0}$,

$$\sup_{\Sigma_1'} |A|^2 \leq r_0^{-2}. \tag{I.3.6}$$

Theorem I.3.3 is used to show that the points of large curvature in an embedded minimal disk all lie on a Lipschitz curve. To be able to discuss this and explain why this follows from the theorem let us introduce some notation for cones.

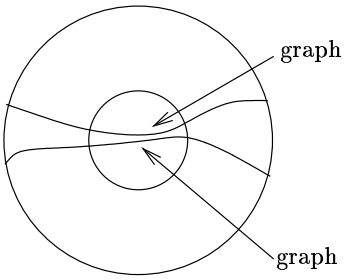


FIGURE 18. Corollary I.3.5: Two sufficiently close components of an embedded minimal disk must each be a graph.

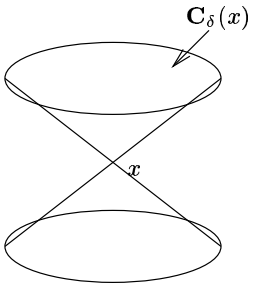


FIGURE 19. The cone $\mathbf{C}_\delta(x)$.

If $\delta > 0$ and $z \in \mathbf{R}^3$, then we denote by $\mathbf{C}_\delta(z)$ the (convex) cone with vertex z, cone angle $(\pi/2 - \arctan\delta)$, and axis parallel to the x_3-axis. That is (see Figure 21):

$$\mathbf{C}_\delta(z) = \{x \in \mathbf{R}^3 \,|\, x_3^2 \geq \delta^2 (x_1^2 + x_2^2)\} + z. \tag{I.3.7}$$

The next direct consequence of Theorem I.3.3 (with Σ_d playing the role of the plane $x_3 = 0$) will be needed when we sketch the proof of Theorem I.0.5 in the next section. This consequence says that the points of large curvature in an embedded minimal disk have what we will in the next section call the cone property. Namely, given a point of large curvature in such a disk, the next corollary asserts that all the other points of large curvature lie within a double convex cone with vertex at the initial point of large curvature. This is the result that will eventually give the regularity of the singular set (the set of points of large curvature). Precisely, this consequence of Theorem I.3.3 is the following:

COROLLARY I.3.8. (Corollary I.1.9 in [**CM6**]). See Figure 20. There exists $\delta_0 > 0$ as follows: Suppose $\Sigma \subset B_{2R}$, $\partial\Sigma \subset \partial B_{2R}$ is an embedded minimal disk containing a 2-valued graph $\Sigma_d \subset \{x_3^2 \leq \delta_0^2 (x_1^2 + x_2^2)\}$ over $D_R \setminus D_{r_0}$ with gradient $\leq \delta_0$. Then each component of $B_{R/2} \cap \Sigma \setminus (\mathbf{C}_{\delta_0}(0) \cup B_{2r_0})$ is a multivalued graph with gradient ≤ 1.

Figure 20 illustrates how this corollary follows from Theorem I.3.3. In this picture, $B_s(y)$ is a ball away from 0 and Σ' is a component of $B_s(y) \cap \Sigma$ disjoint from Σ_d. It follows easily from the maximum principle that Σ' is topologically a disk. Since Σ' is assumed to contain points near Σ_d, then we can let a component

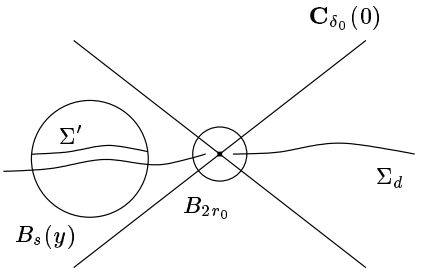


FIGURE 20. Corollary I.3.8: With Σ_d playing the role of $x_3 = 0$, by the one-sided estimate, Σ consists of multivalued graphs away from a cone.

of $B_s(y) \cap \Sigma_d$ play the role of the plane $\{x_3 = 0\}$ in Theorem I.3.3 and the corollary follows.

Note that, since Σ is compact and embedded, the multivalued graphs given by Corollary I.3.8 spiral through the cone. Namely, if a graph did close up, then the graph containing Σ_d would be forced to accumulate into it, contradicting compactness.

It can also be seen from Corollary I.3.8 (see Corollary 6.3 of [**CM8**]) that if Σ_d and Σ are as in Corollary I.3.8, then Σ_d extends in $\Sigma \setminus \mathbf{C}_{\delta_0}(0)$ to a multivalued graph with at least $(\log \rho)^2$ many sheets. Thus Theorem I.1.6 applies.

I.4. Regularity of the Singular Set and Theorem I.0.5

In this section we will indicate how to define the singular set $\mathcal{S}$ in Theorem I.0.5 and show the regularity of $\mathcal{S}$.

First, by a very general compactness argument, we have that for any sequence of surfaces in $\mathbf{R}^3$ (minimal or not), after possibly going to a subsequence, then there is a well-defined notion of points where the second fundamental form of the sequence blows up.

LEMMA I.4.1. Let $\Sigma_i \subset B_{R_i}$, $\partial \Sigma_i \subset \partial B_{R_i}$, and $R_i \to \infty$ be a sequence of (smooth) compact surfaces. After passing to a subsequence, Σ_j, we may assume that for each $x \in \mathbf{R}^3$ either (a) or (b) holds:

(a) $\sup_{B_r(x) \cap \Sigma_j} |A|^2 \to \infty$ for all $r > 0$.
(b) $\sup_j \sup_{B_r(x) \cap \Sigma_j} |A|^2 < \infty$ for some $r > 0$.

PROOF. For $r > 0$ and an integer n, define a sequence of functions on $\mathbf{R}^3$ by

$$\mathcal{A}_{i,r,n}(x) = \min\{n, \sup_{B_r(x) \cap \Sigma_i} |A|^2\}, \tag{I.4.2}$$

where we set $\sup_{B_r(x) \cap \Sigma_i} |A|^2 = 0$ if $B_r(x) \cap \Sigma_i = \varnothing$. Set

$$\mathcal{D}_{i,r,n} = \lim_{k \to \infty} 2^{-k} \sum_{m=0}^{2^k - 1} \mathcal{A}_{i,(1+m2^{-k})r,n}, \tag{I.4.3}$$

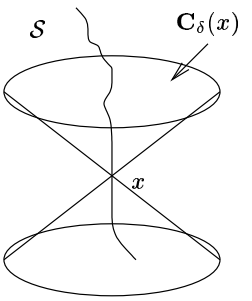


FIGURE 21. It follows from the one-sided curvature estimate that the singular set has the cone property and hence is a Lipschitz curve.

then $\mathcal{D}_{i,r,n}$ is continuous and $\mathcal{A}_{i,2r,n} \geq \mathcal{D}_{i,r,n} \geq \mathcal{A}_{i,r,n}$. Let $\nu_{i,r,n}$ be the (bounded) functionals,

$$\nu_{i,r,n}(\phi) = \int_{B_n} \phi \, \mathcal{D}_{i,r,n} \text{ for } \phi \in L^2(\mathbf{R}^3). \tag{I.4.4}$$

By standard compactness for fixed r, n, after passing to a subsequence, $\nu_{j,r,n} \to \nu_{r,n}$ weakly. In fact (since the unit ball in $L^2(\mathbf{R}^3)$ has a countable basis), by an easy diagonal argument, after passing to a subsequence, we may assume that for all $n, m \geq 1$ fixed $\nu_{j,2^{-m},n} \to \nu_{2^{-m},n}$ weakly. Note that if $x \in \mathbf{R}^3$ and for all m, n with $n \geq |x| + 1$, (identify $B_{2^{-m}}(x)$ with its characteristic function)

$$\nu_{2^{-m},n}(B_{2^{-m}}(x)) \geq n \operatorname{Vol}(B_{2^{-m}}), \tag{I.4.5}$$

then for each fixed $r > 0$, $\sup_{B_r(x) \cap \Sigma_j} |A|^2 \to \infty$. Conversely, if for some $n \geq |x| + 1$, m, (I.4.5) fails at x, then $\sup_j \sup_{B_r(x) \cap \Sigma_j} |A|^2 < \infty$ for $r = 2^{-m-1}$. □

From this lemma we conclude that, for a given sequence of embedded minimal disks (and possibly after passing to a subsequence), there is a well-defined notion of the set of points where the curvatures blow up. To show that this set is in fact a Lipschitz curve we will see below that, as a consequence of the one-sided curvature estimate, the set of such points has what we will call the cone property.

Fix $\delta > 0$. We will say that a subset $\mathcal{S} \subset \mathbf{R}^3$ has the cone property (or the δ-cone property) if $\mathcal{S}$ is closed and nonempty and satisfies:

(1) If $z \in \mathcal{S}$, then $\mathcal{S} \subset \mathbf{C}_\delta(z)$.
(2) If $t \in x_3(\mathcal{S})$ and $\epsilon > 0$, then $\mathcal{S} \cap \{t < x_3 < t + \epsilon\} \neq \varnothing$ and $\mathcal{S} \cap \{t - \epsilon < x_3 < t\} \neq \varnothing$.

Note that (2) just says that each point in $\mathcal{S}$ is the limit of points coming from above and below.

When $\Sigma_i \subset B_{R_i} \subset \mathbf{R}^3$ is a sequence of embedded minimal disks with $\partial\Sigma \subset \partial B_{R_i}$, $R_i \to \infty$ and Σ_j is the subsequence given by Lemma I.4.1 and $\mathcal{S}$ is the set of points where the curvatures of Σ_j blow up (i.e., where (a) in Lemma I.4.1 holds), then (as we indicated above) we will see below that $\mathcal{S}$ has the cone property (after a rotation of $\mathbf{R}^3$). Hence (by the next lemma), $\mathcal{S}$ is in this case a Lipschitz curve which is a graph over the x_3-axis. Note that in the case where Σ_i is a sequence of rescaled helicoids, then $\mathcal{S}$ is simply the x_3-axis.

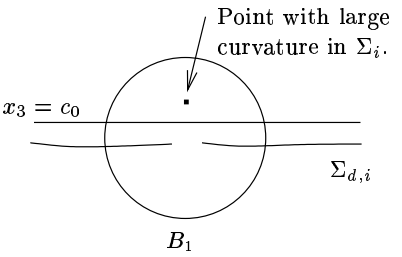


FIGURE 22. Lemma I.4.9: points with large curvature in Σ_i above the plane $x_3 = c_0$ but near the center of the 2-valued graphs $\Sigma_{d,i}$.

LEMMA I.4.6. See Figure 21. If $\mathcal{S} \subset \mathbf{R}^3$ has the δ-cone property, then $\mathcal{S} \cap \{x_3 = t\}$ consists of exactly one point $\mathcal{S}_t$ for all $t \in \mathbf{R}$, and $t \to \mathcal{S}_t$ is a Lipschitz parameterization of $\mathcal{S}$ with

$$|t_2 - t_1| \leq |\mathcal{S}_{t_2} - \mathcal{S}_{t_1}| \leq \sqrt{1 + \delta^{-2}}\,|t_2 - t_1|. \tag{I.4.7}$$

PROOF. Since $\mathcal{S}$ is nonempty, we may after translation assume that $0 \in \mathcal{S}$. By (1), it follows that $\mathcal{S} \cap \{x_3 = t\}$ consists of at most one point for each $t \in \mathbf{R}$. Assume that $\mathcal{S} \cap \{x_3 = t_0\} = \varnothing$ for some t_0. Since $\mathcal{S} \subset \mathbf{R}^3$ is a nonempty closed set and $x_3 : \mathcal{S} \subset \mathbf{C}_\delta(0) \to \mathbf{R}$ is proper, then $x_3(\mathcal{S}) \subset \mathbf{R}$ is also closed (and nonempty). Let $t_s \in x_3(\mathcal{S})$ be the closest point in $x_3(\mathcal{S})$ to t_0. The desired contradiction now easily follows since either $\mathcal{S} \cap \{t_s < x_3 < t_0\}$ or $\mathcal{S} \cap \{t_0 < x_3 < t_s\}$ is nonempty by assumption.

It follows that $t \to \mathcal{S}_t$ is a well-defined curve (from $\mathbf{R}$ to $\mathcal{S}$). Moreover, since

$$\mathcal{S}_{t_2} \subset \{x_3 = t_1 + (t_2 - t_1)\} \cap \mathbf{C}_\delta(\mathcal{S}_{t_1}) \subset B_{\sqrt{1+\delta^{-2}}|t_2 - t_1|}(\mathcal{S}_{t_1}), \tag{I.4.8}$$

(I.4.7) follows. □

Suppose next that Σ_i is as in Theorem I.0.5, that is, $\Sigma_i \subset B_{R_i} = B_{R_i}(0) \subset \mathbf{R}^3$ is a sequence of embedded minimal disks with $\partial\Sigma_i \subset \partial B_{R_i}$ where $R_i \to \infty$ and $\sup_{B_1 \cap \Sigma_i} |A|^2 \to \infty$. Let Σ_j and $\mathcal{S}$ be the subsequence and set, respectively, given by Lemma I.4.1 ($\mathcal{S}$ is the set of points where (a) holds in Lemma I.4.1). In particular, $\mathcal{S}$ is closed by definition and nonempty by the assumption of Theorem I.0.5. From Corollary I.3.8 (cf. the sketch of the proof of Theorem I.0.5 below), it follows that (1) above holds, so to see that $\mathcal{S}$ has the cone property all we need to see is that (2) holds. This follows from the next lemma which relies in part on Theorem I.2.1:

LEMMA I.4.9. [**CM6**, Lemma I.1.10] (See Figure 22.) There exists $c_0 > 0$ as follows: Let $\Sigma_i \subset B_{R_i}$, $\partial\Sigma_i \subset \partial B_{R_i}$ be a sequence of embedded minimal disks with $R_i \to \infty$. If $\Sigma_{d,i} \subset \Sigma_i$ is a sequence of 2-valued graphs over $D_{R_i/C} \setminus D_{\epsilon_i}$ with $\epsilon_i \to 0$ and $\Sigma_{d,i} \to \{x_3 = 0\} \setminus \{0\}$, then

$$\sup_{B_1 \cap \Sigma_i \cap \{x_3 > c_0\}} |A|^2 \to \infty. \tag{I.4.10}$$

SKETCH OF THE PROOF OF THEOREM I.0.5. From all of these results above, we know that if Σ_i is a sequence as in Theorem I.0.5 and Σ_j, $\mathcal{S}$ are given by Lemma I.4.1, then $\mathcal{S}$ has the cone property and hence, by Lemma I.4.1, $\mathcal{S}$ is a Lipschitz graph over the x_3-axis. As mentioned above, it also follows from Theorem I.3.1

together with Bernstein's theorem that, for each $x \in \mathcal{S}$ and each j sufficiently large, there is a 2-valued graph in Σ_j and that this sequence of 2-valued graphs converges (after possibly going to a subsequence) to a plane through x. Since Σ_j is embedded, all of these planes coming from different points $x \in \mathcal{S}$ must be parallel and it now follows from the one-sided curvature estimate that Σ_j consists of multivalued graphs away from $\mathcal{S}$. It is easy to see that there must be at least two such multivalued graphs in the complement of $\mathcal{S}$. That there are not more than two follows from a barrier argument; see [**CM6**, Proposition II.1.3]. This completes the sketch of the proof of Theorem I.0.5. □

Part II. The Proof of the Existence of Multivalued Graphs

Before we proceed let us briefly review the strategy of the proof of our main theorem that every embedded minimal disk is either a graph of a function or part of a double spiral staircase.

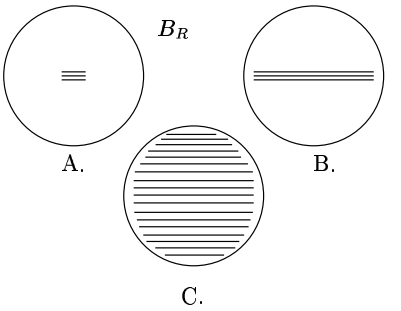


FIGURE 23. Proving Theorem I.0.5. A. Finding a small N-valued graph in Σ. B. Extending it in Σ to a large N-valued graph. C. Extending the number of sheets.

The proof has the following three main steps; see Figure 23:

A. Fix an integer N (the "large" of the curvature in what follows will depend on N). If an embedded minimal disk Σ is not a graph (or equivalently if the curvature is large at some point), then it contains an N-valued minimal graph which initially is shown to exist on the scale of $1/\max|A|$. That is, the N-valued graph is initially shown to be defined on an annulus with both inner and outer radius inversely proportional to $\max|A|$.

B. Such a potentially small N-valued graph sitting inside Σ can then be seen to extend as an N-valued graph inside Σ almost all the way to the boundary. That is, the small N-valued graph can be extended to an N-valued graph defined on an annulus where the outer radius of the annulus is proportional to R. Here R is the radius of the ball in $\mathbf{R}^3$ that the boundary of Σ is contained in.

C. The N-valued graph not only extends horizontally (i.e., tangent to the initial sheets) but also vertically (i.e., transversally to the sheets). That is, once there are N sheets there are many more and, in fact, the disk Σ consists of two multivalued graphs glued together along an axis.

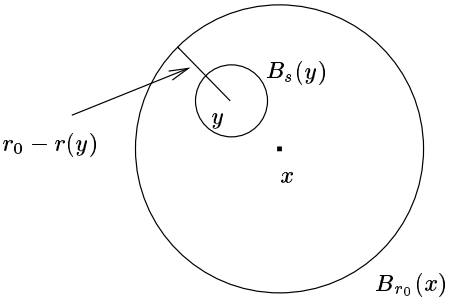


FIGURE 24. Existence of blowup points, that is pairs of points $y \in \Sigma$ and $s > 0$ satisfying (II.0.1).

To describe the existence, i.e., A., of multivalued graphs in embedded minimal disks, we will need the notion of a blowup point. Roughly speaking, a blowup point is a point where the curvature is, up to a fixed constant, the maximum.

To make this precise, let $x \in \Sigma \subset B_{r_0}(x) \subset \mathbf{R}^3$ be a smooth (compact) surface embedded, or just immersed, with $\partial\Sigma \subset \partial B_{r_0}(x)$. Here $B_{r_0}(x)$ is the extrinsic ball of radius r_0 but could as well have been an intrinsic ball in which case the notion of a blowup point below would have to be appropriately changed. Suppose that $|A|^2(x) \geq 4\,C^2\,r_0^{-2}$ for some constant $C > 0$. We claim that there is a "blowup point" $y \in B_{r_0}(x) \cap \Sigma$ and $s > 0$ such that $B_s(y) \subset B_{r_0}(x)$ and

$$\sup_{B_s(y)\cap\Sigma} |A|^2 \leq 4\,C^2\,s^{-2} = 4\,|A|^2(y). \tag{II.0.1}$$

That is, the curvature at y is large (this just means that C should be thought of as a large constant) and is almost (up to the constant 4) the maximum on the ball $B_s(y)$.

PROOF OF CLAIM. That there exists such a point y is easy to see; on $B_{r_0}(x)\cap\Sigma$ set $F(z) = (r_0 - r(z))^2\,|A|^2(z)$ where $r(z) = |z - x|$. Then

$$F(x) \geq 4\,C^2, F \geq 0, \text{ and } F\,|_{\partial B_{r_0}(x)\cap\Sigma} = 0. \tag{II.0.2}$$

Let y be where the maximum of F is achieved and set $s = C/|A|(y)$. One easily checks that y, s have the required properties; see Figure 24. Namely, clearly $|A|^2(y) = C^2\,s^{-2}$ and since y is where the maximum of F is achieved,

$$|A|^2(z) \leq \left(\frac{r_0 - r(y)}{r_0 - r(z)}\right)^2 |A|^2(y). \tag{II.0.3}$$

Since $F(x) \geq 4\,C^2$ it follows from the choice of s that $|r_0 - r(y)| \leq 2\,|r_0 - r(z)|$ for $z \in B_s(y) \cap \Sigma$. Hence, $|A|^2(z) \leq 4\,|A|^2(y)$. Together this gives (II.0.1). □

Returning to the existence of multivalued graphs, then the way we showed Theorem I.3.1 was by combining a blowup result with the extension of multivalued graphs proven in [**CM3**]. This blowup result says that if an embedded minimal disk in a ball has large curvature at a point, then it contains a small (in fact on the scale of 1 over the square root of the curvature) almost flat multivalued graph nearby, that is:

THEOREM II.0.4. [**CM4**, Theorem 0.4] (See Figure 13.) Given $N, \omega > 1$, and $\epsilon > 0$, there exists $C = C(N, \omega, \epsilon) > 0$ as follows: Let $0 \in \Sigma \subset B_R \subset \mathbf{R}^3$ be an embedded minimal disk, $\partial\Sigma \subset \partial B_R$. If $\sup_{B_{r_0} \cap \Sigma} |A|^2 \leq 4\,C^2\,r_0^{-2} = 4\,|A|^2(0)$ for some $0 < r_0 < R$, then there exist $\bar{R} < r_0/\omega$ and (after a rotation) an N-valued graph $\Sigma_g \subset \Sigma$ over $D_{\omega\bar{R}} \setminus D_{\bar{R}}$ with gradient $\leq \epsilon$, and $\mathrm{dist}_\Sigma(0, \Sigma_g) \leq 4\,\bar{R}$.

Note that 0 in Theorem II.0.4 is a blowup point in the above sense.

The result that we needed from [**CM3**] (combining Theorem 0.3 and Lemma II.3.8 there) is Theorem II.0.6 below that allows us to extend the (small) multivalued graphs given by Theorem II.0.4 almost out to the boundary of the "big" ball B_R. In this theorem, by the middle sheet $(\Sigma_g)^M$ of an N-valued graph Σ_g we mean the portion over

$$\{(\rho, \theta) \in \mathcal{P} : r_1 < \rho < r_2 \text{ and } 0 \leq \theta \leq 2\,\pi\}. \tag{II.0.5}$$

THEOREM II.0.6. [**CM3**] (See Figure 14.) Given N_1 and $\tau > 0$, there exist $N, \Omega, \epsilon > 0$ as follows: If $\Omega\,r_0 < 1 < R/\Omega$, $\Sigma \subset B_R$ is an embedded minimal disk with $\partial\Sigma \subset \partial B_R$, and Σ contains an N-valued minimal graph Σ_g over $D_1 \setminus D_{r_0}$ with gradient $\leq \epsilon$ and $\Sigma_g \subset \{x_3^2 \leq \epsilon^2(x_1^2 + x_2^2)\}$, then Σ contains an N_1-valued graph Σ_d over $D_{R/\Omega} \setminus D_{r_0}$ with gradient $\leq \tau$ and $(\Sigma_g)^M \subset \Sigma_d$.

II.1. The Proof of Existence of Small Multivalued Graphs: Theorem II.0.4

We will here describe some of the ideas that go into the proof of the existence of the small multivalued graphs near a point of large curvature. That is, the proof of Theorem II.0.4. In this theorem, Σ is an embedded minimal disk and $0 \in \Sigma$ is a blowup point with scale r_0 satisfying

$$\sup_{B_{r_0} \cap \Sigma} |A|^2 \leq 4\,C^2\,r_0^{-2} = 4\,|A|^2(0). \tag{II.1.1}$$

The theorem then gives a multivalued graph Σ_g contained in Σ defined on the scale r_0 ((II.1.1) says that r_0 is proportional to 1 over the square root of the curvature).

The key step in finding the multivalued graph is to find many large pieces of Σ with a (scale-invariant) quadratic curvature bound (these pieces will be intrinsic sectors; see Figure 27). To do this, we use the upper bound on $|A|^2$ in (II.1.1) to prove that the area of intrinsic balls in Σ grows polynomially and, consequently, get an average curvature bound. This average curvature bound and a curvature estimate for embedded disks (see Proposition II.1.5 below) will give large pieces of Σ with the desired quadratic curvature bound. Using the lower bound on $|A|^2(0)$ in (II.1.1), we show that there are many such pieces so that two must be close together in $\mathbf{R}^3$; embeddedness implies that these are disjoint, hence almost stable, and therefore nearly flat. Piecing together these large flat pieces then gives the desired N-valued graph.

Throughout this section, $\Sigma \subset B_R$ is an embedded minimal disk with $\partial\Sigma \subset \partial B_R$.

Before discussing the main steps in the proof, we will need to recall three facts about minimal surfaces from [**CM4**]. The first is the relationship between area and total curvature of intrinsic balls for disks with nonpositive curvature. The second is a curvature estimate for embedded minimal disks with bounded total curvature. The third is that nearby, but disjoint, minimal surfaces with bounded curvature must in fact be almost stable.

Area and total curvature. The relationship between area and total curvature is particularly simple for disks. Essentially, this is because the first variation of length of a geodesic circle is (up to a constant) given by the total curvature of the disk using the Gauss–Bonnet theorem; this can also be seen using the Jacobi equation for geodesics. Namely, as in [**CM4**, Corollary 1.7], integrating the Jacobi equation (for geodesics) and using $K_\Sigma = -|A|^2/2$ gives

$$4\,(\text{Area}\,(\mathcal{B}_R) - \pi\,R^2) = 2\int_0^R \int_0^t \int_{\mathcal{B}_s} |A|^2 = \int_{\mathcal{B}_R} |A|^2\,(R-r)^2, \tag{II.1.2}$$

where $r(x) = \text{dist}_\Sigma(0,x)$. The second equality in (II.1.2) used two integrations by parts (i.e., $\int_0^R f(t)\,g''(t)\,dt = \int_0^R f''(t)\,g(t)\,dt$ with $f(t) = \int_0^t \int_{\mathcal{B}_s} |A|^2$ and $g(t) = (R-t)^2$).

We will see that (II.1.2) often leads to bounds on the area of the ball $\mathcal{B}_R$. For example, when $\mathcal{B}_R$ is stable, then using $R - r$ (which vanishes on $\partial\mathcal{B}_R$) in the stability inequality (see (II.1.9) below) gives

$$4\,(\text{Area}\,(\mathcal{B}_R) - \pi\,R^2) = \int_{\mathcal{B}_R} |A|^2\,(R-r)^2 \leq \int_{\mathcal{B}_R} |\nabla(R-r)|^2 = \text{Area}\,(\mathcal{B}_R). \tag{II.1.3}$$

Consequently, we get an a priori bound for the area of an intrinsic ball in a stable minimal disk

$$\text{Area}\,(\mathcal{B}_R) \leq 4\pi\,R^2/3. \tag{II.1.4}$$

(This area bound is the starting point in [**CM2**].) We will use two generalizations of this argument below. The first will get a polynomial area bound for embedded minimal disks with bounded curvature. The second generalization will be to bound the area of a 1/2-stable sector (a sector is a specific type of subdomain of an intrinsic ball).

A curvature estimate for embedded disks with bounded total curvature. The following curvature estimate for embedded minimal disks Σ generalizes a result of Schoen and Simon [**CM1**, Theorem 2.5]:

PROPOSITION II.1.5. (Corollary 1.18 in [**CM4**]). Given C_I, there exists C_P so that if

$$\int_{\mathcal{B}_{2s}} |A|^2 \leq C_I, \tag{II.1.6}$$

then

$$\sup_{\mathcal{B}_s} |A|^2 \leq C_P\,s^{-2}. \tag{II.1.7}$$

By (II.1.2), bounds on $\text{Area}(\mathcal{B}_t)/t^2$, $\text{Length}(\partial\mathcal{B}_t)/t$, or $\int_{\mathcal{B}_t} |A|^2$ are equivalent (if we are willing to go to subballs). Therefore, Proposition II.1.5 gives an a priori curvature estimate when any one of these three quantities is bounded. It is important that Σ is an embedded disk; e.g., the catenoid is complete, has finite total curvature, and is not flat.

Nearby disjoint surfaces with $|A|^2 \leq 4$ are nearly stable. See Figure 25. We will see that if two disjoint minimal surfaces with bounded curvature come close enough together, then each of them is 1/2-stable. This 1/2-stability property is a weakening of stability which is still sufficiently strong to imply estimates on the area and curvature of the surfaces.

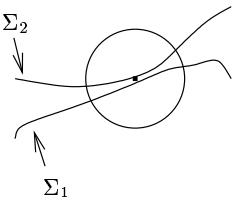


FIGURE 25. Two sufficiently close disjoint minimal surfaces with bounded curvatures must each be nearly stable.

Before making this precise, recall that the linearization of the minimal graph equation is the Jacobi equation (for minimal surfaces)

$$Lu = \Delta u + |A|^2 u = 0. \tag{II.1.8}$$

A domain $\Omega \subset \Sigma$ is said to be stable if $\int \phi L\phi \leq 0$ for every ϕ with compact support in Ω; i.e., if we have the stability inequality

$$\int |A|^2 \phi^2 \leq \int |\nabla \phi|^2. \tag{II.1.9}$$

We will say that Ω is $1/2$-stable if we have the weaker $1/2$-stability inequality

$$\int |A|^2 \phi^2 \leq 2 \int |\nabla \phi|^2. \tag{II.1.10}$$

One useful criterion for stability is that if $u > 0$ and $Lu = 0$ on Ω, then Ω is stable; similarly, when $u > 0$ and Lu/u is small, then Ω is $1/2$-stable (see Section 2 of [**CM4**]).

Using the above criteria for stability, we can now explain why nearby, but disjoint, minimal disks with bounded curvature must be $1/2$-stable. Namely, if two disjoint minimal disks with $|A|^2 \leq 4$ come close at a point, then it is not hard to see that one can be written as a (normal exponential) graph over the other of a function $u > 0$ with $Lu/u \approx 0$ and, consequently, each is $1/2$-stable. This is similar to the case of geodesics. See Lemmas 2.6 and 2.11 of [**CM4**] for the precise statements.

Having these tools at our disposal, we turn next to the main steps in the proof of the existence of the small multivalued graphs near a point of large curvature.

Polynomial area bounds when $|A|^2 \leq 4$. In general, the volume comparison theorem from geometry implies that a surface with bounded curvature has at most exponential area growth (as is the case for hyperbolic space). However, we will see that an embedded minimal disk in $\mathbf{R}^3$ with bounded curvature actually has polynomial area growth. This will be used to prove a doubling property, that is, to find arbitrarily large balls where the area of the double ball increases by at most a bounded factor (see Corollary II.1.21 below for the precise statement). Notice that this kind of doubling occurs for polynomial growth but not for exponential growth, i.e., $\lim_{r\to\infty} \frac{(2r)^p}{r^p} = 2^p < \infty$ while $\lim_{r\to\infty} \frac{e^{2r}}{e^r} = \infty$.

To get this polynomial area bound, we first show that "most of" an embedded minimal disk Σ with $|A|^2 \leq 4$ is nearly stable. More precisely, the next lemma decomposes Σ into a union of disjoint $1/2$-stable domains Ω_j and a remainder with bounded area. Moreover, the lemma also gives a cutoff function with bounded energy which will be used in the $1/2$-stability inequality.

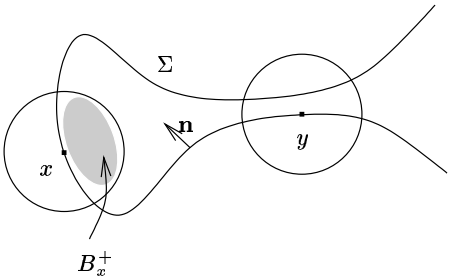


FIGURE 26. The set VB in (II.1.14): $x \in VB$ and $y \in \Sigma \setminus VB$.

LEMMA II.1.11. [**CM4**, Lemma 2.15] There exists C_1 as follows: If $0 \in \Sigma \subset B_{2R}$, $\partial\Sigma \subset \partial B_{2R}$, and $|A|^2 \leq 4$, then there exist disjoint 1/2-stable subdomains $\Omega_j \subset \Sigma$ and a function $\chi \leq 1$ which vanishes on $B_R \cap \Sigma \setminus \cup_j \Omega_j$ so that

$$\text{Area}(\{x \in B_R \cap \Sigma : \chi(x) < 1\}) \leq C_1 R^3, \tag{II.1.12}$$

$$\int_{\mathcal{B}_R} |\nabla \chi|^2 \leq C_1 R^3. \tag{II.1.13}$$

SKETCH OF PROOF. Fix $\rho > 0$ small. Given $x \in \Sigma$, let Σ_x be the component of $B_\rho(x) \cap \Sigma$ with $x \in \Sigma_x$ and let B_x^+ be the component of $B_\rho(x) \setminus \Sigma_x$ which $\mathbf{n}(x)$ points into. Set, see Figure 26,

$$VB = \{x \in B_R \cap \Sigma : B_x^+ \cap \Sigma \setminus \mathcal{B}_{4\rho}(x) = \varnothing\} \tag{II.1.14}$$

and let $\{\Omega_j\}$ be the components of $B_R \cap \Sigma \setminus \overline{VB}$. It follows from the previous subsection that each Ω_j comes equipped with a positive solution u_j of the minimal graph equation and is therefore 1/2-stable for ρ sufficiently small (it is obvious that the function u_j exists locally, but it requires a slight argument to see that it is globally well-defined).

The function χ is a linear cutoff function on the ρ-tubular neighborhood of VB, i.e., set

$$\chi(x) = \begin{cases} 0 & \text{if } x \in VB, \\ \text{dist}_\Sigma(x, VB)/\rho & \text{if } 0 \leq \text{dist}(x, VB) \leq \rho, \\ 1 & \text{otherwise .} \end{cases} \tag{II.1.15}$$

Finally, we use a simple ball counting argument to get (II.1.12) and (II.1.13). Roughly speaking, the definition of VB allows us to cover it by a collection of extrinsic balls $\{B_{2\rho}(x_i)\}$ so that the "half half-balls" $B_{x_i}^+$ are essentially disjoint. Since each of these disjoint "half half-balls" has volume $\approx \rho^3$ and is contained in a fixed ball in $\mathbf{R}^3$, there are at most $C\,(R/\rho)^3$ such balls. This and the curvature bound easily give (II.1.12) and (II.1.13). □

Using the decomposition from Lemma II.1.11 in (II.1.2) (as in (II.1.3)) gives polynomial area bounds for intrinsic balls with bounded curvature [**CM4**, Lemma 3.1]:

PROPOSITION II.1.16. If $0 \in \Sigma \subset B_{2\,R}$, $\partial\Sigma \subset \partial B_{2\,R}$, and $|A|^2 \leq 4$, then

$$\int_0^R \int_0^t \int_{\mathcal{B}_s} |A|^2 \, ds \, dt = 2(\text{Area}(\mathcal{B}_R) - \pi R^2) \leq 6\pi R^2 + 20\, C_1 R^5. \tag{II.1.17}$$

PROOF. Let the constant C_1, the function χ, and the subdomain $\cup_j \Omega_j$ be given by Lemma II.1.11. In particular, the function $\chi(R-r)$ vanishes off of $\cup_j \Omega_j$. Using $\chi(R-r)$ in the 1/2-stability inequality (i.e., in (II.1.10)), the absorbing inequality and (II.1.13) give

$$\int |A|^2 \chi^2 (R-r)^2 \leq 2 \int \left(\chi^2 + 2\chi\, R\, |\nabla \chi| + R^2\, |\nabla \chi|^2\right) \leq 6\, R^2 \int_{\mathcal{B}_R} |\nabla \chi|^2 + 3 \int \chi^2 \leq 6\, C_1\, R^5 + 3\, \text{Area}(\mathcal{B}_R). \tag{II.1.18}$$

On the other hand, combining the area bound (II.1.12) for $\{\chi < 1\}$ and $|A|^2 \leq 4$ gives

$$\int |A|^2 (1-\chi^2)(R-r)^2 \leq 4\, R^2\, \text{Area}(\{x \in B_R \cap \Sigma : \chi(x) < 1\}) \leq 4\, C_1\, R^5. \tag{II.1.19}$$

We see (II.1.17) by using (II.1.18) and (II.1.19) in (II.1.2) to get

$$4\,(\text{Area}\,(\mathcal{B}_R) - \pi\, R^2) = \int |A|^2 (R-r)^2 \leq 10\, C_1\, R^5 + 3\, \text{Area}\,(\mathcal{B}_R). \tag{II.1.20}$$

□

Using the polynomial area growth proven in Proposition II.1.16, it is now standard to find large intrinsic balls with a fixed doubling for area (and hence also for total curvature by (II.1.2)):

COROLLARY II.1.21. (Corollary 3.5 in **[CM4]**). There exists C_2 such that given $\beta, R_0 > 1$, there exists R as follows: If $0 \in \Sigma \subset B_R$ is an embedded minimal disk, $\partial \Sigma \subset \partial B_R$, $|A|^2(0) = 1$, and $|A|^2 \leq 4$, then there exists $R_0 \leq s < R/(2\,\beta)$ with

$$\int_{\mathcal{B}_{3\,s}} |A|^2 \leq C_2\, s^{-2}\, \text{Area}\,(\mathcal{B}_s)\,, \tag{II.1.22}$$

$$\beta^{-10} \int_{\mathcal{B}_{2\,\beta\,s}} |A|^2 \leq C_2\, s^{-2}\, \text{Area}\,(\mathcal{B}_s). \tag{II.1.23}$$

SKETCH OF PROOF. Since the argument is similar, we sketch only the proof of (II.1.22). By (II.1.2), it is easy to see that

$$\int_{\mathcal{B}_{3\,s}} |A|^2 \leq 8\, s^{-2}\, \text{Area}(\mathcal{B}_{4\,s}). \tag{II.1.24}$$

Therefore, to prove (II.1.22), it suffices to find $s \geq R_0$ with

$$\text{Area}(\mathcal{B}_{4\,s}) \leq C_2/8\, \text{Area}(\mathcal{B}_s). \tag{II.1.25}$$

To do this, we use the bounds for the area given by Proposition II.1.16 to get

$$\left(\min_{1 \leq n \leq m-1} \frac{\text{Area}(\mathcal{B}_{4^n\, R_0})}{\text{Area}(\mathcal{B}_{4^{n-1}\, R_0})}\right)^m \leq \frac{\text{Area}(\mathcal{B}_{4^m\, R_0})}{\text{Area}(\mathcal{B}_{R_0})} \leq C\, 4^{5m}\, R_0^3. \tag{II.1.26}$$

Choosing m large so that $C\, R_0^3 \leq 2^m$ and taking the m-th root of (II.1.26) gives

$$\min_{1 \leq n \leq m-1} \frac{\text{Area}(\mathcal{B}_{4^n\, R_0})}{\text{Area}(\mathcal{B}_{4^{n-1}\, R_0})} \leq 2\,(4^5). \tag{II.1.27}$$

Let j be where the minimum in (II.1.27) is achieved, so that we get (II.1.25) with $C_2 = 16\,(4^5)$ and $s = 4^{j-1}\, R_0$. □

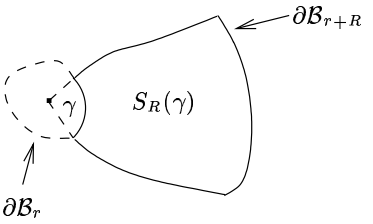


FIGURE 27. An intrinsic sector over a curve γ.

We will not go into why we need both (II.1.22) and (II.1.23) since this is somewhat technical. One consequence of this doubling property given by (II.1.22) is an average curvature bound for the intrinsic ball $\mathcal{B}_{3s}$. Together with the curvature estimate in Proposition II.1.5, this implies that at least "most of" $\mathcal{B}_{3s}$ has a pointwise scale-invariant curvature bound of the form $|A|^2 \leq C\, r^{-2}$. We will see later that much more is true; namely, most of $\mathcal{B}_{3s}$ will be almost stable (and therefore satisfy even better estimates). To make this precise, we next discuss a useful way to subdivide intrinsic balls into intrinsic sectors.

Area and total curvature of $1/2$*-stable sectors.* We next define intrinsic sectors and give estimates for the area and total curvature of a $1/2$-stable intrinsic sector. Given a curve $\gamma \subset \partial\mathcal{B}_s$, define the intrinsic (truncated) sector (see Figure 27) as

$$S = S_R(\gamma) = \{\exp_0(v) : s \leq |v| \leq s + R \text{ and } \exp_0(s\, v/|v|) \in \gamma\}. \tag{II.1.28}$$

(Here $\exp_0 : \mathbf{R}^2 \to \Sigma$ is the exponential map at 0.) The simplest example of a sector is when $\Sigma = \mathbf{R}^2$ and γ is given in polar coordinates as $\{\rho = s,\ \theta_1 \leq \theta \leq \theta_2\}$; in this case, the sector $S_R(\gamma)$ is just $\{s \leq \rho \leq s + R,\ \theta_1 \leq \theta \leq \theta_2\}$. We will often refer to γ as the inner boundary of the sector and the geodesic rays through $\partial\gamma$ as the sides of the sector.

Just as stability led to area bounds for intrinsic balls in (II.1.3), we can bound the area of a $1/2$-stable sector. However, to make the cutoff function compactly supported on a sector, we must also cutoff along the inner boundary and along the sides. This introduces new terms in the upper bound for the area. The next lemma, which is an easy consequence of Lemma II.1.1 and Remark II.1.32 in [**CM3**], gives the resulting estimate (k_g is the geodesic curvature of the curve γ).

LEMMA II.1.29. [**CM3**]. Suppose that S is $1/2$-stable, $\int_\gamma (1 + |k_g|\, s) \leq C_0\, m\, s$, $R > 2\, s$, and for $x \in S$ we have $\sup_{\mathcal{B}_{s/4}(x)} |A|^2 \leq C_0 s^{-2}$. Then for $\Omega > 2$ and $2\, s \leq t \leq 3R/4$,

$$t \int_\gamma k_g \leq \text{Length}(\partial\mathcal{B}_t \cap S) \leq C_3\, (m + R/s)\, t, \tag{II.1.30}$$

$$\int_{S_{R/\Omega} \setminus \mathcal{B}_{\Omega s}} |A|^2 \leq C_1\, R/s + C_2\, m/\log\Omega. \tag{II.1.31}$$

SKETCH OF PROOF. First, as in (II.1.3), we combine the $1/2$-stability inequality and the analog of (II.1.2) for intrinsic sectors to get (II.1.30). The R/s term in (II.1.30) comes from cutting off linearly on a (roughly) s-tubular neighborhood of the sides of the sector (which have length R). Once we have the quadratic area

growth given by (II.1.30), we can then use a radial logarithmic cutoff in the stability inequality to get (II.1.31). □

When we apply Lemma II.1.29, the bound $\sup_{\mathcal{B}_{s/4}(x)} |A|^2 \leq C_0 s^{-2}$ on S will be given by starting with a slightly larger 1/2-stable sector and then coming in from the boundary (using a comparison theorem to guarantee that the required intrinsic balls are contained in the larger sector).

Roughly speaking, Lemma II.1.29 shows that if we can find a large 1/2-stable sector, then we can find subsectors with small average curvature. We next explain how this leads to finding a subsector with small total curvature.

Small total curvature of stable subsectors. Given a 1/2-stable sector S as in Lemma II.1.29 over a long curve γ, we can subdivide γ into subcurves so that one of the sectors over these subcurves has small total curvature. To see this, suppose that $\text{Length}(\gamma) \approx m\, s$ for some m large. If $m/\log\Omega$ is larger than R/s, then (II.1.31) says that the truncated sector has small average curvature

$$\int_{S_{R/\Omega} \setminus \mathcal{B}_{\Omega s}} |A|^2 \leq C\, m/\log\Omega. \tag{II.1.32}$$

If we now subdivide γ into m subcurves, it follows from (II.1.32) that the truncated sector $\hat{S}$ over one of these subcurves has small total curvature

$$\int_{\hat{S}} |A|^2 \leq C/\log\Omega. \tag{II.1.33}$$

The N-valued graph. The next step is to use the 1/2-stable sectors with small total curvature to construct the N-valued graph Σ_g. To see this, suppose that S is as in Lemma II.1.29 and $\hat{S} \subset S$ is a subsector with small total curvature

$$\int_{\hat{S}} |A|^2 \leq \epsilon, \tag{II.1.34}$$

where $\epsilon > 0$ can be chosen small by taking Ω and m large. The small total curvature (and stability) implies that

$$|A|^2 \leq C\, \epsilon\, r^{-2} \text{ on } \hat{S}. \tag{II.1.35}$$

We will use this estimate (II.1.35) to build out Σ_g. For each point x on the inner boundary of $\hat{S}$, let $\gamma_x \subset \Sigma$ be the geodesic leaving x orthogonally to $\partial\hat{S}$. Integrating the curvature bound (II.1.35) along γ_x implies first that, as a curve in $\mathbf{R}^3$, γ_x has small total geodesic curvature and is close to a line segment. Second, it also implies that the unit normal $\mathbf{n}$ to Σ has small oscillation along γ_x. It is now easy to see that as we vary the initial point x on the inner boundary of $\hat{S}$, the geodesics γ_x trace out a multivalued graph Σ_g.

Finally, we review how these steps fit together to prove Theorem II.0.4.

Sketch of the proof of Theorem II.0.4: After rescaling by C/r_0, we can assume that

$$\sup_{B_C \cap \Sigma} |A|^2 \leq 4\, |A|^2(0) = 4. \tag{II.1.36}$$

The key for proving Theorem II.0.4 is to find many large intrinsic sectors with a quadratic curvature bound. To do this, we first fix some large $\bar{m}$ and use Proposition II.1.5 and the lower bound $|A|^2(0) = 1$ in (II.1.36) to get R_0 so that for $R \geq R_0$

$$\text{Length}(\partial\mathcal{B}_R) \geq \bar{m}\, R. \tag{II.1.37}$$

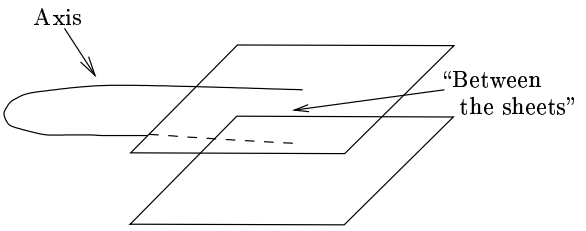


FIGURE 28. The estimate between the sheets; Theorem II.2.1.

Using the upper bound $|A|^2 \leq 4$ in (II.1.36), we can apply Corollary II.1.21 to get $R_3 > R_0$ and many long disjoint curves $\tilde{\gamma}_i \subset \partial \mathcal{B}_{R_3}$ so the sectors over $\tilde{\gamma}_i$ have uniformly bounded total curvature $\int |A|^2$. Proposition II.1.5 then gives a (pointwise) quadratic curvature bound on these sectors.

Now that we have these many disjoint sectors with a quadratic curvature bound (as many as we want by taking $\bar{m}$ large), two must be close together and, hence, also 1/2-stable. Lemma II.1.29 gives a subsector with small total curvature which then must contain the N-valued graph Σ_g (this is done in Corollary II.1.34 of [**CM3**]).

II.2. The Estimate Between the Sheets and the Extension of Multivalued Graphs

In this section, we will give an overview of the proof of Theorem II.0.6 where we extend the multivalued graphs almost to the boundary, corresponding to B in Figure 23. This result is significantly more subtle than the local existence result discussed in the previous section. We only give a very rough outline of the proof and refer to [**CM3**] for the full story.

The key component of the proof of the extension theorem is a curvature estimate "between the sheets" for embedded minimal disks in $\mathbf{R}^3$. We will think of an axis (see Figure 29) for such a disk Σ as a point or curve away from which the surface locally (in an extrinsic ball) has more than one component. With this weak notion of an axis, our estimate is that if one component of Σ is sandwiched between two others that connect to an axis, then the one that is sandwiched has curvature estimates; see Theorem II.2.1. The example to keep in mind is a helicoid and the components are "consecutive sheets" away from the axis.

Let $\gamma_{p,q}$ denote the line segment from p to q and $T_s(\gamma_{p,q})$ its s-tubular neighborhood. A curve γ is *h-almost monotone* if given $y \in \gamma$, then $B_{4\,h}(y) \cap \gamma$ has only one component which intersects $B_{2\,h}(y)$. Our curvature estimate "between the sheets" is:

THEOREM II.2.1. [**CM3**, Theorem I.0.8] (See Figure 29.) There exist $c_1 \geq 4$, $2c_2 < c_4 < c_3 \leq 1$ as follows: Let $\Sigma \subset B_{c_1 r_0}$ be an embedded minimal disk with $\partial\Sigma \subset \partial B_{c_1 r_0}$ and $y \in \partial B_{2 r_0}$. Suppose $\Sigma_1, \Sigma_2, \Sigma_3$ are distinct components of $B_{r_0}(y) \cap \Sigma$ and $\gamma \subset (B_{r_0} \cup T_{c_2 r_0}(\gamma_{0,y})) \cap \Sigma$ is a curve with $\partial\gamma = \{y_1, y_2\}$ where $y_i \in B_{c_2 r_0}(y) \cap \Sigma_i$ and each component of $\gamma \setminus B_{r_0}$ is $c_2 r_0$-almost monotone. Then any component Σ_3' of $B_{c_3 r_0}(y) \cap \Sigma_3$ with y_1, y_2 in distinct components of $B_{c_4 r_0}(y) \setminus \Sigma_3'$ is a graph.

The idea of the proof of Theorem II.2.1 is to show that if this were not the case, then we could find an embedded stable disk that would be almost flat and lie

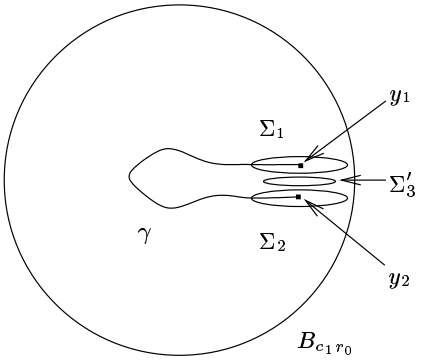


FIGURE 29. y_1, y_2, Σ_1, Σ_2, Σ_3', and γ in Theorem II.2.1.

in the complement of the original disk. In fact, we can choose the stable disk to be sandwiched between the two components as well. The flatness would force the stable disk to eventually cross the axis in the original disk, contradicting that they were disjoint.

The curve γ in Theorem II.2.1 which plays the role of an axis and connects Σ_1 and Σ_2 can in many instances be extended using the maximum principle once it occurs on a given small scale; compare lemma I.0.11 of [**CM3**]. This is used both in the proof of Theorem II.2.1 and in the applications of it.

OVERVIEW OF THE PROOF OF THEOREM II.0.6. The first step (see Part I of [**CM3**]) is to show Theorem II.2.1 when the surface is in a slab; i.e., when $\Sigma \subset \{|x_3| \leq \beta\, h\}$.

The second step (Part II of [**CM3**]) is to show that certain stable disks starting off as multivalued graphs remain multivalued graphs as the outer radius increases. This is needed when we generalize the results of step one to when the surface is not anymore in a slab.

Two facts go into the proof of the extension of stable disks. First, we show that if an almost flat multivalued graph sits inside a stable disk, then the outwardly defined intrinsic sector from a curve which is a multivalued graph over a circle has a subsector which is almost flat; cf. Lemma II.1.29 and the paragraph "Small total curvature of stable subsectors" above. As the initial multivalued graph becomes flatter and the number of sheets in it goes up, the subsector becomes flatter. The second fact is that the separation between the sheets grows sublinearly; see (I.1.2).

The first application of these two facts is to extend the middle sheet as a multivalued graph. This is done by dividing the initial multivalued graph (or curve in the graph that is itself a multivalued graph over the circle) into three parts where the middle sheet is the second part. The idea is then that the first and third parts have subsectors which are almost flat multivalued graphs and the middle part (which has curvature estimates since it is stable) is sandwiched between the two others. Hence its sector is also almost flat.

A thing that adds some technical complications to the above picture is that in the result about almost flat subsectors it is important that the ratio between the size of the initial multivalued graph and how far one can go out is fixed. This is because the estimate for the subsector comes from a total curvature estimate which is in

terms of this ratio (see (II.1.31)) and can only be made small by looking at a fixed large number of rotations for the graph. This forces us to successively extend the multivalued graph. The issue is then to make sure that as we move out in the sector and repeat the argument we have essentially not lost sheets. This is taken care of by using the sublinear growth of the separation between the sheets together with the Harnack inequality and the maximum principle. (The maximum principle is used to make sure that as we try to recover sheets after we have moved out, we don't hit the boundary of the disk before we have recovered essentially all of the sheets that we started with.) The last thing which is used is Theorem 3.36 in [**CM10**]. This is used to guarantee that, as we patch together these multivalued graphs coming from different scales, then the surface that we get is still a multivalued graph over a fixed plane.

The third step (Part III of [**CM3**]) is to generalize the curvature estimate between the sheets to the case where the surface is not anymore in a slab. This uses the extension of the stable graphs from step two.

Finally, using steps one, two, and three we showed Theorem II.0.6 in Part IV of [**CM3**]. □

Part III. The Proof of the One-Sided Curvature Estimate

In [**CM6**, Appendix A] we showed curvature estimates for minimal hypersurfaces in $\mathbf{R}^n$ which on all sufficiently small scales lie on one side of, but come close to, a hyper-plane. Such a scale invariant version of Theorem I.3.3 can (unlike Theorem I.3.3) be proven quite easily by a blowup argument using the minimal surface equation. Moreover, such a scale invariant condition is very similar to the classical Reifenberg property. (After all, a subset of $\mathbf{R}^n$ has the Reifenberg property if it is close on all scales to a hyper-plane.) As explained in Part I (in particular Corollary I.3.8), the significance of Theorem I.3.3 is indeed that it only requires closeness on one scale. On the other hand, this is what makes it difficult to prove and requires us to use the results discussed in the previous parts together with results from [**CM5**].

Let us briefly outline the proof of the one-sided; i.e., Theorem I.3.3. Suppose that Σ is an embedded minimal disk in the half-space $\{x_3 > 0\}$. We prove the curvature estimate by contradiction; so suppose that Σ has low points with large curvature. Starting at such a point, we decompose Σ into disjoint multivalued graphs using the existence of nearby points with large curvature given by corollary III.3.5 of [**CM5**], a blowup argument, and [**CM3**], [**CM4**]. The key point is then to show (see Proposition III.1 below) that we can in fact find such a decomposition where the "next" multivalued graph starts off a definite amount below where the previous multivalued graph started off. In fact, what we show is that this definite amount is a fixed fraction of the distance between where the two graphs started off. Iterating this eventually forces Σ to have points where $x_3 < 0$. This is the desired contradiction.

PROPOSITION III.1. (See proposition III.2.2 in [**CM6**] for the precise statement). See Figure 30. There exists $\delta > 0$ such that if $(0, s)$ satisfies (III.3) and $\Sigma_0 \subset \Sigma$ is the corresponding (to $(0, s)$) 2-valued graph over $D_R \setminus D_s$, then we get (y, t) satisfying (III.3) with $y \in \mathbf{C}_\delta(0) \cap \Sigma \setminus B_{Cs/2}$ and where y is below Σ_0.

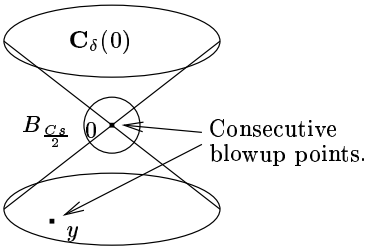


FIGURE 30. Proposition III.1: Two consecutive blowup points satisfying (III.3).

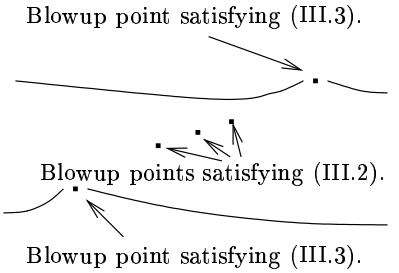


FIGURE 31. Between two consecutive blowup points satisfying (III.3) there are a bunch of blowup points satisfying (III.2).

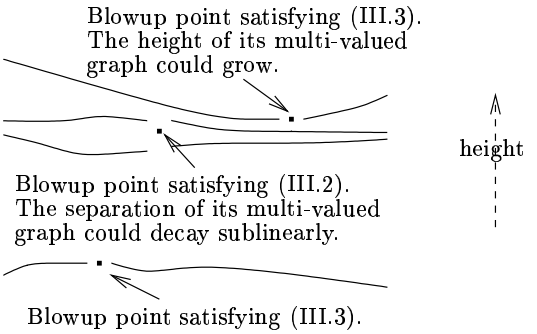


FIGURE 32. Measuring height. Blowup points and corresponding multivalued graphs.

To prove this key proposition (Proposition III.1), we use two decompositions and two kinds of blowup points. The first decomposition uses the more standard blowup points given as pairs (y, s) where $y \in \Sigma$ and $s > 0$ is such that

$$\sup_{\mathcal{B}_{8s}(y)} |A|^2 \leq 4|A|^2(y) = 4\,C_1^2\,s^{-2}. \tag{III.2}$$

The point about such a pair (y, s) is that by **[CM3]**, **[CM4]** (and an argument in section II.2 of **[CM6]** which allows us to replace extrinsic balls by intrinsic ones), then Σ contains a multivalued graph near y starting off on the scale s. (This is assuming that C_1 is a sufficiently large constant given by **[CM3]**, **[CM4]**.) The

second kind of blowup points are the ones where (except for a technical issue) 8 is replaced by some really large constant C, i.e.,

$$\sup_{\mathcal{B}_{Cs}(y)} |A|^2 \leq 4|A|^2(y) = 4\,C_1^2\,s^{-2}. \tag{III.3}$$

The point will then be that we can find blowup points satisfying (III.3) so that the distance between them is proportional to the sum of the scales. Moreover, between consecutive blowup points satisfying (III.3), we can find a bunch of blowup points satisfying (III.2); see Figure 31. The advantage is now that if we look between blowup points satisfying (III.3), then the height of the multivalued graph given by such a pair grows like a small power of the distance, whereas the separation between the sheets in a multivalued graph given by (III.2) decays like a small power of the distance; see Figure 32. Now thanks to the fact that the number of blowup points satisfying (III.2) (between two consecutive blowup points satisfying (III.3)) grows almost linearly, it is the case that even though the height of the graph coming from the blowup point satisfying (III.3) could move up (and thus work against us) the sum of the separations of the graphs coming from the points satisfying (III.2) dominates the other term. Thus the next blowup point satisfying (III.3) (which lies below all the other graphs) is forced to be a definite amount lower than the previous blowup point satisfying (III.3).

Part IV. The Local Case: When Singular Limit Laminations Can Occur

In this part we discuss the differences between the so-called local and global cases. The local case is where we have a sequence of embedded minimal disks in a ball of fixed radius in $\mathbf{R}^3$ — the global case is where the disks are in a sequence of expanding balls with radii tending to infinity. The main difference between these cases is that in the local case we can get limits with singularities. In the global case this does not happen because in fact any limit is a foliation by flat parallel planes (assuming that the curvatures of the sequence are blowing up).

To precisely define the local and global cases, suppose

$$\Sigma_i \subset B_{R_i} = B_{R_i}(0) \subset \mathbf{R}^3$$

with $\partial\Sigma_i \subset \partial B_{R_i}$ is a sequence of (compact) embedded minimal disks and either:

(a) R_i is equal to a finite constant.
(b) $R_i \to \infty$.

Case (a) is what we call the *local case* and (b) is what we refer to as the *global case*; Theorem I.0.5 dealt with the global case. Recall that a surface $\Sigma \subset \mathbf{R}^3$ is said to be properly embedded if it is embedded and the intersection of Σ with any compact subset of $\mathbf{R}^3$ is compact. We say that a lamination or foliation is proper if each leaf is proper.

To explain the difference between the two cases, we consider sequences of minimal disks Σ_i as above where the curvatures blow up, e.g.,

$$\lim_{i\to\infty} \sup_{B_1\cap\Sigma_i} |A|^2 = \infty\,. \tag{IV.1}$$

(Of course, if the curvatures do not blow up, then the Arzela–Ascoli theorem easily gives smooth convergence of a subsequence in either case.) In the global case, Theorem I.0.5 gives a subsequence of the Σ_i converging off of a Lipschitz curve to a

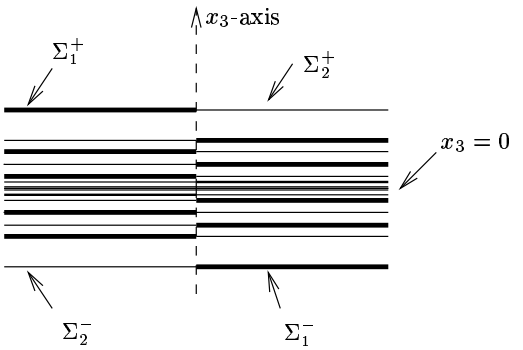


FIGURE 33. A schematic picture of the limit in Theorem IV.2 which is not smooth and not proper (the dotted x_3-axis is part of the limit). The limit contains four multivalued graphs joined at the x_3-axis; Σ_1^+, Σ_2^+ above the plane $x_3 = 0$ and Σ_1^-, Σ_2^- below the plane. Each of the four spirals into the plane.

foliation by parallel planes. In particular, the limit is a (smooth) foliation which is proper. However, we showed in [**CM16**] (see Theorem IV.2 below) that smoothness and properness of the limit can fail in the local case; compare Figure 33.

To illustrate the key issue for the failure of properness, suppose that $|A|^2(0) \to \infty$ as $i \to \infty$. In either the local or global case (see Section I.3), we get a sequence of 2-valued graphs which converges to a minimal graph Σ_0 through 0 (this graph is a plane in the global case). Furthermore, by the one-sided curvature estimate (see Corollary I.3.8), the intersection of Σ_i with a low cone about Σ_0 consists of multivalued graphs for i large. There are now two possibilities:

- The multivalued graphs in this low cone close up in the limit.
- The limits of these multivalued graphs spiral infinitely into Σ_0.

In the first case, where properness holds, the sequence converges to a foliation in a neighborhood of 0. In the second case, where properness fails, the sequence converges to a lamination away from 0 but cannot be extended smoothly to any neighborhood of 0. The proof of properness in the global case is described in Section I.2.

In the local case, Theorem IV.2 constructs a sequence of disks $\Sigma_i \subset B_1$ as above where the curvatures blow up only at 0 and $\Sigma_i \backslash \{x_3\text{-axis}\}$ consists of two multivalued graphs for each i; see (1), (2), and (3). Furthermore (see (4)), $\Sigma_i \setminus \{x_3 = 0\}$ converges to two embedded minimal disks $\Sigma^- \subset \{x_3 < 0\}$ and $\Sigma^+ \subset \{x_3 > 0\}$ each of which spirals into $\{x_3 = 0\}$ and thus is not proper; see Figure 33.

THEOREM IV.2. [**CM16**, Theorem 1] (See Figure 33.) There is a sequence of compact embedded minimal disks $0 \in \Sigma_i \subset B_1 \subset \mathbf{R}^3$ with $\partial\Sigma_i \subset \partial B_1$ and containing the vertical segment $\{(0,0,t) : |t| < 1\} \subset \Sigma_i$ with the following properties:

(1) $\lim_{i\to\infty} |A_{\Sigma_i}|^2(0) = \infty$.
(2) $\sup_i \sup_{\Sigma_i \setminus B_\delta} |A_{\Sigma_i}|^2 < \infty$ for all $\delta > 0$.
(3) $\Sigma_i \setminus \{x_3\text{-axis}\} = \Sigma_{1,i} \cup \Sigma_{2,i}$ for multivalued graphs $\Sigma_{1,i}$ and $\Sigma_{2,i}$.
(4) $\Sigma_i \setminus \{x_3 = 0\}$ converges to two embedded minimal disks $\Sigma^\pm \subset \{\pm x_3 > 0\}$ with $\overline{\Sigma^\pm} \setminus \Sigma^\pm = B_1 \cap \{x_3 = 0\}$. Moreover, $\Sigma^\pm \setminus \{x_3\text{-axis}\} = \Sigma_1^\pm \cup \Sigma_2^\pm$ for multivalued graphs $\Sigma_1^\pm$ and $\Sigma_2^\pm$ which spiral into $\{x_3 = 0\}$ (see Figure 33).

It follows from (4) that $\Sigma_i \setminus \{0\}$ converges to a lamination of $B_1 \setminus \{0\}$ (with leaves Σ^-, Σ^+, and $B_1 \cap \{x_3 = 0\} \setminus \{0\}$) which does not extend to a lamination of B_1. Namely, 0 is not a removable singularity.

The example in Theorem IV.2 was constructed using the Weierstrass representation. Recall that if $\Omega \subset \mathbf{C}$ is a domain, then the classical Weierstrass representation starts from a meromorphic function g on Ω and a holomorphic one-form ϕ on Ω and associates a (branched) conformal minimal immersion $F : \Omega \to \mathbf{R}^3$ by

$$F(z) = \mathrm{Re} \int_{\zeta \in \gamma_{z_0,z}} \left(\frac{1}{2}\,(g^{-1}(\zeta) - g(\zeta)), \frac{i}{2}\,(g^{-1}(\zeta) + g(\zeta)), 1 \right) \phi(\zeta). \tag{IV.3}$$

Here $z_0 \in \Omega$ is a fixed base point and the integration is along a path $\gamma_{z_0,z}$ from z_0 to z. The choice of z_0 changes F by adding a constant. When g has no zeros or poles and Ω is simply connected, then $F(z)$ does not depend on the choice of path $\gamma_{z_0,z}$.

Sketch of proof of Theorem IV.2. We construct a one-parameter family (with parameter $a \in (0, 1/2)$) of minimal immersions F_a by making a specific choice of Weierstrass data $g = \mathrm{e}^{ih_a}$ (where $h_a = u_a + i\,v_a$), $\phi = dz$, and domain Ω_a to use in (IV.3). Namely, for each $0 < a < 1/2$, we define

$$h_a(z) = \frac{1}{a} \arctan \frac{z}{a} \text{ on } \Omega_a = \{(x, y) : |x| \leq 1/2,\ |y| \leq (x^2 + a^2)^{3/4}/2\}. \tag{IV.4}$$

Note that the function h_a is well-defined since Ω_a is simply connected and $\pm i\,a \notin \Omega_a$.

It remains to verify that this sequence of minimal immersions has the desired properties. Properties (1) and (2) in Theorem IV.2 follow easily from calculating the curvature K_a,

$$\mathrm{K}_a(z) = \frac{-|\partial_z h_a|^2}{\cosh^4 v_a} = \frac{-|z^2 + a^2|^{-2}}{\cosh^4 \left(\mathrm{Im} \arctan(z/a)/a\right)}. \tag{IV.5}$$

Properties (3) and (4) as well as the convergence away from 0 follow rather easily by integrating the equations (IV.3). The key point in the proof, and only remaining point, is to show that the immersions $F_a : \Omega_a \to \mathbf{R}^3$ are embeddings. We do this by showing that each vertical line segment in the domain Ω_a is mapped by F_a to curve in a horizontal plane (i.e., a plane where x_3 is constant) and each such curve is a graph over a fixed line segment; see **[CM16]** for the details. □

References

[ChC] J. Cheeger and T. H. Colding, On the Structure of Spaces with Ricci Curvature Bounded Below I, *Jour. of Diff. Geometry* 46 (1997), 406–480.

[CM1] T. H. Colding and W. P. Minicozzi II, Minimal surfaces, Courant Lecture Notes in Math. 4, 1999.

[CM2] T. H. Colding and W. P. Minicozzi II, Estimates for parametric elliptic integrands, *International Mathematics Research Notices*, 2002:6 (2002) 291–297.

[CM3] T. H. Colding and W. P. Minicozzi II, The space of embedded minimal surfaces of fixed genus in a 3-manifold I; Estimates off the axis for disks (math.AP/0210106). To appear in *Annals of Math.*

[CM4] T. H. Colding and W. P. Minicozzi II, The space of embedded minimal surfaces of fixed genus in a 3-manifold II; Multi-valued graphs in disks (math.AP/0210086). To appear in *Annals of Math.*

[CM5] T. H. Colding and W. P. Minicozzi II, The space of embedded minimal surfaces of fixed genus in a 3-manifold III; Planar domains (math.AP/0210141).

[CM6] T. H. Colding and W. P. Minicozzi II, The space of embedded minimal surfaces of fixed genus in a 3-manifold IV; Locally simply connected (math.AP/0210119).

[CM7] T. H. Colding and W. P. Minicozzi II, The space of embedded minimal surfaces of fixed genus in a 3-manifold V; Fixed genus, in preparation.

[CM8] T. H. Colding and W. P. Minicozzi II, Multi-valued minimal graphs and properness of disks, *International Mathematics Research Notices* 2002:21 (2002) 1111–1127.

[CM9] T. H. Colding and W. P. Minicozzi II, On the structure of embedded minimal annuli, *International Mathematics Research Notices*, 2002:29 (2002) 1539–1552.

[CM10] T. H. Colding and W. P. Minicozzi II, Minimal annuli with and without slits, *Jour. of Symplectic Geometry*, 1:1 (2001) 47–61.

[CM11] T. H. Colding and W. P. Minicozzi II, Complete properly embedded minimal surfaces in $\mathbf{R}^3$, *Duke Math. J.* 107:2 (2001), 421–426.

[CM12] T. H. Colding and W. P. Minicozzi II, Convergence of embedded minimal surfaces without area bounds in three manifolds, *C. R. Acad. Sci. Paris* Série I, 327, (1998) 765–770.

[CM13] T. H. Colding and W. P. Minicozzi II, Removable singularities for minimal limit laminations, *C. R. Acad. Sci. Paris* Série, 331, (2000) 465–468.

[CM14] T. H. Colding and W. P. Minicozzi II, Embedded minimal surfaces without area bounds in 3-manifolds. Geometry and topology: Århus (1998), 107–120, Contemp. Math., 258, Amer. Math. Soc., Providence, RI, 2000.

[CM15] T. H. Colding and W. P. Minicozzi II, Disks that are double spiral staircases, *Notices of the AMS*, 50:3 (2003), 327–339.

[CM16] T. H. Colding and W. P. Minicozzi II, Embedded minimal disks: Proper versus nonproper, global versus local, Trans. Amer. Math. Soc. 356 (2004), 283–289.

[CM17] T. H. Colding and W. P. Minicozzi II, An excursion into geometric analysis, Surveys in Differential Geometry, IX (math.DG/0309021).

[MeRo] W. Meeks III and H. Rosenberg, The uniqueness of the helicoid and the asymptotic geometry of properly embedded minimal surfaces with finite topology, preprint.

Courant Institute of Mathematical Sciences and Princeton University, 251 Mercer Street, New York, NY 10012 and Fine Hall, Washington Rd., Princeton, NJ 08544-1000

E-mail address: colding@cims.nyu.edu

Department of Mathematics, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218

E-mail address: minicozz@jhu.edu

Clay Mathematics Proceedings
Volume 2, 2004

Construction of Minimal Surfaces by Gluing Weierstrass Representations

Martin Traizet

0.1. Introduction. In this note I explain the construction of minimal surfaces by a singular perturbation method. We obtain embedded, complete minimal surfaces in $\mathbb{R}^3$, of finite total curvature. The surfaces we obtain may be seen as N parallel planes with small catenoidal necks between them (the planes are perturbed to have logarithmic growth at infinity).

The method we use is inspired from the proof of the Uniqueness of the Riemann Example by W. Meeks, J. Perez, and A. Ros [**4**]. We define quite explicitly the Weierstrass data of our surfaces and we solve the Period Problem using the Implicit Function Theorem.

Before stating a general theorem, I would like to describe in some details the Costa–Hoffman–Meeks family, which we will recover as a particular case. This is also the only case where pictures are available.

The Costa–Hoffman–Meeks family depends on two parameters: an integer $m \geq 2$ and a real $x \geq 1$. Each surface has genus $m-1$, three ends and m vertical planes of symmetry. x is a modulus for the underlying Riemann surface. $m = 2, x = 1$ yields the original Costa surface.

What is relevant for us is the behavior of the family as $x \to \infty$. What we observe on pictures is that for large values of x, the surface looks like three "planes", with one "neck" between the first and second planes (we say this is the neck at *level* one) and m necks between the second and third planes (we say these are the necks at level 2). After suitable scaling, the three "planes" converge when $x \to \infty$ to the horizontal plane $x_3 = 0$, and the necks collapse to $m+1$ distinct points in this plane, which we call their limit *position*. From the symmetries of the surface, it is clear that the top necks converge to the vertices of a regular m-gon and the bottom neck converges to the center of this polygon. In other words, the limit (after suitable scaling) of the Costa–Hoffman–Meeks family when $x \to \infty$ is a 3-sheeted plane with $m+1$ singular points placed at the vertices and center of a regular m-gon.

Under a larger scale, each neck converges, after suitable translation, to a catenoid. We call the radius of the neck of this catenoid the limit *size* of the neck. What is relevant is the ratio between the limit sizes of the necks.

We generalize this situation by allowing more ends and necks. The input data for our construction is the level, position and size of each neck. This is what we call a configuration:

- an integer $N \geq 2$ (number of ends),

©2004 Clay Mathematics Institute

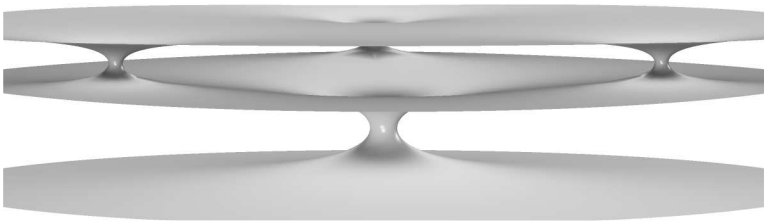


FIGURE 1. The genus 2 Costa–Hoffman–Meeks surface for large value of the conformal parameter. Computer image by J. Hoffman.

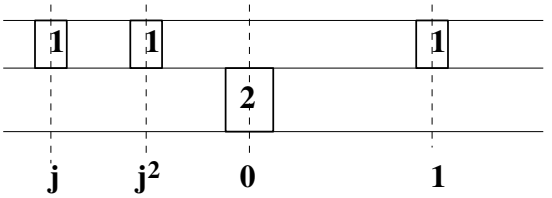


FIGURE 2. Configuration for the genus 2 Costa–Hoffman–Meeks family.

- an integer $n \geq 1$ (number of necks),
- for each $i \in [\![1, n]\!]$, an integer $\ell_i \in [\![1, N-1]\!]$ (level of the i-th neck — here the word level has a combinatorial meaning, as explained above),
- for each $k \in [\![1, N-1]\!]$, a real number $c_k > 0$ (limit size of the necks at level k),
- for each $i \in [\![1, n]\!]$, a complex number p_i (limit position of the i-th neck, identifying $\mathbb{R}^2$ with $\mathbb{C}$).

For example the configuration for the genus 2 Costa–Hoffman–Meeks family would be $N = 3$, $n = 4$, $\ell_1 = 1$, $\ell_2 = \ell_3 = \ell_4 = 2$, $c_1 = 2$, $c_2 = 1$, $p_1 = 0$, $p_2 = 1$, $p_3 = j$, $p_4 = j^2$, where j is the primary cubic root of 1. This data is easier to read on a figure.

Think of the horizontal lines as horizontal planes and the boxes as vertical tubes and you get a topological model for the surface. Our main existence result is the following:

THEOREM 1. *Given a configuration which is* balanced *and* nondegenerate *(these words are explained below) there exists a family of minimal surfaces M_t, $0 < t < \varepsilon$, ε small enough, which we may describe as follows. We may decompose M_t into planar domains and necks:*

$$M_t = \bigcup_{k=1}^{N} \Omega_{k,t} \cup \bigcup_{i=1}^{n} U_{i,t},$$

where

- *$\Omega_{k,t}$ is a graph over an unbounded domain in the horizontal plane. The Gauss curvature of $\Omega_{k,t}$ converges uniformly to 0 when $t \to 0$. At infinity $\Omega_{k,t}$ has a catenoidal end whose logarithmic growth converges when $t \to 0$*

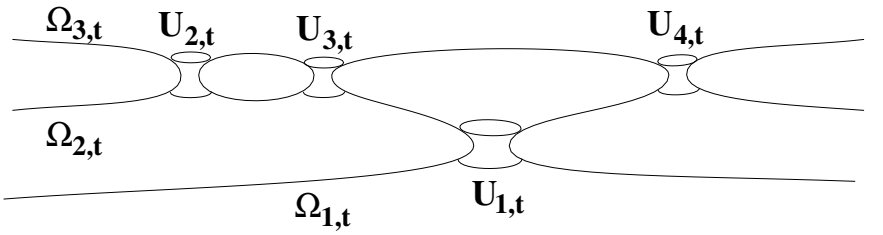


FIGURE 3. Planar domains and necks.

to

$$Q_k = n_{k-1}c_{k-1} - n_k c_k,$$

where n_k is the number of necks at level k, namely, the number of indices i such that $\ell_i = k$. Moreover inside any fixed vertical cylinder of radius R, M_t stays at a bounded distance (depending on R but not on t) from the horizontal plane $x_3 = 2|\log t|(c_1 + \cdots + c_{k-1})$.

- *$U_{i,t}$ is an annulus whose two boundary components are respectively in $\Omega_{\ell_i,t}$ and $\Omega_{\ell_i+1,t}$. There exists vectors $T_{i,t}$ such that $U_{i,t} - T_{i,t}$ converges to a catenoid of size c_{ℓ_i} when $t \to 0$. Moreover, the horizontal part of $t^{-1}T_{i,t}$ converges to p_i when $t \to 0$.*

It follows from this geometrical description that if $Q_1 < Q_2 < \cdots < Q_N$, then M_t is embedded for t small enough. Also M_t converges, after scaling by t, to a single, N-sheeted horizontal plane, with singular points at $p_1, \ldots, p_n$.

0.2. Forces. It is convenient to explain the balancing condition in terms of forces. We use the following notation: $I_k = \{i \in [\![1,n]\!] : \ell_i = k\}$, $n_k = \#I_k$ is the number of necks at level k.

Given $i \in I_k$, we define the force F_i by

$$F_i = \sum_{j\in I_k,\, j\neq i} \frac{2c_k^2}{p_i - p_j} - \sum_{j\in I_{k+1}} \frac{c_k c_{k+1}}{p_i - p_j} - \sum_{j\in I_{k-1}} \frac{c_k c_{k-1}}{p_i - p_j}.$$

For this to make sense we need to assume that $p_j \neq p_i$ whenever $i \neq j$ and $|\ell_i - \ell_j| \leq 1$. We say the configuration is *nonsingular*. A more compact way to write F_i is to define, for $1 \leq i \leq n$ and $1 \leq k \leq N$, the charge $Q_{i,k}$ by

$$Q_{i,k} = \begin{cases} -c_{\ell_i} & \text{if } k = \ell_i, \\ c_{\ell_i} & \text{if } k = \ell_i + 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$F_i = \sum_{k=1}^{N} \sum_{j\neq i} \frac{Q_{i,k}Q_{j,k}}{p_i - p_j}.$$

One may think of p_i as a particle living in an N-sheeted plane and $Q_{i,k}$ as the charge of p_i in the k-th sheet. Then one may interpret $\overline{F_i}$ as an electrostatic force.

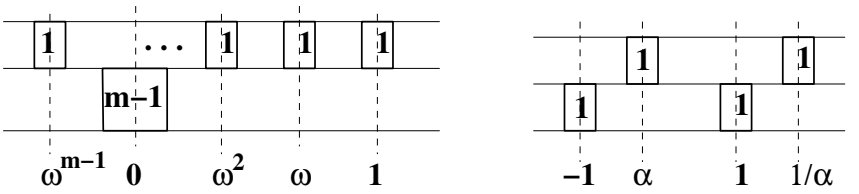


FIGURE 4. Left: configuration for the Costa–Hoffman–Meeks family of genus $m-1$, with $\omega^m = 1$. Right: the Horgan configuration, with $\alpha \in \mathbb{C}^*$.

We say the configuration is *balanced* if $\forall i$, $F_i = 0$. These equations are not independent. The forces satisfy the two equations

$$\sum_{i \in I} F_i = 0 \tag{0–1}$$

and

$$\sum_{i \in I} p_i F_i = \sum_{k=1}^{N} \sum_{i} \sum_{j<i} Q_{i,k} Q_{j,k}. \tag{0–2}$$

A necessary condition for the configuration to be balanced is that the right member of (0–2) be zero. In terms of the neck sizes c_k, this is equivalent to

$$\sum_{k=1}^{N-1} n_k(n_k - 1)c_k^2 - \sum_{k=1}^{N-2} n_k n_{k+1} c_k c_{k+1} = 0. \tag{0–3}$$

This gives one restriction on the neck sizes. Note that the balancing condition is invariant by transformations of the form $p_i \mapsto ap_i + b$, where $a, b \in \mathbb{C}$, namely translation and complex scaling.

We say the configuration is *nondegenerate* if

- the complex matrix $\partial F_i / \partial p_j$ has rank $n-2$ (the largest possible). This means that the only infinitesimal deformations of the configuration are trivial, namely of the form $z \mapsto az + b$.
- There exists k such that

$$\frac{\partial}{\partial c_k}\left(\sum_{k=1}^{N-1} n_k(n_k - 1)c_k^2 - \sum_{k=1}^{N-2} n_k n_{k+1} c_k c_{k+1}\right) \neq 0.$$

0.3. Examples. We have the following classification of balanced configurations with three ends (see [**6**]).

THEOREM 2. *Assume $N = 3$. Without loss of generality we may assume that $n_1 \leq n_2$. Then the only balanced configurations are, up to normalization, the Costa–Hoffman–Meeks configurations and the Horgan configuration.*

The Horgan configuration has a nontrivial free parameter α, so it is degenerate and we cannot tell whether the corresponding minimal surface exists or not. This surface would have genus 2 and three ends. In the most symmetrical case $\alpha = \sqrt{-1}$, it is proven not to exist in [**8**]. Actually the fact that the configuration is balanced explains why it is possible to make really good pictures [**8**] of this nonexisting surface.

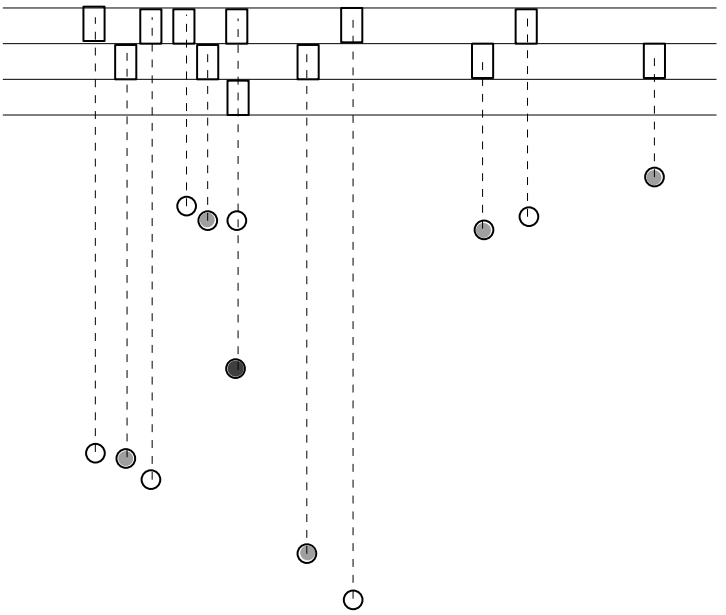


FIGURE 5. A nondegenerate balanced configuration with no symmetries. The position of each neck is represented in the complex plane by a circle. Color indicates level.

This classification of configurations with three ends gives good support to the conjecture that the only embedded minimal surfaces with three ends are the Costa–Hoffman–Meeks surfaces, see [**2**], section 5.2.

On the contrary, the space of minimal surfaces with four or more ends is much richer.

THEOREM 3. [**6**] *There exist embedded minimal surfaces with no symmetries.*

An example is provided by the following configuration.

The sizes of the necks are $c_1 = 23/6$, $c_2 = 9/8$, $c_3 = 1$. The logarithmic growths of the ends are $Q_1 \simeq -3.8333$, $Q_2 \simeq -1.7916$, $Q_3 = -0.375$, $Q_4 = 6$. The catenoidal ends of this example have different vertical axes. This example answers some questions in [**2**], section 5.2.

0.4. Weierstrass representation. All results in this paper are proven using the Weierstrass Representation, which we write in the form

$$x_1(z) + \sqrt{-1}\, x_2(z) = \frac{1}{2}\left(\overline{\int_{z_0}^{z} g^{-1} dh} - \int_{z_0}^{z} g\, dh\right),$$

$$x_3 = \operatorname{Re} \int_{z_0}^{z} dh,$$

where $z \in \Sigma$, Σ is a Riemann surface, g and dh are respectively a meromorphic function (the Gauss map) and a holomorphic 1-form (the height differential) on Σ. The Period Problem asks that x_1, x_2, x_3 be well defined on Σ, which is equivalent

to

$$\overline{\int_\gamma g^{-1}h} - \int_\gamma g\,dh = 0, \qquad \mathrm{Re}\int_\gamma dh = 0,$$

for any cycle γ on Σ.

1. Geometric Meaning of the Forces

In this section we prove that the balancing condition is necessary. Let γ be a curve on a minimal surface. The flux of γ is the homology invariant vector $\mathrm{flux}(\gamma) = \int_\gamma \nu\, ds$ where ν is the conormal to γ. Let $\varphi(\gamma)$ be the horizontal part of $\mathrm{flux}(\gamma)$.

PROPOSITION 1. *Assume that there exists a family of minimal surfaces M_t satisfying the conclusions of Theorem 1. Then*

$$\lim_{t\to 0} t^{-1}\varphi(\partial U_{i,t}) = -2\pi\overline{F_i}.$$

By homology invariance of the flux, this gives $F_i = 0$.

We prove this proposition in the next two sections.

1.1. Limit of the Weierstrass data on $\Omega_{k,t}$. In this section k is fixed. We first need to identify $\Omega_{k,t}$ with a domain in the complex plane. We assume that the Gauss map of $\Omega_{k,t}$ points up. The computations are much similar (up to some complex conjugations) when the Gauss map points down and give the same final result, which makes sense because the flux only depends on the choice of the conormal (we choose the exterior conormal) and not on orientation.

Let $\pi : \Omega_{k,t} \to \mathbb{C}$ be the projection to the horizontal plane followed by scaling by t. Since the Gauss map is close to (0,0,1), the map π is close to be conformal and preserves orientation. Using quasi-conformal mappings, it is possible to prove that $\Omega_{k,t}$ is conformal to a domain which is close to $\pi(\Omega_{k,t})$. This is quite intuitive so we omit the details (see [**7**]).

Let I be the set of indices i such that $\ell_i = k-1$ or $\ell_i = k$. For $i \in I$, let γ_i be the component of $\partial U_{i,t}$ which is in $\Omega_{k,t}$, namely, the top component if $\ell_i = k-1$ and the bottom component if $\ell_i = k$. Then $\pi(\gamma_i)$ converges when $t \to 0$ to the point p_i. So $\Omega_{k,t}$ is conformal to a domain which converges, when $t \to 0$, to $\mathbb{C} \setminus \{p_i : i \in I\}$. We identify $\Omega_{k,t}$ with this domain.

LEMMA 1. *Let (g_t, dh_t) be the Weierstrass data of M_t. On $\Omega_{k,t}$, seen as a domain in $\mathbb{C}$, we have*

$$\lim_{t\to 0} dh_t = \sum_{i\in I} \frac{Q_{i,k}}{z-p_i}dz, \qquad \lim_{t\to 0} t g_t dh_t = -2dz.$$

PROOF. Since $U_{i,t}$ converges to a catenoid, we have

$$\lim_{t\to 0}\int_{\gamma_i} dh_t = 2\pi\sqrt{-1}\,Q_{i,k} \tag{1–1}$$

and

$$\int_{\gamma_i} |dh_t| \le C$$

for some constant C. Write $dh_t = f_t dz$. Let V_ε be the set of $z \in \mathbb{C}$ which are at distance at least ε from the points p_i, $i \in I$. Then for t small enough, any $z \in V_\varepsilon$ is at distance at least $\varepsilon/2$ from $\partial\Omega_{k,t}$. By the Cauchy theorem,

$$\forall z \in V_\varepsilon, \quad |f_t(z)| = \left| \frac{1}{2\pi i} \int_{\partial\Omega_{k,t}} \frac{f_t(w)}{w - z} dw \right| \leq \frac{C}{\pi\varepsilon}. \tag{1–2}$$

By Montel's Theorem, we may extract a subsequence, still denoted f_t, which converges on V_ε. By a diagonal process, we may extract a subsequence which converges on V_ε for all ε. Let $f = \lim f_t$. By equation (1–2), for any small $\varepsilon > 0$ and any large R, we have

$$\int_{|z-p_i|=\varepsilon} |f(z)dz| \leq 2C, \quad \int_{|z|=R} |f(z)dz| \leq 2C.$$

This implies that $f(z)dz$ has at most a simple pole at p_i and at ∞. By equation (1–1) the residue at p_i is $Q_{i,k}$. This gives the first statement. To prove the second statement, write

$$dz \sim d\pi = \frac{t}{2}\left(\overline{g_t^{-1} dh_t} - g_t dh_t\right) \sim -\frac{t}{2} g_t dh_t. \qquad \square$$

1.2. Limit of the flux. By a standard computation, the horizontal part of the flux is

$$\varphi(\gamma) = \frac{\sqrt{-1}}{2}\left(\overline{\int_\gamma g^{-1} dh} + \int_\gamma g\, dh\right).$$

When γ is a closed curve,

$$\overline{\int_\gamma g^{-1} dh} - \int_\gamma g\, dh = 0.$$

This gives, for $i \in I$,

$$\begin{aligned}
\lim_{t\to 0} t^{-1}\varphi(\gamma_i) &= \lim_{t\to 0} \sqrt{-1}\, \overline{\int_{|z-p_i|=\varepsilon} t^{-1} g_t^{-1} dh_t} = \lim_{t\to 0} \sqrt{-1}\, \overline{\int_{|z-p_i|=\varepsilon} (dh_t)^2 (t g_t dh_t)^{-1}} \\
&= -\frac{\sqrt{-1}}{2} \overline{\int_{|z-p_i|=\varepsilon} \left(\sum_{j\in I} \frac{Q_{j,k}}{z - p_j}\right)^2 dz} \\
&= -\pi \overline{\operatorname{Res}_{p_i} \left(\sum_j \frac{Q_{j,k}}{z - p_j}\right)^2} \\
&= -2\pi \sum_{j\neq i} \frac{Q_{i,k} Q_{j,k}}{\bar{p}_i - \bar{p}_j}.
\end{aligned}$$

To obtain Proposition 1 we simply have to add the two terms coming from the two components of $\partial U_{i,t}$, which amounts to a sum on k.

REMARK 1. *We assumed, in the definition of a configuration, that necks at the same level have the same limit size. It is possible to prove that this condition is in fact necessary (see section 2.5).*

2. Proof of Theorem 1

In this section, we prove Theorem 1, but we omit most technical details and some computations. A fully detailed proof will appear in [**6**]. The idea of the proof is as follows.

We define the Weierstrass data (Σ, g, dh) depending on the parameter $t > 0$ (which is the same as t in the statement of Theorem 1) and some other parameters. The Riemann surface and the Gauss map are defined by explicit formulae. The height differential dh is defined in a more abstract way by prescribing its residues and periods.

The key point is that when $t \to 0$, we can compute explicitly the limit of dh. Roughly speaking, when $t \to 0$, the Riemann surface Σ degenerates into a Riemann surface with nodes, whose parts have genus zero, so that the limit of dh can be explicitly computed.

Using our explicit formulae for the Weierstrass data when $t \to 0$, we compute the limit of the periods of the Weierstrass data. After suitable renormalization, we prove that each equation extends smoothly to $t = 0$. The Period Problem boils down to the balancing condition when $t = 0$. We solve the period problem for t close to 0 using the Implicit Function Theorem. This is where we use the nondegeneracy condition.

2.1. The Riemann surface and the Gauss map. Consider N copies of the complex planes, labeled $\mathbb{C}_1, \cdots, \mathbb{C}_N$. For each $i \in [\![1, n]\!]$ consider some complex parameters a_i, b_i, $\alpha_i \neq 0$, $\beta_i \neq 0$. Recall that I_k is the set of indices i such that $\ell_i = k$. Define the Gauss map g on the disjoint union $\mathbb{C}_1 \cup \cdots \cup \mathbb{C}_N$ by

$$g(z) = \begin{cases} t g_k(z) & \text{if } z \in \mathbb{C}_k, \ k \text{ odd}, \\ (t g_k(z))^{-1} & \text{if } z \in \mathbb{C}_k, \ k \text{ even}, \end{cases}$$

and

$$g_k(z) = \sum_{i \in I_k} \frac{\alpha_i}{z - a_i} + \sum_{i \in I_{k-1}} \frac{\beta_i}{z - b_i}.$$

To define the Riemann surface Σ, we identify pairs of points to create necks. Consider some $i \in I_k$. We need to create a neck connecting $\mathbb{C}_k$ and $\mathbb{C}_{k+1}$. Think of $a_i \in \mathbb{C}_k$ and $b_i \in \mathbb{C}_{k+1}$ as the bottom and top position of the neck (these parameters are closely related to p_i). Define local complex coordinates $v_i = 1/g_k$ and $w_i = 1/g_{k+1}$ in a neighborhood of a_i and b_i respectively.

We want to identify pairs of points z in a neighborhood of a_i and z' in a neighborhood of b_i which have the same Gauss map, so that g is well defined in the quotient. Hence the rule to identify points is (when, say, k is odd)

$$g(z) = g(z') \iff \frac{t}{v_i(z)} = \frac{w_i(z')}{t} \iff v_i(z) w_i(z') = t^2.$$

So the definition of Σ is as follows. Consider a fixed, small enough $\varepsilon > 0$. For each $i \in [\![1, n]\!]$, remove the disks $|v_i| \leq t^2/\varepsilon$ and $|w_i| \leq t^2/\varepsilon$. Identify the points z and z' such that

$$\frac{t^2}{\varepsilon} < |v_i(z)| < \varepsilon, \qquad \frac{t^2}{\varepsilon} < |w_i(z')| < \varepsilon, \qquad v_i(z) w_i(z') = t^2.$$

This defines a Riemann surface Σ. This is the conformal model for the minimal surface we want to construct. By construction g is a well-defined meromorphic

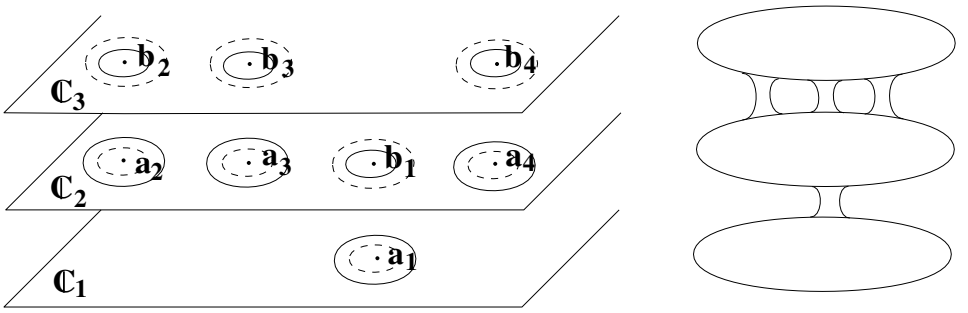


FIGURE 6. Left: definition of Σ. Right: topological picture of $\widehat{\Sigma}$.

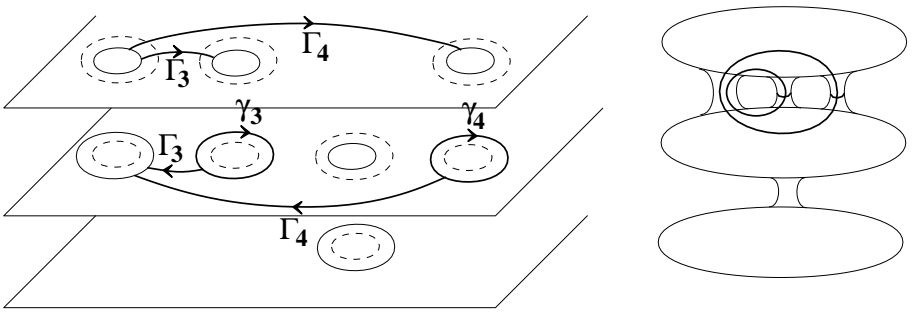


FIGURE 7. Canonical homology basis (genus 2).

function on Σ. Let $\widehat{\Sigma} = \Sigma \cup \{\infty_1, \dots, \infty_N\}$ be the compactification of Σ, where ∞_k is the point at infinity in $\mathbb{C}_k$. We have $g(\infty_k) = 0$ if k is odd and $g(\infty_k) = \infty$ if k is even.

2.2. The height differential. We need to define a meromorphic 1-form dh with simple poles at $\infty_1, \dots \infty_N$ because we want catenoidal ends. Let me recall some standard complex analysis. Let $\widehat{\Sigma}$ be a compact Riemann surface of genus G.

- A canonical homology basis of $\widehat{\Sigma}$ is a set of $2G$ closed curves γ_i, Γ_i, $i \in [\![1, G]\!]$ such that γ_i intersects Γ_i with intersection number 1 and all other intersection numbers are zero.
- The space of holomorphic 1-forms on $\widehat{\Sigma}$ has complex dimension G. An isomorphism with $\mathbb{C}^G$ is given by integration along the curves $\gamma_1, \dots, \gamma_G$.
- The space of meromorphic 1-forms with simple poles at given points $q_1, \dots, q_N$ has complex dimension $G + N - 1$. One may prescribe the integrals along the curves $\gamma_1, \dots \gamma_G$ and the residues at the poles, with the only condition that the sum of the residues be zero.

In our case, it follows from the topological picture that $\widehat{\Sigma}$ has genus $n - N + 1$. We define a canonical basis as in Figure 7.

Here the curve γ_i is a small circle around a_i with the negative orientation. It is homologous in Σ to a small circle around b_i with the positive orientation. We think of γ_i as a curve around the i-th neck. Let $i_0(k) = \min I_k$ and $J = \{i : i > i_0(\ell_i)\}$. Then $\{\gamma_i, \Gamma_i\}_{i \in J}$ is a canonical homology basis of $\widehat{\Sigma}$.

The height differential dh is defined as the unique meromorphic 1-form on Σ with simple poles at $\infty_1, \dots \infty_N$ such that

$$\int_{\gamma_i} dh = 2\pi\sqrt{-1}\, r_i$$

for all $i \in J$, and

$$\operatorname{Res}_{\infty_k} dh = -Q_k$$

for all $k \in [\![1, N]\!]$, where r_i are positive real numbers and Q_k are real numbers such that $\sum Q_k = 0$. Geometrically r_i is the size of the i-th neck and Q_k is the logarithmic growth of the catenoidal end ∞_k.

If $i \notin J$, we *define* r_i by $\int_{\gamma_i} dh = 2\pi\sqrt{-1}\, r_i$. It easily follows from the Residue Theorem that

$$\sum_{i \in I_k} r_i - \sum_{j \in I_{k-1}} r_j = -Q_k. \tag{2–1}$$

2.3. Limit of dh when $t \to 0$. When $t = 0$, Σ is the disjoint union $\Sigma_0 = \mathbb{C}_1 \cup \dots \cup \mathbb{C}_N$. We admit the following lemma, see [**5**] or [**6**], whose proof relies on algebraic geometry results by Fay [**1**] and Masur [**3**].

LEMMA 2. *When $t \to 0$, dh converges to a meromorphic 1-form on Σ_0 which has simple poles at a_i and b_i, $1 \le i \le n$ and at ∞_k, $1 \le k \le N$. The convergence is uniform on compact subsets of $\Sigma_0 \setminus \{a_i, b_i\}$. Moreover, dh depends analytically on all parameters, including at $t = 0$.*

Since $\int_{\gamma_i} dh = 2\pi\sqrt{-1}\, r_i$ and γ_i may be seen as a negative circle around a_i or a positive circle around b_i, we see that the residues at a_i and b_i are respectively $-r_i$ and r_i. Hence

$$\lim_{t \to 0} dh = \sum_{i \in I_k} \frac{-r_i}{z - a_i} dz + \sum_{i \in I_{k-1}} \frac{r_i}{z - b_i} dz \quad \text{on } \mathbb{C}_k.$$

2.4. Zeros of dh. By standard minimal surface theory, dh needs a zero at each zero or pole of g (except the ends), with the same multiplicity. In this section we solve these equations. This will determine the value of the parameters α_i, β_i as functions of the other parameters.

We make the simplifying (but generic) assumption that all the zeros of g_k are simple (including ∞_k). Recall that $n_k = \#I_k$ is the number of necks at level k. Let $\zeta_{k,j}$, $1 \le j \le n_k + n_{k-1} - 1$ be the zeros of g_k in $\mathbb{C}_k$. Let

$$\mathcal{F}_{j,k} = \frac{dh}{dz}(\zeta_{j,k});$$

$\mathcal{F}_{j,k}$ is an analytic function of all parameters. We first solve $\mathcal{F} = 0$ when $t = 0$. Then we solve $\mathcal{F} = 0$ for t close to 0 using the Implicit Function Theorem.

- An obvious solution to $\mathcal{F} = 0$ when $t = 0$ is to take $\alpha_i = -r_i$ and $\beta_i = r_i$ for all i. Indeed, this gives $dh = g_k dz$ so dh and g_k have the same zeros in $\mathbb{C}_k$. The solution is not unique but since we only care about existence we may choose the solution which is most convenient. In fact any other solution is proportional to that one, and choosing this particular solution is a matter of normalization.

- It is easy to see that the partial differential of $\mathcal{F}$ with respect to the variables α_i, β_i, $1 \le i \le n$ is surjective. By the Implicit Function Theorem, for t close to 0, there exists α_i, β_i depending analytically on t and all remaining parameters (namely a_i, b_i, r_i and Q_k) such that $\mathcal{F} = 0$.

2.5. Γ-periods of dh. We now start to solve the period problem. Note that $\int_{\gamma_i} dh$ is pure imaginary by definition of dh. In this section we solve the equations $\operatorname{Re} \int_{\Gamma_i} dh = 0$. This will determine the value of the parameters r_i.

LEMMA 3. *Consider some k, $1 \le k \le N-1$. Let $i_0 = i_0(k) = \min I_k$. Given $i \in I_k$ such that $i > i_0$, we have*

$$\operatorname{Re} \int_{\Gamma_i} dh = 2(r_i - r_{i_0}) \log t + \textit{bounded analytic terms.}$$

PROOF. We may decompose B_i as a path from $v_i = t$ to $v_{i_0} = t$, contained in $\mathbb{C}_k$, followed by a path from $w_{i_0} = t$ to $w_i = t$, contained in $\mathbb{C}_{k+1}$. Using Lemma 2, we have

$$\begin{aligned} \int_{v_i=t}^{v_{i_0}=t} dh &\simeq \int_{v_i=t}^{v_{i_0}=t} \sum_{i \in I_k} \frac{-r_i}{z-a_i} dz + \sum_{i \in I_{k-1}} \frac{r_i}{z-b_i} dz \\ &\simeq -r_{i_0} \log(v_{i_0}^{-1}(t) - a_{i_0}) + r_i \log(v_i^{-1}(t) - a_i) \\ &\simeq (r_i - r_{i_0}) \log t. \end{aligned}$$

The other integral is evaluated in the same way. This computation is of course not correct because the convergence of dh is not uniform on the path of integration. This may be corrected using a Laurent series expansion of dh in the annular region $t^2/\varepsilon < |v_i| < \varepsilon$ (see details in [**6**]). □

Let $\mathcal{F}_i$ be the renormalized period of dh along Γ_i, namely

$$\mathcal{F}_i = \frac{1}{\log t} \operatorname{Re} \int_{\Gamma_i} dh.$$

The problem is that $\mathcal{F}_i$ is not differentiable with respect to t at $t = 0$. We solve this problem by writing $t = \exp(-1/\tau^2)$ where τ is a real number in a neighborhood of 0. Now

$$\mathcal{F}_i = 2(r_i - r_{i_0}) - \tau^2 \operatorname{analytic}\left(\exp(-1/\tau^2)\right)$$

is a *smooth* function of τ, with $\mathcal{F}_i|_{\tau=0} = 2(r_i - r_{i_0})$. From now on our main parameter is τ instead of t. Note that this change of variable works because we shall never take derivatives with respect to t in our use of the Implicit Function Theorem.

- When $\tau = 0$, $\mathcal{F}_i = 0$ gives that $r_i = r_{i_0}$ which means that all necks at level k have the same size (see Remark 1). Let $c_k = r_{i_0(k)}$. By equation (2–1) we have

  $$-n_k c_k + n_{k-1} c_{k-1} = Q_k,$$

  which determines c_k as a function of $Q_1, \dots, Q_N$ by induction on k. In fact, the above equation determines an isomorphism $(c_1, \dots c_{N-1}) \mapsto (Q_1, \dots, Q_N)$ from $\mathbb{R}^{N-1}$ to $\{(Q_1, \dots, Q_N) : \sum Q_k = 0\}$. In the introduction, Q_k was defined as a function of c_k, whereas here c_k is defined as a function of Q_k, which is still a free parameter.

- The partial differential of $(\mathcal{F}_i)_{i\in J}$ with respect to $(r_i)_{i\in J}$ at $\tau=0$ is clearly an isomorphism. By the Implicit Function Theorem, for τ close to 0, there exist r_i depending smoothly on τ and all remaining parameters, such that $\mathcal{F}_i=0$.

2.6. Horizontal Γ-periods. We continue with the period problem. The idea is the same as in the previous section, namely, compute the leading term of the period when $\tau\to 0$, prove that the renormalized period extends smoothly to $\tau=0$, solve when $\tau=0$ and use the Implicit Function Theorem. However, to avoid making the paper too long, I will skip all computations from now on. My goal is mostly to show how the balancing and nondegeneracy conditions are used to solve the period problem. See [**6**] for the details.

We define the *horizontal period* along a cycle c by

$$P(c)=\frac{1}{2}\left(\overline{\int_c g^{-1}dh}-\int_c g dh\right).$$

In this section we solve the equations $P(\Gamma_i)=0$. This determines the value of the parameters b_i. We find, for $i\in I_k$, k odd,

$$P(\Gamma_i)=\frac{t^{-1}}{2}(b_{i_0}+\overline{a_{i_0}}-b_i-\overline{a_i})+o(t^{-1})$$

and minus the conjugate of this formula when k is even. An obvious solution when $t=0$ is $b_i=-\overline{a_i}$. Using the implicit function theorem, for τ close to 0, there exists b_i depending smoothly on τ and the remaining parameters (namely a_i and Q_k) such that $P(\Gamma_i)=0$.

2.7. Horizontal γ-periods. The last equations we have to solve are $P(\gamma_i)=0$ for all $i\in J$ and $P(\delta_k)=0$ for all $k\in[\![1,N]\!]$, where δ_k is a small circle around ∞_k. Using the Residue Theorem, these equations are equivalent to $P(\gamma_i)=0$ for all $i\in[\![1,n]\!]$. We find, for $i\in I_k$, k odd,

$$P(\gamma_i)=2\pi\sqrt{-1}\,t\left(2\sum_{j\in I_k,j\neq i}\frac{r_ir_j}{a_i-a_j}-\sum_{j\in I_{k-1}\cup I_{k+1}}\frac{r_ir_j}{a_i+\overline{a_j}}\right)+o(t)$$

and minus the conjugate of this formula if k is even. We define the renormalized period $\mathcal{F}_i=t^{-1}P(\gamma_i)$. This is a smooth function of τ and the remaining parameters a_i and Q_k.

- To solve $\mathcal{F}=0$ when $\tau=0$, we use the given configuration. Take

$$a_i \;=\; \begin{cases} p_i & \text{if } i\in I_k,\ k \text{ odd},\\ -\bar{p}_i & \text{if } i\in I_k,\ k \text{ even}.\end{cases}$$

  The equation above gives

$$\mathcal{F}_i=2\pi\sqrt{-1}\,(-1)^{k+1}\overline{F_i}=0,$$

  where F_i is the force defined in the introduction. Note that at this point we do not know yet that p_i is the asymptotic position of the neck. We only use the configuration as a solution to a set of algebraic equations.
- By the nondegeneracy, the matrix $\partial F_i/\partial p_j$ has an invertible minor of size $n-2$. Hence there exist indices i_1,i_2 and j_1,j_2 such that the partial differential of $(\mathcal{F}_i)_{i\neq i_1,i_2}$ with respect to $(a_j)_{j\neq j_1,j_2}$ is an isomorphism. By

the Implicit Function Theorem, for τ in a neighborhood of 0, there exists a_j, depending smoothly on τ and Q_k, such that $\mathcal{F}_i = 0$ for all $i \neq i_1, i_2$.

We still need to prove that $P(\gamma_{i_1}) = P(\gamma_{i_2}) = 0$. It turns out that the periods satisfy some relations. Using the Residue Theorem and Riemann bilinear relations (applied to the forms $g^{-1}dh$ and gdh) we find respectively

$$P(\gamma_{i_1}) + P(\gamma_{i_2}) = 0,$$

$$\operatorname{Re}\left(P(\gamma_{i_1}) \int_{\Gamma_{i_1}} g^{-1}dh + P(\gamma_{i_2}) \int_{\Gamma_{i_2}} g^{-1}dh\right) = 0.$$

We need one more real equation, so let

$$\mathcal{G} = \operatorname{Im}\left(\sum_i P(\gamma_i) \int_{\Gamma_i} g^{-1}dh\right).$$

We find that $\mathcal{G}$ extends smoothly to $\tau = 0$ with

$$\mathcal{G}|_{\tau=0} = 2\pi \sum p_i F_i = 2\pi \sum_k \sum_{i<j} Q_{i,k} Q_{j,k}.$$

By the second statement of the nondegeneracy hypothesis and the Implicit Function Theorem, for τ in a neighborhood of 0, there exist Q_k, depending smoothly on τ, such that $\mathcal{G} = 0$. Together with the above equations, this implies that $P(\gamma_{i_1}) = P(\gamma_{i_2}) = 0$. This concludes the proof of Theorem 1, or rather, that the period problem can be solved. We still have to prove that the family of minimal surfaces M_t given by this Weierstrass data has all the claimed geometric properties. In particular, we have to prove that p_i is, up to scaling, the asymptotic position of the neck, that the necks converge to catenoids, and prove embeddedness. This is quite easy to do using our explicit formula for the Weierstrass data when $t \to 0$.

REMARK 2. *We see that we had to perturb the logarithmic growths of the ends. In fact we already knew, from the Costa–Hofmann–Meeks example, that this was necessary. This shows that there are no further relations amongst the periods and that it is necessary to introduce the function $\mathcal{G}$.*

Also we can in fact prescribe the value of $N-2$ logarithmic growths. This gives a family depending on $N-1$ real parameters, the conjectured dimension for the space of embedded minimal surfaces with N ends, modulo rigid motions.

References

[1] J. D. Fay, *Theta functions on Riemann surfaces*, Lecture Notes in Mathematics **352**, Springer, 1973.

[2] D. Hoffman, H. Karcher, "Complete embedded minimal surfaces of finite total curvature", pp. 5–93 in *Geometry V*, Encyclopedia of Math. Sci. **90**, edited by R. Osserman, Springer, 1997.

[3] H. Masur, "The extension of the Weil–Petersson metric to the boundary of Teichmüller space", *Duke Math. J.* **43** (1976), 623–635.

[4] W. H. Meeks, J. Perez, A. Ros: "Uniqueness of the Riemann minimal example" *Invent. Math.* **133**:1 (1998), 107–132.

[5] M. Traizet, "Adding handles to Riemann minimal examples", *J. Inst. Math. Jussieu* **1**:1 (2002), 145–174.

[6] M. Traizet, "An embedded minimal surface with no symmetries", *J. Diff. Geom.* **60**:1 (2002), 103–153.

[7] M. Traizet, "A balancing condition for weak limits of families of minimal surfaces", to appear in *Commentarii Math. Helv.*

[8] M. Weber, "On the Horgan minimal non-surface", *Calc. Var.* **7** (1998), 373–379.

Université de Tours, 37200 Tours, France
E-mail address: martin@gargan.math.univ-tours.fr

Clay Mathematics Proceedings
Volume 2, 2004

Global Problems in Classical Minimal Surface Theory

William H. Meeks III

1. Introduction

Our goal here is to present some of the basic theory and conjectures in the classical theory of minimal surfaces in $\mathbb{R}^3$ which have a global or topological emphasis. In Section 2 we discuss some of the basic definitions and theorems needed to understand the unsolved problems that appear in Section 3; we include in Section 2 the part of the theory that deals with the structure of the ends and the conformal structure of a properly embedded minimal surface. Section 3 is devoted to the discussion of many of the outstanding conjectures in the global theory of minimal surfaces. Many of these problems appear in print for the first time here. We refer the interested reader to the author's book [**29**] for a long list of unsolved problems that dates back to 1978; of course, some of the problems listed in the book have been solved.

The only theorem, Theorem 2.12, that appears here and does not appear elsewhere is the algebraic flux lemma for parabolic manifolds with boundary (see the related Conjecture 3.3 in Section 3). Partial unpublished results related to this theorem were obtained by Collin, Kusner, Meeks and Rosenberg during their research contained in [**13**]. An important application of this flux result to classical minimal surface theory is that associated to every properly immersed minimal surfaces M in $\mathbb{R}^3$ is a vector $F(M) \in (\mathbb{R}^+ \cup \infty)^3$ whose coordinates correspond to the fluxes arising from the ordered coordinate functions x_1, x_2, x_3 of M. Theoretical results of this type in the past have had a significant impact on the global theory of proper minimal surfaces in $\mathbb{R}^3$ and hopefully this result will have important applications as well.

2. Basic Definitions and Conformal Structure

We first define a minimal surface.

DEFINITION 1. An isometric immersion $f\colon M^2 \to \mathbb{R}^3$ of a Riemannian surface $(M^2, \langle\,,\rangle)$ is *minimal* if the coordinate functions f_1, f_2, f_2 are harmonic functions on M^2.

We will be interested in noncompact minimal surfaces which are complete in the following sense.

This research was supported by NSF grant DMS-0104044.

©2004 Clay Mathematics Institute

DEFINITION 2. An immersed surface $M \subset \mathbb{R}^3$ is *complete* if geodesics on M can be extended indefinitely or, equivalently, if M is a complete metric space with respect to the intrinsic distance function defined to be the infimum of the lengths of smooth curves on M joining a pair of points in M.

In principle we are interested in understanding the space of all complete embedded minimal surfaces in $\mathbb{R}^3$. All known examples of such surfaces satisfy the stronger hypothesis given in the next definition. Recall that a surface M has more than one end if it contains a smooth compact subdomain such that the complement of the interior of this domain in M has more than one noncompact component.

DEFINITION 3. An immersion $f\colon M \to \mathbb{R}^3$ is *proper* if for any compact ball B, $f^{-1}(B)$ is compact in M. Let $\mathcal{P}$ denote the space of all properly embedded connected minimal surfaces in $\mathbb{R}^3$ and let $\mathcal{M} \subset \mathcal{P}$ be the subset of examples with more than one end. The topology on $\mathcal{P}$ is the topology of smooth convergence on compact subsets of $\mathbb{R}^3$.

One of the outstanding problems in the global theory of complete embedded minimal surfaces in $\mathbb{R}^3$ is to decide if all such surfaces are in fact proper. By a theorem of Nadirashvili [**45**], there exist complete immersed minimal surfaces in $\mathbb{R}^3$ which are contained inside a ball and so are not proper in $\mathbb{R}^3$ (see also [**14**]).

It is clear that every smooth properly embedded minimal surface M has *locally bounded* Gaussian curvature in the sense that its Gaussian curvature in balls is bounded. In [**37**], Meeks and Rosenberg prove that if M is a complete embedded minimal surface in $\mathbb{R}^3$ with locally bounded Gaussian curvature, then the closure of M in $\mathbb{R}^3$ has the structure of a $C^{1,\alpha}$-lamination. In the same paper, they also give a structure theorem for minimal laminations which, together with curvature estimates from [**8**] and [**30**], can be used to prove the following theorem.

THEOREM 2.1. [**30, 37**] *A complete embedded minimal surface M of finite genus in $\mathbb{R}^3$ which does not have exactly one limit end is proper if and only if it has locally bounded Gaussian curvature.*

The conformal structure on a complete minimal surface influences many related global properties. The following definition introduces a key vocabulary word for these discussions.

DEFINITION 4. A Riemannian manifold M with nonempty boundary is *parabolic* if every bounded harmonic function on M is determined by its boundary values.

DEFINITION 5. Given a Riemannian manifold M and a point $p \in \mathrm{Int}(M)$, one can define the *harmonic* or *hitting* measure μ_p of an interval $I \subset \partial M$ as the probability that a Brownian path, beginning at p, exits for the first time the boundary ∂M somewhere on the interval I.

Instead of defining the harmonic measure μ_p of $I \subset \partial M$ in terms of probability and Brownian motion, one can also define it as follows. Consider a compact exhaustion $I \subset \partial M_1 \subset M_1 \subset M_2 \subset \cdots$ of M. Let $h_n\colon M_n \to [0,1]$ be the bounded harmonic function with boundary values 1 on $\mathrm{Int}(I)$ and 0 on $\partial M_n - I$. Since h_n is an increasing sequence of harmonic functions on M bounded by the constant function 1, h_n has a unique limit harmonic function h. In this case $\mu_p(I) = h(p)$.

The following useful proposition is an elementary consequence of the definition of *harmonic* or *hitting* measure.

PROPOSITION 2.2. *Suppose M is a Riemannian manifold with nonempty boundary. The following are equivalent:*

(1) *M is parabolic;*
(2) *Bounded harmonic functions on M are determined by their boundary values;*
(3) *For some $p \in Int(M)$, the measure μ_p is full on ∂M, i.e., $\int_{\partial M} \mu_p = 1$;*
(4) *Given any $p \in Int(M)$ and any bounded harmonic function $f\colon M \to \mathbb{R}$, $f(p) = \int_{\partial M} f(x)\mu_p$;*
(5) *There exists a proper positive superharmonic function on M.*

The property of a Riemannian manifold with boundary being parabolic is closely related to the following notion of recurrence for Brownian motion.

DEFINITION 6. A Riemannian manifold M (without boundary) is *recurrent* or *recurrent for Brownian motion* if and only if for any given point $p \in M$, almost all continuous paths $\alpha\colon [0,\infty) \to M$ with $\alpha(0) = p$ are dense in M.

It is known that $\mathbb{R}^n$ is recurrent for Brownian motion if and only if $n \leq 2$.

LEMMA 2.3. *A connected Riemannian manifold M is recurrent if and only if for any nonempty open set $O \subset\neq M$, $M - O$ is parabolic.*

PROOF. Suppose $O \subset\neq M$ is a nonempty open subset. Let C be a nonempty component of $M - O$ and $p \in \mathrm{Int}(C)$. Since almost all Brownian paths beginning at p are dense in M, almost all Brownian paths beginning at p must enter O. But, in order to enter O, such a path must cross ∂C which means that the hitting measure μ_p on ∂C is full and so Proposition 2.2 implies that C is parabolic.

We now prove the converse statement. Suppose that for any nonempty open set $O \subset\neq M$, $M - O$ is parabolic. Let $p, q \in M$ and let $\alpha \subset M$ be a Brownian path starting at p. For any open ball $B(\varepsilon, q)$ centered at q, $M - B(\varepsilon, q)$ is parabolic and so, with probability 1, the path α will enter the closed ball $\overline{B}(\varepsilon, p)$. Since ε is arbitrary, the closure of α in M is all of M. □

Recent work of Meeks and Rosenberg [**37**], [**38**] proves:

THEOREM 2.4. *An $M \in \mathcal{P}$ of finite topology is conformally diffeomorphic to a finitely punctured Riemann surface. In particular, such a minimal surface is recurrent for Brownian motion.*

In order to understand further generalizations of the above theorem, we now discuss the topological notion of "ends" of a surface.

DEFINITION 7. Suppose M is a noncompact connected manifold. The space of ends of M, denoted by $\mathcal{E}(M)$, is the set of equivalence classes of proper arcs $\alpha\colon [0,\infty) \to M$ where α_1 is equivalent to α_2 if for every smooth compact subdomain $C \subset M$, α_1 and α_2 intersect the same component of $M - \mathrm{Int}(C)$ in a noncompact set.

A basis for the topology of $\mathcal{E}(M)$ is defined as follows. For each compact set $C \subset M$, define the basis open set $B(C) \subset \mathcal{E}(M)$ to be those equivalence classes of proper arcs in M which have representatives contained in $M - C$. Another way to define $\mathcal{E}(M)$, and its topology, is to consider $\mathcal{E}(M)$ to be the inverse limit of complements of compact subsets of M under the ordering of inclusion.

It can be shown that $\mathcal{E}(M)$ is a totally disconnected compact Hausdorff space that embeds as a compact subspace of the unit interval (this is not difficult to prove once one knows that it is true but I do not have a reference for it). Conversely, every compact totally disconnected subset C of the unit interval corresponds to the space of ends of a noncompact surface; namely, consider C to lie on an interval in S^2, then $S^2 - C$ has C as its space of ends. It is also interesting to note that two connected genus zero surfaces are homeomorphic if and only if their spaces of ends are homeomorphic.

DEFINITION 8. An end $e \in \mathcal{E}(M)$ is a *simple* end of M if it is an isolated point in $\mathcal{E}(M)$. Note that e is simple if and only if there exists a representative proper arc $\alpha \in e$ and a proper subdomain $W \subset M$ containing α such that W is homeomorphic to $S^1 \times [0, \infty)$ or to $S^1 \times [0, \infty)$ connected sum with an infinite number of tori where the n-th connected sum occurs at the point $(1, n) \in S^1 \times [0, \infty)$. In the first case we refer to e as an *annular* end and in the second case we refer to e as a *simple end of infinite genus.*

DEFINITION 9. An end $e \in \mathcal{E}(M)$ is a *limit* end of $\mathcal{E}(M)$ if it is not a simple end. In other words, e is a limit end if it is a limit point of $\mathcal{E}(M)$. As in the previous definition, a limit end has *genus zero* if it can be represented by a proper domain $W \subset M$ with compact boundary and the genus of W is zero. If a limit end e does not have genus zero, we say that e has infinite genus; in this case every proper subdomain with compact boundary representing e has infinite genus.

DEFINITION 10. The *limit tangent plane* at infinity of a properly embedded minimal surface M in $\mathbb{R}^3$ is the plane passing through the origin whose normal vector equals the normal vector of some end of a minimal surface $\Sigma \subset (\mathbb{R}^3 - M)$ with compact boundary and finite total curvature; see [**3**] for further details on existence and uniqueness of the limit tangent plane at infinity when $M \in \mathcal{M}$. We say that the limit tangent plane at infinity is *horizontal* if it is the x_1x_2-plane.

The following theorem by Frohman and Meeks [**17**], called the ordering theorem, is the starting point for describing the geometry and topology of minimal surfaces in $\mathcal{M}$.

THEOREM 2.5. *Suppose $M \in \mathcal{M}$ and has horizontal limit tangent plane at infinity. Then the ends of M can be linearly ordered geometrically by their relative heights over the x_1x_2-plane. Furthermore, this ordering is a topological ordering in the following sense. If M is properly isotopic to a properly embedded minimal surface M' with horizontal limit tangent plane at infinity, then the associated ordering of the ends of M' either agrees with or is opposite to the ordering coming from M.*

DEFINITION 11. Consider the ordering on the ends $\mathcal{E}(M)$. The end $e_T \in \mathcal{E}(M)$ is the *top* end of M if it is the unique end in $\mathcal{E}(M)$ that is maximal in the ordering. The end $e_B \in \mathcal{E}(M)$ is the *bottom* end of M if it is minimal in the ordering of the ends. An end of M that is neither the top nor the bottom end of M is called a *middle* end of M.

DEFINITION 12. A surface $M \subset \mathbb{R}^3$ has *quadratic area growth* if there exists a positive constant c such that for all large positive R, the area of M inside the ball $B(R)$ of radius R centered at the origin is less than cR^2.

By the proof of the ordering theorem [**17**], a middle end of a properly embedded minimal surface M can be represented by a proper subdomain $W \subset M$ with compact boundary such that W "lies between two catenoids." This means that W is contained in a neighborhood S of the x_1x_2-plane, S being topologically a slab, whose width grows logarithmically with the distance from the origin.

One of the fundamental results in the global theory of properly embedded minimal surfaces is that the middle ends of an $M \in \mathcal{M}$ are never limit ends. This is shown by first proving that if W is a properly immersed minimal surface with compact boundary and contained between two catenoids, then W has quadratic area growth [**13**]. By the monotonicity formula for area [**48**], every subend of a representative of a limit end has limiting area growth at least πR^2, and so a limit end never has a representative with quadratic area growth. Thus:

THEOREM 2.6. [**13**] *If $M \in \mathcal{M}$, then a limit end of M must be a top or bottom end of M. In particular, M can have at most two limit ends.*

COROLLARY 1. *A properly embedded minimal surface in $\mathbb{R}^3$ has a countable number of ends.*

Consider a middle end e of $M \in \mathcal{M}$. As we said above, M can be represented by a proper subdomain $W \subset M$ contained in a topological slab S whose width grows at most logarithmically. Furthermore, by Theorem 2.6, we may assume that W has one end. After removing a large round open ball B from S which contains ∂R, $(S - B) - W$ consists of two components, C_1, C_2. The end e is called *odd* if each of the components C_1, C_2 intersects one of the two boundary components ∂S; otherwise, e is called *even*. Note that middle ends which are annular are asymptotic to the end of a plane or a catenoid and so are always odd. It is shown in [**13**] that if e is odd, then area growth of W is $(2n-1)\pi R^2$; if e is even, then area growth of W is $2n\pi R^2$, where n is a positive integer.

Theorem 2.6 plays a crucial role in the proof of the following topological classification theorem of properly embedded minimal surfaces in $\mathbb{R}^3$ by Frohman and Meeks [**16, 18**].

THEOREM 2.7. *Suppose $M_1, M_2 \in \mathcal{P}$. There exists an orientation preserving diffeomorphism $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ with $f(M_1) = M_2$ if and only if there exists a homeomorphism $F\colon M_1 \to M_2$ that preserves the ordering on the ends of M_1 and M_2 and the eveness-oddness invariant of each of the middle ends. Thus, if M_1 and M_2 are homeomorphic and have at most two ends, then there exists a diffeomorphism $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ with $f(M_1) = M_2$.*

Motivated by the above result and work of Collin, Hauswirth and Rosenberg [**12**], the author conjectures that properly embedded constant mean curvature $H = 1$ surfaces M in hyperbolic three-space $\mathbb{H}^3$ are unknotted; such surfaces are called Bryant surfaces. When M_1, M_2 are homeomorphic Bryant surfaces with finite topology, the author has shown that M_1 and M_2 are properly ambiently isotopic in $\mathbb{H}^3$; the similar theorem for minimal surfaces in $\mathbb{R}^3$ was proved by Meeks and Yau [**43**].

CONJECTURE 2.1. (Meeks) [Topological Uniqueness Theorem for Bryant Surfaces] If M_1, M_2 are homeomorphic properly embedded surfaces of constant mean curvature 1 in $\mathbb{H}^3$, then M_1 and M_2 are properly ambiently isotopic.

The proof that a properly immersed minimal surface with compact boundary and contained between two catenoids must have quadratic area growth follows from an analysis using certain "universal superharmonic functions" defined on $\mathbb{R}^3$ together with the divergence theorem.

DEFINITION 13. A function $f\colon D \to \mathbb{R}$ defined on a domain $D \subset \mathbb{R}^3$ is called a *universal superharmonic* function if its restriction to any minimal surface in D is superharmonic.

Examples of universal superharmonic functions on all of $\mathbb{R}^3$ include coordinate functions such as x_1 or the function $-{x_1}^2$. In the proof of the quadratic area growth property used in the proof of Theorem 2.6 one uses the universal superharmonic function $\ln(\sqrt{x_1^2+x_2^2}) - x_3 \tan^{-1}(x_3) + \frac{1}{2}\ln(1+x_3^2)$ on a certain region of $\mathbb{R}^3$.

Recall by Proposition 2.2 that a Riemannian surface M with boundary is parabolic if and only if there exists a proper positive superharmonic function on M. We now use this defining property of parabolicity and the universal superharmonic function $\ln(\sqrt{x_1^2+x_2^2}) - x_3^2$ on a certain region of $\mathbb{R}$ to prove the following basic conformal result for proper minimal surfaces in a closed halfspace. Also see [**25**] for some results related to and motivated by the next theorem.

THEOREM 2.8. [**13**] *If M is a properly immersed minimal surface in $\mathbb{R}^3$, then $M(+) = \{(x_1,x_2,x_3) \in M \mid x_3 \geq 0\}$ is parabolic.*

PROOF. Assume that $M(+)$ is connected; the general case can be obtained by proving each component of $M(+)$ is parabolic. For each positive integer n define $M(n) = \{(x_1,x_2,x_3) \in M \mid 0 \leq x_3 \leq n\}$ and let $(M(n,*) = \{(x_1,x_2,x_3) \in M(n) \mid 1 \leq x_1^2 + x_2^2\}$. Let h_n be the restriction of the universal superharmonic function $\ln(\sqrt{x_1^2+x_2^2}) - x_3^2$ to $M(n,*)$ and note that $h_n\colon M(n,*) \to \mathbb{R}$ is proper and bounded from below. Hence, $M(n,*)$ is parabolic. Since $M(n)$ is the union of a compact surface with $M(n,*)$, $M(n)$ is also parabolic.

For n large choose a $p \in \mathrm{Int}(M(n))$; note that $x_3(p)$ is positive and after a homothety of $\mathbb{R}^3$ we may assume $x_3(p) = 1$. Let $\partial(n)$ denote the part of $\partial M(n)$ at x_3-height n and let $\partial(0)$ denote the part of ∂M_n at height zero. Since $M(n)$ is parabolic and $x_3|M(n)$ is a bounded harmonic function on $M(n)$, for the hitting measure $\mu_p(n)$ on $\partial M(n)$, we have by Proposition 2.2

$$1 = x_3(p) = \int_{\partial M(n)} x_3(x)\mu_p(n) = n \int_{\partial(n)} \mu_p(n).$$

Hence, $\int_{\partial(n)} \mu_p(n) = \frac{1}{n}$ and since $M(n)$ is parabolic, $\int_{\partial(0)} \mu_p(n) = 1 - \frac{1}{n}$. Hence, for all positive integers n,

$$\int_{\partial M(+)} \mu_p = \int_{\partial(0)} \mu_p \geq \int_{\partial(0)} \mu_p(n) = 1 - \frac{1}{n}.$$

Therefore, $\int_{\partial M(+)} \mu_p = 1$, which proves that $M(+)$ is parabolic. □

A slight modification of the proof of the above theorem shows that the theorem also holds if M is allowed to have boundary; this generalization can sometimes be useful. Also, the following corollary is an important consequence of the above theorem.

COROLLARY 2. *If M is a properly immersed minimal surface in $\mathbb{R}^3$ and M intersects some plane in a compact set, then M is recurrent for Brownian motion.*

The next theorem from [**13**] gives an interesting sufficient topological condition for a properly embedded minimal surface to be recurrent.

THEOREM 2.9. *If $M \in \mathcal{M}$ has two limit ends, then there exists a proper collection $P = \{P(n) \mid n \in \mathbb{Z}\}$ of parallel planes such that each plane in P intersects M transversely in a compact set and such that every middle end of M is represented uniquely for some $n \in \mathbb{Z}$ by the intersection of M with the slab $S(n)$ between plane $P(n)$ and plane $P(n+1)$. In particular, M is recurrent for Brownian motion.*

We now describe a useful analytic property, called the Liouville property, that implies the constancy of certain harmonic functions. Since the complex plane is recurrent, the next theorem implies the usual Liouville Theorem.

THEOREM 2.10. *(Liouville Theorem) If M is a recurrent Riemannian surface and $h\colon M \to \mathbb{R}^+$ is a positive harmonic function, then h is constant.*

PROOF. Suppose h has at least two different values $a, b \in \mathbb{R}^+, a < b$. Choose a regular value c, $a < c < b$, for h. Then $W = h^{-1}((0,c]) = M - h^{-1}((c,\infty))$ is parabolic by Lemma 2.3. Thus, $h|W$ is a bounded harmonic function on W with constant boundary value c. Since W is a parabolic Riemannian manifold, Proposition 2.2 implies $h|W$ is constant which implies h is constant. This contradiction proves the Liouville Theorem. □

We now give the proof that the complex plane is recurrent for Brownian motion.

THEOREM 2.11. *The complex plane $\mathbb{C}$ is recurrent for Brownian motion. In particular, positive harmonic functions on $\mathbb{C}$ are constant.*

PROOF. It suffices to prove that almost all Brownian paths in $\mathbb{C}$ beginning at the point $2 \in \mathbb{C}$ have the point $0 \in \mathbb{C}$ in their closure. Clearly to prove this it suffices to show that for any small $\varepsilon > 0$, in the region $A(\varepsilon) = \{z \mid |z| \geq \varepsilon\}$, $\int_{\partial A(\varepsilon) = S^1(\varepsilon)} \mu_2(\varepsilon) = 1$, where $\mu_2(\varepsilon)$ is the hitting measure on $A(\varepsilon)$ based at the point 2; equivalently, $A(\varepsilon)$ is parabolic.

Consider the domain $A(\varepsilon, n) = \{z \mid \varepsilon \leq |z| \leq n\}$ and the harmonic function $\ln|z|$ on $A(\varepsilon, n)$. By part 4 of Proposition 2.2, we can evaluate $\ln(2)$ in terms of the hitting measure $\mu_2(\varepsilon, n)$ on $\partial A(\varepsilon, n)$:

$$\ln(2) = \int_{\partial A(\varepsilon,n)} \ln(x)\mu_2(\varepsilon,n) = \ln(\varepsilon)\int_{S^1(\varepsilon)} \mu_2(\varepsilon,n) + \ln(n)\int_{S^1(n)} \mu_2(\varepsilon,n).$$

Therefore, $\ln(\varepsilon)\int_{S^1(\varepsilon)} \mu_2(\varepsilon,n) = \ln(2) - \ln(n)\int_{S^1(n)} \mu_2(\varepsilon,n))$, which implies that $\lim_{n\to\infty} \int_{S^1(n)} \mu_2(\varepsilon,n) = 0$ and so $\lim_{n\to\infty} \int_{S^1(\varepsilon)} \mu_2(\varepsilon,n) = 1$. Since

$$\int_{S^1(\varepsilon)} \mu_2(\varepsilon) \geq \int_{S_1(\varepsilon)} \mu_2(\varepsilon,n),$$

we have $\int_{S^1(\varepsilon)} \mu_2(\varepsilon) = 1$, which proves the theorem. □

Meeks and Rosenberg have conjectured that if h is a coordinate function on a properly embedded minimal surface M in $\mathbb{R}^3$, then, except for a countable number of integral curves, an integral curve $\alpha\colon \mathbb{R} \to M$ of ∇h satisfies $h \circ \alpha\colon \mathbb{R} \to \mathbb{R}$ is a diffeomorphism. A particular consequence of this conjecture is that the flux of a

coordinate function h of M is an invariant of h where we define the $F(\nabla h, h^{-1}(t)) = \int_{\partial(h^{-1}(-\infty,t])} \nabla h \cdot \nu = \int_{\partial(h^{-1}(-\infty,t])} |\nabla h|$, the invariance being that this function is independent of t. We will now prove a basic result on the invariance of flux.

Given a Riemannian manifold N with boundary, $\alpha \subset \partial N$ a portion of the boundary and X a smooth vector field on N, then the flux of X across α is defined to be the indefinite integral $F(X, \alpha) = \int_\alpha X \cdot \nu$ where ν is the outward pointing vector with normal to N along α. In the case X is the gradient of a function on N and α is contained in a level set of this function, then $X \cdot \nu$ is pointwise always nonpositive or nonnegative and hence, $F(X, \alpha) = \pm \int_\alpha |X|$ takes on a well-defined value in $[-\infty, \infty]$ even when α is noncompact. When we think of α as an oriented curve, then we always consider in the level set case that the flux is positive.

If N is compact and u and v are harmonic functions on N, then the double Green's formula implies that

$$F(u\nabla v, \partial N) = F(v\nabla u, \partial N).$$

We will use this double Green's formula in the proof of the following theorem.

THEOREM 2.12. *(Algebraic Flux Lemma) Suppose M is a parabolic Riemannian manifold with boundary and $h\colon M \to [0,1]$ is a nonconstant harmonic function with ∂M consisting of $\partial_0 = h^{-1}(0)$ and $\partial_1 = h^{-1}(1)$. If for some $t_0 \in [0,1]$, $F(\nabla h, h^{-1}(t_0))$ is finite, then the flux $F(\nabla h, h^{-1}(t))$ is a finite positive number independent of t.*

PROOF. Without loss of generality, we may assume that $t_0 = 1$ and $t = 0$. Let $M(1) \subset M(2) \subset \cdots$ be a smooth compact exhaustion of M where $M(k)$ contains the points of M of distance at most k from some fixed point of M. The boundary of $M(k)$ separates into three pieces: $\partial_0(k) = \partial M(k) \cap \partial_0, \partial_1(k) = \partial M(k) \cap \partial_1$, and $\partial_*(k) = \partial M(k) - [\partial_0(k) \cup \partial_1(k)]$. Let $u(k)$ be the harmonic function on $M(k)$ with boundary values 1 on $\partial_1(k)$ and 0 on $\partial M(k) - \partial_1(k)$. Since the $u(k)$ are bounded by the constant 1 on compact subsets of M and increasing, the $u(k)$ converge to a bounded harmonic function which has the same boundary values as h. By assumption, M is parabolic and so on compact subsets of M, $u(k) \to h$ as $k \to \infty$.

Now let $v(k) = h|M(k)$. By the double Green's formula above, we have $F(\nabla h, \partial_1(k)) = F(\nabla u(k), \partial_1(k)) + F(h\nabla u(k), \partial_*(k))$. This formula together with the divergence theorem applied to $u(k)$ yields

$$F(\nabla h, \partial_1(k)) + F(\nabla u(k), \partial_0(k)) = F((h-1)\nabla u(k), \partial_*(k)).$$

Since $h-1$ is nonpositive and $\nabla u(k) \cdot \nu \le 0$ on $\partial_*(k)$, we have $F((h-1)\nabla u(k), \partial_*(k))$ nonnegative. Therefore

$$|F(\nabla h, \partial_1(k))| \ge |F(\nabla u(k), \partial_0(k))|,$$

and since $|F(\nabla h, \partial_1(k))| \le |F(\nabla h, \partial_1)|$, which is finite by assumption, there is a uniform bound for the $|F(\nabla u(k), \partial_0(k))|$. Since $\nabla u(k)$ converges uniformly to ∇h on compact subsets of M,

$$\lim_{k\to\infty} F(\nabla u(k), \partial_0(k)) = F(\nabla h, \partial_0)$$

is finite. Thus, taking limits of the inequalities $|F(\nabla u(k), \partial_0(k))| \le |F(\nabla h, \partial_1(k))|$, we obtain $|F(\nabla h, \partial_0)| \le |F(\nabla h, \partial_1)|$.

Now consider the function $\tilde{h} = 1 - h$. Note that $\tilde{\partial}_0 = \tilde{h}^{-1}(0) = h^{-1}(1) = \partial_1, \tilde{\partial}_1 = \tilde{h}^{-1}(1) = h^{-1}(0)$ and $|F(\nabla \tilde{h}, \tilde{\partial}_1)| = |F(\nabla h, \partial_0)|$ is finite. By our previous

calculations with $\tilde{h}$ replacing h, we get $|F(\nabla\tilde{h}, \tilde{\partial}_0)| \leq |F(\nabla\tilde{h}, \tilde{\partial}_1|$ which implies the inequality $|F(\nabla h, \partial_1)| \leq |F(\nabla h, \partial_0)|$. Hence, $F(\nabla h, \partial_0) = -F(\nabla h, \partial_1)$ which completes the proof of the theorem. □

The following corollaries are immediate consequences of Theorems 2.1, 2.8, 2.9 and 2.12.

COROLLARY 3. *If M is a properly immersed minimal surface in $\mathbb{R}^3$, then there is a well-defined vector $F(M) \in (\mathbb{R}^+ \cup \infty)^3$ with k-th coordinate equal to the flux of the k-th coordinate of M.*

COROLLARY 4. *If $M \in \mathcal{P}$ has two limit ends and horizontal limit tangent plane at infinity, then the flux $x_3(F(M))$ is finite.*

It seems reasonable to expect that if M is a parabolic Riemannian manifold with boundary and $h\colon M \to [0,1]$ is a nonconstant harmonic function with $\partial M = h^{-1}(\{0,1\})$, then almost-all integral curves of ∇h begin at points of $h^{-1}(0)$ and end at points of $h^{-1}(1)$. However, as pointed out to me by A. Eremenko, a theorem of Stephenson [**50**] implies that one can construct such a harmonic function $h\colon M \to [0,1]$ where M is the closed unit disk with a closed subset of measure zero removed from its boundary, such that an open set of integral curves of ∇h neither begins on $h^{-1}(0)$ nor ends on $h^{-1}(1)$. In fact it is not difficult to construct a function $h\colon D \to \mathbb{R}$ with the property that for any two regular values, $t_1 < t_2, h^{-1}([t_1, t_2])$ is parabolic and there exists an open set of integral curves of ∇h which neither begin at $-\infty$ nor end at $+\infty$. Also, A. Eremenko informed me of the following related positive result: If M is a Riemann surface which is recurrent for Brownian motion, then for every nonconstant holomorphic function $u + iv$, the harmonic function u tends to infinity along almost every gradient line of u. This result follows from Theorem X.14 on page 439 of [**53**].

3. Mostly New Conjectures

In this section the author will present about twenty-five fundamental conjectures in the classical theory of minimal surfaces. For the most part these conjectures are motivated by the author's own research and are not widely known except to classical minimal surface specialists. Hopefully, the presentation of these problems and suggestions for a plan of attack on solving them will speed up their solution and stimulate further interest in this beautiful subject. In the statement of each conjecture the author has included a suggested expected time frame for a solution; only time will tell how accurate this time frame is. The author has listed in the statement of each conjecture the principal researchers to whom the conjecture might be attributed. These conjectures are listed in order according to the author's interest in them or by his personal feeling of their general importance or deepness. For these discussions we again let $\mathcal{P}$ denote the set of all properly embedded connected minimal surfaces in $\mathbb{R}^3$ and let $\mathcal{M} \subset \mathcal{P}$ denote the subset with more than one end. We say that $f\colon M \to \mathbb{R}^3$ is a complete injective minimal immersion if M is a surface and f is a one-to-one minimal immersion of M to $\mathbb{R}^3$ which is complete with respect to the pulled back metric.

CONJECTURE 3.1. [Properness Conjecture (Calabi, Meeks); time frame 100 years] If $f\colon M \to \mathbb{R}^3$ is a complete injective minimal immersion, then $M \in \mathcal{P}$.

The author has an outline for a possible proof of properness in the finite topology case. The author, Pérez and Ros have conjectured [**30**] that if a complete embedded minimal $M \subset \mathbb{R}^3$ has finite genus, then M has bounded Gaussian curvature (see also, [**37**]). It follows from work in [**30**] and [**37**] that if such an M has locally bounded Gaussian curvature and finite genus, then M is properly embedded. The suggested time frame for a solution to the finite genus case of the above conjecture is three years. The author's suggestion for an attack on the finite genus case is to employ a blowup argument in conjunction with the imminent proof of the next conjecture.

CONJECTURE 3.2. [Genus-zero Conjecture (Meeks–Pérez–Ros); time frame 2 years] If $M \in \mathcal{P}$ has genus 0, then M is a plane, a catenoid, a helicoid or a Riemann example.

There is a related conjecture of Meeks–Pérez–Ros, which we call the Asymptotic Finite Genus Conjecture, that deals with the expected asymptotic geometry of the ends of an $M \in \mathcal{P}$ with finite genus. From the work of Meeks–Rosenberg [**37**], it is known that each simple end of such an M is asymptotic to the end of a plane, a catenoid or a helicoid. The related conjecture states that if M has an infinite number of ends and finite genus, then M has two limit ends, each asymptotic to a Riemann-example-type limit end. What this means is that, after a rotation of M, for each limit end e of M there is a representative $E \subset M$ with compact boundary, horizontal planar ends and one limit end such that E converges exponentially quickly to a limit end of a fixed Riemann example as a function of the third coordinate function. The solution of this conjecture would complete the desired description of the asymptotic geometry of finite genus surfaces in $\mathcal{M}$. The suggested time frame for solving this related conjecture is also 2 years. For related partial results and related theorems, see [**30**], [**31**], [**32**], [**33**] and [**34**].

CONJECTURE 3.3. [Flux Conjecture (Meeks–Rosenberg); time frame 80 years] Suppose $M \in \mathcal{P}$ and $h\colon M \to \mathbb{R}$ is a nonconstant coordinate function on M. Consider the set I of integral curves of ∇h. Then there exists a countable set $C \subset I$ such that for any integral curve $\alpha \in I - C$, $h|\alpha\colon \alpha \to \mathbb{R}$ is a diffeomorphism. Here we consider $\alpha\colon \mathbb{R} \to M$ to be, after a choice of $p \in \alpha$, a curve $\alpha(t)$ with $\alpha(0) = p$ and $\alpha'(t) = \nabla h(\alpha(t))$.

By recent work on finite genus minimal surfaces [**11, 30, 31, 37**] the Flux Conjecture holds for $M \in \mathcal{P}$ of finite genus. One could weaken the hypothesis in the above conjecture that "except for a countable number of integral curves, h restricted to an integral curve of ∇h is a diffeomorphism with $\mathbb{R}$" to the hypothesis that "for almost-all integral curves of ∇h, h restricted to an integral curve is a diffeomorphism with $\mathbb{R}$". This weaker version of the Flux Conjecture is, for technical reasons, more likely to be solved with a suggested time frame of only five years for its solution. This weaker conjecture, via Stokes theorem, has as a consequence the following algebraic version of the Flux Conjecture: Given a coordinate function $h\colon M \to \mathbb{R}$ on an $M \in \mathcal{P}$, the flux of ∇h across a level set of h is independent of the level set. This algebraic flux lemma, which was proved in Section 2 as Corollary 3, implies that the flux of a fixed coordinate function of an $M \in \mathcal{P}$ is an invariant of M. The author feels that this result will likely have important theoretical consequence.

Meeks and Wolf have proven the following Conjecture under the additional hypothesis that the boundary of M is compact and M has dimension two.

CONJECTURE 3.4. [Geometric Flux Conjecture (Meeks); time frame 5 years] If M is a parabolic manifold with boundary and $h\colon M \to [0,1]$ is a nonconstant harmonic function with $\partial M = h^{-1}(\{0,1\})$, and the flux of ∇h across $h^{-1}(0)$ is finite, then almost all integral curves of ∇h are compact arcs with end points in ∂M.

CONJECTURE 3.5. [Liouville Conjecture (Meeks, Sullivan); time frame 50 years] If $M \in \mathcal{P}$ and $h\colon M \to \mathbb{R}$ is a positive harmonic function, then h is constant.

This conjecture is closely related to work in [**13**], [**35**]. For example, from the discussion in the previous section we know that if $M \in \mathcal{P}$ has finite topology, then M is recurrent which implies M satisfies the Liouville Conjecture. Also, from the previous section, we know that the above conjecture holds if M is recurrent for Brownian motion and so is true if $M \in \mathcal{M}$ has two limit ends. It follows from Theorem 2.4 and the results in [**31**] that the Liouville Conjecture holds for $M \in \mathcal{P}$ of finite genus; in fact, these surfaces are recurrent for Brownian motion. The author also conjectures:

CONJECTURE 3.6. [Multiple-End Recurrency Conjecture (Meeks); time frame 5 years] If $M \in \mathcal{M}$, then M is recurrent for Brownian motion.

Assuming that one can prove the above conjecture, one would reduce the proof of the Liouville Conjecture to the case where M has infinite genus and one end. Although every triply-periodic minimal surface satisfies the Liouville Conjecture, triply-periodic minimal surfaces are never recurrent. Thus, in the above conjecture the assumption that $M \in \mathcal{M}$, not $M \in \mathcal{P}$, is a necessary one. It should be noted that the previous two conjectures need the hypothesis of embeddedness since there exist properly immersed minimal surfaces with two ends which admit bounded nonconstant harmonic functions [**13**].

CONJECTURE 3.7. [Finite Topology Conjecture (Hoffman and Meeks); time frame 100 years] An orientable surface M of finite topology with genus g and k ends, $k \neq 2$, occurs in $\mathcal{P}$ if and only if $g \geq k-2$.

A recent theorem of Traizet [**52**] gives many examples of properly embedded minimal surfaces of finite topology which satisfy the above inequality. A recent paper of Weber and Wolf [**54**] states that if M is an orientable surface with genus g and k ends, $k \geq 3$, then M properly minimally immerses in $\mathbb{R}^3$ with planar horizontal ends if $g \geq k-2$. These examples of Weber and Wolf are almost certainly embedded, however, they can only prove that these surfaces are embedded outside of a compact set. There is experimental computer evidence that every orientable surface with finite genus and one end properly minimally embeds in $\mathbb{R}^3$ as a minimal surface of finite type (see [**1**], [**2**]). Recent work of Meeks, Pérez and Ros [**32**] shows that for each positive genus g, there exists an upper bound $k(g)$ on the number of ends of an $M \in \mathcal{M}$ with finite topology and genus g. Results of Collin [**11**] and Schoen [**47**] imply that the only examples in $\mathcal{M}$ with finite topology and two ends are catenoids. Results of Collin [**11**] and López–Ros [**26**] imply that if M has finite topology, genus zero and at least two ends, then M is a catenoid.

The Finite Topology Conjecture of Hoffman and Meeks proposes the precise topological conditions under which a noncompact orientable surface of finite topology would properly minimally embed in $\mathbb{R}^3$. What about the case where the noncompact orientable surface M has infinite topology (either infinite genus or an

infinite number of ends)? By Theorem 2.6, such an M can have at most one or two limit ends. Meeks, Pérez and Ros [**32**] prove that such an M cannot have one limit end and finite genus. The following conjecture is nothing more than the claim that these restrictions are the only ones.

CONJECTURE 3.8. [Infinite Topology Conjecture (Meeks); time frame 25 years] A noncompact orientable surface of infinite topology occurs in $\mathcal{P}$ if and only if it has at most one or two limit ends and in the case it has one limit end it also has infinite genus.

CONJECTURE 3.9. [One-ended Conjecture (Meeks and Rosenberg); time frame 20 years] For every nonnegative integer g, there exists a unique nonplanar $M \in \mathcal{P}$ with genus g and one end.

There are some partial results on the above conjecture. Meeks and Rosenberg have shown that every example $M \in \mathcal{P}$ of finite genus and one end has a special analytic representation which makes M into a "surface of finite type" (see [**37**]). In the case of genus zero, Meeks and Rosenberg proved that the plane and the helicoid are the only genus zero examples. Based on an earlier computational proof of the existence of a helicoid with a handle by Hoffman, Karcher and Wei [**20**], Hoffman, Weber and Wolf [**21**] have given a rigorous mathematical proof of the existence of the genus-one helicoid. Work in progress by Martin and Weber [**27**] indicates that this genus-one helicoid is unique. Other computational work in [**1**], [**2**] and [**51**] indicate the conjecture is true for genus 2, 3, 4 and 5.

The next conjecture appears in [**29**]. Although there are a number of partial results, the conjecture still remains unsolved.

CONJECTURE 3.10. [Convex Curve Conjecture (Meeks); time frame 30 years] Two convex Jordan curves in parallel planes cannot bound a compact minimal surface of positive genus.

In fact in [**29**], Meeks conjectures that if $D = \{D_1, \ldots, D_n\}$ is a finite collection of convex pairwise disjoint disks in the x_1x_2-plane and Γ is a convex Jordan curve in a parallel plane, then $\{\Gamma, \partial D_1, \ldots, \partial D_n\}$ does not bound a compact minimal surface of positive genus. The suggested time frame for solving this more general conjecture is 50 years.

There are some partial results on the Convex Curve Conjecture under the assumption of some symmetry on the curves (see [**41**], [**46**], [**47**]). Also, the results in [**41**] and [**42**] indicate that the Convex Curve Conjecture probably holds in the more general case where the two convex planar curves do not necessarily lie in parallel planes but rather lie on the boundary of their convex hull; in this case the planar Jordan curves are called extremal. Recent results by Ekholm, White and Wienholtz [**15**] show that every compact orientable minimal surface that arises as a counterexample to the convex curve conjecture is embedded and that for a fixed pair of extremal convex planar curves there is a bound on the genus of such a minimal surface.

CONJECTURE 3.11. [Isometry Conjecture 1 (Meeks); time frame 20 years] If $M \in \mathcal{P}$, then intrinsic isometries of M extend to ambient isometries of $\mathbb{R}^3$.

This Isometry Conjecture 1 is known if $M \in \mathcal{P}$ has more than one end (see [**4**]). Results of Meeks and Rosenberg [**37**] and [**39**] imply that the isometry conjecture

can only fail if M has one end and infinite genus. One way to prove the conjecture would be to prove that if $M \in \mathcal{P}$ has one end and infinite genus, then there exists a plane in $\mathbb{R}^3$ that intersects M in a set that contains a simple closed curve.

A similar argument might also give a proof of the following related conjecture.

CONJECTURE 3.12. [Isometry Conjecture 2 (Meeks); time frame 20 years] If $M \in \mathcal{P}$ and M is nonsimply connected, then this surface is the unique isometric minimal immersion of M into $\mathbb{R}^3$ up to ambient isometry.

CONJECTURE 3.13. [Connected Graph Conjecture (Meeks); time frame 8 years] A minimal graph with zero boundary values over a proper, possibly disconnected, domain in $\mathbb{R}^n$ can have at most two nonplanar components. If the graph has sublinear growth, then such a graph is connected.

This conjecture came out of work by Meeks and Rosenberg on the uniqueness of the helicoid (see [**37**]). They proved, under the additional hypothesis of gradient estimates, that such a graph can only have a finite number of nonplanar components. Spruck [**49**] has given some related results and Li and Wang [**24**] have recently proven finiteness of the number of nonflat components without assuming gradient estimates.

CONJECTURE 3.14. [Scherk Uniqueness Conjecture (Meeks); time frame 5 years] If $M \in \mathcal{P}$ and, in balls $B(R)$ of radius R, $\text{Area}(M \cap B(R)) < 2\pi R^2$, then M is a Scherk singly-periodic minimal surface, a catenoid or a plane.

A related conjecture on the uniqueness of Scherk's doubly-periodic minimal surfaces was recently solved by Lazard-Holly and Meeks [**23**]; they proved that if $M \in \mathcal{P}$ is doubly-periodic and the quotient surface has genus zero, then M is one of Scherk's doubly-periodic minimal surfaces. One approach for solving the general conjecture is first to prove the following conjecture on the uniqueness of the limit tangent cone of M, from which it follows by an unpublished work of Meeks and Ros that M has two Alexandrov-type planes of symmetry. From these planes of symmetry one can describe the Weierstrass representation of M, which hopefully would be useful in completing the proof of the conjecture. Much of the interest in the previous conjecture arises from the question of the role that Scherk surfaces play in desingularizing two intersecting minimal surfaces (see Kapoleous [**22**]).

CONJECTURE 3.15. [Unique Limit Tangent Cone Conjecture (Meeks); time frame 4 years] If $M \in \mathcal{P}$ has quadratic area growth, then $\lim_{t\to\infty} \frac{1}{t}M$ exists and is a cone. Furthermore, if M has area growth $2\pi R^2$ in balls of radius R centered at the origin, then the limit tangent cone is the union of two planes or one plane of multiplicity two passing through the origin.

CONJECTURE 3.16. [Quadratic Area Growth Conjecture (Meeks); time frame 9 years] $M \in \mathcal{P}$ has quadratic area growth if and only if there exist a double cone C (modeled on $x_3^2 = \lambda(x_1^2 + x_2^2)$) that intersects M in a compact set.

It follows from computations in [**13**] that $M \in \mathcal{P}$ has quadratic area growth if M intersects the union of the negative end of a catenoid and a positive vertical cone in a compact set. If the conclusion of the previous unique limit tangent cone conjecture holds for an $M \in \mathcal{P}$ with quadratic area growth, then for such an M there exists a double cone that intersects M in a compact set. Hence, the validity of the unique limit tangent cone conjecture would give one of the implications in the quadratic area growth conjecture.

CONJECTURE 3.17. [Maximum Principle at Infinity Conjecture (Meeks and Rosenberg); time frame 1 year] If M_1, M_2 are two disjoint properly immersed minimal surfaces with boundary, then

$$d(M_1, M_2) = \min\{d(M_1, \partial M_2), d(M_2, \partial M_1)\}.$$

The maximum principle at infinity has been proved in many special cases, usually with important applications (see for example, [**34**], [**35**] and [**38**]). The above conjecture represents the most general result of its type. A deep consequence of this maximum principle at infinity is that if $M \in \mathcal{P}$ has absolute Gaussian curvature less than or equal to one, then the open 1-neighborhood $N_1(M)$ of the normal bundle to M embeds in $\mathbb{R}^3$ under the exponential map. From this result it follows that such an M has cubical area growth. Meeks and Rosenberg believe they have found a proof of Conjecture 3.17.

Meeks and Rosenberg have reduced the proof of the maximum principle at infinity to the following related conjecture.

CONJECTURE 3.18. [Stable Minimal Surface Conjecture (Meeks and Rosenberg); time frame 1 year] If M is a complete, orientable, stable minimal surface with boundary, then, for any $\varepsilon > 0$, $M(\varepsilon) = \{p \in M \mid d(p, \partial M) \geq \varepsilon\}$ is a parabolic surface.

Hopefully, this basic stability conjecture will be proven shortly.

CONJECTURE 3.19. [Standard Middle End Conjecture (Meeks); time frame 20 years] If $M \in \mathcal{M}$ and $E \subset M$ is a simple end representative for a middle end of M, then E is C^0-asymptotic to the end of a plane or catenoid. In particular, if M has two limit ends, then each middle end is C^0-asymptotic to a plane.

By the work in [**13**], $\lim_{t\to\infty} \frac{1}{t}E$ is a plane P passing through the origin with positive integer multiplicity. Also, if M has two limit ends, then, for such a middle end representative E, a sequence of translates of E has a positive finite number of parallel planes for a limit; in this case, it remains to prove there is only one plane in this limit.

Gulliver and Lawson [**19**] proved that if Σ is a stable orientable minimal surface with compact boundary that is properly embedded in the punctured unit ball in $\mathbb{R}^3$, then its closure is an embedded surface. If Σ is not stable, then the corresponding result is not known.

CONJECTURE 3.20. [Isolated Singularities Conjecture (Gulliver and Lawson); time frame 8 years] The closure of a properly embedded minimal surface in the punctured ball $B - \{(0,0,0)\}$ is a surface.

CONJECTURE 3.21. [Cone Recurrence Conjecture (Meeks); time frame 10 years] If $M \in \mathcal{P}$ intersects a cone in a compact set, then M is recurrent for Brownian motion.

This conjecture is related to the following.

CONJECTURE 3.22. [Disjoint Cone Conjecture (Meeks); time frame 10 years] If $M \in \mathcal{M}$ does not have two limit ends, then M is disjoint from a cone.

CONJECTURE 3.23. [Finite Intersection Implies Recurrence Conjecture (Meeks); time frame 10 years] If $M \in \mathcal{M}$ intersects some plane transversely in a finite number of connected curves, then M is recurrent for Brownian motion.

This was proven by Meeks and Rosenberg [**35**] in the case where M is a properly immersed minimal surface with finite topology.

In [**10**] and [**7**], Colding and Minicozzi derive curvature estimates, compactness and regularity results for a sequence of properly embedded minimal disks $D(i)$ inside the open unit ball B. They prove that a subsequence of the $D(i)$ converges to a $C^{1,\alpha}$-minimal lamination $\mathcal{L}$ and that the convergence is smooth outside of a proper sequence of Lipschitz curves $S(\mathcal{L})$; nearby the singular set $S(\mathcal{L})$ the curvature of the limiting subsequence of disks blows up. Actually their theorem applies to sequences of embedded minimal surfaces in any Riemannian three-manifold which are "uniformly locally simply connected".

CONJECTURE 3.24. [Singular Curve Conjecture (Meeks and Weber); time frame 3 years] Any embedded open co-compact analytic arc α in $\mathbb{R}^3$ is the singular set $S(\mathcal{L})$ of some Colding–Minicozzi limit lamination $\mathcal{L}$ in a neighborhood of α.

The author [**28**] has proven that $S(\mathcal{L})$ consists of $C^{1,1}$-curves orthogonal to the leaves of $\mathcal{L}$ which means that they are C^1 and the unit tangent field is a Lipschitz vector field in B. This is shown by actually proving that $S(\mathcal{L})$ consists of a locally finite collection of integral curves of the unit normal vector field to $\mathcal{L}$ which is a Lipschitz vector field in B. An important consequence of the work in [**28**] is that the metric spaces $D(i)$ converge in a natural sense to a connected metric space structure on $\mathcal{L}$; this result is called the Lamination Metric Theorem and is useful in applications. Recently Meeks and Weber [**40**] have been able to construct a sequence of "bent helicoids" in $\mathbb{R}^3$ that have a limit foliation $\mathcal{L}$ with $S(\mathcal{L})$ the unit circle in the x_1x_2-plane; the above conjecture is motivated by their construction.

Recently Meeks, Pérez and Ros [**31**] proved that if $M \in \mathcal{P}$ has finite genus, then M is recurrent for Brownian motion. S. Morales [**44**] has shown that the embeddedness assumption is a necessary hypothesis in the above conjecture, even in the simply-connected case, by proving the existence of a proper conformal minimal immersion of the open disk into $\mathbb{R}^3$.

One of the outstanding problems in three-manifold theory is to prove that if $G \subset \mathrm{Diff}(S^3)$ is a finite group acting freely on the three-sphere S^3, then G is conjugate to a subgroup of the orthogonal group $O(4) \subset \mathrm{Diff}(S^3)$. If $G \subset \mathrm{Diff}(S^3)$ is a finite group which acts nonfreely on S^3, then it is known that G is conjugate to a subgroup of the orthogonal group $O(4) \subset \mathrm{Diff}(S^3)$. A positive solution of the following variant of a conjecture of Pitts–Rubinstein would solve the just described "Spherical Space Form Problem". The Pitts–Rubinstein conjecture motivates the study of properly embedded minimal surfaces in $\mathbb{R}^3$ with finite genus for the following reason: understanding the geometry of examples in $\mathcal{P}$ gives, by blowup arguments, a good understanding of the geometry of compact embedded minimal surfaces in Riemannian 3-manifolds near points of large Gaussian curvature.

CONJECTURE 3.25. [Pitts–Rubinstein Conjecture; time frame 20 years] For every compact three-manifold M^3 with finite fundamental group, there exists a Riemannian metric $\langle\,,\rangle$ which is generic in the sense that on the universal cover $(\tilde{M}^3, \widetilde{\langle\,,\rangle})$ every compact embedded minimal surface in M^3 has no Jacobi vector fields and such that for any fixed genus g there is a bound on the stability index of embedded compact minimal surfaces of genus g in $(\tilde{M}^3, \widetilde{\langle\,,\rangle})$.

The author has made the following related conjecture to the Pitts–Rubinstein Conjecture that would give necessary and sufficient conditions for a given generic Riemannian three-manifold to satisfy the conclusion of their conjecture.

CONJECTURE 3.26. [Finite Genus Finite Index Conjecture (Meeks); time frame 5 years] A compact Riemannian three-manifold M^3 with generic metric satisfies the conclusions of the Pitts–Rubinstein Conjecture if and only if there are no stable embedded minimal two-spheres in M^3.

References

[1] A. I. Bobenko. Helicoids with handles and Baker-Akhiezer spinors. *Math. Z.*, (1):9–29, 1998.

[2] A. I. Bobenko and M. Schmies. Computer graphics experiments for helicoids with handles. Personal communication.

[3] M. Callahan, D. Hoffman, and W. H. Meeks III. The structure of singly-periodic minimal surfaces. *Invent. Math.*, 99:455–481, 1990.

[4] T. Choi, W. H. Meeks III, and B. White. A rigidity theorem for properly embedded minimal surfaces in $\mathbb{R}^3$. *J. of Differential Geometry*, 32:65–76, 1990.

[5] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold I; estimates off the axis for disks. Preprint.

[6] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold II; multi-valued graphs in disks. Preprint.

[7] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold III; planar domains. Preprint.

[8] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold IV; locally simply-connected. Preprint.

[9] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold V; fixed genus. Preprint.

[10] T. H. Colding and W.P. Minicozzi II. Convergence of embedded minimal surfaces without area bounds in three-manifolds. *C. R. Acad. Sci. Paris Sér. I Math.*, 327(8):765–770, 1998.

[11] P. Collin. Topologie et courbure des surfaces minimales de $\mathbb{R}^3$. *Annals of Math.*, 145(1):1–31, 1997.

[12] P. Collin, L. Hauswirth, and H. Rosenberg. The geometry of finite topology surfaces properly embedded in hyperbolic space with constant mean curvature one. *Annals of Math.*, 623–659, 2001.

[13] P. Collin, R. Kusner, W. H. Meeks III, and H. Rosenberg. The geometry, conformal structure and topology of minimal surfaces with infinite topology. Preprint.

[14] P. Collin and H. Rosenberg. Notes sur la démonstration de N. Nadirashvili des conjectures de Hadamard et Calabi-Yau. *Bull. Sci. Math.*, 123(7):563–575, 1999.

[15] T. Ekholm, B. White, and D. Wienholtz. Embeddedness of minimal surfaces with total curvature at most 4π. *Annals of Math.*, 155:209–234, 2002.

[16] C. Frohman and W. H. Meeks III. The topological classification of minimal surfaces in $\mathbb{R}^3$. Preprint.

[17] C. Frohman and W. H. Meeks III. The ordering theorem for the ends of properly embedded minimal surfaces. *Topology*, 36(3):605–617, 1997.

[18] C. Frohman and W. H. Meeks III. The topological uniqueness of complete one-ended minimal surfaces and Heegaard surfaces in $\mathbb{R}^3$. *J. of the Amer. Math. Soc.*, 10(3):495–512, 1997.

[19] R. Gulliver and H. B. Lawson. The structure of minimal hypersurfaces near a singularity. *Proc. Symp. Pure Math.*, 44:213–237, 1986.

[20] D. Hoffman, H. Karcher, and F. Wei. The genus one helicoid and the minimal surfaces that led to its discovery. In *Global Analysis and Modern Mathematics*. Publish or Perish Press, 1993. K. Uhlenbeck, editor, 119–170.

[21] D. Hoffman, M. Weber, and M. Wolf. The existence of the genus-one helicoid. Preprint.

[22] N. Kapouleas. Complete embedded minimal surfaces of finite total curvature. *J. Differential Geom.*, 47(1):95–169, 1997.

[23] H. Lazard-Holly and W. H. Meeks III. The classification of embedded doubly–periodic minimal surfaces of genus zero. *Invent. Math.*, 143:1–27, 2001.

[24] P. Li and J. Wang. Finiteness of disjoint minimal graphs. Preprint.

[25] F. J. Lopez and J. Pérez. Parabolicity and Gauss map of minimal surfaces. *Indiana J. of Math.* (to appear).
[26] F. J. Lopez and A. Ros. On embedded complete minimal surfaces of genus zero. *J. of Differential Geometry*, 33(1):293–300, 1991.
[27] F. Martin and M. Weber. The uniqueness of the genus-one helicoid. Work in progress.
[28] W. H. Meeks III. The regularity of the singular set in the Colding and Minicozzi lamination theorem. To appear in Duke Journal of Math.
[29] W. H. Meeks III. *Lectures on Plateau's Problem.* Instituto de Matematica Pura e Aplicada (IMPA), Rio de Janeiro, Brazil, 1978.
[30] W. H. Meeks III, J. Pérez, and A. Ros. The geometry of minimal surfaces of finite genus I; curvature estimates and quasiperiodicity. Preprint.
[31] W. H. Meeks III, J. Pérez, and A. Ros. The geometry of minimal surfaces of finite genus II; nonexistence of one limit end examples. Preprint.
[32] W. H. Meeks III, J. Pérez, and A. Ros. The geometry of minimal surfaces of finite genus III; bounds on the topology and index of classical minimal surfaces. Preprint.
[33] W. H. Meeks III, J. Pérez, and A. Ros. The geometry of minimal surfaces of finite genus IV; Jacobi fields and uniqueness for small flux. Work in progress.
[34] W. H. Meeks III, J. Pérez, and A. Ros. Uniqueness of the Riemann minimal examples. *Invent. Math.*, 131:107–132, 1998.
[35] W. H. Meeks III and H. Rosenberg. Maximum principles at infinity with applications to minimal and constant mean curvature surfaces. Preprint.
[36] W. H. Meeks III and H. Rosenberg. Minimal surfaces of finite topology. Preprint.
[37] W. H. Meeks III and H. Rosenberg. The uniqueness of the helicoid and the asymptotic geometry of properly embedded minimal surfaces with finite topology. Preprint.
[38] W. H. Meeks III and H. Rosenberg. The geometry and conformal structure of properly embedded minimal surfaces of finite topology in $\mathbb{R}^3$. *Invent. Math.*, 114:625–639, 1993.
[39] W. H. Meeks III and H. Rosenberg. The geometry of periodic minimal surfaces. *Comment. Math. Helvetici*, 68:538–578, 1993.
[40] W. H. Meeks III and M. Weber. Existence of bent helicoids and the geometry of the singular set in the Colding-Minicozzi lamination theorem. Preprint.
[41] W. H. Meeks III and B. White. Minimal surfaces bounded by convex curves in parallel planes. *Comment. Math. Helvetici*, 66:263–278, 1991.
[42] W. H. Meeks III and B. White. The space of minimal annuli bounded by an extremal pair of planar curves. *Communications in Analysis and Geometry*, 1(3):415–437, 1993.
[43] W. H. Meeks III and S. T. Yau. The topological uniqueness of complete minimal surfaces of finite topological type. *Topology*, 31(2):305–316, 1992.
[44] S. Morales. On the existence of a proper minimal surface in $\mathbb{R}^3$ with the conformal type of a disk. To appear in GAFA.
[45] N. Nadirashvili. Hadamard's and Calabi-Yau's conjectures on negatively curved and minimal surfaces. *Invent. Math.*, 126(3):457–465, 1996.
[46] A. Ros. Embedded minimal surfaces: forces, topology and symmetries. *Calc. Var.*, 4:469–496, 1996.
[47] R. Schoen. Uniqueness, symmetry, and embeddedness of minimal surfaces. *J. of Differential Geometry*, 18:791–809, 1983.
[48] L. Simon. Lectures on geometric measure theory. In *Proceedings of the Center for Mathematical Analysis*, volume 3, Canberra, Australia, 1983. Australian National University.
[49] J. Spruck. Two-dimensional minimal graphs over unbounded domains. Preprint.
[50] K. Stephenson. Concerning the Gross star theorem. In *Complex analysis, Joensuu*, volume 1351, pages 328–338. Springer Lecture Notes in Math., 1987.
[51] M. Traizet. The genus 2 helicoid. Personal Communication.
[52] M. Traizet. An embedded minimal surface with no symmetries. *J. Differential Geometry*, 60(1):103–153, 2002.
[53] M. Tsuji. *Potential theory in modern function theory.* Maruzen, Tokyo, 1959.
[54] M. Weber and M. Wolf. Teichmuller theory and handle addition for minimal surfaces. *Annals of Math.*, 156:713–795, 2002.

Mathematics Department, University of Massachusetts, Amherst, MA 01003
E-mail address: bill@gang.umass.edu

Clay Mathematics Proceedings
Volume 2, 2004

Minimal Surfaces of Finite Topology

William H. Meeks III and Harold Rosenberg

1. Introduction

This manuscript is based on a series of five lectures we gave at the Clay/MSRI minimal surface summer school at MSRI in the summer of 2001. These lectures covered our recent proof of the uniqueness of the helicoid and plane as the only properly embedded simply-connected minimal surfaces in $\mathbb{R}^3$. This theorem depends on recent curvature estimates for embedded minimal disks in Riemannian three-manifolds and regularity and compactness result for limits of such disks by Colding and Minicozzi (see [**4, 3, 5**]). Bill Minicozzi also gave five lectures on their results at the conference. We include in this manuscript much of the basic theory and results in classical minimal surface theory that are relevant to our lectures. At the end of the manuscript we give a sketch of the proof of the main theorem in [**30**] on the uniqueness of the helicoid.

Generally, we will discuss the geometry of finite topology properly embedded minimal surfaces M in $\mathbb{R}^3$. M of finite topology means M is homeomorphic to a compact surface $\widehat{M}$ (of genus k and empty boundary) minus a finite number of points $p_1, \ldots, p_k \in \widehat{M}$, called the punctures. A neighborhood E of a puncture in M is called an end of M. We will choose the ends sufficiently small so they are topologically $S^1 \times [0,1)$. We remark that $\widehat{M}$ is orientable since M is properly embedded in $\mathbb{R}^3$.

The simplest examples (discovered by Meusnier in 1776) are the helicoid and catenoid (and a plane of course). It was only in 1982 that another example was discovered. In his thesis at IMPA, Celso Costa discovered the Weierstrass representation of a complete minimal surface modeled on a 3-punctured torus. He observed the three ends of this surface were embedded; one top catenoid-type end, one bottom catenoid-type end, and a middle planar type end [**9**]. Subsequently, Hoffman and Meeks [**18**] proved this example is embedded and they constructed related embedded examples for every positive genus and also with three ends.

In 1993, Hoffman, Karcher and Wei [**17**] discovered the Weierstrass data of a complete minimal surface of genus one and one annular end. Computer pictures suggested this surface is embedded and the end is asymptotic to an end of a helicoid. Hoffman, Weber and Wolf [**20**] have announced a proof that there is such an

This research was supported by NSF grant DMS-0104044.

©2004 Clay Mathematics Institute

embedded surface. Moreover, computer evidence suggests that one can add many handles to a helicoid.

For many years, the search went on for simply connected examples other than the plane and helicoid. One goal of these notes is to sketch our proof that there are no such examples. For details we refer the reader to our paper [**30**].

THEOREM 1.1. *A properly embedded simply-connected minimal surface in* $\mathbb{R}^3$ *is either a plane or a helicoid.*

In the last decade, it was established that the unique 1-connected example is the catenoid. First we proved such an example is transverse to a foliation of $\mathbb{R}^3$ by planes [**33**] and then Pascal Collin [**7**] proved this implies it is a catenoid. More generally, Collin proved a conjecture of Hoffman and Meeks, usually referred to as the generalized Nitsche conjecture, that a properly embedded minimal surface of finite topology and more than one end has finite total curvature. Recently, Colding and Minicozzi [**6**] have given a proof of Collin's theorem as a consequence of their curvature estimates for embedded minimal disks.

There is an important difference between properly embedded minimal surfaces M with one end and those with more than one end. The latter surfaces have the property that one can find planar or catenoid ends in their complement; this limits the surface to regions of space where it is more accessible to analysis. Clearly the helicoid admits no such end in its complement. To find planar and catenoidal ends in the complement of an M with at least two ends, one solves Plateau problems in appropriate regions of space and passes to limits to obtain complete stable minimal surfaces. Then the stable surface has finite total curvature by [**11**], and hence has standard ends. This construction is carried out in Section 2.

In addition to proving the uniqueness of the helicoid, we can also describe the asymptotic behavior of any properly embedded minimal annulus A in $\mathbb{R}^3$, A diffeomorphic to $S^1 \times [0,1)$. In [**30**] we prove that either A has finite total Gaussian curvature and is asymptotic to the end of a plane or catenoid, or A has infinite total Gaussian curvature and is asymptotic to the end of a helicoid. In fact, if A has infinite total curvature, we prove that A has a special conformal analytic representation on the punctured disk D^* which makes it into a minimal surface of "finite type" (see [**16**], [**39**] and [**40**]). In this case the stereographic projection of the Gauss map $g\colon D^* \to \mathbb{C} \cup \{\infty\}$ has finite growth at the puncture in the sense of Nevanlinna. Since a nonplanar properly embedded minimal surface in $\mathbb{R}^3$ with finite topology and one end always has infinite total curvature and one annular end, such a surface always has finite type.

THEOREM 1.2. *If* M *is a properly embedded nonplanar minimal surface with finite genus* k *and one end, then* M *is asymptotic to a helicoid. In fact,* M *is a minimal surface of finite type, which means, after a possible rotation of* M *in* $\mathbb{R}^3$, *that:*

1. M *is conformally equivalent to a compact Riemann surface* $\overline{M}$ *punctured at a single point* p_∞;
2. *If* $g\colon M \to \mathbb{C} \cup \{\infty\}$ *is the stereographic projection of the Gauss map, then* dg/g *is a meromorphic* 1*-form on* $\overline{M}$ *with a double pole at* p_∞;
3. *The holomorphic* 1*-form* $dx_3 + i dx_3^*$ *extends to a meromorphic 1-form on* $\overline{M}$ *with a double pole at* p_∞ *and with zeroes at each pole and zero of* g *of*

the same order as the zero or pole of g. The meromorphic function g has k zeroes and k poles counted with multiplicity.

A consequence of the above theorem is that the moduli space of properly embedded one-ended minimal surfaces of genus k is an analytic variety; we conjecture that this variety always consists of a positive finite number of points. Theorem 1.2 and the main theorem in [**7**] have the following corollary:

COROLLARY 1. *If M is a properly embedded minimal surface in $\mathbb{R}^3$ of finite topology, then each annular end of M is asymptotic to the end of a plane, a catenoid or a helicoid. In particular, the Gaussian curvature of M is uniformly bounded.*

The corollary above demonstrates the strong geometric consequences that finite topology has for a properly embedded minimal surface. The validity of the following "bounded curvature conjecture" would show that the hypotheses of Theorems 1.1 and 1.2 can be weakened by changing "proper" to "complete", since by Theorem 4.6, a complete embedded minimal surface of bounded curvature is proper.

CONJECTURE 1. *Any complete embedded minimal surface in $\mathbb{R}^3$ with finite genus has bounded Gaussian curvature.*

Our paper is organized as follows. In Section 2 we show how to construct ends of planes or catenoids in the complement of a properly embedded minimal surface with more than one end. In Section 3 we explain some earlier work of ours dealing with minimal surfaces in nonsimply-connected flat three-manifolds, including our finite total curvature theorem. In Section 4 we establish properties of minimal laminations of $\mathbb{R}^3$. We show that a minimal lamination consists of either one leaf, which is a properly embedded minimal surface, or if there is more than one leaf in the lamination, then there are planar leaves. The set of planar leaves P is closed and each limit leaf is planar. Furthermore, in each open slab or half-space in the complement of P there is at most one leaf of the lamination; such a leaf has unbounded curvature, is proper in the slab, and each plane in the slab separates this leaf in exactly two components. Furthermore, we prove that if the lamination has more than one leaf, then each leaf of finite topology is a plane.

In Section 5 we begin the study of a properly embedded simply-connected minimal surface M, which we will always assume is not a plane. The starting point is the theorem of Colding and Minicozzi concerning homothetic blow-downs of M. They prove that any sequence of homothetic scalings of M by $\lambda(i)$, with the scalings converging to zero, has a subsequence $\lambda(i)M$ that converges to a minimal lamination $\mathcal{L}$ of $\mathbb{R}^3$. Here, the convergence is smooth except along a (possibly disconnected) Lipschitz curve $S(\mathcal{L})$ that meets each leaf in isolated points. The existence of $\mathcal{L}$ depends on the simply-connected assumption on M; for example, if N is a properly embedded triply-periodic minimal surface, then no sequence of homothetic blow-downs of N can converge to a lamination. Also notice that if M is a vertical helicoid, then M has a unique blow-down which is a foliation by horizontal planes and the singular set is a vertical line. In Section 5 we prove that a homothetic blow-down $\mathcal{L}$ of M is a foliation of $\mathbb{R}^3$ by parallel planes and that the singular set of nonsmooth convergence to $\mathcal{L}$ is a possibly disconnected Lipschitz curve $S(\mathcal{L})$; actually this result is proved by Colding and Minicozzi in their limit

lamination theorem. In Section 6 we explain other important properties of a blow-down of M and give a sketch of the proof of how $\mathcal{L}$ is used to prove that M is a helicoid (see [**30**] for details).

2. Minimal Surfaces with More Than One End

A technique used often to study a complete minimal surface M in a flat three-manifold N is the construction of finite total curvature minimal surfaces $\Sigma \subset N$, with $\partial\Sigma$ compact and nonempty, Σ noncompact, such that $\partial\Sigma \subset N$ and $\mathrm{Int}(\Sigma) \cap M = \varnothing$. Such surfaces Σ trap M in smaller regions of N which help make the geometry of M understandable. This construction, in the just described context, was carried out for a first time by Meeks and Yau [**36**] in their proof of topological uniqueness of minimal surfaces of finite topology in $\mathbb{R}^3$. We now present several examples of this technique.

First we would like to explain how Σ can be obtained. Let Ω be a complete region of N, whose boundary is a good barrier for solving the least-area Plateau problem (this theory was developed by Meeks and Yau [**35**]). This means $\partial\Omega = C$ is a 2-dimensional variety, smooth except along an analytic one-dimensional variety, such that

(1) C is mean convex at the smooth points, i.e., the mean curvature vector at such points, points into Ω (the zero vector points into Ω), and
(2) at a nonsmooth point of C, the angle between the smooth faces of C, at the point, is less than or equal to π (measured in Ω).

Under the above hypotheses on Ω, Meeks and Yau [**35**] proved that any smooth embedded 1-cycle Γ in $\partial\Omega$ that is null homologous in Ω is the boundary of a compact least-area surface Σ_Γ in Ω, and Σ_Γ is smooth and embedded. The idea for the construction of Σ is to solve the Plateau problem in N by taking a limit of embedded surfaces with boundary Γ whose areas converge to the infimum of all possible areas. Then one checks that such a minimizing sequence can be constructed to stay in Ω. The mean convexity (and angle condition) implies that surfaces leaving Ω will increase area when crossing $\partial\Omega$. Then one works (considerably) to extract a subsequence that converges to a smooth embedded surface Σ.

One can also use geometric measure theory to obtain an oriented Σ_Γ [**42**]. Again, we are assuming $C = \partial\Omega$ is a good barrier and $\Gamma \subset \partial\Omega$ a smooth 1-cycle (i.e., a collection of disjoint smooth Jordan curves). If Γ bounds an oriented 2-chain in Ω, then Γ bounds a smooth embedded orientable surface Σ_Γ in Ω which minimizes area among all orientable 2-chains in Ω with boundary Γ. If Γ is a $\mathbb{Z}_2$-boundary in Ω, then Γ bounds a smooth embedded least-area surface in the same relative $\mathbb{Z}_2$-homology class. If Γ bounds an orientable (immersed) surface of genus n in Ω, then Σ_Γ can be chosen of genus at most n and of least-area in its homotopy class [**13**] [**34**].

Now we will discuss how the least-area compact minimal surfaces Σ_Γ can converge to finite total curvature, noncompact, minimal surfaces Σ. We assume M orientable, A an end of M, $A \subset \partial\Omega$, and Γ a smooth Jordan curve on A, not homologous to zero in Ω. Let $A_1 \subset A_2 \subset \cdots$ be an increasing sequence of compact submanifolds of A, which exhausts A, and $\partial A_i = \Gamma \cup \Gamma_i$. By our previous discussion of how one can solve the Plateau problem in Ω using geometric measure theory, we know there exists a least-area smooth embedded surface $\Sigma(i)$ in Ω such that $\partial\Sigma(i) = \Gamma \cup \Gamma_i$ and $\Sigma(i)$ is $\mathbb{Z}_2$-homologous to A_i rel ∂A_i. Since A_i is orientable

and $\Sigma(i) \cup A_i$ is $\mathbb{Z}_2$-homologous to zero, $\Sigma(i)$ is also orientable. Since Γ is not homologous to zero in Ω, $\Sigma(i)$ can be chosen connected.

Now we will show a subsequence of the $\Sigma(i)$ converges to a stable embedded minimal surface Σ with $\partial\Sigma = \Gamma$. First observe that there are uniform local area bounds for the family $\Sigma(i)$. For if $B \subset \Omega$ is a ball of radius r, ∂B transverse to $\Sigma(i)$, then $\partial B \cap \Sigma(i)$ is a 1-cycle on ∂B that bounds (mod 2) a 2-chain on ∂B of area at most $2\pi r^2$. Since $\Sigma(i)$ minimizes area bounded by ∂A_i (in the $\mathbb{Z}_2$-homology class), we conclude $B \cap \Sigma(i)$ has area at most $2\pi r^2$. Similarly if B is a ball centered at a point of $\partial\Sigma(i)$, then the area of $\Sigma(i) \cap B$ is at most the area of ∂B.

Now let $B(r) \subset \Omega$. By the curvature estimates of Schoen [**41**], after choosing a possibly smaller r, each component of $\Sigma(i) \cap B(r)$ that intersects $B(\frac{r}{2})$ can be expressed as a graph, of small gradient, over a plane P_i in $B(r)$, passing through the center of the ball, and P_i does not depend on the component. By the uniform area estimates, $\Sigma(i) \cap B(\frac{r}{2})$ contains a bounded number of components independent of i and hence there is a bounded number of associated graphs. Suppose for the moment that for every i, $\Sigma(i) \cap B(\frac{r}{2})$ contains one component. Choose a subsequence of the P_i to converge to a plane P through the center of the ball. Then the standard compactness theorem for minimal graphs implies a subsequence of the graphs $\Sigma(i) \cap B(\frac{r}{2})$ converge to a minimal graph over $P \cap B(\frac{r}{2})$. When $\Sigma(i) \cap B(\frac{r}{2})$ has more than one component, we apply the preceding argument to each component and obtain a (uniformly bounded) finite number of graphs over $P \cap B(\frac{r}{2})$, to which the subsequence of $\Sigma(i) \cap B(\frac{r}{2})$ converges.

Now Ω has a countable basis of balls B_n where for every n and subsequence $\Sigma(i_j)$ of $\Sigma(i)$, the $\Sigma(i_j) \cap B_n$ have a convergent subsequence in B_n. Suppose the subsequence $\Sigma(i_j) \cap B_1$ converges in B_1. Then the associated sequence of graphs in $B_2 \cap \Sigma(i_j)$ has a subsequence converging in $B_2 \cup B_1$. Continue in this manner from B_1 to B_{i+1} and take a diagonal subsequence. This yields a subsequence of the $\Sigma(i)$ that converges to a smooth minimal surface Σ with $\partial\Sigma = \Gamma$. It is not hard to see that Σ is embedded and stable (since it's a limit of least-area embedded surfaces). Also the boundary regularity theorem of Hardt and Simons [**15**] implies Σ is smooth along Γ. Finally, the theorem of Fisher-Colbrie [**11**] implies that Σ has finite total curvature.

In particular, this technique yields:

LEMMA 2.1. *Let M be a properly embedded minimal surface in $\mathbb{R}^3$ with more than one end. Then there is an end of a catenoid or of a plane in the complement of M.*

PROOF. Let Γ be a smooth Jordan curve on M that separates M into two noncompact components, one of which we denote by A. M separates $\mathbb{R}^3$ into two connected components and Γ cannot be homologous to zero in both components; let Ω be a component such that Γ is not homologous to zero in Ω.

By our previous discussion, there is a finite total curvature embedded minimal surface Σ_Γ in Ω with $\partial\Sigma_\Gamma = \Gamma$. More precisely, $C = \partial\Omega = M$ is a minimal surface, hence a good barrier for solving the Plateau problem. Let $A_i \subset A_{i+1}$ be a compact exhaustion of A with $\partial A_i = \Gamma \cup \Gamma_i$. Let Σ_i be an embedded least-area minimal surface in Ω, $\mathbb{Z}_2$-homologous to A_i, with $\partial\Sigma_i = \partial A_i$. As before, a subsequence of Σ_i converges to Σ_Γ.

Now it may be that $\Sigma_\Gamma \subset M$ (if it touches M at one interior point, then since it's on one side of M at this point, it is contained in M). In this case at least one

end of M is of finite total curvature, so asymptotic to a planar end or catenoid end B. By the maximum principle at infinity [**32**], the distance between the ends of M is strictly positive. So B can be translated into Ω to be disjoint from M. Similarly, if $\mathrm{Int}(\Sigma_\Gamma) \subset \mathrm{Int}(\Omega)$, then the ends of Σ_Γ and a strictly positive distance from M so the conclusion of the lemma is clear. □

There is a slight refinement of this lemma which is useful.

LEMMA 2.2. *Let B be a ball in $\mathbb{R}^3$ and A_1, A_2 properly embedded minimal surfaces which are noncompact with ∂A_1, ∂A_2 smooth Jordan curves such that $B \cap (A_1 \cup A_2) = \partial A_1 \cup \partial A_2$ and $A_1 \cap A_2 = \varnothing$. Let Δ be the annulus on ∂B, bounded by $\partial A_1 \cup \partial A_2$ and let Ω be the connected component of $\mathbb{R}^3 - (A_1 \cup A_2 \cup \Delta)$ disjoint from B. Then there is an end of a plane or a catenoid in the interior of Ω. Moreover, ∂A_1 is the boundary of a smooth embedded surface Σ in Ω and outside of a larger ball $\tilde{B}$ containing B, Σ is a finite total curvature minimal surface that separates ends of A_1 and A_2, i.e., any path from A_1 to A_2 in $\mathbb{R}^3 - \tilde{B}$, meets Σ.*

PROOF. Let $\Gamma = \partial A_1$ and consider Ω, with $\partial\Omega = A_1 \cup \Delta \cup A_2$. If $\partial\Omega$ were a good barrier for solving the Plateau problem, then the construction of $\Sigma = \Sigma_\Gamma$ proceeds exactly as in the previous lemma. However, ∂B is not mean convex with respect to Ω, so one first changes the Riemannian metric of $\mathbb{R}^3$ near ∂B. □

3. Periodic Minimal Surfaces

The theory of periodic minimal surfaces in $\mathbb{R}^3$ deals with minimal surfaces in $\mathbb{R}^3$ that are invariant under a group Γ of isometries of $\mathbb{R}^3$ that acts properly discontinuously. Up to finite index subgroups, Γ can be chosen to be a lattice L in $\mathbb{R}^3$ (the triply-periodic case), a rank two group of translations that we may assume are a lattice $L \subset \mathbb{R}^3 \times \{0\}$ (the doubly-periodic case) or L is a cyclic group generated by a screw motion symmetry S_θ, a rotation around the x_3-axis by angle θ composed with a nontrivial translation by a vector on the x_3-axis (the singly-periodic case). Frequently, the singly-periodic case is considered as two cases depending on whether or not θ is zero; the $\theta = 0$ case occurs when S_θ is a pure translation. The geometry of a periodic minimal surface in $\mathbb{R}^3$ can usually be described in terms of the geometry of its quotient surface M in the flat three-manifold $\mathbb{R}^3/\Gamma$. Thus, triply-periodic minimal surfaces are the minimal surfaces in a flat three-torus, doubly-periodic minimal surfaces are the minimal surfaces in $\mathbb{T} \times \mathbb{R}$ where $\mathbb{T}$ is a flat two-torus, and one-periodic minimal surfaces is the study of minimal surfaces in $\mathbb{R}^3/S_\theta$ where S_θ is a screw motion symmetry. Note that if S_θ is a pure translation, then $\mathbb{R}^3/S_\theta$ is the flat three-manifold $S^1 \times \mathbb{R}^3$ for some circle S^1.

LEMMA 3.1. *Suppose M is a nonflat properly embedded minimal surface in a complete flat orientable three-manifold $N = \mathbb{R}^3/\Gamma$. Then the induced map*

$$i_*\colon \pi_1(M) \to \pi_1(N) = \Gamma$$

is surjective. In particular, M separates N^3.

PROOF. By elementary covering space theory, if $i_*\colon \pi_1(M) \to \pi_1(N)$ is not surjective, then M has at least two lifts of M to the universal cover $\mathbb{R}^3$ of N. However, by the strong half-space theorem [**19**], these lifts must be parallel planes which contradicts the assumption that M is not flat. □

It follows from the above lemma that nonflat properly embedded periodic minimal surfaces in $\mathbb{R}^3$ are precisely the lifts of nonflat minimal surfaces in complete flat three-manifolds.

The periodic minimal surfaces easiest to understand are the triply-periodic minimal surfaces which correspond to compact embedded orientable minimal surfaces in a flat $\mathbb{T}^3$. Such a surface $M \subset \mathbb{T}^3$ separates $\mathbb{T}^3$ into two regions R_1, R_2. Since $\pi_1(\partial R_i) \to \pi_1(R_i)$ is onto by Lemma 3.1, elementary three-manifold theory implies that R_1 and R_2 are handlebodies; equivalently, M is a Heegaard surface in $\mathbb{T}^3$. It is known that Heegaard surfaces in $\mathbb{T}^3$ of the same genus are isotopic and so any two triply-periodic minimal surfaces of the same genus in a flat $\mathbb{T}^3$ are ambiently isotopic.

Since the Gauss map $G\colon M \to S^2$ makes sense in a flat torus, the Gauss–Bonnet Theorem gives the following basic result.

THEOREM 3.2. *(Gauss–Bonnet Theorem) If $M_n \subset \mathbb{T}^3$ is a minimal surface of genus n, then the Gauss map of M is a meromorphic function of degree $(n-1)$. In particular, the genus of M is never two and if the genus of M is three, then M is hyperelliptic.*

Another related result is the following version of Abel's theorem [**25**].

THEOREM 3.3. *(Abel's Theorem) Suppose $M_n \subset \mathbb{T}^3 = \mathbb{R}^3/L$ is a minimal surface of genus n and $G\colon M \to S^2$ is the Gauss map. The sum map $\tilde{G}\colon M_n \to \mathbb{T}^3$ is defined by $\tilde{G}(v) = \Sigma_{p \in G^{-1}(v)} p$, where the summation is done in the abelian group $\mathbb{T}^3 = \mathbb{R}^3/L$. Then $\tilde{G}$ is a constant function.*

COROLLARY 2. *Suppose $M_3 \subset \mathbb{T}^3 = \mathbb{R}^3/L$ is an embedded minimal surface of genus 3. After a translation of M_3 in $\mathbb{T}^3$, $M_3 = -M_3$. The zeroes of Gaussian curvature of M_3 are the elements of the abelian group of order two.*

The first general result on the topology of doubly-periodic is the following theorem of Meeks [**26**]. Recently, Meeks and Rosenberg [**29**] have generalized the next theorem to $M \times \mathbb{R}$, where M is any compact Riemannian surface.

THEOREM 3.4. *If $\Sigma \subset \mathbb{T}^2 \times [0,\infty), \mathbb{T}^2$ a flat torus is a properly embedded noncompact minimal surface with compact boundary and more than one end, then Σ has linear area growth in $\mathbb{T}^2 \times [0,t]$ for t large. In fact, there exist positive constants c_1, c_2 such that*

$$c_1 e(\Sigma) \leq \mathit{Area}(\Sigma \cap (\mathbb{T}^2 \times [0,t])) \leq c_2 e(\Sigma)$$

where $e(\Sigma)$ is the number of ends of Σ. Here c_1 depends only on Σ. In particular, Σ has a finite number of ends and linear area growth.

If $M \subset \mathbb{T}^2 \times \mathbb{R}$ is a properly embedded minimal surface of finite topology, then each end of M is an annulus, which is a cyclic fundamental group. Hence, after taking a finite cover of $\mathbb{T}^2 \times \mathbb{R}$, one may assume that $M \cap \mathbb{T}^2 \times [0,\infty)$ and $M \cap \mathbb{T}^2 \times (-\infty, 0]$ each have more than one end. Theorem 3.4 immediately implies that M has extrinsic, and hence intrinsic, linear area growth. A straightforward application of the Gauss–Bonnet Theorem and the first variation formula of area, implies:

COROLLARY 3 (Meeks–Rosenberg [**31**]). *If $M \subset \mathbb{T}^2 \times \mathbb{R}$ is a properly embedded minimal surface of finite topology, then*

$$C(M) = \int_M |K| dA = 2\pi\chi(M).$$

It should be noted that Meeks and Rosenberg [**29**] have been able to prove the similar corollary in the case where $\mathbb{T}^2$ is replaced by any compact Riemannian surface with nonpositive curvature. Meeks and Rosenberg, after proving the above finite total curvature theorem, described the asymptotic properties of the doubly-periodic minimal surfaces and some surprising rigidity results (see [**22**]). These rigidity results eventually were useful in proving the following deep result.

THEOREM 3.5 (Lazard-Holly and Meeks [**22**]). *If M is a properly embedded nonflat doubly-periodic minimal surface in $\mathbb{R}^3$ with a genus zero quotient surface, then M is one of Scherk's doubly-periodic minimal surfaces.*

The theory of singly-periodic minimal surfaces is substantially more complex than the doubly-periodic situation. For example, by considering a Scherk doubly-periodic minimal surface to be one-periodic, one can obtain properly embedded minimal planar domains in $S^1 \times \mathbb{R}^3$ which have an infinite number of ends. Another difficulty is that the Gauss map of a surface $M \subset \mathbb{R}^3, S_\theta$ is not well-defined for $\theta \neq 0$. In spite of these problems one can still prove the following strong result.

THEOREM 3.6 (Meeks and Rosenberg [**35**]). *Suppose $M \subset \mathbb{R}^3/S_\theta$ is a properly embedded minimal surface of finite topology. Then:*

(1) *M is conformally diffeomorphic to a closed Riemann surface $\bar{M}$ punctured in points;*
(2) *M can be defined in terms of meromorphic data on $\bar{M}$;*
(3) *M has finite total curvature.*

The proof of Theorem 3.4 and Theorem 3.6 both use the technique of trapping surfaces discussed in Section 2, to restrict the ends of the minimal surfaces to small regions of the flat three-manifolds where their geometry can be analyzed. Also see [**24**] for a proof that a properly embedded minimal surface Σ in $\mathbb{R}^3/S_\theta$ has a finite number of ends when the rotational part of S_θ has order at least 3.

4. Minimal Laminations of R^3

A closed set $\mathcal{L}$ in $\mathbb{R}^3$ is called a minimal lamination if $\mathcal{L}$ is the union of pairwise disjoint connected complete injectively immersed minimal surfaces. Locally we require that there are $C^{1,\alpha}$ coordinate charts $f\colon D \times (0,1) \to \mathbb{R}^3, 0 < \alpha < 1$, with $\mathcal{L}$ in $f(D \times (0,1))$ the image of the $D \times \{t\}$, t varying over a closed subset of $(0,1)$. The minimal surfaces in $\mathcal{L}$ are called the leaves of $\mathcal{L}$.

The leaves L of a minimal lamination $\mathcal{L}$ are smooth (even analytic), and if K is a compact set of L which is a limit leaf of $\mathcal{L}$, then the leaves of $\mathcal{L}$ converge smoothly to L over K; the convergence is uniform in the C^k-topology for any k.

Our work will depend upon the following (very important) curvature estimates of Colding and Minicozzi [**4**], which we will refer to as the curvature estimates $\mathcal{C}$. There exists an $\varepsilon > 0$ such that the following holds. Let $y \in \mathbb{R}^3$, $r > 0$ and $\Sigma \subset B_{2r}(y) \cap \{x_3 > x_3(y)\} \subset \mathbb{R}^3$ be a compact embedded minimal disk with

$\partial\Sigma \subset \partial B_{2r}(y)$. For any connected component Σ' of $B_r(y) \cap \Sigma$ with $B_{\varepsilon r}(y) \cap \Sigma' \neq \varnothing$, one has $\sup_{\Sigma'} |A_{\Sigma'}|^2 \leq r^{-2}$.

A consequence of these curvature estimates is the following. Let Σ be any compact smooth surface passing through the origin with boundary contained in the boundary of the ball $B(1)$ of radius one centered at the origin. There is an ε and a constant c such that if D is an embedded minimal disk in $B(1)$, disjoint from Σ, and with boundary contained in the boundary of $B(1)$, then in $B(\varepsilon)$, the curvature of D is bounded by c. This can be seen by homothetically expanding Σ; the ε depends on the norm of the second fundamental form of Σ in the ball $B(1/2)$. In our applications Σ will be a stable minimal disk for which one always has a bound on the norm of the second fundamental form in $B(1/2)$ by curvature estimates for stable surfaces.

In this section we will prove a general structure theorem that explains some of the geometric properties that hold for a minimal lamination $\mathcal{L}$ of $\mathbb{R}^3$. A properly embedded minimal surface is the simplest example of a minimal lamination. The only known examples of minimal laminations of $\mathbb{R}^3$ with more than one leaf are closed sets of parallel planes in $\mathbb{R}^3$ and we conjecture that these are the only ones. In fact, we will prove that in the case $\mathcal{L}$ has more than one leaf, then every leaf of $\mathcal{L}$ with finite topology is a plane.

Every leaf L of a minimal lamination $\mathcal{L}$ of $\mathbb{R}^3$ has locally bounded Gaussian curvature in the sense that the intersection of L with any ball has Gaussian curvature bounded from below by a constant that only depends on the ball. The reason that the curvature is locally bounded is that the intersection of $\mathcal{L}$ with a closed ball is compact and the Gaussian curvature function is continuous.

LEMMA 4.1. *Suppose M is a complete connected embedded minimal surface in $\mathbb{R}^3$ with locally bounded Gaussian curvature. Then one of the following holds:*

(1) *M is properly embedded in $\mathbb{R}^3$;*
(2) *M is properly embedded in an open half-space of $\mathbb{R}^3$ with limit set the boundary plane of this half-space;*
(3) *M is properly embedded in an open slab of $\mathbb{R}^3$ with limit set consisting of the boundary planes.*

PROOF. Let x_n be any sequence of points in M, converging to some x in $\mathbb{R}^3$. Since M has locally bounded curvature, there is a $\delta = \delta(x)$ such that for n sufficiently large, M is a graph F_n over the disk $D_\delta(x_n)$ in the tangent plane to M at x_n, of radius δ and centered at x_n. Moreover each such local graph F_n has bounded geometry.

Choose a subsequence of the x_n so that the tangent planes to M at the subsequence converge to some plane P at x. Then the F_n of this subsequence will be graphs (for n large) over the disk D of radius $\delta/2$ in P centered at x. By compactness of minimal graphs, a subsequence of the F_n will converge to a minimal graph F_∞ over $D, x \in F_\infty$.

Notice that F_∞ at x does not depend on the subsequence of the x_n. If $y_n \in M$ is a sequence converging to x with the tangent planes of M at y_n converging to a plane Q at x, then $P = Q$ and the local graphs G_n of M at y_n converge to F_∞ as well. If this were not the case, F_∞ and G_∞ would cross each other near x (i.e., $x \in F_\infty \cap G_\infty$ and the maximum principle implies there are points of $F_\infty \cap G_\infty$ near x where they meet transversely). Now F_∞ is the uniform limit of the graphs F_n and G_∞ is the uniform limit of the graphs G_n, so near a point of transverse

intersection of F_∞ and G_∞ we would have F_i intersecting G_j transversely for i, j large. This contradicts M embedded. Notice also that each F_n is disjoint from F_∞; this follows by the same reasoning as above. Thus we have a local lamination of the closure $\bar{M}$ of M near x.

Each point $y \in \partial F_\infty$ is also an accumulation point of M so there is a limit graph $F_\infty(y)$ over a disk of radius $\delta(y)$ centered at y. By uniqueness of limits, $F_\infty(y) = F_\infty$ where they intersect. Thus F_∞ may be continued analytically to obtain a complete minimal surface in $\bar{M}$. The lamination $\mathcal{L}$ is obtained by taking the closure of all the limit surfaces so obtained.

Next we will prove that any limit leaf of $\mathcal{L}$ is a plane.

Let L be a limit leaf and $\widehat{L}$ the universal covering space of L. The exponential map of L is a local diffeomorphism and there is a normal disk bundle ν over $\widehat{L}$, of varying radius, that submerses in $\mathbb{R}^3$. Give ν the flat metric induced by the submersion; $\widehat{L}$ is the zero section of ν.

Let $\widehat{D}$ be a compact simply-connected domain of $\widehat{L}$, D its projection into L. Each point of D has a neighborhood that is a uniform limit of (pairwise disjoint) local graphs of M. The usual holonomy construction allows one to lift these local graphs along the lifting of paths in D to obtain $\widehat{D}$ as a uniform limit of pairwise-disjoint embedded minimal surfaces E_n in ν.

It is known that any compact domain F (here $F = \widehat{D}$) that is a limit of disjoint minimal domains E_n is stable. Here is a proof. If F were unstable, the first eigenvalue λ_1 of the stability operator L of the minimal surface F (here $L = \Delta - 2K$) is negative. Let $\vec{n}$ denote the unit vector field along F in ν and f the corresponding eigenfunction of $\lambda_1, L(f) + \lambda_1 f = 0, f > 0$ in F and $f = 0$ on ∂F.

Consider the variation of F: $F(t) = \{x + tf(x)\vec{n}(x) \mid x \in F\}$. The first variation $\dot{H}(0)$ of the mean curvature of $F(t)$ at $t = 0$ is given by $L(f)$. Since $\lambda_1 < 0$, and $f(x) > 0$ for $x \in \mathrm{Int}(F)$, it follows that the mean curvature vector of $F(t)$, for t small, points away from F, that is, $\langle \vec{H}_t(x), \vec{n}(x) \rangle > 0$.

Now for t_0 small, choose n large so that $E(n)$ is close enough to F so there is a a nonempty intersection of $F(t_0)$ and $E(n)$. As t decreases from t_0 to 0, the $F(t)$ go from $F(t_0)$ to F. So there will be a smallest positive t so that $D(t)$ has a nonempty intersection with $E(n)$. Let $y \in F(t) \cap E(n)$. Near y, $E(n)$ is on the mean convex side of $F(t)$. Since $E(n)$ is a minimal surface, this is impossible.

Then by [**10**] or [**12**], $\widehat{L}$ is a plane, hence L as well, and each limit leaf of $\mathcal{L}$ is a plane. □

REMARK 4.2. F. Xavier [**44**] proved that a complete nonflat immersed minimal surface of bounded curvature in $\mathbb{R}^3$ is not contained in a half-space.

LEMMA 4.3. *Suppose M is a complete connected embedded minimal surface in $\mathbb{R}^3$ with locally bounded curvature. If M is not proper and P is a limit plane of M, then, for any $\varepsilon > 0$, the closed ε-neighborhood of P intersects M in a connected set.*

Idea of Proof: If the ε-neighborhood of P intersects M in more than one component, one constructs a stable minimal surface between them and then, using the following topological Lemma 4.4, one obtains a contradiction. We refer the reader to [**30**] for details of the proofs of Lemmas 4.3 and 4.4.

LEMMA 4.4. *Suppose M and N are smooth connected manifolds of the same dimension such that N is simply-connected and M may have boundary. If $\pi\colon M \to$*

N is a proper submersion onto its image and $\pi|\partial M$ is injective on each boundary component of M, then π is injective. In particular, if M is a smooth immersed surface with boundary in $\mathbb{R}^3$, the projection $\pi\colon M \to \mathbb{R}^2$ to the x_1x_2-plane is a proper submersion onto its image and $\pi|\partial M$ is injective, then M is a graph over $\pi(M) \subset \mathbb{R}^2 \times \{0\}$.

LEMMA 4.5. *If M is a complete embedded minimal surface in $\mathbb{R}^3$ with finite topology and locally bounded curvature, then M is properly embedded in $\mathbb{R}^3$.*

PROOF. Suppose now that M has finite topology and lies in the upper half-space of $\mathbb{R}^3$ with limit set the x_1x_2-plane P. If M has bounded curvature in some ε-neighborhood of P, then it was proved above that M is proper in this neighborhood and has a plane in its closure. This is impossible by the half-space theorem. It remains to prove that M has bounded curvature in an ε-neighborhood of P.

Arguing by contradiction, assume that M does not have bounded curvature. In this case, there is an annular end $E \subset M$ whose Gaussian curvature is not bounded in the slab $S = \{(x_1, x_2, x_3) \mid 0 \leq x_3 \leq 1\}$. After a homothety of M, we may assume that ∂E is contained in the ball B_0 of radius one centered at the origin. Since M has locally bounded curvature, the part of E inside $B(0)$ has bounded curvature.

Since $E \cap S$ does not have bounded curvature, there exists a sequence $p(1), \ldots, p(i), \ldots$ in $E \cap S$ with $\|p(i)\| \geq i$ and $|K(p(i))| \geq i$. After possibly rotating M around the x_3-axis and choosing a subsequence, we may assume that the sequence $(5/\|p(i)\|) \cdot p(i)$ converges to the point $(5, 0)$ in the x_1x_2-plane. Let B be the ball of radius one in the x_1x_2-plane and centered at $(5, 0)$. Notice that there is no compact connected minimal surface with one boundary curve in B_0 and the other boundary curve in B (pass a catenoid between B_0 and B). Using the convex hull property of a compact minimal surface, it is easy to check that $[(5/\|p(i)\|)E] \cap B$ consists only of simply-connected components which are disjoint from the boundary of $(5/\|p(i)\|)E$. The curvature estimates $\mathcal{C}$ imply that, as $i \to \infty$, the curvature of $(5/\|p(i)\|)E$ at $(5/\|p(i)\|) \cdot p(i)$ converges to 0. But the Gaussian curvature at such points approaches $-\infty$ as $i \to \infty$. This contradiction proves the lemma. □

The next theorem follows immediately from the previous lemmas.

THEOREM 4.6. *Suppose $\mathcal{L}$ is a minimal lamination in $\mathbb{R}^3$. If $\mathcal{L}$ has one leaf, then this leaf is a properly embedded surface in $\mathbb{R}^3$. If $\mathcal{L}$ has more than one leaf, then $\mathcal{L}$ consists of the disjoint union of a nonempty closed set of parallel planes $\mathcal{P} \subset \mathcal{L}$ together with a collection of complete minimal surfaces of unbounded Gaussian curvature that are properly embedded in the open slabs and half-spaces of $\mathbb{R}^3 - \mathcal{P}$ and each of these open slabs and half-spaces contains at most one leaf of $\mathcal{L}$. In this case every plane, parallel to but different from the planes in $\mathcal{P}$, intersects at most one of the leaves of $\mathcal{L}$ and separates such a leaf into two components. Furthermore, in the case $\mathcal{L}$ contains more than one leaf, the leaves of $\mathcal{L}$ of finite topology are planes.*

REMARK 4.7. Recently, Meeks, Perez and Ros [**27**] have been able to prove that the finite genus leaves in a minimal lamination $\mathcal{L}$ of $\mathbb{R}^3$ with more than one leaf are planes.

5. Homothetic Blow-downs of M

Given a sequence $\lambda(i) \in \mathbb{R}^+$, $\lambda(i) \to 0$, the sequence $M(i) = \lambda(i)M$ is uniformly-locally-simply-connected in the sense of Colding and Minicozzi. This is because given any ball $B \subset \mathbb{R}^3$, every boundary component of every component of $M(i) \cap B$ bounds a disk in $M(i)$, which is contained in B by the convex ball property; hence, every component of $M(i) \cap B$ is simply-connected. The main regularity theorem of [**4**] implies that a subsequence $M(i_j)$ of the $M(i)$ converges to a minimal lamination $\mathcal{L}$ of $\mathbb{R}^3$ such that for any leaf L of $\mathcal{L}$, the convergence of $M(i_j)$ to L is smooth except at isolated points. It follows from the curvature estimates $\mathcal{C}$ (see Section 4) that if $p \in L$ is a point of smooth convergence, then $M(i_j)$ converges smoothly to leaves of $\mathcal{L}$ in a neighborhood of p in $\mathbb{R}^3$. Recall that $S(\mathcal{L})$ is the singular set of convergence of the $M(i_j)$ to $\mathcal{L}$. Consider the leaf $L(0)$ of $\mathcal{L}$ that passes through the singular point in $S(\mathcal{L})$ which is the origin in $\mathbb{R}^3$; the origin is always a singular point if M is not a plane. We will now prove that $L(0)$ is a limit leaf of $\mathcal{L}$.

Suppose, after a possible rotation of M, that the tangent space to $L(0)$ at the origin is the x_1x_2-plane. There exists an $\varepsilon > 0$ so that the component of $\{(x_1, x_2, x_3) \in L(0) \mid x_1^2 + x_2^2 \leq \varepsilon^2\}$ containing the origin is a graph of small gradient over the disk $\{(x_1, x_2, 0) \mid x_1^2 + x_2^2 \leq \varepsilon^2\}$. Let $D(\varepsilon)$ denote this component and let C denote its boundary. Let $C(\varepsilon) = \{(x_1, x_2, x_3) \mid |x_3| \leq 1 \text{ and } x_1^2 + x_2^2 = \varepsilon^2\}$. Suppose ε is chosen sufficiently small so that $C(\varepsilon) \cap S(\mathcal{L}) = \varnothing$, and C is contained in the interior of $C(\varepsilon)$. Let δ be a positive number, and $C(\varepsilon, \delta)$ equal points of $C(\varepsilon)$ at most a vertical distance δ from C. For δ sufficiently small and for i_j large, $C(\varepsilon, \delta) \cap M(i_j)$ consists of curves and the tangent plane to $M(i_j)$ along these curves converges to that of $D(\varepsilon)$ along C.

We now show that for some fixed δ sufficiently small and for all i_j sufficiently large, $M(i_j) \cap C(\varepsilon, \delta)$ does not contain a closed curve. Suppose to the contrary $M(i_j) \cap C(\varepsilon, \delta)$ contains a closed curve $\alpha(i_j, \delta)$. Since the curve $\alpha(i_j, \delta)$ is a vertical graph over a circle in the x_1x_2-plane, Rado's theorem [**21**] implies that the disk $D(i_j, \delta) \subset M(i_j)$ bounded by $\alpha(i_j, \delta)$ is a vertical graph. Letting δ tend to zero, the disks $D(i_j, \delta)$ converge smoothly to $D(\varepsilon)$ since $\alpha(i_j, \delta)$ converge to C. Then these disks have uniform curvature estimates near the origin. But the curvature of $M(i_j)$ blows up at the origin, which contradicts the curvature estimates $\mathcal{C}$ (the connected component of $M(i_j)$ containing the origin, in a fixed ball centered at the origin, is a minimal disk disjoint from $D(i_j, \delta)$). Hence, for i_j large, $M(i_j) \cap C(\varepsilon, \delta)$ contains some arcs that begin at the bottom boundary component of $C(\varepsilon, \delta)$ and end at the top boundary component of $C(\varepsilon, \delta)$ and, as $i_j \to \infty$, these arcs are multisheeted graphs of order going to infinity. In particular, any homotopically nontrivial closed curve γ in $C(\varepsilon, \delta)$ intersects $M(i_j)$ for i_j large. By compactness of such a γ, γ intersects some leaf of $\mathcal{L}$ and so $L(0)$ is a limit leaf of $\mathcal{L}$. This same proof shows that if $L \in \mathcal{L}$ is any leaf of $\mathcal{L}$ that intersects $S(\mathcal{L})$, then L is a limit leaf of $\mathcal{L}$. Theorem 4.6 implies $L(0)$ is a horizontal plane; the same argument shows that any leaf of $\mathcal{L}$ that intersects $S(\mathcal{L})$ is also a horizontal plane.

We now prove that $\mathcal{L}$ is a foliation of $\mathbb{R}^3$ by horizontal planes. We first check that a neighborhood of $L(0)$ is foliated by horizontal planes in $\mathcal{L}$. Let $\mathcal{P} \subset \mathcal{L}$ be the set of planes in $\mathcal{L}$. Notice that since $\mathcal{L}$ intersects $C(\varepsilon, \delta)$ in closed curves near $C(\varepsilon, \delta) \cap L(0)$ and every homotopically nontrivial curve in $C(\varepsilon, \delta)$ intersects $\mathcal{L}$, $\mathcal{L} \cap C(\varepsilon, \delta)$ must be a foliation of closed curves near $C(\varepsilon, \delta) \cap L(0)$. Therefore,

a neighborhood of $L(0)$ is foliated by leaves in $\mathcal{L}$, all of which are limit leaves and hence planes. In particular, a neighborhood of $L(0)$ is foliated by planes in $\mathcal{P}$. This same argument shows that every plane in $\mathcal{P}$ that intersects $S(\mathcal{L})$ must have a neighborhood foliated by planes in $\mathcal{P}$. Since $S(\mathcal{L})$ is nonempty, the curvature estimates $\mathcal{C}$ imply that every plane in $\mathcal{P}$ intersects $S(\mathcal{L})$. To see this suppose P is a plane in $\mathcal{P}$ that is disjoint from $S(\mathcal{L})$. In this case, there exists a sequence of disks $D(r(i_j))$ with fixed center in P of radii $r(i_j)$, $r(i_j) \to \infty$, and lifts $\tilde{D}(r(i_j)) \subset M(i_j)$ which are small graphs over $D(r(i_j))$ and converge to P. But there are points on $M(i_j)$ near the origin in $\mathbb{R}^3$ where the curvature is blowing up as $i \to \infty$, a uniformly bounded distance from the almost flat larger and larger disks $\tilde{D}(r(i_j))$, a situation which contradicts the curvature estimates $\mathcal{C}$. This proves $P \cap S(\mathcal{L}) \neq \varnothing$. Hence, we conclude that $\mathcal{P}$ is an open subset in $\mathbb{R}^3$. Since $\mathcal{P}$ is a nonempty closed subset of $\mathbb{R}^3$ and $\mathbb{R}^3$ is connected, $\mathcal{P} = \mathbb{R}^3$.

The natural continuous parametrization of the leaves of the foliation of $\mathbb{R}^3$ by the parallel planes in $\mathcal{L}$ induces a natural parametrization of $S(\mathcal{L})$. The curvature estimates $\mathcal{C}$ immediately imply that the parametrized curve $S(\mathcal{L})$ is a continuous Lipschitz curve with Lipschitz constant that only depends on a universal curvature estimate. This discussion completes the proof of the following theorem, which is one of the main results of Colding and Minicozzi [**4**].

THEOREM 5.1. *Suppose M is a properly embedded nonflat simply-connected minimal surface in $\mathbb{R}^3$ and $\lambda(i) \in \mathbb{R}^+$ with $\lambda(i) \to 0$. Then a subsequence of the surfaces $\lambda(i)M$ converges to a foliation $\mathcal{L}$ of $\mathbb{R}^3$ by parallel planes. Furthermore, this convergence is smooth outside of a Lipschitz curve $S(\mathcal{L})$, possibly disconnected, that passes through the origin. The component of the Lipschitz curve $S(\mathcal{L})$ that passes through the origin is contained in a double cone $C(\mathcal{L})$ with vertex at the origin, axis orthogonal to $\mathcal{L}$ and with aperture independent of M.*

6. Uniqueness of the Helicoid

Given M and a sequence $\lambda(i) \in \mathbb{R}^+$, $\lambda(i) \to 0$, we proved in the previous section that a subsequence, $M(i_j) = \lambda(i_j)M$, converges to a foliation $\mathcal{L}$ of $\mathbb{R}^3$ by parallel planes. Furthermore, the convergence $M(i_j) \to \mathcal{L}$ is smooth except along a Lipschitz curve $S(\mathcal{L})$ and the connected component of $S(\mathcal{L})$ that passes through the origin is contained in a double cone $C = C(\mathcal{L})$ with cone point at the origin and axis orthogonal to $\mathcal{L}$. Furthermore, the aperture of the cone C only depends on the Lipschitz constant of S which, in turn, depends only on curvature estimates. In this section we will assume that: $\mathcal{L}$ is a foliation by horizontal planes. Let E be the compact vertical cylinder centered along the x_3-axis and such that $E \cap C = E \cap \partial C = \partial E$, which consists of the circles $S_+ = \partial C \cap \{(x_1, x_2, 1)\}$ and $S_- = \partial C \cap \{(x_1, x_2, -1)\}$. After a possible fixed homothety of M, we may assume that there is only one component of $S(\mathcal{L})$ inside E. For i large, the intersection of $M(i)$ with E consists of $n(i)$ α curves which spiral from the bottom of E to the top of E; they are almost horizontal and multisheeted graphs.

The next theorem states that, for i large, the number $n(i)$ of α curves is equal to two, and that the limit foliation $\mathcal{L}$ is transverse to M. In particular, it follows that the Gauss map of M misses the two vertical vectors, $(0, 0, \pm 1)$, because $\mathcal{L}$ is a foliation by horizontal planes. Since the Gauss map of a complete nonflat minimal surface can miss at most four points on S^2, [**14**], [**37**], it then follows that at most two foliations of $\mathbb{R}^3$ by parallel planes can arise as limits of homothetic shrinkings

of M. In fact we prove in [**30**] that if M gives rise to two different foliations as blow-downs, then there would be a third such blow-down, a contradiction which proves the uniqueness of $\mathcal{L}$. Moreover, the transversality enables us to prove $S(\mathcal{L})$ is connected.

THEOREM 6.1. *For any positive sequence $\lambda(i) \to 0$, there is a subsequence $\lambda(i_j)$ so that $M(j) = \lambda(i_j)M$ converges to a linear foliation $\mathcal{L}$ of $\mathbb{R}^3$ with connected singular set $S(\mathcal{L})$. Moreover, M is transverse to $\mathcal{L}$, $\mathcal{L}$ is the unique blow-down of M, and the sheeting number of the associated α curves is two.*

Uniqueness of the homothetic blow-down $\mathcal{L}(M)$ implies strong asymptotic convergence properties for M outside the cone C associated to $\mathcal{L}(M)$.

Definition 6.1. Let $\mathcal{H}$ be a solid hyperboloid of revolution with boundary asymptotic to the boundary of the cone C and let $K = \mathcal{H} \cap M$. We define $\mathcal{W}$ to be the closure of $\mathbb{R}^3 - \mathcal{H}$.

We now give the statement and proof of a descriptive theorem on the asymptotic behavior of M, whose proof follows immediately from Theorem 6.1.

THEOREM 6.2. *After a possible homothety of M, $M \cap \mathcal{W}$ is a multisheeted graph over its projection onto the x_1x_2-plane with two simply-connected components, $M(1), M(2)$, each with one boundary component which is a proper arc. After choosing an orientation of M, for any divergent sequence of points in $M(1)$, the sequence of unit normal vectors at these points converges to $(0,0,1)$ and for any divergent sequence of points in $M(2)$ the unit normals converge to $(0,0,-1)$. In particular, the multisheeted graphs $M(1), M(2)$ have asymptotically zero gradient and sublinear growth. Note that $K = M - [\mathrm{Int}(\mathrm{M}(1)) \cup \mathrm{Int}(\mathrm{M}(2))]$ is a strip in M.*

PROOF. Fix a divergent sequence of points $\{p(i)\} \in M \cap \mathcal{W}$ and consider the associated sequence $(1/\|p(i)\|)M = M(i)$ of homothetic scalings of M. If the normal line of M at $p(i)$ does not converge to a vertical line, then the points $q(i) = p(i)/\|p(i)\|$ must have a subsequence $q(i_j)$ that converges to a point $q \in S^2 - \mathrm{Int}(C)$. Since $\mathcal{L}(M)$ is a foliation by horizontal planes, any convergent subsequence of $M(i_j)$ must have q as a singular point. But $q \notin \mathrm{Int}(C)$ which contradicts the property that the singular set of convergence is contained in $\mathrm{Int}(C) \cup \{(0,0,0)\}$. This contradiction proves that outside some compact subset of $M \cap \mathcal{W}$, the Gauss map of M is bounded away from the horizontal. Hence, after a homothety of M, we may assume that the Gauss map on $M \cap \mathcal{W}$ is bounded away from the horizontal and is asymptotic to the vertical. By Theorem 3.1, $\partial(M \cap \mathcal{W})$ has two boundary curves and hence $M \cap \mathcal{W}$ consists of two components which are multisheeted graphs over their projections to the x_1x_2-plane. Since M separates $\mathbb{R}^3$, these two multisheeted graphs have opposite orientations. The theorem now follows from these observations. □

The following theorem from [**30**] is used in conjunction with Theorem 6.2 to show that not only is $\mathcal{L}$ transverse to M but that each plane in $\mathcal{L}$ intersects M in a connected proper arc, as is the case of the helicoid.

THEOREM 6.3. *Suppose that G is a minimal graph of a function f defined on a proper domain $\mathcal{D}$ of $\mathbb{R}^n$. Assume f has zero boundary values and the gradient of f is bounded. Then there are at most a finite number of components of $\mathcal{D}$ where f is nonconstant.*

It is a conjecture of Meeks that when $n = 2$ the graph G described above can have at most two nonplanar components independent of any gradient bound and in the sublinear growth case only one nonplanar component.

The above theorem is proved in [**30**]. Recently Spruck [**43**] made some progress on the above conjecture of Meeks by proving that in dimension two there can be at most two components in the bounded gradient case. Li and Wang [**23**] have also made some progress on the conjecture of Meeks by proving that in all dimensions there can be only a finite number of components; it remains to prove that the number of these components is two when $n = 2$.

In Theorem 6.2 we described a decomposition of M into three components—a smooth proper strip K and two disks $M(1), M(2)$, each having one noncompact boundary component and such that the projection π of $M(1) \cup M(2)$ to the x_1x_2-plane is a proper submersion onto its image. This special decomposition of M is used to prove that the holomorphic function $h = x_3 + ix_3^*: M \to \mathbb{C}$ is a conformal diffeomorphism (for details, see [**30**]). This result follows from the proof of the nonexistence of certain asymptotic curves for h.

Definition 6.2. An integral curve $\gamma: [0, \infty) \to M$ of ∇x_3 or $-\nabla x_3$ is an asymptotic curve with asymptotic value $\gamma(\infty) \in \mathbb{R}$ if $\lim_{t \to \infty} x_3(\gamma(t)) = \gamma(\infty)$.

PROPOSITION 6.4. *If none of the integral curves of ∇x_3 or $-\nabla x_3$ are asymptotic curves, then $h = x_3 + ix_3^*: M \to \mathbb{C}$ is a conformal diffeomorphism.*

The proof of Proposition 6.4 appears in [**30**] and we include a sketch here. Using elementary properties of integral curves, one first checks that if there are no asymptotic curves on M, then $h: M \to \mathbb{C}$ is a conformal diffeomorphism with its image. Next using the fact proved in [**8**] or [**28**] that the intersection of M with a closed half-space in $\mathbb{R}^3$ is parabolic, it is easy to check that h must be onto. This completes our sketch of the proof.

Proposition 6.4 reduces the proof of showing $h: M \to \mathbb{C}$ is a conformal diffeomorphism to demonstrating that ∇x_3 and $-\nabla x_3$ have no asymptotic integral curves. In [**30**] we prove that the existence of an asymptotic integral curve yields an infinitely disconnected graph G contained in $M(1) \cup M(2)$ with boundary values in a fixed horizontal plane, which is impossible by Theorem 6.3. A first step towards proving the nonexistence of asymptotic curves is the following lemma.

LEMMA 6.5. *If W is a component of $x_3^{-1}((-\infty, 0])$ with an infinite number of boundary curves, then W is conformally diffeomorphic to the closed unit disk D with a closed countable set E removed from ∂D. The set E has a single limit point $*$ and $*$ is a limit of points in E on both sides of $*$ on ∂D. Also, $K \cap W$, considered to be a subset of D, is noncompact with $*$ as its unique limit point in ∂D, where K is the strip in M defined in Definition 6.1.*

Note that in the above lemma there is no mention of asymptotic integral curves; the lemma is used later to prove the nonexistence of asymptotic curves.

It turns out that there are two kinds of asymptotic integral curves. The first type, which we call *simple*, have the following property: Suppose α is an asymptotic integral curve for ∇x_3, α is contained in a component W of ${x_3}^{-1}(C - \infty, \alpha(\infty))]$ and, considered as a proper arc in W, α corresponds to an isolated end of W. Using the above lemma, one first shows that none of the integral curves of ∇x_3 or $-\nabla x_3$ are simple. This result and further analysis is then used to prove that there are no

asymptotic curves which are not simple, which completes the proof of the following theorem.

THEOREM 6.6. *There are no asymptotic curves for ∇x_3 or $-\nabla x_3$. In particular, $h = x_3 + ix_3^*: M \to \mathbb{C}$ is a conformal diffeomorphism.*

Every conformal minimal immersion $f: \Sigma \to \mathbb{R}^3$ has an analytic representation in terms of a holomorphic 1-form dh, where $h = x_3 + ix_3^*$, and the stereographic projection $g: \Sigma \to \mathbb{C} \cup \{\infty\}$ of its Gauss map $G: \Sigma \to S^2$. Namely, assuming $f(p_0) = (0,0,0)$, then f can be recovered from dh and g by integration:

$$f(p) = \mathbb{R}e \int_{p_0}^{p} \left(\frac{1}{2}\left(\frac{1}{g} - g\right) dh, \frac{1}{2}\left(\frac{1}{g} + g\right) i\, dh, dh \right).$$

The above representation is called the Weierstrass representation of $f: \Sigma \to \mathbb{R}^3$. If the surface in question is a vertical helicoid, then we can take $\Sigma = \mathbb{C}$, $g(z) = e^{\alpha z}$ for some $\alpha \in i\mathbb{R}$, and $dh = dz$.

By Theorem 6.6 and elementary covering space theory, we know that, after a possible rotation, a properly embedded simply-connected minimal surface M in $\mathbb{R}^3$ can be parametrized by $\mathbb{C}$ with $x_3 = \mathbb{R}e(z)$ and $g(z) = e^{H(z)}$ for some entire function $H(z)$ on $\mathbb{C}$. It remains to prove that $H(z) = \alpha z$ for some $\alpha \in i\mathbb{R} - \{0\}$.

If $H(z) = a_n z^n + \cdots + a_0$ is a nonconstant polynomial, then $g(z) = e^{H(z)}$ has an essential singularity at infinity. In this case there exists a sequence of points $p(i)$ in $\mathbb{C}$ with $|p(i)| \to \infty$ and $|g(z)| \to 1$. Since the Gaussian curvature can be expressed in $\mathbb{C}$ coordinates by

$$K = -16\left(\frac{|g||g'|}{(1+|g|^2)^2}\right)^2 = -16|H'|^2\left(\frac{|g|^2}{(1+|g|^2)^2}\right)^2,$$

then $K(p(i)) \to -\infty$ if $H(z)$ is not a polynomial of degree one. If $H(z) = az + b$, then, after a conformal affine change of coordinates w, $H(w) = iw$ and $dw = cdz$ for some $c \in \mathbb{C}$, which means that M is an associate surface to a vertical helicoid. But such a surface is well-known to be embedded only if it is actually a vertical helicoid (see, for example, [**38**]). This gives a proof of the following lemma.

LEMMA 6.7. *A complete minimal surface M defined as $f: \mathbb{C} \to \mathbb{R}^3$ with $x_3 = \mathbb{R}e(z)$ and $g(z) = e^{H(z)}$, where $H(z)$ is a nonconstant polynomial, has bounded nonzero Gaussian curvature if and only if $H(z)$ is linear. In particular, by Theorems 6.1, 6.6, if M is embedded, has bounded curvature and $H(z)$ is a polynomial, then M is a vertical helicoid.*

The above lemma reduces the proof of showing M is a helicoid to proving that $H(z)$ is a linear function. We first sketch the proof that $H(z)$ is a polynomial.

If $H(z)$ is not a polynomial, then for some latitude α of S^2, $G^{-1}(\alpha)$ consists of an infinite number of proper arcs on M where $G: M \to S^2$ is the Gauss map. After choosing the solid hyperboloid $\mathcal{H}$ large enough, $G^{-1}(\alpha)$ is contained in $\mathcal{H}$. Suppose that an infinite number of ends of $G^{-1}(\alpha)$ have unbounded positive x_3-coordinates. Since some blow-down of M converges to a foliation by horizontal planes with singular set S, there exists a sequence of minimal surfaces $M(i)$ obtained from M by first translating by a divergent sequence of vectors $-v(i) = (x_1(i), x_2(i), x_3(i))$, followed by homotheties by multiplication by $\lambda(i) \in \mathbb{R}^+$ such that the sequence of surfaces $M(i) = \lambda(i)(M - v(i))$ converges with multiplicity one to a properly embedded simply-connected minimal surface $\hat{M}$ of bounded nonzero curvature with $\mathcal{L}(\hat{M})$ a

foliation by horizontal planes. However, only a finite number of the infinite number of proper arcs ${G_i}^{-1}(\alpha)$ can be contained in the corresponding solid hyperboloid $\hat{\mathcal{H}}$ for $\hat{M}$. By applying an appropriately defined new sequence of homotheties to the $M(i)$, the new surfaces again converge to a foliation of $\mathbb{R}^3$ by horizontal planes with a singular curve which passes through the origin but is not contained in the vertical double cone which curvature estimates demand. However, $\tau(i)$ is chosen so that $\tau(i)G_i^{-1}(\alpha)$ intersects the unit circle in the x_1x_2-plane, which means that the singular curve of convergence is not contained in a vertical double cone. This contradiction proves one can choose an $\alpha \subset S^2$, so that $G^{-1}(\alpha)$ consists of a finite number of proper arcs on M, which implies $H(z)$ is a polynomial.

Lemma 6.7 now implies that M is a helicoid if it has bounded curvature. However, in the proof that $H(z)$ is polynomial, we produced a surface $\hat{M}$ of bounded curvature with $\mathcal{L}(\hat{M})$ being a foliation by horizontal planes. Hence, $\hat{H}(z)$ is polynomial of degree one. But, making appropriate choices in the construction of $\hat{M}$, one checks that $\text{degree}(H(z)) = \text{degree}(H_i(z))$ is at most $\text{degree}(\hat{H}(z)) = 1$. Hence, $H(z)$ has degree one, which, by Lemma 6.7, proves in the general case that M is a vertical helicoid. This completes our sketch of the proof of Theorem 1.1.

References

[1] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold I; estimates off the axis for disks. Preprint.

[2] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold II; multi-valued graphs in disks. Preprint.

[3] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold III; planar domains. Preprint.

[4] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold IV; locally simply-connected. Preprint.

[5] T. H. Colding and W. P. Minicozzi II. The space of embedded minimal surfaces of fixed genus in a 3-manifold V; fixed genus. Preprint.

[6] T. H. Colding and W. P. Minicozzi II. Complete properly embedded minimal surfaces in $\mathbb{R}^3$. *Duke Journal*, 107(2):421–426, 2001.

[7] P. Collin. Topologie et courbure des surfaces minimales de $\mathbb{R}^3$. *Annals of Math. 2nd Series*, 145–1:1–31, 1997.

[8] P. Collin, R. Kusner, W. H. Meeks III, and H. Rosenberg. The geometry, conformal structure and topology of minimal surfaces with infinite topology. Preprint.

[9] C. Costa. Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends. *Bull. Soc. Bras. Mat.*, 15:47–54, 1984.

[10] M. do Carmo and C. K. Peng. Stable minimal surfaces in $\mathbb{R}^3$ are planes. *Bulletin of the AMS*, 1:903–906, 1979.

[11] D. Fischer-Colbrie. On complete minimal surfaces with finite Morse index in 3-manifolds. *Invent. Math.*, 82:121–132, 1985.

[12] D. Fischer-Colbrie and R. Schoen. The structure of complete stable minimal surfaces in 3-manifolds of non-negative scalar curvature. *Comm. on Pure and Appl. Math.*, 33:199–211, 1980.

[13] M. Freedman, J. Hass, and P. Scott. Least area incompressible surfaces in 3-manifolds. *Invent. Math.*, 71:609–642, 1983.

[14] H. Fujimoto. *Value distribution theory of the Gauss map of surfaces in* $\mathbf{R}^m$. Vieweg, 1993.

[15] R. Hardt and L. Simon. Boundary regularity and embedded minimal solutions for the oriented Plateau problem. *Annals of Math.*, 110:439–486, 1979.

[16] L. Hauswirth, J. Pérez, and P. Romon. Embedded minimal ends of finite type. *Transactions of the AMS*, 353:1335–1370, 2001.

[17] D. Hoffman, H. Karcher, and F. Wei. The genus one helicoid and the minimal surfaces that led to its discovery. In *Global Analysis and Modern Mathematics*. Publish or Perish Press, 1993. K. Uhlenbeck, editor, p. 119–170.

[18] D. Hoffman and W. H. Meeks III. Complete embedded minimal surfaces of finite total curvature. *Bulletin of the AMS*, 12:134–136, 1985.
[19] D. Hoffman and W. H. Meeks III. The strong halfspace theorem for minimal surfaces. *Invent. Math.*, 101:373–377, 1990.
[20] D. Hoffman, M. Weber, and M. Wolf. The existence of the genus-one helicoid. Preprint.
[21] H. B. Lawson, Jr. *Lectures on Minimal Submanifolds*. Publish or Perish Press, Berkeley, 1971.
[22] H. Lazard-Holly and W. H. Meeks III. The classification of embedded doubly–periodic minimal surfaces of genus zero. *Invent. Math.*, 143:1–27, 2001.
[23] P. Li and J. Wang. Finiteness of disjoint minimal graphs. Preprint.
[24] W. H. Meeks III. The geometry and topology of singly-periodic minimal surfaces. to appear in the Asian J. of Math.
[25] W. H. Meeks III. The theory of triply-periodic minimal surfaces. *Indiana Univ. Math. J.*, 39(3):877–936, 1990.
[26] W. H. Meeks III. The geometry, topology, and existence of periodic minimal surfaces. *Proceedings of Symposia in Pure Math.*, 54:333–374, 1993. Part I.
[27] W. H. Meeks III, J. Pérez, and A. Ros. The geometry of minimal surfaces of finite genus I; curvature estimates and quasiperiodicity. Preprint.
[28] W. H. Meeks III and H. Rosenberg. Maximum principles at infinity with applications to minimal and constant mean curvature surfaces. Preprint.
[29] W. H. Meeks III and H. Rosenberg. The theory of minimal surfaces in $M \times \mathbb{R}$. Preprint.
[30] W. H. Meeks III and H. Rosenberg. The uniqueness of the helicoid and the asymptotic geometry of properly embedded minimal surfaces with finite topology. Preprint.
[31] W. H. Meeks III and H. Rosenberg. The global theory of doubly periodic minimal surfaces. *Invent. Math.*, 97:351–379, 1989.
[32] W. H. Meeks III and H. Rosenberg. The maximum principle at infinity for minimal surfaces in flat three-manifolds. *Comment. Math. Helvetici*, 65:255–270, 1990.
[33] W. H. Meeks III and H. Rosenberg. The geometry and conformal structure of properly embedded minimal surfaces of finite topology in $\mathbb{R}^3$. *Invent. Math.*, 114:625–639, 1993.
[34] W. H. Meeks III, L. Simon, and S. T. Yau. The existence of embedded minimal surfaces, exotic spheres and positive Ricci curvature. *Annals of Math.*, 116:221–259, 1982.
[35] W. H. Meeks III and S. T. Yau. The existence of embedded minimal surfaces and the problem of uniqueness. *Math. Z.*, 179:151–168, 1982.
[36] W. H. Meeks III and S. T. Yau. The topological uniqueness of complete minimal surfaces of finite topological type. *Topology*, 31(2):305–316, 1992.
[37] X. Mo and R. Osserman. On the Gauss map and total curvature of complete minimal surfaces and an extension of Fujimoto's Theorem. *J. of Differential Geometry*, 131(2):343–355, 1990.
[38] L. Rodriguez and H. Rosenberg. Some remarks on complete simply connected minimal surfaces meeting the planes $x_3 =$ constant transversally. *J. Geom. Anal.*, 7(2):329–342, 1997.
[39] L. Rodriguez and H. Rosenberg. Minimal surfaces in $\mathbb{R}^3$ with one end and bounded curvature. *Manuscripta Math.*, 96(1):3–7, 1998.
[40] H. Rosenberg. Minimal surfaces of finite type. *Bull. Soc. Math. France*, 123:351–354, 1995.
[41] R. Schoen. *Estimates for Stable Minimal Surfaces in Three Dimensional Manifolds*, volume 103 of *Annals of Math. Studies*. Princeton University Press, 1983.
[42] L. Simon. Lectures on geometric measure theory. In *Proceedings of the Centre for Mathematical Analysis* 3, Australian National University, Canberra, Australia, 1983.
[43] J. Spruck. Two-dimensional minimal graphs over unbounded domains. Preprint.
[44] F. Xavier. Convex hulls of complete minimal surfaces. *Math. Ann.*, 269:179–182, 1984.

Mathematics Department, University of Massachusetts, Amherst, MA 01003
E-mail address: bill@gang.umass.edu
E-mail address: hrose@noos.fr

Clay Mathematics Proceedings
Volume 2, 2004

Constructions of Minimal Surfaces by Gluing Minimal Immersions

Nikolaos Kapouleas

Abstract. This is a survey where various gluing constructions for constant mean curvature and minimal surfaces are discussed. The general ideas and various important technical issues are explained, both in the general framework and in the specific aspects of the various constructions. Analogies and differences between the various constructions are pointed out. The constructions discussed are gluing constructions for Delaunay or Wente surfaces, connected sum constructions for minimal surfaces, desingularizing the curve of intersection constructions for minimal surfaces in a Riemannian three-manifold, and a doubling construction for the Clifford torus in the round three-sphere.

Contents

1. The General Outline of a Gluing Construction 490
2. The Initial Surfaces for Gluing Delaunay Surfaces 491
3. Forces and Balancing 492
4. Sufficient Conditions for the Delaunay Gluing 493
5. Remarks on the Theorem 494
6. Compact Surfaces by Gluing Delaunay Surfaces 495
7. Gluing Wente Tori 496
8. Constructions of Minimal Surfaces 497
9. Connected Sum Constructions 497
10. Remarks on Yang's Theorem 498
11. The Motivation for Desingularizing Constructions and Scherk's Singly Periodic Surfaces 499
12. Highly Symmetric Desingularizing Constructions 500
13. Desingularizing the Intersections of Compact Surfaces 501
14. Comparison with the Delaunay Gluing Construction 502
15. The Case of Noncompact Complete Surfaces with Closed Curve of Intersection 504
16. Applications and Open Problems 505
17. Doubling Constructions 507
18. The General Framework for Understanding the Linearized Equation 509
19. The Obstructions and the Extended Substitute Kernel $\widetilde{\mathcal{K}}$ 511
20. Resolving the Difficulties and the Geometric Principle 513

©2004 Clay Mathematics Institute

21. The Constant Mean Curvature Constructions 516
22. Yang's Connected Sum Construction 517
23. The Desingularizing Constructions 518
24. The Doubling of the Clifford Torus Construction 522
References 522

The subject of these lectures is constructions for constant mean curvature and minimal two-dimensional surfaces by singular perturbation methods. These methods have been among the most successful in Geometric Partial Differential Equations. In particular, they have been applied extensively and with great success in Gauge Theories. The particular kind of methods discussed here relates more closely to the work of R. Schoen on the construction of certain metrics of constant scalar curvature [Sc2], and also the work on gluing Delaunay surfaces [K1]. They evolved further and were systematized in the work on gluing Wente tori [K3]. These talks emphasize this systematic approach, and its application to constructions for minimal surfaces. In the first part I review the main theorems which have been proved, and in the second part I discuss important aspects of the proofs. Presenting all these constructions together helps to illuminate the underlying ideas by pointing out the similarities and idiosyncrasies of the various cases. All gluing constructions share a general set-up which I describe now as it applies to surfaces:

1. The General Outline of a Gluing Construction

In the first step of the construction, simpler surfaces which satisfy the geometric condition — $H \equiv 1$ or $H \equiv 0$ for these talks — are combined to construct an initial surface which is more complicated and which satisfies the condition approximately. More precisely, the construction of the initial surface depends on parameters. When certain of these parameters, which I call τ, are small enough, then $H - 1$ or H is small in appropriate norms. When $\tau \to 0$, the initial surface has a singular limit, similar to the limit the Delaunay surfaces have, as discussed in the next section.

The second step involves small perturbations of the initial surface. These can be described as graphs of a section of the normal bundle, or, when there is a global unit normal ν and since the codimension is one, of a function on the surface. I denote by f the section or the function, by $X : M \to N$ the immersion of the initial surface M in the ambient manifold N, and by $X_f : M \to N$ the graph of f. Let H and H_f be the mean curvature of X and X_f. It follows then that

$$H_f = H + \mathfrak{L}f + \mathfrak{Q}_f,$$

where $\mathfrak{L} = \frac{1}{2}(\Delta + |A|^2 + \mathrm{Ric}(\nu, \nu))$ is the linearized operator, and $\mathfrak{Q}_f$ is a quadratic and higher order expression in f and its derivatives, with coefficients involving the geometric invariants of X. X_f is the desired perturbation when $H_f \equiv 1$ or $H_f \equiv 0$. Assuming that the smallness of f allows us to ignore for a moment the higher order term $\mathfrak{Q}_f$, the equation reduces to

$$\mathfrak{L}f = -H_{\mathrm{err}},$$

where $H_{\mathrm{err}} := H - 1$ in the constant mean curvature case, and $H_{\mathrm{err}} := H$ in the minimal surface case. Assuming that the linear operator $\mathfrak{L}$ has an inverse $\mathfrak{L}^{-1}$,

and by taking $f = u + v$, where $u = -\mathcal{L}^{-1} H_{\text{err}}$, that is $\mathcal{L}u + H_{\text{err}} = 0$, the original equation reduces to

$$v = -\mathcal{L}^{-1} \mathcal{Q}_{u+v}.$$

Finding then a solution v, amounts to finding a fixed point for the map $v \to -\mathcal{L}^{-1}\mathcal{Q}_{u+v}$.

To accomplish this, appropriate norms and spaces have to be defined so that H_{err}, and then u, have small norm, when τ is close to 0. When v is appropriately restricted to have small norm, $\mathcal{Q}_{u+v}$, and then $\mathcal{L}^{-1}\mathcal{Q}_{u+v}$, should have even smaller norm because of the quadratic nature of $\mathcal{Q}_{u+v}$. The corresponding estimates have to be appropriately uniform in the parameters. The existence of the solution v can be established by employing some appropriate fixed point theorem. The strongest results can be obtained by using the contraction mapping theorem, but the estimates required are the most demanding. An often preferable alternative is the Schauder fixed point theorem [GT], which minimizes the required estimates.

This general idea of the construction is very similar to the proof of the inverse function theorem, the main difference being that the limiting configuration when $\tau \to 0$, is singular. Because of this the method is called a singular perturbation method. Although this approach works as described in some cases (for example [Sm]), in the constructions under discussion a major difficulty is the presence of small eigenvalues of the linearized operator $\mathcal{L}$. These are inherited from the non-trivial kernels $\mathcal{L}$ has on the simpler surfaces, out of which the initial surfaces are constructed. These small eigenvalues make it impossible to define and estimate the inverse of $\mathcal{L}$ in a straightforward way as above. Dealing with this difficulty is a major issue and will be discussed in detail in the second part of these lectures starting with Section 18. The basic idea is the introduction of various modifications of the initial surface controlled by some new parameters. The surface constructed satisfies the geometric condition when both f and the value of the parameters are correctly chosen by a modified application of the fixed point theorem.

2. The Initial Surfaces for Gluing Delaunay Surfaces

The first construction I discuss is a gluing construction for constant mean curvature surfaces in the Euclidean three-space [K1]. In this construction Delaunay surfaces are "fused" together along regions which approximate round spheres. Although the constructions in [Sc2] and [K1] are the earliest, the theorems proven can serve as prototypes of the general kind of theorem one can expect.

The Delaunay surfaces are the rotationally invariant surfaces of constant mean curvature, and they were first studied by Delaunay [D]. By scaling appropriately, their mean curvature can be assumed to be the same as that of the unit sphere, which I take to be 1. The Delaunay surfaces are singly-periodic and they form a one-parameter family. The embedded ones are called unduloids and the rest nodoids. They can be parametrized by a parameter τ which is defined by $\tau := 2\pi(r - r^2)$ for the unduloids and $\tau := 2\pi(-r - r^2)$ for the nodoids, where r is the smallest radius of the meridian circles. The physical significance of τ is explained in the next section. τ is positive for the unduloids and negative for the nodoids.

When $\tau \to 0$ both the unduloids and nodoids tend to a sequence of round spheres touching externally at points on the axis. The Delaunay surfaces, with the exception of the cylinder, consist of alternating positive and negative Gauss curvature regions, one of each in a fundamental domain under the translational period.

The positive curvature regions tend to the round sphere minus two antipodal points as $\tau \to 0$. The convergence is uniform in all norms on compact subsets of the sphere minus the two points. The negative curvature regions collapse to one point each as $\tau \to 0$. To recover the negatively curved regions in the limit, a magnification by a factor $1/\tau$ can be applied, and then they converge to the catenoid as $\tau \to 0$. The convergence is again uniform on compact subsets of the catenoid in all norms. In the unduloids the outward normal of the spherical region extends to the outward (away from the axis) normal of the catenoidal region, while in the nodoids to the inward normal.

The initial surfaces are constructed by combining Delaunay surfaces and round spheres. The construction is codified by a graph which consists of vertices, edges, rays, and weights assigned to the edges and rays. The vertices are points which correspond to the round spheres used in the construction of the initial surfaces. The edges are line segments of even integer length whose endpoints are vertices. They correspond to pieces of Delaunay surfaces used to connect the round spheres corresponding to their endpoints. The length of the edges is chosen even, because the period of the Delaunay surfaces tends to 2 as $\tau \to 0$. When the length of the edge is $2k$, the corresponding Delaunay piece consists of $k+1$ regions of positive Gauss curvature, two of which are positioned at the extremes, while the other $k-1$ together with k regions of negative Gauss curvature are in between. The "extreme" regions of positive Gauss curvature, are "centered" at points at distance approximately $2k$ from each other; I call this distance the length of the Delaunay piece. The rays are half-lines with a vertex as the endpoint; each corresponds to a Delaunay end, attached to the round sphere corresponding to the endpoint. Finally the weights are nonvanishing real numbers whose function is to determine the parameters of the Delaunay pieces: Each Delaunay piece has parameter $\tau = \hat{\tau} w$, where w is the weight of the corresponding edge or ray, and $\hat{\tau}$ is a parameter of the construction.

The construction of the initial surface is relatively straightforward. The edges, rays, the weights, and $\hat{\tau}$, determine the Delaunay pieces to be used. They have to be placed appropriately, so that the positively curved regions positioned at the extremes, are centered at the corresponding vertex of the graph. Since the lengths of the Delaunay pieces differ from those of the corresponding edges, to achieve this, small changes in the graph are required, as discussed in Section 4. By removing small discs from the round spheres then, and transiting from the sphere to the positively curved region of the Delaunay, the initial surface is constructed. The transition is realized by the use of cut-off functions, and introduces annuli on the initial surface where $H-1$ is supported. Moreover $H-1 = O(\hat{\tau})$ in all norms, and the supporting annuli are independent of $\hat{\tau}$ by construction. Not all initial surfaces constructed this way can be corrected to constant mean curvature by a small perturbation. Understanding the obstruction is facilitated by the balancing formula discussed in the next section.

3. Forces and Balancing

The presentation of the balancing formula can be motivated, by recalling that a surface M of constant mean curvature (cmc for short), in the Euclidean space, can be considered as a fluid interface in equilibrium, where the only forces are the surface tension and the pressure jump from one side of the surface to the other.

Following this description one can consider a "virtual cut", that is an embedded surface transverse to the cmc surface and which separates the whole physical system into two parts. The force exerted by one part to the other is then

$$F = \int_{M\cap\hat{M}} \boldsymbol{\eta} - 2\int_{\hat{M}} \boldsymbol{\nu}, \tag{1}$$

where it is assumed that the pressure "outside" the cmc surface is 0, the pressure "inside" and the surface tension are measured in suitable physical units, $\hat{M}$ is the intersection of the "virtual cut" with the region enclosed by the cmc surface, $M\cap\hat{M}$ is the curve of intersection of the cmc surface with the virtual cut, $\boldsymbol{\eta}$ is the unit conormal of $M\cap\hat{M}$, tangent to the cmc surface and pointing in the direction of the part which is exerting the force, and $\boldsymbol{\nu}$ is the unit normal to $\hat{M}$ and likewise pointing in the direction of the part which is exerting the force.

In the case M is a minimal surface in the Euclidean space, and $\hat{M}$ is a "virtual cut" as before, the only force is due to surface tension, and then the force exerted by one part to the other is given by

$$F = \int_{M\cap\hat{M}} \boldsymbol{\eta}. \tag{2}$$

In more generality, by considering the first variation of the area, it follows [Si, Sc1] that

$$\int_M H\boldsymbol{\nu}\cdot Y = \frac{1}{2}\int_{\partial M} \boldsymbol{\eta}\cdot Y, \tag{3}$$

where M is a compact surface in a three-dimensional Riemannian manifold, ν is the unit normal, H the mean curvature, $\boldsymbol{\eta}$ is the conormal to M pointing away from M at ∂M, and Y is a Killing field of the ambient manifold. By subtracting the first variation of the volume [KKS] it follows that

$$\int_M (H-1)\boldsymbol{\nu}\cdot Y = \frac{1}{2}\int_{\partial M} \boldsymbol{\eta}\cdot Y - \int_{\hat{M}} \boldsymbol{\nu}\cdot Y, \tag{4}$$

where $\hat{M}$ is such that $M\cup\hat{M}$ encloses an open region of the ambient manifold, ν on M is the inward normal while on $\hat{M}$ it is the outward normal, and the conormal η points away from M as before. In the framework of the previous paragraph and equation (1), if two virtual cuts are considered separating a bounded part of the system, (3) implies that the total force on this bounded part vanishes, as it should, since the system is in equilibrium.

Consider now "virtual cuts" separating the two ends of an unduloid and with $\hat{M}$ connected and bounded. The corresponding force is independent of the particular "virtual cut", because the union of two such "cuts" separates a bounded part of the system on which the total force has to vanish. Properly interpreted these statements are valid for a nodoid as well. By using a cut through the meridian circle of least radius, it is straightforward to calculate that the force is in the direction of the axis and its signed size is τ — attractive for unduloids and repulsive for nodoids.

4. Sufficient Conditions for the Delaunay Gluing

The first condition I assume is called the "balancing" condition: This condition asserts that at each vertex p of the given graph

$$F_p := \sum_e w_e \vec{v}_e = 0$$

holds, where e is any edge or ray with p as an endpoint, w_e is the weight corresponding to e, and $\vec{v}_e$ is the unit vector parallel to e and pointing away from p. Assuming that the Delaunay pieces of the initial surface have their axes parallel to the corresponding edges and rays, it follows that the resultant force $\hat{F}_p$ exerted through a virtual cut enclosing a neighborhood of the vertex satisfies $\hat{F}_p = \hat{\tau} F_p$. If $\hat{F}_p \neq 0$, then any appropriate small perturbation of the initial surface would still have a nonvanishing force through the same virtual cut, contradicting the possibility of finding a small perturbation of constant mean curvature: The persistence of the nonvanishing of the force for small perturbations becomes clear by choosing the virtual cut so it passes through the meridian circles of least radius on the Delaunay pieces. These radii are of order $\hat{\tau}$, so "small perturbation" should, and in the construction does, mean that the changes in the vicinity of these circles, including the change of the force, is small compared to $\hat{\tau}$, ensuring the persistence of the nonvanishing.

Besides the balancing condition, a "flexibility" condition is needed. The reason is that the period of the Delaunay surfaces is not exactly 2, although it tends to 2 in the singular limit. Hence the length of the edges has to change somewhat, but there are cases where the graph is too rigid, as in item 5 of the next section, and this is not possible. The flexibility condition stipulates that small changes in the lengths of the edges can be arbitrarily prescribed, and the resulting graph depends continuously on them.

Finally, an "unbalancing" condition is needed. The reason is that the process of the construction introduces changes in the resultant forces, and a way to ensure that their vanishing in the final surface can be arranged, is to be able to prescribe any small resultant force on each vertex. This resultant force can be used then to cancel the change introduced during the process of the construction. The unbalancing and flexibility conditions have to hold simultaneously, so that any small change of lengths and resultant forces can be prescribed simultaneously. A version of the main theorem is then the following:

THEOREM A ([K1]). *Given any graph as above, if $|\hat{\tau}|$ is small enough depending on the graph, the corresponding initial surface can be perturbed to satisfy the constant mean curvature condition, provided the graph satisfies the balancing, unbalancing, and flexibility conditions.*

5. Remarks on the Theorem

1. The ends of the cmc surface constructed decay exponentially to Delaunay ends which may have slightly different parameters — axis and force — than the ones implied by the corresponding ray.

2. This theorem and work by Meeks [M] motivated results [KKS; KK] establishing that all embedded complete cmc surfaces share many of the features of the surfaces constructed, for example their ends decay exponentially to Delaunay ends.

3. The estimates in [K1] of the decay of the solution between successive positive or negative Gauss curvature regions are not optimal. It was necessary to substantially strengthen them in [K3], and this was achieved by expanding and refining the general approach. By following the new approach and using the stronger estimates, both the results and the proofs in [K1] can be substantially strengthened. For example, the embeddedness of many more surfaces can be asserted, belonging to a large class of finite topological types.

4. If the original graph is symmetric with respect to a group of isometries, then the same symmetries can be imposed on the whole construction. The unbalancing and flexibility conditions have to be modified then, to require prescribing only resultant forces and length changes which are consistent with the symmetries imposed.

5. It is not possible to construct a closed cmc surface by a direct application of Theorem A, because there is no rayless graph satisfying all the conditions: Although it is easy to arrange balancing and unbalancing, it is not possible to arrange flexibility also. For example, consider the graph whose vertices are the vertices of an equilateral triangle together with its center, and whose edges are all the possible line segments: the three sides of the triangle and the three radii connecting the center to the other vertices. After imposing the maximal symmetry, there are only two weights to be prescribed, and both balancing and unbalancing can be arranged by fixing one of them and then choosing appropriately the other. Flexibility though does not hold, because once the length of the sides is determined, there is no freedom to choose the length of the radii. With some extra work however [K2], closed cmc surfaces of any genus $g \geq 3$ can be constructed, as discussed shortly.

6. Most finite topological types of surfaces can be realized as complete cmc surfaces by this construction [K1] and its extension in [K2]. For most of these topological types there are no other examples of cmc surfaces known, even today.

7. Since the construction in [K1], it was generally believed that a modification where minimal n-noids instead of round spheres are used to replace some of the vertices of the graph, should be possible. This was proven recently by Mazzeo and Pacard [MP] in the case the graph has a single vertex. The minimal n-noids are minimal surfaces with n ends asymptotic to catenoidal ends, so the surfaces constructed transit from such an end to the catenoidal region of a Delaunay end.

8. It is fairly easy to extend the construction to complete cmc surfaces of unbounded geometry, that is, injectivity radius 0 and unbounded second fundamental form [K1].

6. Compact Surfaces by Gluing Delaunay Surfaces

As I have already mentioned, Theorem A cannot be used as it stands to produce closed cmc surfaces. The obstruction though is rather mild, as in the example discussed in item 5 above. Returning to that example, notice that after constructing the graph so that it is balanced, the construction has the free parameter $\hat{\tau}$. If the number of fundamental domains in the Delaunay piece replacing the sides of the triangle is fixed, then the appropriate length for the Delaunay piece replacing the radii is determined, and it is a continuous function L of $\hat{\tau}$. The change of L as $\hat{\tau}$ varies, can be magnified by choosing the number of fundamental domains large enough. By making the total change longer than the periods of the Delaunay surfaces involved, the existence of some value of $\hat{\tau}$ such that there is a Delaunay piece fitting the radii exactly, can be asserted. This "period closing" approach, by varying $\hat{\tau}$ and choosing long Delaunay pieces, allows the construction of cmc closed surfaces of any genus $g \geq 3$ [K2], as in the next theorem. The graphs mentioned consist of vertices and edges of arbitrary length, to which nonzero real numbers called weights are assigned as before. They have to satisfy the balancing and unbalancing conditions, but the flexibility condition is appropriately modified [K2].

THEOREM B ([K2]). *For each $g \geq 3$ there are infinitely many graphs, which codify the construction of closed cmc surfaces of genus g, where the edges of the graphs correspond to long enough Delaunay pieces.*

While it is a celebrated theorem of Hopf [H] that the only genus 0 closed cmc surface is the round sphere, and of Alexandrov [A] that the only embedded closed cmc surface is again the round sphere, it is open what is true when the "closed" condition is replaced by "compact with boundary a round planar circle". A modification of the above construction [K2], gives the only examples known, of compact cmc surfaces with boundary a round planar circle. These examples can be again of any genus $g \geq 3$.

7. Gluing Wente Tori

The first examples of closed cmc surfaces which were not round spheres were certain cmc tori constructed by H. Wente [W]. Following the construction of closed cmc surfaces of any genus $g \geq 3$ [K2], the only case still open was for closed cmc surfaces of genus 2. This motivated the construction of closed cmc surfaces of any genus $g \geq 2$, by fusing g Wente tori [K3,4]. To describe this construction I first give a brief description of the Wente tori. There is a one-parameter family of cmc cylinders I call the Wente cylinders. There is a strong analogy with the Delaunay cylinders, but many differences as well. In particular the translational periodicity of the Delaunay cylinders is replaced by rotational periodicity.

The parameter of the family of the Wente cylinders cannot be defined by considering the force through a meridian as in the Delaunay case, because this force vanishes, as implied by the symmetries of the surface. A parameter τ can be defined however [K3], so that the regions of positive and negative Gauss curvature demonstrate similar behavior, to the ones of the Delaunay surfaces, as $\tau \to 0$. In this case τ takes positive values only. The fundamental domain of a Wente cylinder consists of a positively curved region and a negatively curved region, as in the Delaunay case. The boundary of this fundamental region consists of two topological circles, which are planar curves resembling the figure 8. The lines normal to the planes of these figures 8, and passing through the self-intersection of each figure 8, intersect at a distance 1/2 from the planes. I call the point of intersection the center of the Wente surface. Each of the normal lines is a line of reflectional symmetry. The angle between the two normal lines varies continuously with τ, and is of order $\sqrt[4]{\tau}$ for small τ. Clearly then the angle is a rational multiple of π for infinitely many choices of τ, and then the Wente cylinder immersion factors through a torus.

As $\tau \to 0$ the regions of positive Gauss curvature tend to a round sphere with a single point removed. The convergence is uniform in all norms on compact subsets of the sphere minus the point; more precisely the Wente surfaces differ from the sphere by $O(\tau)$ in all norms, as in the Delaunay case. The negatively curved regions tend to a point. Magnified by a factor of $1/\tau$ they tend to an Enneper surface, with uniform convergence again on compact subsets of the Enneper surface. An important difference between the Wente and the Delaunay cylinders, is that the set $\{K = 0\}$ where the Gauss curvature vanishes, is connected in the former case, while in the latter it has a finite number of components in each fundamental domain. The part of $\{K = 0\}$ in a fundamental domain has length of order $\sqrt[4]{\tau}$ in the former case, and $\sqrt{\tau}$ in the latter.

The construction of the initial surfaces is by "fusing" g Wente tori. For simplicity the construction in [K3] is done under maximal symmetry. The initial surface consists of the following: First, a round sphere from which g small geodesic discs have been removed; their centers form a regular g-gon inscribed in an equatorial circle. Second, g copies of a Wente torus from which most of a single positively curved region has been removed, so that only a small neighborhood of its boundary is left. Each of these Wente tori is placed so that both the removed region and the whole torus are reflectionally symmetric, with respect to the plane, and one of the radii, of the regular g-gon. Moreover the removed region fits the round sphere as well as possible, and its boundary is close to the corresponding vertex of the regular g-gon. Third, g transitional annuli between the above, on which $H_{\text{err}} = H - 1$ is supported. Since the sphere and the Wente torus over the transition regions are close in order τ as $\tau \to 0$, H_{err} is of order τ in all norms as well.

THEOREM C ([K3]). *For each $g \geq 2$ the above constructed surfaces can be perturbed to give closed cmc surfaces of genus g, provided τ is small enough.*

As mentioned before, to correct these initial surfaces, I had to expand and refine the general philosophy, in order to produce more precise estimates than the ones in [K1]. This general philosophy is discussed in the second part starting with Section 18.

8. Constructions of Minimal Surfaces

Various gluing constructions for minimal surfaces have been developed. The constructions I will discuss can be classified into three groups:

1. Connected sum constructions.
2. Desingularizing curves of intersection constructions.
3. Doubling constructions.

In all these constructions small "necks" or "handles" are introduced to a given configuration of minimal surfaces, to produce more complicated minimal surfaces. In the limit where the introduced necks tend to vanishing size, the given configuration of minimal surfaces is recovered. In the connected sum constructions the number of necks is independent of their size, while in the other two cases the number of necks tends to infinity as their size tends to 0. In constructions of the second kind, the given configuration consists of minimal surfaces intersecting along curves, and the "necks" are placed along the curves of intersection. In constructions of the third kind, the given configuration consists of two — or perhaps more — sheets of a given minimal surface, and the necks are used to connect the sheets.

9. Connected Sum Constructions

In order for a connected sum construction to be carried out, a model for the connecting bridge is needed. The obvious model for this is the catenoid. A natural class of minimal surfaces to attempt connected sum constructions, is the class of complete minimal surfaces of finite total curvature in Euclidean three-space. This seems the most interesting class since the connected sum of embedded such surfaces cannot be embedded in general. A general theorem in this context has been proven by S. D. Yang:

Assume a finite collection of surfaces in the class above given, and on them a finite collection of chosen points together with a choice of one side for each point.

The construction requires that at each chosen point, a small neighborhood of the point is removed, and a small truncated catenoid is attached. Each catenoid can be attached to two different points in which case it serves as a bridge connecting two surfaces, or at a single point in which case it adds a small catenoidal end to the surfaces in consideration. A graph can be used to codify the way the surfaces are connected, with the vertices corresponding to the minimal surfaces, and the catenoidal bridges corresponding to edges either connecting two vertices, or attached to a single vertex.

An important issue in this construction is the details of the construction of the initial surface. The simplest approach is to transit from the catenoidal bridge to the corresponding minimal surface by using cut-off functions, after the catenoid is placed in an optimal position relative to the minimal surface. It turns out that the resulting initial error is optimized in the norms used, by arranging the transition in the vicinity of the meridian circles of the catenoid with radius of order $\sqrt{\tau}$, where τ is the size of the corresponding catenoid. After constructing carefully the initial surface in this way, Yang proves that it can be perturbed to minimality under certain conditions:

THEOREM D ([Y1, Y2]). *The initial surface constructed can be perturbed to minimality provided that the catenoidal bridges are small enough, the graph corresponding to the construction is a tree, and the given minimal surfaces have no exceptional bounded Jacobi fields.*

Exceptional bounded Jacobi fields are defined to be bounded solutions of the linearized equation which are not generated by a translation. In some sense this holds for a generic surface, and there are many examples of surfaces where it holds as well [PR].

10. Remarks on Yang's Theorem

1. The nonexistence of exceptional bounded Jacobi fields is a nondegeneracy condition. It is often the case, that in general constructions, such a nondegeneracy condition has to be assumed.
2. The asymptotics of some of the ends have to change during the construction. Yang had to study carefully the force exerted by the various possible kinds of ends to the rest of the minimal surface, and prove that it is always possible to prescribe a required change of force. This, together with a study of certain changes in the catenoidal neck, imply effectively that the condition analogous to the "unbalancing" condition of the Delaunay gluing is always satisfied in this set-up.
3. Yang assumes that the graph is a tree, to avoid obstructions arising from difficulties in fitting the surfaces together. It would be interesting to relax the tree condition on the graph and understand the corresponding obstructions.
4. The sizes of the necks are parameters of the construction. In the limit as they tend to 0, the original minimal surfaces are recovered.
5. This construction is not suitable as it stands for closed minimal surfaces in the round three-sphere, because in this case the force exerted by the neck cannot be balanced by changing the force of an end, since a closed surface has no ends.
6. For cmc surfaces a similar set-up can be developed, where the role of the bridges can be played either by catenoids, or by Delaunay pieces attaching through their catenoidal regions [MPP].

11. The Motivation for Desingularizing Constructions and Scherk's Singly Periodic Surfaces

It has been observed that for some discrete families of surfaces parametrized by the genus, a sequence of handles seems to concentrate along a curve as the genus tends to infinity. This holds for some minimal surfaces in the round sphere [Ka3], for the Costa–Hoffman–Meeks surfaces, which were the first complete embedded minimal surfaces of finite total curvature besides the plane and the catenoid to be found [Co1-2, HM1-3], and for many of the surfaces found in this spirit later [HM2,3,5, W1,2, Ka2, HKM, WW, HK], by using a global version [Os1,2] of the Enneper–Weierstrass representation. Hoffman and Meeks studied this phenomenon carefully [HM4], and they discovered that if the surfaces are magnified appropriately so that the necks are of fixed size and position, in the limit they converge to a classical singly periodic minimal surface, sometimes called Scherk's fifth surface [S]. This motivated Hoffman and Meeks to ask the question whether it might be possible to develop a gluing construction where neighborhoods of curves of intersection of minimal surfaces are replaced by sequences of necks modeled after the Scherk surfaces. It turns out that this is indeed possible. Before discussing these constructions I briefly review Scherk's singly periodic surfaces.

The most symmetric of these surfaces is given by the equation

$$\sin x = \sinh y \sinh z, \tag{1}$$

where $O.xyz$ is a given Cartesian coordinate system. The whole family is given by the equation [Ni]

$$\cos^2\alpha \cosh \frac{y}{\cos\alpha} - \sin^2\alpha \cosh \frac{z}{\sin\alpha} = \cos x, \tag{2}$$

where the parameter α takes values in $(0, \pi/2)$. In the literature these surfaces (or their scaled versions) are referred to as Scherk's fifth surfaces, Scherk's singly periodic surfaces, or Scherk-towers [Ni, HM4, DHKW, Ka1], but in these lectures they will be simply referred to as Scherk surfaces.

It is clear that these surfaces are periodic in the direction of the x-axis, and the coordinate planes and the planes $x = n\pi$ for $n \in \mathbb{Z}$ are planes of symmetry. The complement of a large enough cylindrical neighborhood of the x-axis consists of four "wings". Each wing decays exponentially fast to an asymptotic half-plane making an angle α with the xy-plane. The two "top" wings—in $\{z > 0\}$—are joined to the two "bottom" wings, through a sequence of handles, one handle in each slab $\{(2n+1)\pi < x < (2n+3)\pi\}$ for $n \in \mathbb{Z}$. Alternatively, the two "right" wings — in $\{y > 0\}$ — are joined to the two "left" wings, through a sequence of handles, one handle in each slab $\{2n\pi < x < 2(n+1)\pi\}$ for $n \in \mathbb{Z}$. The x-axis is the "axis" of the surface.

The Gauss map maps the part of the surface in a slab $\{n\pi < x < (n+1)\pi\}$ to a hemisphere $\mathbb{S}^2 \cap \{x > 0\}$ or $\mathbb{S}^2 \cap \{x < 0\}$. The whole surface equipped with the pullback metric is then isometric to an infinite covering of $\mathbb{S}^2 \backslash \{(0, \pm\cos\alpha, \pm\sin\alpha)\}$, consisting of a sequence of hemispheres connected as follows: Consider the four disjoint open arcs, on the equator of the sphere, with endpoints $(0, \pm\cos\alpha, \pm\sin\alpha)$. Each hemisphere connects to the previous one through a pair of opposite arcs, and to the next through the other two arcs.

12. Highly Symmetric Desingularizing Constructions

By "highly symmetric" constructions I mean constructions where enough symmetry is imposed so that the number of handles in a fundamental region is finite and independent of their size, or equivalently, their total number. A highly symmetric construction is substantially easier compared to a general construction, because most of the difficulties idiosyncratic to desingularizing curves of intersection constructions are avoided. The first desingularizing construction in the spirit of the previous section was a highly symmetric construction carried out by M. Traizet [T1-2]. In Traizet's construction the minimal surfaces being desingularized are finitely many planes in Euclidean three-space, all parallel to a given direction. The initial surface is constructed by replacing the lines of intersection with Scherk surfaces (scaled so that the period is small and the same for all of them), whose axis is the line of intersection. The construction is such that the asymptotic half-planes can be paired in such a way that the planes in each pair intersect in a planar strip bounded by two straight lines parallel to the axes. The wings then are truncated and fused over the middle of this strip. The exponential decay of the wings to the asymptotic half-planes implies that an exponentially small initial error can be arranged. The philosophy of the proof and the estimates were modeled after the construction in [K1].

The next construction I discuss was developed independently [K5], and involves desingularizing a finite collection of coaxial catenoids and planes in Euclidean three-space. The basic idea of the construction is to scale the suitable Scherk surface to a suitably small size, rotate it to an appropriate position, and then "bend" it, so that its axis "bends" to the circle of intersection in consideration, and the wings match roughly the catenoids or planes approaching to intersect at this circle. It is assumed that no circles of intersection belong to more than two of the catenoids and planes, so that the number of approaching minimal pieces is four, like the number of the wings. This way neighborhoods of the circles of intersection are replaced with a sequence of handles modeled after the appropriate Scherk surface suitably deformed to fit the given configuration. The scaling factor, which controls the size of the necks, is inversely proportional to the number of handles prescribed to replace the circle of intersection. Reflectional symmetry with respect to certain planes through the common axis of the catenoids is imposed through the whole construction. The angle between two subsequent planes varies with the number of necks used, so that in a single fundamental domain the number of necks is finite and independent of their total number. Therefore this construction is highly symmetric also.

The main difference with Traizet's construction is that the error in the initial surface is much larger, and actually to even control it enough, so that the construction is possible, requires a subtler approach. This is so because the main source of the error is the "bending" of the Scherk surface, while in Traizet's case there is no bending. Moreover, the "bent" wing is not appropriate to be used far away from the circle of intersection. This is because a "bent" wing would decay to a "bent" asymptotic half-plane, which amounts to a cone, which is not a minimal surface. The idea for managing this step is to construct the bent wing more carefully: First, the bent half-plane is replaced by a suitable catenoid. Next, the bent wing is constructed as a graph over the catenoid, by transplanting the function which gives the Scherk wing as a graph over the asymptotic half-plane. This way the error close to

the circle of intersection is of the same order as the error due to the bending, while further away it decays exponentially fast.

The only obstruction to this construction is the existence of nontrivial rotationally invariant functions in the kernel for the Dirichlet problem for the linearized operator on the annuli which are obtained when the circles of intersection are removed from the catenoids:

THEOREM E. *Assuming that the condition on the kernel above is satisfied, and that the angle between subsequent planes of symmetry through the axis is small enough, where the number of necks per fundamental domain has been prescribed, the initial surface can be perturbed to minimality.*

The condition on the kernel is generically satisfied, so that in the space of all configurations of a certain number of coaxial catenoids and planes, the set of configurations for which the construction works, is open and dense with respect to a natural topology. In this construction, as usual, the asymptotics of the ends may have to change. Actually, one asymptotic catenoidal end for each given catenoid, and the height of each planar end, can be prescribed. The asymptotics of the other end of each catenoid may have to change, and the asymptotics of each planar end may have to change to those of an end of a small catenoid. Notice that there are as many free parameters as ends, where surfaces which differ by a homothety are counted as different.

13. Desingularizing the Intersections of Compact Surfaces

I consider now a general configuration of intersecting minimal surfaces immersed in a three-dimensional Riemannian manifold. To avoid the extra difficulties implied otherwise, I assume that the intersections I intend to desingularize are transverse and contain no triple or boundary points. The surfaces may have boundary which I fix through the construction for simplicity, although obvious changes in the statements and proofs of the theorems would accommodate other boundary conditions as well. The following notation is adopted for future ease of reference: $\boldsymbol{X} : \mathcal{W} \to \mathcal{N}$ denotes the given minimal immersion, where $\mathcal{W}$ is an abstract surface, usually consisting of many components corresponding to the given surfaces, and $\mathcal{N}$ is a Riemannian three-dimensional manifold. To describe the intersections to be desingularized, I assume given a smooth closed curve $\mathcal{C}$ embedded in the interior of $\mathcal{W}$. $\mathcal{C}$ is assumed to have finitely many components and there is a factorization

$$\boldsymbol{X}|_{\mathcal{C}} = \underline{\boldsymbol{\gamma}} \circ \underline{\boldsymbol{P}}$$

where $\underline{\boldsymbol{P}}$ is some two-to-one covering $\underline{\boldsymbol{P}} : \mathcal{C} \to \underline{\mathcal{C}}$, $\underline{\mathcal{C}}$ is some abstract closed one-manifold, and $\underline{\boldsymbol{\gamma}} : \underline{\mathcal{C}} \to \mathcal{N}$ is an immersion. Finally, let $\widehat{\mathcal{W}}$ be the surface $\mathcal{W}$ after being "cut" through $\mathcal{C}$. This means that there is some $\widehat{\mathcal{C}} \subset \partial\widehat{\mathcal{W}}$, and a smooth map $\boldsymbol{Q} : \widehat{\mathcal{W}} \to \mathcal{W}$, such that the restriction of $\boldsymbol{Q}$ to $\widehat{\mathcal{W}} \setminus \widehat{\mathcal{C}}$ is a diffeomorphism onto $\mathcal{W} \setminus \mathcal{C}$, and the restriction to $\widehat{\mathcal{C}}$ is a two-to-one covering of $\mathcal{C}$. Notice then that $\widehat{\mathcal{C}}$ covers $\underline{\mathcal{C}}$ four-to-one, corresponding to the fact that there are four minimal pieces bordering the curve of intersection.

I first discuss the case of compact surfaces in which the statement of the theorem is much simpler. The curves then $\mathcal{C}$, $\widehat{\mathcal{C}}$, and $\underline{\mathcal{C}}$ have to be closed as well. The first observation to be made, is that unlike in the previous section where the angle between the intersecting surfaces was constant, now it varies along the curves of intersection. Although the basic idea of the construction of the initial surfaces

remains as before, the changing angle has to be accommodated by varying the parameter of the Scherk surface used along the curve of intersection. More precisely, for each component of the curve of intersection, a suitable covering of the component by the axis of the Scherk surfaces is defined. This map is extended to a neighborhood of the axis, by mapping the perpendicular planes of the axis, to the images under the exponential map of the perpendicular plane to the curve in the tangent space. This map incorporates the twisting of the minimal surfaces along the curve, and the desired scaling which also varies along the curve. By changing appropriately the parameter of the Scherk surface, slowly along the axis, and using these maps, the desired handles modeled after the Scherk surfaces, are introduced along the curve of intersection.

The construction of the wings requires again the construction of an appropriate minimal surface which plays the role of the asymptotic half-plane the way a catenoid did in the previous special case. This requires solving a partial differential equation. It is an indication of the difficulties involved in this construction, that, unlike all the other constructions, one has to solve a partial differential equation and produce involved estimates in order to construct even the initial surface. I will give more details on the construction of the wings later. I state now the theorem in the compact case:

THEOREM F ([K6, K7]). *Assuming given a configuration as above, where $\mathcal{W}$ is compact, and $\mathcal{C}$ is embedded in the interior of $\mathcal{W}$, and the intersections implied along $\underline{\mathcal{C}}$ are transverse, the initial surfaces can be constructed so that they can be perturbed to minimality, provided that the number of handles per component of $\underline{\mathcal{C}}$ is uniformly large enough, and the following conditions hold:*

1. *The kernel for the linearized operator $\mathfrak{L} = \Delta + |A|^2 + \mathrm{Ric}(\nu, \nu)$ on $\mathcal{W}$, with Dirichlet conditions on $\partial\mathcal{W}$, is trivial (unbalancing condition).*
2. *The kernel for the linearized operator $\mathfrak{L}$ on $\widehat{\mathcal{W}}$, with Dirichlet conditions on $\partial\widehat{\mathcal{W}}$, is trivial (flexibility condition).*

14. Comparison with the Delaunay Gluing Construction

To start with, there are certain parallels with the Delaunay gluing construction which are worth pointing out. The way the Delaunay gluing is guided by the given graph, the desingularization construction is guided by the given configuration of minimal surfaces, as described by $\boldsymbol{X}$ and $\mathcal{C}$. The vertices of the graph in the Delaunay gluing, correspond to the components of the curve of intersection, or more precisely, to the restriction of $\boldsymbol{X}$ to the components of $\mathcal{C}$, or equivalently to that of $\boldsymbol{X} \circ \boldsymbol{Q}$ to the components of $\widehat{\mathcal{C}}$. The round spheres substituting for the vertices in the Delaunay gluing, correspond to the Scherk surfaces substituting for the components of $\underline{\mathcal{C}}$. The compact components of $\widehat{\mathcal{W}}$ correspond to the edges of the graph, and the non-compact components — if any; none in the compact case — to the rays of the graph. The Delaunay surfaces substituting the edges and the rays, correspond to appropriate perturbations of $\boldsymbol{X}$ restricted to the interior of $\widehat{\mathcal{W}}$.

Recall that the balancing and unbalancing conditions for the Delaunay gluing were motivated by considering the resultant force on a neighborhood of each sphere substituting for a vertex. Consider in analogy a neighborhood of a small interval along $\underline{\mathcal{C}}$ containing a small number of necks. Because the size of the Scherk necks and the length of the interval are small, it is reasonable to ignore the "bending" and

the "twisting", and consider the necks scaled in the Euclidean space instead. Consider a cylindrical neighborhood of the axis such that the surface is approximately orthogonal to the top and bottom faces which are orthogonal to the axis. The force then on the part of the surface in the interior of the neighborhood is

$$F = \int_{L_+} \boldsymbol{\eta} + \int_{L_-} \boldsymbol{\eta} + \sum_{i=\pm 1,\pm 2} \int_{L_i} \boldsymbol{\eta},$$

where $L_\pm$ are the parts of the boundary of the surface in the top and bottom face of the cylinder, and L_i ($i = \pm 1, \pm 2$) are the parts of the boundary of the surface in the four different wings, where L_i is in the wing roughly opposing L_{-i}. This being a heuristic argument, it is reasonable to approximate the wing with the asymptotic half-plane to which it decays, and then $\int_{L_i} \boldsymbol{\eta} = \ell \vec{\eta}_i$, where ℓ is the length of the cylinder, and $\vec{\eta}_i$ is perpendicular to the axis and parallel to the corresponding asymptotic half-plane, and pointing into the half-plane from its boundary line. Similarly, $\int_{L_\pm} \boldsymbol{\eta}$ are approximately parallel to the axis. The vanishing of F then amounts to

$$\sum_{i=\pm 1,\pm 2} \vec{\eta}_i = 0 \quad \text{and} \quad \int_{L_+} \boldsymbol{\eta} + \int_{L_-} \boldsymbol{\eta} = 0.$$

The first condition is automatically satisfied, because $\eta_i = -\vec{\eta}_{-i}$. The second condition turns out to relate the change of scale with the other changes of geometry along the axis for the Scherk necks. Therefore although it dictates the correct scaling along the curve of intersection for the Scherk surfaces, it does not impose any conditions to the initial configuration.

Although "balancing" is automatic as described, "unbalancing" and "flexibility" are not. The flexibility condition can be taken to require that any small perturbation of the restriction $\boldsymbol{X} \circ \boldsymbol{Q}|_{\widehat{\mathcal{C}}}$, can be extended to a minimal small perturbation of $\boldsymbol{X} \circ \boldsymbol{Q}$ on $\widehat{\mathcal{W}}$. This is in agreement with the analogy where the edges of the graph codifying the gluing of the Delaunay surfaces correspond to the components of $\widehat{\mathcal{W}}$, and the vertices to the components of the curve of intersection: The flexibility condition in the Delaunay case amounts to small changes in the position of the vertices being accommodated by perturbing slightly the whole graph, and here small changes of the immersion on the components of $\widehat{\mathcal{C}}$ being accommodated by a small change of the whole configuration. To reduce this condition to the one stated in the theorem, note that since only small perturbations are involved, the perturbed immersions can be described as graphs over the old ones, and then the equation of minimality can be linearized as usual. The condition then amounts to the solvability of the Dirichlet problem for $\mathfrak{L}$ on $\widehat{\mathcal{W}}$ for data on $\widehat{\mathcal{C}}$, which because of the compactness of $\widehat{\mathcal{W}}$ amounts to the condition stated in the theorem.

The "unbalancing" condition should clearly require the ability to prescribe the resultant force F mentioned in the discussion of the balancing condition above. Prescribing the component along the axis, which as discussed before relates to the change of scaling from the top face of the cylinder to the bottom one, amounts to appropriately modifying how the scaling factor varies along the intersection curve. Although this motivates some of the details of the construction, it gives no obstruction at the level of the given configuration. The creation though of a resultant force perpendicular to the axis amounts to prescribing $\sum_{i=\pm 1,\pm 2} \vec{\eta}_i$. This is equivalent to prescribing a small angle between $\vec{\eta}_i$ and $-\vec{\eta}_{-i}$ for $i = 1, 2$. This would destroy

the smoothness of $\boldsymbol{X}$ as one crosses from one side of $\mathcal{C}$ to the opposite. The angles prescribed are small and vary along $\mathcal{C}$.

The idea for arranging this is as follows: Consider small changes of the immersions of the curves of intersection $\underline{\mathcal{C}}$. These determine then changes of $\boldsymbol{X}$ on $\mathcal{C}$ and of $\boldsymbol{X} \circ \boldsymbol{Q}$ on $\widehat{\mathcal{C}}$. By the flexibility condition then, these changes can be extended to obtain a minimal small perturbation of $\boldsymbol{X} \circ \boldsymbol{Q}$ on $\widehat{\mathcal{W}}$. The opposing conormals then will not be exactly opposite to each other, but they will make an angle which hopefully can be prescribed by choosing appropriately the original change of the immersion of $\underline{\mathcal{C}}$. The perturbed immersions can be described as graphs over the original immersions, and by linearizing, the problem reduces to whether the given angle change $\boldsymbol{\theta}$, defined on $\mathcal{C}$, is in the range of the operator $\mathcal{J}$, which is defined as follows: Given Dirichlet data $\boldsymbol{\chi}$ on $\mathcal{C}$, solve the Dirichlet problem with vanishing data on ∂W and $\boldsymbol{\chi}$ on $\mathcal{C}$, for the linearized operator $\mathcal{L}$ on $\mathcal{W}$, to obtain $\boldsymbol{\chi}$ defined on $\mathcal{W}$, continuous on $\mathcal{W}$, and smooth on $\mathcal{W} \setminus \mathcal{C}$. (It would be more precise to transplant the data to $\widehat{\mathcal{C}}$ and solve the Dirichlet problem there, but this description seems more immediate). The one-sided derivative of $\boldsymbol{\chi}$ in the direction of the conormal of $\mathcal{C}$ is different then on the two sides of $\mathcal{C}$; the difference, which I call $\mathcal{J}\boldsymbol{\chi}$, gives at the linearized level the angle between the graphs of $\boldsymbol{\chi}$ on the two sides of $\mathcal{C}$.

To summarize, the linear operator $\mathcal{J}$ was defined, and the unbalancing condition was reduced to the requirement that for every $\boldsymbol{\theta}$ there is a $\boldsymbol{\chi}$ such that $\mathcal{J}\boldsymbol{\chi} = \boldsymbol{\theta}$. By choosing the spaces for $\boldsymbol{\theta}$ and $\boldsymbol{\chi}$ appropriately and using the compactness of $\mathcal{W}$, this formulation turns out to be equivalent to the triviality of the kernel of $\mathcal{J}$: Clearly if $\boldsymbol{\chi}$ is in the kernel of $\mathcal{J}$, then its extension as above to $\mathcal{W}$ has continuous derivatives across $\mathcal{C}$, so it is smooth on $\mathcal{W}$, and hence it is in the kernel of the Dirichlet problem for $\mathcal{L}$ on $\mathcal{W}$. The converse is also clearly true; that is, a nontrivial element of the kernel of the Dirichlet problem for $\mathcal{L}$ on $\mathcal{W}$ restricted to $\mathcal{C}$ gives a nontrivial element of the kernel of $\mathcal{J}$. The unbalancing condition reduces therefore to the requirement that the kernel of the Dirichlet problem for $\mathcal{L}$ on $\mathcal{W}$ be trivial, as formulated already in the statement of the theorem.

15. The Case of Noncompact Complete Surfaces with Closed Curve of Intersection

The difference between the case we now consider and the previous case is the presence of unbounded ends which complicates the analysis somewhat. Such difficulties have been faced in earlier constructions [K5,Y1-2], and dealing with them is a technical issue. It is natural to impose various decay conditions at infinity for the solutions. This would seem to further restrict the class of possible constructions, compared to the compact case. The opposite is true however, provided that modifications of the asymptotics of the ends are allowed. This makes the statement of the theorem rather technical, but the intuition is quite clear. The following modifications are needed:

First, the norms measuring the sizes of the solutions to the inhomogeneous linear equation, and also the sizes of the inhomogeneous terms, need to incorporate the appropriate decay at infinity. The solutions and the nonlinear terms, and also $\mathcal{L}^{-1}\boldsymbol{\Omega}_{u+v}$, have to be small in these norms, and so decay appropriately fast, so that the approach outlined in Section 1 can be completed. This step has similarities to the treatment of the ends in earlier constructions.

Second, the flexibility and unbalancing conditions have to be modified appropriately. The flexibility condition amounts as before to the ability to solve the Dirichlet problem for $\mathfrak{L}$ on $\widehat{\mathcal{W}}$ with any given data on $\widehat{\mathcal{C}}$. The solutions have to be in the allowed space, that is to decay appropriately at infinity, modulo the Jacobi fields corresponding to the allowed changes of the asymptotics. Similarly, the unbalancing condition amounts to the operator $\mathfrak{J}$ being onto modulo a finite dimensional space of data on $\mathcal{C}$ which can be realized by using again the Jacobi fields corresponding to allowed changes of asymptotics. The precise formulation of these conditions, which is rather technical, will be presented in [K8].

For an illustration I discuss the following example: Consider a collection of catenoids whose axes are parallel but different, and close enough to each other so that the curves of intersection are closed. By making the usual conformal change of the metric to $\frac{1}{2}|A|^2 g$, the catenoids become isometric to the round sphere with two antipodal points removed. Since the linearized operator then can be changed to $\Delta + 2$ which has a three-dimensional kernel on the sphere, the argument in Section 14 can be modified to show that $\mathfrak{J}$ is not onto. This can be remedied by allowing the appropriate change of asymptotics. In particular changes corresponding to rotations of the ends around a horizontal axis have to be allowed. Because of this the surfaces desingularizing these intersecting catenoids are not necessarily embedded.

16. Applications and Open Problems

It seems very interesting to generalize the theorem to even more general settings. Recall that the conditions assumed up to now are that the curves of intersection are closed, the intersections are transverse, and that there are no triple points. I briefly discuss now the implications the removal of each of these conditions would have.

Allowing open curves is quite interesting and may have interesting applications. The main difficulty in understanding this situation is that the angle between the intersecting surfaces is not guaranteed any more to be bounded away from 0 and π. This makes necessary a more careful study of the analysis of the construction, but does not seem to be out of reach of the current technology.

There are many examples of configurations of minimal surfaces, especially in space forms, which one would like to desingularize, but there are points of intersection where the intersecting surfaces are tangent to each other. The curves of intersection at such a point have a well-known behavior, intersecting at equal angles. The main difficulty is that the angle of intersection degenerates again, but unlike the previous case where this happens at infinity, now it happens at an interior point. It should be possible to understand the conditions required to carry out the construction in such a setting and prove a general theorem.

Generically, at a triple point one gets three distinct tangent planes to the given minimal surfaces. This point then can be thought of as a triple point of the intersection curve as well, in the sense that there will be three curve pieces passing through it. When these pieces belong to different components of the curve of intersection, a possible approach is to desingularize some components of the curve of intersection first, consider then the resulting configuration of minimal surfaces, and desingularize again. When these pieces connect along the curve of intersection the Scherk model for the necks fails and a new model is needed. There is some hope that minimal surfaces desingularizing three distinct planes, and which decay

asymptotically to Scherk singly periodic surfaces along the intersection lines and away from the triple point, will be found and understood well, so that they can serve as models for desingularizing a configuration of minimal surfaces with triple points of this kind.

I discuss now various possible applications. Theorem F is surprisingly general for a singular perturbation construction. In many cases the conditions — flexibility and unbalancing — are generically satisfied, for example when the given surfaces have nontrivial boundary. In other cases further work is required to demonstrate that these conditions are not too restrictive, for example for closed surfaces in the round three-sphere or the flat three-torus, where the Killing fields of the ambient space make the kernel on the surface nontrivial. In order to avoid the implied violation of the unbalancing condition one has to consider constructions with enough symmetry to trivialize the kernel.

A case which demonstrates the power of the theorem refers to certain configurations of minimal surfaces in the round three-sphere: The class of examples of closed minimal surfaces in the round three-sphere is fairly limited [L, Ka3, PRu, KPS]. A fairly general construction is possible [KY]: Consider a finite number of square lattices $L_1 \subset L_2 \subset \cdots \subset L_k$ and the corresponding surfaces M_i $(i = 1, \ldots, k)$ constructed as described in Section 17 below. By applying the methods developed in [K3] to understand the Jacobi fields, it is proved that the flexibility and unbalancing conditions are satisfied for the collection of these closed minimal surfaces M_i, and then by desingularizing many new families of closed minimal surfaces are produced. Note that at the vicinity of a point common to all the lattices, the M_i's resemble coaxial catenoids of — very — different sizes perturbed away from rotational invariance. In that neighborhood then the picture is similar to the picture in [K5], and the generality of the construction is comparable (although the parameters are discrete) to the one in [K5].

I discuss now a possible application which is of a more theoretical character. The first problem in the Minimal Submanifolds section (problem 88) of the list of open problems proposed by S.-T. Yau in 1982 [Y], requires to prove that any three-dimensional manifold must contain an infinite number of immersed minimal surfaces. Theorem F seems to be a significant step in the right direction, because it reduces the problem to finding a single configuration of intersecting minimal surfaces in each manifold which satisfies the conditions. The required conditions would be relaxed if one uses more general versions of the theorem along the lines just discussed. Note also that one can probably ignore the flexibility and unbalancing conditions by assuming that the metric of the manifold is "generic", or "bumpy", as it is sometimes called. The theorem then would assert that there are infinitely many immersed minimal surfaces in a "generic" Riemannian manifold, or at least in some general class of Riemannian manifolds. If this is successful, it seems not much harder to prove the existence of infinitely many embedded minimal surfaces, since desingularizing all intersections gives embedded surfaces, at least in the compact case.

Another open question is whether there exist complete, properly embedded in E^3, minimal surfaces of infinite genus, with more than one end. Similarly, for periodic surfaces, where the genus and the ends are counted in the quotient space [CHM]. Various desingularizing constructions have been proposed to produce such examples. Consider for example desingularizing the configuration — proposed to me

by Meeks and by Ros — of a catenoid and its reflection across a plane parallel to the axis and missing the waist. Or, for the periodic case, the construction proposed in [CHM]. These constructions are particular cases where the curve of intersection is not closed, as discussed earlier.

17. Doubling Constructions

These constructions are motivated by work of Karcher, Pinkall, and Sterling [KPS], and Pitts and Rubinstein [PRu]. In [KPS] minimal surfaces are constructed resembling roughly an equatorial sphere in $\mathbb{S}^3(1)$ which has been "doubled" and the two sheets have been connected by necks arranged at the vertices of a Platonic solid, with the corresponding symmetry imposed. The examples constructed this way are finitely many because the Platonic solids are finitely many and the size of the neck is determined by the neck configuration (their number and positions).

In [PRu] certain constructions are discussed. The size of the necks used can be arbitrarily small and the genus then tends to infinity, while the surfaces tend to a limit varifold. These constructions are highly symmetrical. Some of the constructions have a limit varifold which is a minimal surface counted with multiplicity 2; I call such constructions "doubling" constructions.

The constructions above motivate a question: Under what conditions is it possible to carry out doubling constructions based on a given minimal surface? By a doubling construction I mean a construction where the new minimal surface consists of the following: First, a region that approximates, in some norm involving derivatives as well, two copies of the given surface from which a (large) number of small discs has been removed. The total area of the discs is small compared to the area of the given surface. I call the collection of the centers of these discs the "neck configuration" of the doubling construction. Second, small approximate catenoids having waist small compared to the radius of the circles, truncated and perturbed appropriately to connect smoothly to the boundary circles of the previous region, and positioned around the points of the "neck configuration", with axes approximately perpendicular to the original surface. These constructions can be considered as singular perturbation constructions, where the size of the necks is the parameter τ which has to be small enough for the construction to work. The number of the necks is supposed to tend to infinity as $\tau \to 0$. The limit varifold as $\tau \to 0$ is the given minimal surface counted with multiplicity 2. I consider understanding which "neck configurations" are possible a major part of the question.

Although I hope that a doubling construction of similar generality to the desingularizing construction in theorem F is possible, it seems that a construction of such generality may take some time to develop. It seems prudent therefore, to concentrate in the beginning on doubling constructions which have enough symmetry imposed, as was done in the case of the desingularizing constructions I already discussed in Section 12. The simplest possible construction is a doubling construction for the Clifford torus, proposed in [PRu], whose neck configuration consists of the points of a square lattice on the torus. This construction by singular perturbation methods has been developed in collaboration with S. D. Yang [KY]. As I have mentioned already, an extra motivation for this construction, is that the surfaces constructed can be combined to configurations to which theorem F can be applied, to produce a large class of closed minimal surfaces in the round three-sphere.

The parameter of this construction is m, where m^2 is the number of points in the square lattice neck configuration, that is, the number of necks used; the construction works when m is large enough. I call the size of the catenoidal neck τ. The group of symmetries imposed on the construction is the group of isometries of the three-sphere, preserving the square lattice as a set. Consider the geodesics of the torus, reflections with respect to which preserve the square lattice, but do not contain points of the lattice. They subdivide the torus into squares each of which contains a single point of the lattice. Let the vertices of one such square be A_i $(i = 1, 2, 3, 4)$, the sides $E_1 = A_1A_2, \ldots, E_4 = A_4A_1$, each of length $\sqrt{2}\pi/m$, and let p be the point of the lattice at the center of the square. Let Π be the nearest point projection of the round three-sphere minus the two circles furthest from the Clifford torus, onto the Clifford torus, and $W_i = \Pi^{-1}(E_i)$ $(i = 1, 2, 3, 4)$. The W_i's are contained in open hemispheres, and W_i intersects W_{i-1} —where $W_0 = W_4$ —along $\Pi^{-1}(A_i)$, a quarter of a great circle which is their common boundary. Reflections with respect to the great spheres containing the W_i's are symmetries of the construction and map p to its nearest neighbors in the lattice. The W_i's enclose the region R of the points of the sphere which have p as their closest point among the points of the lattice.

It is important to construct the initial surfaces carefully enough so the initial error can be managed. A fundamental domain of the initial surface can be constructed inside R. It consists of three regions. The first region is constructed by truncating a catenoid at infinity, scaling it by a factor so that the radius of the waist is τ, and placing it inside R by using normal coordinates so that its center is at p, and its axis is perpendicular to the Clifford torus. Using cut-off functions there is a smooth transition to the second region. The second region is constructed as an approximately rotationally invariant region by solving an appropriate ordinary differential equation. Close to the W_i's there is a transition to the third region using cut-off functions again. The third region is constructed as a graph over the Clifford torus, and intersects the W_i's orthogonally, so that by reflecting across their supporting spheres, and by applying the rest of the symmetries, the fundamental domain can be extended to a smooth surface. The initial surface constructed this way has nonzero mean curvature everywhere, but the main term is supported close to the W_i's. By "balancing" considerations I will present soon, it follows that for the construction to work,

$$\tau \approx m^{-1}e^{-m^2/4\pi} \tag{1}$$

has to hold, that is the ratio of the two sides of the equation has to be bounded away from 0 and ∞ uniformly as $m \to 0$.

It is instructive also to compare with the situation in Euclidean three-space, where the analogous construction, that is a doubling construction for the plane, fails. I use the same symbols for the analogous objects in Euclidean space. The Clifford torus is replaced by a plane, on which a square lattice is defined, whose nearby points are distance $\sqrt{2}\pi/m$ apart. The E_i's can be defined as before, and they are the sides of a square, the center of which is a point p of the lattice. The W_i's are strips on planes perpendicular to the plane of the lattice, and reflections with respect to the planes on which the W_i's lie carry p to its nearest neighbors on the lattice. The W_i's enclose the region R of points closer to p than to the other points of the lattice.

Let Y be the Killing field in the Euclidean space generating the translations in a direction perpendicular to the plane. In the case of the sphere Y is the Killing field generating rotations which rotate the great circle C_p perpendicular to the torus at p, while keeping the perpendicular to C_p great circle $C_p^{\perp}$ pointwise fixed. In both cases Y is normalized to have unit maximum length. Let ∂_0 be the waist of the catenoidal neck on the minimal surface to be constructed, and ∂_1 be the component of the intersection of the surface with $\bigcup_{i=1}^{4} W_i$ to which Y is pointing from the other component. By equation (3) in Section 3 it follows that

$$\int_{\partial_1} \boldsymbol{\eta} \cdot Y = \int_{\partial_0} \boldsymbol{\eta} \cdot Y,$$

where $\boldsymbol{\eta}$ is the conormal, tangent to the surface, and pointing in the direction of Y on ∂_0, and outside of R on ∂_1. In the Euclidean case the left-hand side clearly vanishes because Y is vertical and $\boldsymbol{\eta}$ horizontal, while the right-hand side is approximately $2\pi\tau$, a contradiction which demonstrates the impossibility of the construction. In the spherical case however Y is not perpendicular to $\boldsymbol{\eta}$ on ∂_1 anymore. Using the geometry of the sphere, and assuming that the height — that is the signed distance from the Clifford torus — at ∂_1 does not oscillate too much and is close to what it would be for a Euclidean catenoid, assumptions consistent with the construction, the left-hand side can be calculated to be approximately equal to

$$4\frac{\pi^2}{m^2}\tau \log \frac{\pi}{\sqrt{2}m\tau}.$$

The right-hand side is still approximately equal to $2\pi\tau$. The equation can be solved for τ then to produce equation (1) above. I state now the theorem:

THEOREM G ([KY]). *For each large enough m, there is a τ in the range specified by (1), such that the initial surface whose construction was outlined above can be perturbed to a minimal surface M_m. Moreover if $m_1, \ldots, m_k$ are large enough integers, and m_i divides m_{i+1}, so that the corresponding lattices satisfy $L_1 \subset L_2 \subset \cdots \subset L_k$, the configuration of the minimal surfaces $M_{m_1}, \ldots, M_{m_k}$, and their intersections, satisfy the conditions of theorem F, and then can be desingularized to give embedded minimal surfaces in the sphere.*

18. The General Framework for Understanding the Linearized Equation

I start now the second part of these lectures, in which I discuss the general set-up and the main ideas of the detailed construction and the proofs. The fundamental idea of all gluing constructions has already been discussed in Section 1. It was pointed out then, that the invertibility of the linearized operator is an issue relating to the obstructions in these constructions. In this section I discuss a general framework for understanding the linear equation, which helps to pinpoint the difficulties, and prepares the grounds for their resolution.

The initial surfaces constructed can be subdivided into "standard regions" and "transition regions". The standard regions, perhaps after scaling appropriately, tend to simple geometric objects as $\tau \to 0$, the convergence being uniform on compact subsets of the limit. I will refer to these simple geometric objects as the "standard limit objects". The standard regions have been appropriately — depending on τ — truncated, so that they approximate the corresponding piece of the limit object

closely. I call the standard regions scaled or unscaled, depending on whether they have to be scaled or not, before taking the limit. The transition regions are the connected components of the complement of the union of the standard regions. The geometry of the transition regions often diverges substantially from that of the limit standard objects. In the constructions under consideration the transition regions always contain annuli, each of which is a neighborhood of a boundary circle shared between the transition region and a standard region. The existence of these annuli is important for the analysis of the linearized operator as described in Section 19.

I call the union of a standard region with the transition regions bordering it an "extended standard region". Under the approach presented, to understand the (inhomogeneous) linearized equation on the initial surface, one understands it first on the extended standard regions. By combining the solutions on the extended standard regions, an approximate solution on the whole surface is obtained. The approximate solution is corrected to a solution by iterating. The advantage of this "semi-localization" to the extended standard regions, is that since each of them contains a single standard region, the obstructions to the invertibility of the linearized operator are simpler to understand. This is especially true in the Wente gluing case. In general, these obstructions approximate the ones on the corresponding standard limit object, while the transition regions do not contribute any. There are two main difficulties in implementing this approach: The first difficulty originates from the existence of these obstructions to the invertibility of the linearized operator. The second difficulty is that the semi-local solutions have to decay along some of the transition regions. This decay is needed for two reasons: first, to ensure that the iteration above converges, by ensuring that the error of combining the semi-local solutions is small compared to them; second, to ensure that in regions of high curvature the solution is smaller than in other regions, because otherwise the nonlinear terms could dominate the linear ones.

More systematically, the following set-up is used: The linear equation to solve for u is

$$\mathfrak{L}u = E \tag{1}$$

for some given E, and where appropriate boundary conditions on ∂M are imposed when $\partial M \neq \varnothing$. By using cut-off functions a decomposition $E = \sum_S E_S$ can be arranged, where for each standard region S, E_S is supported on the corresponding extended standard region $\tilde{S}$. The linearized equation to understand on each $\tilde{S}$ has then the form

$$\mathfrak{L}u_S = E_S, \tag{2}$$

where Dirichlet boundary conditions are imposed for u_S on $\partial\tilde{S} \setminus \partial M$, besides the already imposed conditions on $\partial\tilde{S} \cap \partial M$. It is convenient before proceeding to carry out an extra step to ensure that E_S vanishes on the transition regions: Because the transition regions carry no kernel, the linearized equation can be solved without obstructions on an appropriate neighborhood T' in $\tilde{S}$ of the transition region, and the solution appropriately smoothed out in $S \cap T'$, to give u'_S smoothly defined on $\tilde{S}$, supported on T', satisfying $\mathfrak{L}u'_S = E_S$ on the transition regions, and the appropriate boundary conditions on $\partial\tilde{S}$. Moreover u'_S should inherit the appropriate decay from the decay of E_S. It is enough to understand then the equation $\mathfrak{L}u''_S = E_S - \mathfrak{L}u'_S$ for u''_S, and take $u_S = u'_S + u''_S$. By renaming u''_S and $E_S - \mathfrak{L}u'_S$ into u_S and E_S, equation (2) is reduced to the case where E_S vanishes on the transition regions.

To understand now equation (2), finite dimensional spaces of sections defined on $\tilde{S}$, and supported on S, $\mathcal{K}(S)$ and $\overline{\mathcal{K}}(S)$, have to be specified, with the following properties: $\mathcal{K}(S)$, which I call the substitute kernel on S, is constructed so that any E_S can be corrected by using an element $w_S \in \mathcal{K}(S)$, so that the obstructions are avoided and a solution u_S on $\tilde{S}$ can be found, satisfying the boundary conditions and

$$\mathcal{L}u_S = E_S + w_S, \tag{3}$$

where the appropriate norms of u_S and w_S are controlled by the norm of E_S.

$\overline{\mathcal{K}}(S)$ is defined by

$$\overline{\mathcal{K}}(S) = \{\mathcal{L}v : v \in \underline{\mathcal{K}}(S)\},$$

where $\underline{\mathcal{K}}(S)$ is an appropriately defined space of sections on $\tilde{S}$ vanishing on $\partial\tilde{S}\setminus\partial M$ and satisfying the required boundary conditions on $\partial\tilde{S}\cap\partial M$. The role of $\overline{\mathcal{K}}(S)$ is in ensuring the decay of the solution and is constructed to satisfy the following property: Given any solution u_S, there is a $v_S \in \underline{\mathcal{K}}(S)$, whose norm is controlled by the norm of u_S, such that $u_S + v_S$ decays as required.

Combining, it follows that for each E_S there are $\overline{w}_S \in \mathcal{K}(S) \oplus \overline{\mathcal{K}}(S)$ and $\overline{u}_S$, whose norms are controlled by the norm of E_S, such that $\overline{u}_S$ decays as required and satisfies the required boundary conditions, and

$$\mathcal{L}\overline{u}_S = E_S + \overline{w}_S, \tag{4}$$

where $\overline{u}_S = u_S + v_S$ and $\overline{w}_S = w_S + \mathcal{L}v_S$. I call $\overline{\mathcal{K}}(S) \oplus \mathcal{K}(S)$ the extended substitute kernel on S. In other words by solving modulo the extended substitute kernel, semi-local solutions of the appropriate size and decay can be obtained. The semi-local solutions can be combined then as suggested before, by smoothing them and iterating for the error introduced, to a global solution u modulo the global extended substitute kernel

$$\widetilde{\mathcal{K}} = \bigoplus_S \mathcal{K}(S) \oplus \overline{\mathcal{K}}(S),$$

which therefore satisfies

$$\mathcal{L}u = E + \tilde{w}, \tag{5}$$

for some $\tilde{w} \in \widetilde{\mathcal{K}}$. Note that the elements of $\mathcal{K}(S)\oplus\overline{\mathcal{K}}(S)$ have been trivially extended to the whole initial surface, which is possible because they were constructed to vanish on the transition regions.

19. The Obstructions and the Extended Substitute Kernel $\widetilde{\mathcal{K}}$

The simplest case for understanding the obstructions to the invertibility of the linear operator is when the standard limit object $\hat{S}$ is a closed surface with finitely many points removed. The linear operator $\mathcal{L}$ is self-adjoint, and has discrete spectrum and finite-dimensional kernel on $\hat{S}$. Killing fields of the ambient space always induce elements of the kernel, as it follows from the following lemma which has other important applications as well:

LEMMA. *Suppose M is a constant mean curvature surface or a minimal surface, immersed in a Riemannian three-manifold which has a Killing field Y. Then the projection then $Y^\perp$ of the Killing field to the normal direction on the surface satisfies the linearized equation $\mathcal{L}Y^\perp = 0$.*

The proof is by considering the flow of the surface by Y. The nearby surfaces satisfy the same condition and can be considered as graphs over the original surface. By linearizing, the result follows. A corollary of the lemma is that if the surface is closed, then $Y^{\perp}$ is in the kernel of the linearized operator. It turns out that the kernel on the standard limit objects is usually generated in this way by the translational Killing fields of the Euclidean space. It can be checked [K1,3,5-9] by using fundamental properties [Ch], that the transition regions do not contribute to the lower spectrum of $\mathfrak{L}$ on the extended standard region $\tilde{S}$ corresponding to $\hat{S}$. The low eigenvalues of $\mathfrak{L}$ on $\tilde{S}$ approximate then the ones of $\mathfrak{L}$ on $\hat{S}$, and the corresponding eigenfunctions of $\tilde{S}$ the ones of $\hat{S}$, in any norm on S. The only small eigenvalues of $\mathfrak{L}$ on $\tilde{S}$ then, are the ones with corresponding eigenfunctions close to some $Y^{\perp}$, where Y is a translational Killing field. Then defining $\mathcal{K}(S)$ is very simple: Just use any space of sections whose projection to the space $\{Y^{\perp} : Y \text{a translational Killing field}\}$, has an inverse of bounded norm. The sections in $\mathcal{K}(S)$ can clearly be chosen to have support in the vicinity of chosen points in S, if so desired.

Consider now the case where the standard limit object is not bounded: A special feature of the constructions under discussion is their two-dimensionality. Because of this, a conformal change of the metric changes the Laplacian simply by a factor. It turns out that a useful conformal change is the one of the metric g to $h := |A|^2 g/2$, or appropriate variations of this to accommodate for possible zeroes of $|A|$. Such a change on a minimal surface in the Euclidean space amounts to pulling-back the metric of the round sphere by the Gauss map. The standard limit objects corresponding to the scaled standard regions are such minimal surfaces, and this change compactifies them, or at least compactifies their fundamental domain, as in the case of the Scherk surfaces. The linear equation $\mathfrak{L}u = E$ gets simplified to $\mathfrak{L}_h u = 2E/|A|^2$, where in the limit $\mathfrak{L}_h = \Delta_h + 2$, and so the norms of the coefficients of the linear operator are uniformly bounded. When the standard limit object with respect to the new metric h, is isometric to a closed surface minus points, the same approach as in the previous case can be used to define $\mathcal{K}(S)$. This covers all the cases under discussion except for the case of theorem F, which is more complicated and will be discussed later.

The fundamental property of $\underline{\mathcal{K}}(S)$ is that the low harmonics of the restriction of its elements to certain components of ∂S can be prescribed at will. Note the analogy with the elements of $\mathcal{K}(S)$ whose projection to the space spanned by the eigenfunctions of small eigenvalue can be prescribed at will. This fundamental property of $\underline{\mathcal{K}}(S)$ can be used to ensure the required decay of the solutions as follows: Let C_S be one of the components of ∂S on which the low harmonics of the elements of $\underline{\mathcal{K}}(S)$ can be prescribed. There is an annulus A then, contained in the transition region with $\partial A = C_S \cup C_{\partial}$; in the simplest case A is a component of the transition region and $C_{\partial} \subset \partial \tilde{S} \setminus \partial M$. To summarize, if u_S is as in equation (3) in Section 18 and $v_S \in \underline{\mathcal{K}}(S)$, then $u_S + v_S$ has the following properties on A: first, $\mathfrak{L}(u_S + v_S) = 0$; second, by choosing v_S appropriately, the low harmonics of its restriction to C_S can be prescribed; third, in the simplest case where $C_{\partial} \subset \partial \tilde{S}$, it vanishes on C_{∂}.

It turns out that in the constructions under consideration, there is a conformal change of the metric g to a metric χ, under which the geometry of the annulus approximates that of a standard cylinder, and the linear operator transforms to

one which is approximated by Δ_χ. $u_S + v_S$ is approximately harmonic then on the approximate metric cylinder (A, χ). Recall that a harmonic function defined on a cylinder, vanishing at one boundary circle and with vanishing low harmonics on the other, decays exponentially at the rate of the next harmonic, as can be seen by using separation of variables and Fourier decomposition on the meridian circles. By comparing the two set-ups, $u_S + v_S$ can be arranged to decay exponentially at a rate close to the rate of the next harmonic as well. This ensures the desired decay because either the cylinder is long enough, or enough low harmonics are prescribed.

I discuss now how to construct $\underline{\mathcal{K}}(S)$. The simplest case is when one can take $\overline{\mathcal{K}}(S) = \mathcal{K}(S)$. $\underline{\mathcal{K}}(S)$ is constructed then as the span of $f - u_f$, where f is an eigenfunction of small eigenvalue for $\mathfrak{L}$ on $\tilde{S}$ with the usual boundary conditions, and u_f is the solution u_S as in equation (3) of the previous section, with $E_S = \mathfrak{L}f$. In this case E_S is allowed not to vanish on the transition regions, but clearly $\mathfrak{L}(f - u_f) \in \mathcal{K}(S)$ and so does vanish. That the $\underline{\mathcal{K}}(S)$ has the fundamental property required above, that is that the low harmonics on certain components of ∂S can be prescribed, is easy to check when true because $f - u_f$ can be approximated on the whole S, and hence on ∂S as well, by some $Y^\perp$. In the case when this procedure is not adequate to define $\underline{\mathcal{K}}(S)$, for example, when there are no eigenfunctions of small eigenvalue but certain decay needs to be arranged, the construction of $\underline{\mathcal{K}}(S)$ is based on the geometric principle which I discuss in the next section.

It should be emphasized that the conformal changes of g to h and χ discussed above are for convenience only. The obstructions to solving the linearized equation on the scaled standard regions can be alternatively understood by using weighted Sobolev spaces, and $\overline{\mathcal{K}}(S)$ can be constructed then in a similar way as above. The decay on each A can also be established without a conformal change of the metric once the lower harmonics on C_S can be prescribed. In constructions in higher dimensions where the conformal changes are not available, it would be necessary to proceed this way. In such a case C_S would be a higher dimensional manifold and A would be topologically C_S times an interval. Note also that the transition regions can have more complicated geometry, as in the case of the Wente tori for example, which is discussed in more detail in Section 21.

20. Resolving the Difficulties and the Geometric Principle

At this stage of the discussion, where the global solution of the linear equation is constructed modulo the elements of $\widetilde{\mathcal{K}}$, the difficulties have only been bypassed and not resolved. More precisely, the general approach outlined in Section 1 has to be modified to take into account that the linear equations involved are not solved precisely but modulo $\widetilde{\mathcal{K}}$: u, z, and $\tilde{w}_1, \tilde{w}_2 \in \widetilde{\mathcal{K}}$ can be found, as described in the previous section, so that

$$\mathfrak{L}u = -H_{\text{err}} + \tilde{w}_1, \qquad \mathfrak{L}z = -Q_{u+v} + \tilde{w}_2, \tag{1}$$

where u and z satisfy the appropriate boundary conditions, if any, and v is given. By finding then a fixed point v of the map $v \to z$ as in Section 1, and taking $f = u + v$, a solution to the nonlinear equation modulo $\widetilde{\mathcal{K}}$ is obtained, since then

$$H_f = c + \tilde{w}, \tag{2}$$

where $c = 1$ in the constant mean curvature case, $c = 0$ in the minimal surface case, and $\tilde{w} = \tilde{w}_1 + \tilde{w}_2 \in \widetilde{\mathcal{K}}$. Therefore it only remains to find a way for correcting in the direction of the elements of the extended substitute kernel.

The well-known idea for dealing with this is to introduce appropriate modifications, controlled by parameters, in the original construction of the initial surface. The mean curvature of the initial surface changes then as a function of the parameters. This changes the inhomogeneous term of the linearized equation, and then the original $\tilde{w}_1$ used in (1) changes to some $\tilde{w}_1 + \hat{w}$, where $\hat{w} \in \widetilde{\mathcal{K}}$ depends on the parameters. The construction should be such that any small element of $\widetilde{\mathcal{K}}$ can be prescribed as $\hat{w}$, by suitable choice of the parameters. The same construction of f as before leads to (2), but with $\tilde{w} = \tilde{w}_1 + \hat{w} + \tilde{w}_2$ instead. Since $\hat{w}$ can be prescribed, it can be arranged to cancel $\tilde{w}_1 + \tilde{w}_2$ exactly, so that for that value of the parameters and the corresponding f, $H_f = c$ holds as desired. Note that $\tilde{w}_1 + \tilde{w}_2$ may depend on the newly introduced parameters, but should satisfy the same bounds as before uniformly in the new parameters. At the technical level, finding the correct value of the parameters can be incorporated in the fixed point argument used to find the fixed point v. Therefore, the construction of the desired minimal or constant mean curvature surface, as a graph over one of the modified initial surfaces, is achieved.

The central question now, is how to modify the original initial surface, so that $\hat{w}$ can be prescribed. It turns out that in all these constructions the following principle, which I call the "geometric principle", applies: Creating $\hat{w}$ is equivalent to "relocating" various pieces of the initial surface relative to each other. Therefore, the modifications introduced to the initial surface, amount to repositioning various pieces of the initial surface relative to each other. This repositioning sometimes can be done just by geometric construction, but other times it requires solving a partial differential equation, as I discuss shortly.

To understand the creation of those components of $\hat{w}$ which correspond to a substitute kernel, it is very helpful to use the formulas of Section 3. Recall that in the constructions under consideration an eigenfunction of small eigenvalue on an extended spherical region $\tilde{S}$ can be approximated on the corresponding standard region S — where the initial surface is close to the standard limit object — by $Y^{\perp}$. Assuming that H_{err} vanishes on the transition regions, the projection of H_{err} to the eigenfunction can be calculated then approximately, by calculating the projection to $Y^{\perp}$, which is $\int_{S'} H_{\text{err}} \cdot Y^{\perp}$, where S' is any domain satisfying $S \subset S' \subset \tilde{S}$. By the formulas of Section 3, $\int_{S'} H_{\text{err}} \cdot Y^{\perp}$ equals integrals relating only to $\partial S'$. When Y is a translation in the Euclidean space, these integrals give the corresponding component of the resultant force exerted on S' by the rest of the system. This is an important point because it explains how the balancing and unbalancing conditions come into the process of the construction and proof: The balancing condition amounts to the orthogonality of the initial H_{err} to the eigenfunctions of small eigenvalue, without which it would be hopeless to expect that the error can be corrected in this direction. The unbalancing condition amounts to the ability to create H_{err} in the direction of the eigenfunctions of small eigenvalue, for those standard regions that have the resultant force on them prescribed by the condition. Creating H_{err} in these directions is clearly equivalent to creating the corresponding components of $\hat{w}$.

Creating the components of $\hat{w}$, which correspond to eigenfunctions on other extended standard regions, rather than the ones covered by the unbalancing condition, is often done by repositioning the standard regions relative to each other. In [K1] where the Delaunay surfaces were involved, this was done by simple geometric construction. The standard regions in the Delaunay pieces are topologically annuli. At each such standard region some "bending" or "change of parameter" is introduced, resulting in different forces exerted at the two boundaries of the standard region, and at the same time repositioning the other standard regions. In [K3,4] the standard regions which carry approximate kernel are discs topologically. Placing the various standard regions to their new positions, determines also the new position of the boundaries of the transition regions. The modification of the transition regions can be determined then, by solving the Dirichlet problem on the transition regions for the given data on the boundary, so that the resulting transition regions satisfy the geometric condition exactly. The resulting surface is continuous but not smooth across these boundaries. The simplest approach is to smooth it out by using cut-off functions. This introduces new mean curvature $\hat{H}$, which can be arranged to be supported on the standard regions at the vicinity of their boundaries. This modification of the initial surface can be done sometimes at the linearized level, that is instead of specifying a new initial surface, a section of the normal bundle $\hat{f}$ is specified, and $\mathfrak{L}\hat{f}$ plays the role of the new mean curvature $\hat{H}$, and is supported on the standard regions at the vicinity of their boundaries. This simplification is possible only when there are no ends whose asymptotics change while the initial surface is being modified, as in the case of gluing Wente tori for example. A careful analysis of the modification introduced to the transition region allows us to calculate the change of the forces exerted, and hence the amount of the components of $\hat{w}$ created. This analysis of the modification can often be greatly facilitated by using the lemma in Section 19, to appropriately modify the boundary data to other ones easier to handle, by adding to the linearized solution some appropriate $Y^{\perp}$.

I discuss now the creation of those components of $\hat{w}$ which do not correspond to a substitute kernel, and also the definition of the corresponding sections of the $\underline{\mathcal{K}}(S)$'s. In this case some repositioning of various pieces of the surface is introduced. Let v' be the linearized version of the section which expresses the old initial surface as a graph over the new one, or, more generally, an initial surface piece which has no repositioning of the kind in consideration, as a graph over the extended standard region $\tilde{S}$ in consideration. The difficulty in using v' as an element of $\underline{\mathcal{K}}(S)$ is that it does not satisfy the required boundary conditions. To correct this, v' is replaced by $v := v' + v''$, where v'' is the solution of $\mathfrak{L}v'' = 0$ on $\tilde{S}$, subject to appropriate boundary conditions so as to ensure that v satisfies the required boundary conditions. v is taken then to be an element of $\underline{\mathcal{K}}(S)$, and the creation of the corresponding component of $\hat{w}$ has already been arranged by the above construction. The difficulty is that estimating v' and v'' requires often a significant amount of work, and usually is based on applying the lemma in Section 19, which allows changing the boundary data to easier ones to handle. Finally, it is usually straightforward to show that the corresponding component of $\hat{w}$ can be prescribed by varying the corresponding parameter because of the way v, and hence $\mathfrak{L}v$, was defined.

21. The Constant Mean Curvature Constructions

I proceed now to discuss the relation of the general method with the particular details of each construction. I first discuss the constant mean curvature constructions. In the approach adopted the standard limit objects are taken to be the round sphere minus points, the catenoid, and the Enneper surface. The first one corresponds to the unscaled standard regions, and the last two correspond to the scaled standard regions. When the h metric is used, they all become isometric to the round sphere minus points. On the standard limit objects the linearized operator changes to $\mathfrak{L}_h = \Delta_h + 2$, and the kernel is spanned by the first harmonics of the sphere, which are the coordinates of the Gauss map. The transition regions in the Delaunay case are annuli which in an appropriately defined χ metric as discussed in Section 19, are long standard cylinders of length $|\log\tau| + O(1)$ [K1]. Moreover the standard regions in the Delaunay case are separated by such transition regions. A Wente torus or cylinder on the other hand, contains a single transition region, which shares a single boundary circle with each standard region. A χ metric can again be defined in which each neighborhood of a boundary circle is an approximate standard cylinder of length $\frac{1}{4}|\log\tau| + O(1)$. The rest of the transition region is periodic with fundamental domain a region whose geometry has a regular limit in the χ metric as $\tau \to 0$ and has four boundary circles: Two of them are shared with two cylinders as above, and two are shared with the adjacent fundamental regions. Understanding the Dirichlet problem on the transition regions for various boundary data is a major part of [K3].

I outline now some more aspects of the construction in [K3]: Recall from Section 7, that on this construction certain symmetries have been imposed. The effect of these symmetries is to reduce the number of small eigenvalues — counted with multiplicity — on the extended standard regions: Most of them have two small eigenvalues, but the extended standard region which contains the central sphere where the fusion occurs, has none, and there are a few who have only one. $\overline{\mathcal{K}}(S) = \mathcal{K}(S)$ in all cases except when S is the central standard region. When S is the central standard region where the fusion of the Wente tori takes place, $\overline{\mathcal{K}}(S)$ is one-dimensional. To describe the parameter which controls the corresponding modification, recall first the construction of the initial surface from Section 7. Each Wente torus was placed relative to the central sphere so that one of its positively curved regions matched the central sphere well. Given the symmetries, there is only one degree of freedom in positioning the Wente torus relative to the sphere, which corresponds to translations along the radius of the sphere connecting its center with the corresponding vertex of the g-gon. The parameter is taken to be the signed magnitude of a translation of the Wente torus from its original position along this radius. It is possible then to define $\underline{\mathcal{K}}(S)$ and $\overline{\mathcal{K}}(S)$ by a somewhat complicated analysis of the v' and v'' as proposed in the previous Section 20, and also to ensure that the corresponding component of $\hat{w}$ can be prescribed by varying this parameter.

It remains to describe how to prescribe the other components of $\hat{w}$. These correspond to the small eigenvalues of the other extended standard regions. Although it is clear from the discussion in Section 20 that this should be achieved by appropriately relocating the standard regions relative to each other, it is highly nontrivial to implement this in a way that the components of $\hat{w}$ can be prescribed independently. The reason is that the geometry of the transition region is such that too much interaction between the standard regions is allowed. A rather systematic

approach has to be developed: Under yet another conformal change of the metric the geometry of the Wente torus becomes that of a standard cylinder where pairs of standard regions are placed along the cylinder. Some of the meridians of the cylinder lie in the middle of the transition region and on them the Gauss curvature changes sign. Between two such meridians lies a pair of standard regions. The relocation of the standard regions is organized so that the standard regions "before" such a meridian get repositioned relative to the ones "after" the meridian. This repositioning is either by a translation, or by a rotation around the "center" of the Wente cylinder, in a way consistent with the symmetries of the construction. Since the construction uses Wente tori actually, and not Wente cylinders, this repositioning induces a period problem, which is dealt with by allowing the creation of an appropriate force or torque through these meridians independent of the particular meridian.

By using an exponential decay imposed along the tori away from the central region where the fusion occurs, and carefully organizing the relative motions, the force and torque through the meridians separating the consecutive pairs of standard regions can be prescribed. For a certain type of translational repositioning, this requires further repositioning of the adjacent to the separating meridian standard regions, so as to ensure fast enough decay away from the meridian. This step makes heavy use of the lemma in Section 19 to estimate the modifications on the standard regions by changing appropriately the Dirichlet data. Also the Dirichlet to Neumann or its inverse problem on "half" the transition region is understood in certain cases, by comparing with that for the Laplacian on a half-cylinder. This allows the estimation of some of the boundary integrals involved. Finally one more force relating to the interaction of the two standard regions inside a pair can be prescribed. This together with the prescription of the forces and torques on the meridians, turns out to be equivalent to prescribing the components of $\hat{w}$ as desired. The whole work can be realized at the linearized level and makes heavy use of the symmetries of the Wente torus. I refer to the paper [K3] for more details.

22. Yang's Connected Sum Construction

In this case the unscaled standard regions are the given minimal surfaces, minus small discs centered at the points where the small truncated catenoids are attached. The corresponding standard limit objects are the given minimal surfaces minus the points marked for the attachment of the catenoids. The scaled standard regions are the catenoidal bridges truncated at some distance from their waist, and the corresponding standard limit object is the catenoid.

On the standard limit objects the kernel is three-dimensional, because of the nonexistence of bounded exceptional Jacobi fields on the minimal surfaces, and the properties of the catenoid. The dimension of each $\mathcal{K}(S)$ is 3 then also. In order to create the corresponding components for $\hat{w}$ as discussed in Section 20, the asymptotics of one of the ends of each minimal surfaces are modified: Yang proves that it is always possible to find an end which can be modified, so that the force it exerts on the rest of the minimal surface changes in a prescribed way. The catenoids are modified so that the axis and scale of the two ends differ by prescribed amounts.

$\overline{\mathcal{K}}(S)$ is two-dimensional when S is a catenoidal bridge, one-dimensional when S is an attached catenoidal end, and 0-dimensional otherwise. The corresponding

parameters introduce a translation of each attached catenoid, in the normal direction to the surface it is attached, relative to the surface. The decay which is ensured is from the gluing region towards the waist of the catenoid. This is similar to the treatment of decay and construction of $\overline{\mathcal{K}}(S)$, in the case S is the fusion region in the Wente gluing case.

23. The Desingularizing Constructions

The unscaled standard limit objects in this case are the components of $\widehat{\mathcal{W}}$ immersed into $\mathcal{N}$ by $\boldsymbol{X} \circ \boldsymbol{Q}$. In the case of highly symmetric constructions the scaled standard limit objects are singly periodic Scherk surfaces with the appropriate symmetries imposed and whose parameter is determined by the angle between the intersecting minimal surfaces. In the general case the scaled standard limit objects are again singly periodic Scherk surfaces. To obtain a limit, a point of the curve of intersection has to be fixed, and then the limit is taken of the surfaces enlarged around this point. The parameter of the Scherk surface obtained this way depends on the point chosen, reflecting the changing angle of intersection along the curve.

The unscaled standard regions are the components of $\widehat{\mathcal{W}}$ with small tubular neighborhoods of $\widehat{\mathcal{C}}$ removed, and immersed into $\mathcal{N}$ by $\boldsymbol{X} \circ \boldsymbol{Q}$. The scaled standard regions are built around small neighborhoods of the curves of intersection, and they approximate, close to a point of the curve of intersection, a Scherk surface. The parameter of this Scherk surface and the scaling factor vary along the curve of intersection in the general case, but in the highly symmetric case they are constant. The transition regions are annuli which, close to one boundary circle, approximate tubular neighborhoods of $\widehat{\mathcal{C}}$ in $\widehat{\mathcal{W}}$ immersed by $\boldsymbol{X} \circ \boldsymbol{Q}$, and close to the other they approximate a wing each of a Scherk surface, whose parameter varies along the boundary circle.

The metric on the extended standard region can be conformally changed to h, compactifying this way the region in the highly symmetric case, which is the case I discuss first. Since the corresponding limit object is a minimal surface, the h metric on it is the standard metric on the sphere pulled-back by the Gauss map. Recall the description of this metric from Section 11. The standard limit object is then, because of the symmetries, isometric to the union of a finite number of hemispheres connected by arcs along the boundary equators. It is not hard to see that because of the reflectional symmetries, the kernel of the modified linearized operator $\mathfrak{L}_h = \Delta_h + 2$ is two-dimensional, and is generated by the translations perpendicular to the axis of the Scherk surface. The corresponding $\mathcal{K}(S)$ is then two-dimensional, while $\overline{\mathcal{K}}(S)$ turns out to be four-dimensional, corresponding to the four wings of the Scherk surface. More precisely, to ensure decay along the wings and the transition regions away from the Scherk surfaces, the ability to prescribe a constant on each boundary circle of the transition regions bordering the Scherk surface is needed. To achieve the creation of the required $\hat{w}$, six new parameters for each Scherk surface are introduced, controlling the corresponding modification of the initial surfaces. Two of these correspond to the elements of $\mathcal{K}(S)$ and they amount to the small angles introduced between opposing wings of the Scherk surface by appropriately deforming the Scherk surface. To ensure the wings still match with the catenoidal pieces in the construction of [K5, Theorem E], the same angles have to be introduced between the corresponding catenoidal pieces. This is easy by modifying appropriately the catenoidal pieces successively [K5]. The remaining four

parameters, corresponding to elements of $\overline{\mathcal{K}}(S)$, are angles of bending introduced so that each wing of the Scherk surface fails to match the catenoidal piece by this angle. Correcting this mismatch would amount approximately to changing the surface to the graph of an element of $\underline{\mathcal{K}}(S)$.

As mentioned already, one of the main difficulties in generalizing the approach in the highly symmetric case, to the general case, is the fact that the various parameters are not constant, but vary along the curves of intersection. Another important new feature is that the dimensions of $\mathcal{K}(S)$ and $\overline{\mathcal{K}}(S)$ tend to infinity, as the number of handles replacing each topological circle of intersection tends to infinity. This reflects the fact that the Scherk surface given by Equation (2) in Section 11, when quotiented out by a translation $z \to z+2k\pi$, has the number of its small eigenvalues for $\mathfrak{L}_h = \Delta_h + 2$ tend to infinity as $k \to \infty$. By small eigenvalues I mean the ones with absolute value less than ϵ, where ϵ is some given small positive constant. It is very important to understand the corresponding eigenfunctions. The following lemma gives the information needed:

LEMMA. *Given $\epsilon > 0$ there is $\delta > 0$ such that if f is an eigenfunction on a Scherk surface $\mathcal{S}$, satisfying $\mathfrak{L}_h f = \lambda f$ for some λ with $|\lambda| \leq \delta$, and f is invariant under some translation $z \to z + 2k\pi$, then there is a sequence of vectors $\{a_i \in \mathbb{R}^3\}_{i=1}^k$, such that*

$$\|f - Y^{\perp}\|_{L^2} \leq \epsilon \|f\|_{L^2}, \qquad \|b\|_{\ell^2} \leq \epsilon \|f\|_{L^2},$$

where Y on $\{2(i-1)\pi \leq z < 2i\pi\}$ is the translational Killing field induced by a_i, the L^2 norms are taken over $\mathcal{S} \cap \{0 \leq z \leq 2k\pi\}$, $\|b\|_{\ell^2} := \left(\sum_{i=1}^k b_i^2\right)^{1/2}$, $b_i = a_i - a_{i-1}$, and $a_0 := a_k$.

In other words the eigenfunctions of small eigenvalue approximate the translational Killing fields locally, but the translation involved is allowed to vary slowly from fundamental domain to the next. This result extends to cases where the parameter of the Scherk surface varies slowly, as it does for the surfaces used in the construction. One further issue is the dual character of various data along the curves of intersection: They can be discrete, taking values for each handle, or continuously varying, as when they are considered as boundary data for a partial differential equation, for example. Keeping these difficulties in mind I proceed to outline the construction and proof:

First the construction of the modified initial surfaces has to be carried out depending on the appropriate parameters. The number of the parameters depends on the number of handles used to desingularize the intersecting curves. There are seven parameters for each handle used, similar to the six parameters of the highly symmetric constructions. The difference now is that they vary from handle to handle, and moreover there is an extra parameter corresponding to the translations in the direction of the axis of the Scherk surface. The set of allowed values for these parameters is determined by requiring the following: First, the ℓ^2 norm of the values for the handles used to desingularize one of the components of the curve of intersections is $\leq Cm^{-1/2}$, where $C > 0$ is an appropriate constant depending on the given configuration, and m is the number of the handles. Second, the ℓ^2 norm of the differences of the values between successive handles is $\leq Cm^{-3/2}$.

These discrete parameters have to be converted to smooth data along the curves of intersection, which smooth data approximate the discrete data at each handle. The first of these parameters, $\mu_i \in \mathbb{R}$, where i refers to the corresponding handle,

determines the change of scale from handle to handle, where by scale I mean the factor by which the Scherk surface is homothetically reduced before replacing the curve of intersection. Converting these data to a smooth scale function along the curve involves solving an appropriate Ordinary Differential Equation. The scale function determines then the position of each handle along the curve by integrating. The position of the handles varies substantially — in the same order as the length of the intersection curve — depending on the values of the μ_i's. The next two parameters, $\theta_i \in \mathbb{R}^2$, determine the angle desired between the two pairs of opposing wings in the vicinity of each handle. By smoothing, using the position data already determined, and decomposing the $\mathbb{R}^2$ vector appropriately along the normals of the intersecting surfaces, an angle change $\boldsymbol{\theta}$ smoothly varying along $\mathcal{C}$ is determined, which in turn determines a $\boldsymbol{\chi} = \mathcal{S}\mathcal{J}\boldsymbol{\theta}$ where $\mathcal{J}$ is as in Section 14, and $\mathcal{S}$ is a smoothing operator needed to ensure the appropriate smoothness for $\boldsymbol{\chi}$. The other four parameters $\varphi_i \in \mathbb{R}^4$, determine an angle $\boldsymbol{\varphi}$ smoothly varying along $\widehat{\mathcal{C}}$ — recall that $\widehat{\mathcal{C}}$ covers the curve of intersection $\underline{\mathcal{C}}$ four-to-one.

$\boldsymbol{\chi}$ on $\mathcal{C}$ determines a perturbation of the original immersion of the curve of intersection $\underline{\mathcal{C}}$. This perturbation of the curve of intersection is used in the placement of the Scherk surfaces along the curve of intersection, where a map from the axis of the Scherk surface to each component, is already specified by the data given. The determination of the direction of the wings of the Scherk surface, and hence also the changing parameter of the Scherk surface along the curve, uses the normal one-sided derivatives of $\boldsymbol{\chi}$ along $\mathcal{C}$, to incorporate the introduction of the prescribed angle between the opposing wings, and also $\boldsymbol{\varphi}$, to produce the prescribed bending of the wings relative to the inner core of the Scherk surfaces. This way the immersions of the Scherk surfaces replacing the curves of intersection are determined, up to a certain distance from the axis.

Beyond this distance the wings have to be viewed as graphs over the asymptotic half-planes, to avoid diverging from minimality too much, as was already done in the construction in [K5]. Together with the immersion of the Scherk surfaces close to the axis, immersions of the boundaries of the asymptotic half-planes are automatically determined as well. These amount to a perturbation of $\boldsymbol{X} \circ \boldsymbol{Q}$ on $\widehat{\mathcal{C}}$. This perturbation can be used as Dirichlet data to determine a new perturbation $\boldsymbol{Y}$ of $\boldsymbol{X} \circ \boldsymbol{Q}$ on $\widehat{\mathcal{W}}$, as a graph over the original one, satisfying the minimal surface equation. This immersion plays the role of the asymptotic half-planes, and the immersions of the wings are determined as appropriate graphs over it. The varying scaling factor along the curve, and the varying parameter of the Scherk surface, which have already been determined, are needed to carry out this step. The wings constructed this way are matched smoothly with the Scherk cores already constructed, by a straightforward use of cut-off functions. Away from the Scherk cores they decay to $\boldsymbol{Y}$, to which they are matched eventually, again by a straightforward use of cut-off functions. This completes the construction of the modified initial surfaces.

As seen by the above outline, the construction of the initial surfaces is rather involved, with many steps, some of which require solving Partial Differential Equations. Many estimates have to be produced along the way in order to ensure that the resulting surfaces are smooth enough and appropriately close to the Scherk models used. In particular, the mean curvature on the initial surfaces can be written as $H = \hat{H} + \bar{H}$, where $\hat{H}$ is the mean curvature produced by the modifications of the

initial surface prescribed by the parameters μ_i, θ_i, and φ_i. $\hat{H}$ vanishes when all these parameters vanish, and more precisely on each handle the following "almost holds":

$$(2)\quad \|\hat{H}-\hat{H}_{linear}\| \leq C(|\mu_i|^2+|\theta_i|^2+|\varphi_i|^2+\sum_{i'=i\pm 1}(|\mu_i-\mu_{i'}|+|\theta_i-\theta_{i'}|+|\varphi_i-\varphi_{i'}|)),$$

where i refers to the handle on which the norm is taken, i' to the adjacent handles, and $\hat{H}_{linear}$ is linear in the parameters μ_i, θ_i, and φ_i. The actual estimate is a slight modification of (2), reflecting the non-local definition of $\boldsymbol{\chi}$, but this does not affect its usage.

$\bar{H}$ is of order $1/m$ on each handle. Moreover $\bar{H}$ on subsequent handles changes little; the change which I denote by $\boldsymbol{\partial}^d\bar{H}$ is of order $1/m^2$ on the average in ℓ^2-norm. Solving the linearized equation on the extended standard region containing one of the Scherk cores substituting for a component of the curve of intersection is a two-step process: First, corresponding to each handle of the desingularizing Scherk core, consider the Scherk surface $\mathcal{S}$ of the same parameter, as the one used in the vicinity of this handle. To this Scherk surface, the mean curvature H of the handle in consideration can be transplanted, to provide a periodic E which on each handle approximates H. This is possible because the change of H from handle to handle is small, and so H itself is approximately periodic. The solution of the linearized equation can be attempted then, where the periodicity of $\mathcal{S}$ and E is imposed on the solution as well. This linearized problem is similar to the one in [K5] discussed above. The fundamental domain in the h-metric is a sphere and the kernel is spanned by the three coordinate functions. Solving then as before, a solution u is obtained satisfying $\mathfrak{L}u = -H + \tilde{w}$, where $\tilde{w}$ belongs to a seven-dimensional space and its coordinates approximate the parameters μ_i, θ_i, and φ_i corresponding to the handle of the desingularizing Scherk core in consideration. Moreover u decays exponentially away form the Scherk core.

These solutions can be transplanted back to the corresponding handles of the desingularizing Scherk cores, and be combined to provide an approximate solution $u_{\mathcal{S}}$ on the extended standard region, satisfying

$$(3)\qquad \mathfrak{L}u_{\mathcal{S}} = E_{\mathcal{S}} - H + \tilde{w}_{\mathcal{S}},$$

where $\tilde{w}_{\mathcal{S}} = \hat{w}_{\mathcal{S}} + \bar{w}_{\mathcal{S}}$ belongs to a space of dimension $7m'$, where m' is the number of handles used in desingularizing the component of the intersection curve in consideration, $\hat{w}_{\mathcal{S}}$ has components prescribed by the given parameters, $\bar{w}_{\mathcal{S}}$ has bounds similar to the ones of $\bar{H}$, and $E_{\mathcal{S}}$ is the error due to transplanting and combining of the solutions, and is of order $1/m^2$ on each handle in an ℓ^2-average sense.

To correct for $E_{\mathcal{S}}$ the equation has to be solved on the extended standard region at once. By using a variation of the lemma in this section, and a similar lemma for the low harmonics on the circles which form the common boundary of the Scherk core and the transition regions adjacent to it, the equation

$$\mathfrak{L}u'_{\mathcal{S}} = E_{\mathcal{S}} + \tilde{w}'_{\mathcal{S}}$$

admits a solution $u'_{\mathcal{S}}$ of order $1/m$ in the ℓ^2-average, which decays exponentially along the wings as needed, and where $\tilde{w}'_{\mathcal{S}}$ is of the same order, and moreover the norm of its change $\boldsymbol{\partial}^d\tilde{w}'_{\mathcal{S}}$ from handle to handle is $\leq \epsilon/m^2$, for given ϵ, in the ℓ^2-average. This last estimate is important because it ensures that $\tilde{w}'_{\mathcal{S}}$ can be controlled and prescribed in the construction, and hence a fixed point theorem can

be used to find the desired surface, as in the other constructions. For more details, and for the many technical issues which have to be resolved and which are not mentioned here, I refer the reader to [K6-8].

24. The Doubling of the Clifford Torus Construction

In this case the unscaled standard regions approximate two copies of the Clifford torus from each of which m^2 small discs have been removed. The corresponding limit object is the Clifford torus covered twice. The scaled standard regions are the catenoidal necks, and the standard limit objects are catenoids. When the symmetries are taken into account there is only one neck in each fundamental domain. It turns out that when S is the fundamental domain of the Clifford torus, $\mathcal{K}(S) = \overline{\mathcal{K}}(S)$ is one-dimensional, while for the catenoidal neck they are both 0-dimensional because of the symmetries.

References

[A] A. D. Alexandrov. Uniqueness theorems for surfaces in the large. *V. Vestnik Leningrad Univ.*, 13, No. 19:5–8, 1958. Translation in *Am. Math. Soc. Translations* (Series 2), **21**, 412–416.

[Ch] I. Chavel. *Eigenvalues in Riemannian Geometry.* Academic Press, 1984.

[CHM] M. Callahan, D. Hoffman, and W. H. Meeks III. The structure of singly-periodic minimal surfaces. *Invent. math.*, 99:455–481, 1990.

[Co1] C. Costa. Imersões mínimas em $\mathbb{R}^3$ de gênero um e curvatura total finita. Ph.D. thesis, IMPA, Rio de Janeiro, Brazil, 1982.

[Co2] C. Costa. Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends. *Bull. Soc. Bras. Mat.*, 15:47–54, 1984.

[D] D. Delaunay. Sur la surface de revolution dont la courbure moyenne est constante. *J. Math. Pures et Appl.*, 6:309–320, 1841.

[DHKW] U. Dierkes, S. Hildebrandt, A. Küster, and O. Wohlrab. *Minimal Surfaces I.* Springer-Verlag, 1992.

[GT] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order.* 2nd Edition, Springer-Verlag, 1983.

[H] H. Hopf. Lectures on differential geometry in the large. Technical report. Notes by John Gray, Stanford University.

[HK] D. Hoffman and H. Karcher. Complete embedded minimal surfaces of finite total curvature. *Encyclopaedia of Mathematical Sciences*, 90:5–93, 1997.

[HKM] D. Hoffman, H. Karcher, and W. H. Meeks III. One-parameter families of embedded complete minimal surfaces of finite topology. In preparation.

[HM1] D. Hoffman and W. H. Meeks III. A complete embedded minimal surface with genus one, three ends and finite total curvature. *J. Differential Geometry*, 21:109–127, 1985.

[HM2] D. Hoffman and W. H. Meeks III. A variational approach to the existence of complete embedded minimal surfaces. *Duke Journal of Math.*, 57:877–894, 1988.

[HM3] D. Hoffman and W. H. Meeks III. Embedded minimal surfaces of finite topology. *Ann. of Math.*, 131:1–34, 1990.

[HM4] D. Hoffman and W. H. Meeks III. Limits of minimal surfaces and Scherk's fifth surface. *Arch. Rat. Mech. Anal.*, 111 (2):181–195, 1990.

[HM5] D. Hoffman and W. H. Meeks III. Minimal surfaces based on the catenoid. *Amer. Math. Monthly, Special Geometry Issue*, 97:702–730, 1990.

[K1] N. Kapouleas. Complete constant mean curvature surfaces in euclidean three-space. *Ann. of Math. (2)*, 131:239–330, 1990.

[K2] N. Kapouleas. Compact constant mean curvature surfaces in euclidean three-space. *J. Differential Geometry*, 33:683–715, 1991.

[K3] N. Kapouleas. Constant mean curvature surfaces constructed by fusing Wente tori. *Invent. math.*, 119:443–518, 1995.

[K4] N. Kapouleas. Constant mean curvature surfaces in euclidean spaces. *Proceedings of the International Congress of Mathematicians, Zürich, Switzerland 1994*, pages 481–490, 1995.

[K5] N. Kapouleas. Complete embedded minimal surfaces of finite total curvature. *J. Differential Geometry*, 47:95–169, 1997.

[K6] N. Kapouleas. On desingularizing the intersections of minimal surfaces (research announcement). Preprint.

[K7] N. Kapouleas. On desingularizing the intersections of minimal surfaces I: the compact case. In preparation.

[K8] N. Kapouleas. On desingularizing the intersections of minimal surfaces II. In preparation.

[K9] N. Kapouleas. *The gluing method for surfaces.* In preparation.

[Ka1] H. Karcher. Embedded minimal surfaces derived from Scherk's examples. *Manuscripta Math.*, 62:83–114, 1988.

[Ka2] H. Karcher. Construction of minimal surfaces. *Surveys in Geometry, University of Tokyo*, pages 1–96, 1989.

[Ka3] H. Karcher. The triply periodic minimal surfaces of Alan Schoen and their constant mean curvature companions. *Manuscripta Math.*, 64:291–357, 1989.

[KK] N. Korevaar and R. Kusner. The global structure of constant mean curvature surfaces. *Invent. Math.*, 114:311–332, 1993.

[KKS] N. Korevaar, R. Kusner, and B. Solomon. The structure of complete embedded surfaces with constant mean curvature. *J. Differential Geometry*, 30:465–503, 1989.

[KPS] H. Karcher, U. Pinkall, and I. Sterling. New minimal surfaces in S^3. *J. Differential Geometry*, 28:169–185, 1988.

[KY] N. Kapouleas and S. D. Yang. Minimal surfaces in the round three-sphere. In preparation.

[L] H. B. Lawson and Jr. Complete minimal surfaces in S^3. *Ann. of Math.*, 92:335–374, 1970.

[M] W. H. Meeks III. The geometry and topology of embedded surfaces of constant mean curvature. *J. Diff. Geom.*, 27:539–552, 1988.

[MP] R. Mazzeo and F. Pacard. Constant mean curvature surfaces with Delaunay ends. *Comm. Anal. Geom.*, 9:169–237, 2001.

[MPP] R. Mazzeo, F. Pacard, and D. Pollack. Connected sums of constant mean curvature surfaces in euclidean 3 space. *J. Reine Angew. Math*, 536:115–165, 2001.

[MR] S. Montiel and A. Ros. Schrödinger operators associated to a holomorphic map. *in Global Differential Geometry and Global Analysis (Berlin, 1990), Lecture Notes in Mathematics*, 1481:147–174, 1990.

[Ni] J. C. C. Nitsche. *Lectures on minimal surfaces.* Cambridge university press, 1989.

[Os1] R. Osserman. Global properties of minimal surfaces in E^3 and E^n. *Annals of Math.*, 80 (2):340–364, 1964.

[Os2] R. Osserman. *A survey of minimal surfaces.* Dover publications, New York, 2nd edition, 1986.

[PR] J. Pérez and A. Ros. The space of properly embedded minimal surfaces with finite total curvature. *Indiana Univ. Math. J.*, 45:177–204, 1996.

[PRu] J. Pitts and J. H. Rubinstein. Equivariant minimax and minimal surfaces in geometric three-manifolds. *Bull. A.M.S.*, 19:302–309, 1988.

[S] H. F. Scherk. Bemerkungen über die kleinste Fläche innerhalb gegebener Grenzen. *J. r. angew. Math.*, 13:185–208, 1835.

[Sc1] R. Schoen. Uniqueness, symmetry, and embeddedness of minimal surfaces. *J. Differential Geometry*, 18:791–809, 1983.

[Sc2] R. Schoen. The existence of weak solutions with prescribed singular behavior for a conformally invariant scalar equation. *Comm. Pure Appl. Math.*, 41:371–392, 1988.

[Si] J. Simons. Minimal varieties in riemannian manifolds. *Ann. of Math. (2)*, 88:62–105, 1968.

[Sm] N. Smale. A bridge principle for minimal and constant mean curvature submanifolds in $\mathbb{R}^N$. *Invent. math.*, 90:505–549, 1987.

[T1] M. Traizet. Construction de surfaces minimales en recollant des surfaces de Scherk. *PhD thesis*, 1995.

[T2] M. Traizet. Construction de surfaces minimales en recollant des surfaces de Scherk. *Ann. Inst. Fourier, Grenoble*, 46:1385–1442, 1996.

[W] H. C. Wente. Counterexample to a conjecture of H. Hopf. *Pacific Jour. of Math.*, 121:193–243, 1986.

[W1] M. Wohlgemuth. Higher genus minimal surfaces by growing handles out of a catenoid. *Manuscripta Math.*, 70:397–428, 1991.

[W2] M. Wohlgemuth. Vollständige Minimalflächen höheren Geschlechts und endlicher Totalkrümmung. *PhD thesis, University of Bonn*, 1993.

[WW] M. Weber and M. Wolf. Teichmüller theory and handle addition for minimal surfaces. Preprint.

[Y] Shing-Tung Yau. *Seminar on Differential Geometry*. Princeton University Press, 1982.

[Y1] S. D. Yang. On a connected sum construction for minimal surfaces. *PhD thesis, Brown University*, 1997.

[Y2] S. D. Yang. A connected sum construction for complete minimal surfaces of finite total curvature. *Comm. Anal. Geom.*, 9:115–167, 2001.

Department of Mathematics, 151 Thayer Street, Box 1917, Brown University, Providence, RI 02912

E-mail address: nicos@math.brown.edu

Clay Mathematics Proceedings
Volume 2, 2004

The Conformal Theory of Alexandrov Embedded Constant Mean Curvature Surfaces in $\mathbb{R}^3$

Rafe Mazzeo, Frank Pacard, and Daniel Pollack

Abstract. We first prove a general gluing theorem which creates new nondegenerate constant mean curvature surfaces by attaching half-Delaunay surfaces with small necksize to arbitrary points of any nondegenerate CMC surface. The proof uses the method of Cauchy data matching of Mazzeo and Pacard. In the second part of this paper, we develop the consequences of this result and (at least partially) characterize the image of the map which associates to each complete, Alexandrov-embedded CMC surface with finite topology its associated conformal structure, which is a compact Riemann surface with a finite number of punctures. In particular, we show that this "forgetful" map is surjective when the genus is zero. This proves in particular that the CMC moduli space has a complicated topological structure. These latter results are closely related to those in Kusner's paper in this same volume.

1. Introduction

In this paper we consider the class of surfaces $\Sigma \subset \mathbb{R}^3$ which have constant mean curvature (CMC) equal to 1, are complete and of finite topology. We shall assume that these surfaces are properly immersed, and in fact satisfy the stronger condition that they are *Alexandrov embedded*. This last condition means that the immersion $\varphi : \Sigma \hookrightarrow \mathbb{R}^3$ extends to a proper immersion $Y \hookrightarrow \mathbb{R}^3$ of a three-manifold Y whose boundary is the surface Σ, $\partial Y = \Sigma$. We remark that, except for the sphere, no compact CMC surface, including Wente tori, ever satisfy this condition. The space of all surfaces of this type, of genus g with k ends, will be denoted $\mathcal{M}_{g,k}$. To be definite, we do not identify elements of this space which differ by Euclidean motions.

The simplest examples of surfaces of this type are the rotationally invariant Delaunay surfaces. Up to Euclidean motions, these are parametrized by a single "necksize" parameter τ, and will be denoted D_τ. We discuss these at greater length later. A remarkable theorem of Meeks [**13**] states that each end of an Alexandrov embedded surface is cylindrically bounded, and using this Korevaar, Kusner and Solomon [**6**] proved that each end is in fact strongly convergent to some Delaunay surface, and thus each end has an associated asymptotic necksize parameter. This

Mazzeo was supported by the NSF under grant DMS-9971975, and Pollack under grant DMS-9704515. All three authors were supported at MSRI by the NSF under grant DMS-9701755.

©2004 Clay Mathematics Institute

strong control on the asymptotic geometry of these surfaces makes it possible to understand the rudimentary structure of the corresponding moduli spaces, and in [**7**] it is proved that $\mathcal{M}_{g,k}$ is always a locally real analytic space of virtual dimension $3k$. Somewhat remarkably, this dimension only depends on the number of ends and not the genus. Furthermore, if $\Sigma \in \mathcal{M}_{g,k}$ satisfies a certain analytic nondegeneracy condition, which we explain below, then the moduli space is a smooth real analytic manifold of dimension $3k$ in a neighborhood of the point Σ.

To describe this nondegeneracy condition, recall that the Jacobi operator L_Σ of a CMC surface Σ is the linearization of the mean curvature operator at Σ. Surfaces which are $\mathcal{C}^2$ close to Σ may be parametrized as normal graphs, that is, as the images of the map

$$\Sigma \ni x \longmapsto x + w(x)\nu(x)$$

where ν is the unit normal to Σ at x and w is the (scalar) displacement function. With respect to this representation, L_Σ takes the simple form $\Delta_\Sigma + |A_\Sigma|^2$. Functions in the nullspace of L_Σ are called Jacobi fields.

DEFINITION 1. *The surface $\Sigma \in \mathcal{M}_{g,k}$ is said to be nondegenerate if it has no nontrivial Jacobi fields which decay at all ends of Σ.*

It is natural to investigate the structure of the moduli spaces $\mathcal{M}_{g,k}$ in greater detail. Two basic questions are: first, for which values of g and k is $\mathcal{M}_{g,k}$ nonempty, and second, when does $\mathcal{M}_{g,k}$ contain a nondegenerate element? We now discuss briefly what is known about these questions.

That $\mathcal{M}_{g,1}$ is empty for all g is due to Meeks [**13**]. Also, $\mathcal{M}_{g,2}$ is empty unless $g = 0$, and in this case this space contains only the Delaunay surfaces. As we indicate later, Delaunay surfaces are always nondegenerate. (As a check on dimensions, note that there is a one-dimensional family of the Delaunay surfaces with fixed axis which satisfy an additional "positioning" normalization determining their translational location along this axis. The group of Euclidean motions is six dimensional, but the rotations about the fixed axis act trivially on these surfaces. Thus there is a $3 \cdot 2 = 6$-dimensional family of surfaces with 2 ends, as predicted.) To proceed further it is necessary to use transcendental methods, in particular, analytic gluing constructions of these surfaces. The first such surfaces with $k \geq 2$ were constructed by Kapouleas [**4**], but while his method gives elements in $\mathcal{M}_{g,k}$ for infinitely many values of g and k, it provides little geometric or analytical control on the surfaces themselves, and more specifically it seems hard to determine whether the surfaces he constructs satisfy the nondegeneracy condition. More recently, the first two authors [**8**] introduced a new method to handle these gluing constructions which involves matching Cauchy data across the gluing interfaces. This technique has many advantages over previous methods: it involves considerably fewer technicalities (and thus makes it possible to approach more complicated geometric problems of this type), it provides very good control of the resulting surfaces, and most importantly, one may often prove that the surfaces obtained in this way are nondegenerate. The main result of [**8**] is that if M is a complete *minimal* surface of finite total curvature in $\mathbb{R}^3$, of genus g with k (asymptotically catenoidal, but not planar) ends, Alexandrov-embedded and satisfying an analogous nondegeneracy condition, then it is possible to obtain a nondegenerate CMC surface Σ of genus g with k ends by gluing half-Delaunay surfaces onto the boundaries of a truncation of M. This construction yields the existence of nondegenerate elements

in $\mathcal{M}_{g,k}$ for infinitely many values of g and k. This result is a basic ingredient in the nice result of Große-Brauckman, Kusner and Sullivan [**3**], who prove that modulo rigid motion $\mathcal{M}_{0,3}$ is a three-dimensional ball; in particular this space is connected. (Unfortunately, their method does not prove that every element in this space is nondegenerate, although this is quite likely true.) As a further development of this new gluing method, a connected sum construction for compact CMC surfaces with boundary is given in [**10**].

Beyond these theorems, very little else is known about these moduli spaces. It is of interest to determine even the most basic properties of their topological structure. For example, it is even unknown whether these spaces are ever disconnected. It is also unknown whether the natural Lagrangian structure [**7**] on these moduli spaces can be used in any significant way.

The present paper has two main parts which we describe now in turn. In the first, which occupies the bulk of the paper, we establish a new gluing theorem:

THEOREM 1. *Let $\Sigma \in \mathcal{M}_{g,k}$ be nondegenerate. Then for any point $p \in \Sigma$ there is a one-parameter family of nondegenerate CMC surfaces $\Sigma_\tau(p) \in \mathcal{M}_{g,k+1}$ obtained by gluing a half-Delaunay surface D_τ, with τ sufficiently small, to Σ at p.*

The geometry of Σ away from the point p is perturbed very little in this construction, and in fact as $\tau \to 0$, $\Sigma_\tau(p)$ converges, on compact subsets of $\mathbb{R}^3 - \{p\}$, to the union of the initial surface Σ and an infinite family of mutually tangent spheres of radius 2 arranged along a ray normal to Σ at p. We actually prove a slightly stronger theorem: we show that if $\Sigma \in \mathcal{M}_{g,k}$ is nondegenerate and $p \in \Sigma$, then there are two distinct one-parameter families of nondegenerate CMC surfaces $\Sigma_\tau^\pm(p)$ with $k+1$ ends; these are obtained by gluing half of either an embedded Delaunay surface (an unduloid) or an immersed Delaunay surface (a nodoid) with very small neck. The surfaces $\Sigma_\tau^-(p)$ obtained by gluing a nodoid on to Σ are no longer Alexandrov-embedded, but we see that they behave very much like Alexandrov-embedded CMC surfaces. (As an aside, this property of nodoids behaving like Alexandrov-embedded CMC surfaces is only true when the necksize is small; the paper [**9**] shows that this fails rather dramatically when nodoids with large necksizes are considered.) We remark also that this construction was anticipated and motivated by some numerical and computer graphic studies carried out (and brought to our attention) several years ago by Große-Brauckmann. There are nice illustrations of these surfaces at `http://www.gang.umass.edu/cmc/`.

An immediate consequence of this gluing theorem is:

COROLLARY 1. *If $\mathcal{M}_{g,k}$ contains a nondegenerate element, then for any $k' > k$, the moduli space $\mathcal{M}_{g,k'}$ also contains a nondegenerate element.*

The second part of this paper shows how Theorem 1 can be used to obtain some information about the global structure of these CMC moduli spaces. To describe these results, recall from [**6**] that a complete Alexandrov-embedded CMC surface of finite topology is conformally equivalent to the complement of a finite number of points in a compact Riemann surface. Thus any $\Sigma \in \mathcal{M}_{g,k}$ is conformally equivalent to a punctured compact Riemann surface $\overline{\Sigma} - \{p_1, \ldots, p_k\}$. This suggests that it might be useful to consider the 'forgetful map'

$$\mathcal{F}_{g,k} = \mathcal{F} : \mathcal{M}_{g,k} \longrightarrow \mathcal{T}_{g,k}.$$

By definition, $\mathcal{T}_{g,k}$ is the Teichmüller space of conformal structures on a surface of genus g with k punctures, and this map is defined by sending $\Sigma \in \mathcal{M}_{g,k}$ to the marked conformal structure $[\Sigma] \in \mathcal{T}_{g,k}$ determined by its (Alexandrov) embedding. In other words, this map forgets the CMC embedding of this surface and only retains its conformal structure.

To begin:

THEOREM 2. *For each g and k, the map $\mathcal{F}_{g,k}$ is real analytic.*

This statement requires clarification. As already indicated, $\mathcal{M}_{g,k}$ is a locally real analytic space. What this means is that for any $\Sigma \in \mathcal{M}_{g,k}$ there is an open finite-dimensional real analytic manifold Y in the space of all Alexandrov-embedded surfaces near to Σ (in an appropriate topology) which contains a neighborhood $\mathcal{U}$ of Σ in $\mathcal{M}_{g,k}$, such that $\mathcal{U}$ is closed in Y and is given as the zero set of a real analytic function in Y. When Σ is nondegenerate, then Y can be taken as the neighborhood $\mathcal{U}$ in $\mathcal{M}_{g,k}$ itself. To say that $\mathcal{F}$ is analytic on $\mathcal{M}_{g,k}$ means that it has a real analytic extension to all such real analytic manifolds Y.

One consequence of the real analyticity of these spaces and maps is that $\mathcal{M}_{g,k}$ is stratified by open real analytic manifolds S_j such that on each of these strata $\mathcal{F}$ has constant rank.

We are not claiming that $\mathcal{F}$ is a proper mapping, and indeed, it is clear (from any of the gluing constructions) that this is false. In another paper in this volume [**5**], Kusner shows that there is a suitable modification of the forgetful map which is proper.

Recall next that the dimension of $\mathcal{M}_{g,k}$ around nondegenerate points is $3k$, while on the other hand, $\mathcal{T}_{g,k}$ is $6g - 6 + 2k$-dimensional. This suggests that $\mathcal{F}$ might conceivably be surjective. This is most likely false, in general, but our next theorems address this issue:

THEOREM 3. *Let $g = 0$; then for any $k \geq 3$, the map $\mathcal{F}_{0,k}$ is surjective.*

Kusner has obtained a different proof of this result [**5**] using his theorem about the properness of $\mathcal{F}$.

THEOREM 4. *Fix $g \geq 1$ and suppose that $\mathcal{M}_{g,k_0}$ contains a nondegenerate element for some $k_0 \geq 3$. Then for any $k \geq k_0$, the image of $\mathcal{F}_{g,k}$ contains an analytic submanifold of codimension $d_{g,k}$ which is uniformly bounded as $k \to \infty$. In other words, the codimension of the image of $\mathcal{F}_{g,k}$ is bounded as $k \to \infty$ for each g.*

To explain the statement of this last theorem, recall the stratified structures of these spaces and maps. The image $\mathcal{I}_{g,k}$ of $\mathcal{F}_{g,k}$ is again a stratified real analytic space, and we say that the maximum dimension of any one of these strata in the image is the dimension of $\mathcal{I}_{g,k}$.

Proceeding further, we note that the space $\mathcal{T}_{g,k}$ has rather nontrivial topology. In fact, there is a natural mapping

$$F' : \mathcal{T}_{g,k} \longrightarrow \mathcal{C}_{g,k}$$

which carries the marked Riemann surface $[(\overline{\Sigma}; p_1, \ldots, p_k)]$ to the associated element in the configuration space of k distinct ordered points on a surface of genus g. The fundamental group and cohomology ring of this configuration space have been

studied intensively, see [**1**], and are known to be rather complicated. It is (barely) conceivable that the image of $F' \circ \mathcal{F}_{g,k}$ does not see any of this topology, but this is not the case.

THEOREM 5. *When $g = 0$ and $k \geq 3$, the map*

$$(F' \circ \mathcal{F}_{g,k})_* : \pi_1(\mathcal{M}_{0,k}) \longrightarrow \pi_1(\mathcal{C}(0,k))$$

is an epimorphism. If $g \geq 1$ and $\mathcal{M}_{g,k_0}$ contains a nondegenerate element for some $k_0 \geq 3$, then for any $k \geq k_0$, the image of the fundamental group $\pi_1(\mathcal{M}_{g,k})$ under the homomorphism $(F' \circ \mathcal{F}_{g,k})_*$ contains a finitely generated group with an increasing number of generators as $k \to \infty$.

We refer to the final section of this paper for a more detailed statement.

The end-to-end gluing construction of Ratzkin [**12**], also implies the topological nontriviality of $\mathcal{M}_{g,k}$, but gives less information than is obtained here.

These results together constitute the first and simplest steps of a more detailed investigation of the topology of the moduli spaces $\mathcal{M}_{g,k}$.

The outline of the rest of the paper is as follows. Sections 2 through 4 contain the proof of the gluing result, Theorem 1. More specifically, §2 contains the analysis of CMC deformations of half-Delaunay surfaces and §3 contains the analysis of CMC deformations of $\Sigma - D$, where Σ is any element in $\mathcal{M}_{g,k}$ and D is a small geometric disk in Σ. These results are then combined in §4, where it is shown how to perform the Cauchy data matching across the gluing interface, and hence how to produce a new CMC surface with $k+1$ ends. Finally, in §5 we undertake the analysis of the forgetful map and develop its properties and derive the theorems stated above concerning the image of $\mathcal{M}_{g,k}$.

The authors wish to thank Matthias Weber and Rob Kusner for a number of useful conversations. The first author also wishes to thank Ralph Cohen and Gunnar Carlsson for setting him straight on the topology of configuration spaces.

2. The Geometry and Analysis of Delaunay Surfaces

In this section we recall the family of Delaunay surfaces D_τ, and review some of their basic geometric and analytic properties. We focus particularly on their behavior in the singular limit as $\tau \to 0$.

2.1. Definition and first properties. To find all CMC surfaces in $\mathbb{R}^3$ which are rotationally invariant about an axis, one is quickly led to the ODE which the generating curve for such surfaces must satisfy. This ODE has a first integral, and from this one can see that all of its solutions are periodic. Thus we obtain a family of periodic, rotationally invariant surfaces D_τ. With a particular normalization described in the next section, the parameter τ lies in $(-\infty, 0) \cup (0, 1]$; when $\tau > 0$, the surface D_τ is embedded, while when $\tau < 0$, it is only immersed. There is a geometric limit, as $\tau \to 0^{\pm}$, consisting of an infinite arrangement of spheres, each tangent to the next, arranged along an axis. We shall not describe this more carefully, but instead refer to [**8**] for details of the preceding discussion as well as the material in the remainder of this section.

The parametrization given by this cylindrical coordinate description is not very tractable analytically, but fortunately it turns out that there is a parametrization

which is much easier to use. This is the isothermal parametrization given by

$$X_\tau : \mathbb{R} \times S^1 \ni (s,\theta) \longmapsto \tfrac{1}{2}\left(\tau\, e^{\sigma(s)}\, \cos\theta, \tau\, e^{\sigma(s)}\, \sin\theta, \kappa(s)\right), \tag{2–1}$$

where the functions σ and κ are described as follows (see §3 of [**8**]). For any $\tau \in (0,1]$, the function σ is defined to be the unique smooth nonconstant solution of the ODE

$$(\partial_s \sigma)^2 + \tau^2 \cosh^2 \sigma = 1, \qquad \partial_s \sigma(0) = 0, \qquad \sigma(0) = -\operatorname{arccosh} 1/\tau,$$

while, for any $\tau \in (-\infty, 0)$, the function σ is defined to be the unique smooth nonconstant solution of the ODE

$$(\partial_s \sigma)^2 + \tau^2 \sinh^2 \sigma = 1, \qquad \partial_s \sigma(0) = 0, \qquad \sigma(0) = \operatorname{arcsinh} 1/\tau.$$

Again, the definition of κ differs according to whether τ is positive or negative. If $\tau \in (0,1]$, then we define the function κ by

$$\partial_s \kappa = \tau^2\, e^{\sigma}\, \cosh\sigma, \qquad \kappa(0) = 0,$$

while if $\tau < 0$, we define the function κ by

$$\partial_s \kappa = \tau^2\, e^{\sigma}\, \sinh\sigma, \qquad \kappa(0) = 0.$$

Observe that when $\tau > 0$, κ is monotone increasing, and hence X_τ is an embedding, whereas when $\tau < 0$, this is no longer true and the surfaces are only immersed. The Delaunay surfaces D_τ are known as unduloids and nodoids in these two cases, respectively. The extreme element in the family of unduloids is D_1, the cylinder of radius $1/2$. The limit of D_τ, either as $\tau \nearrow 0$ or as $\tau \searrow 0$, is an infinite union of tangent spheres of radius 1, arrayed along a common axis.

As noted above, these surfaces are all periodic. When $\tau > 0$, there is a unique $t_\tau > 0$ such that $\tau \cosh t_\tau = 1$, and if we define

$$s_\tau := \frac{1}{2} \int_0^{t_\tau} \frac{dt}{\sqrt{1 - \tau^2 \cosh^2 t}}, \tag{2–2}$$

then it is not hard to see that σ has period $8\, s_\tau$. On the other hand, when $\tau < 0$, we define $t_\tau > 0$ by the equation $\tau \sinh t_\tau = -1$ and define

$$s_\tau := \frac{1}{2} \int_0^{t_\tau} \frac{dt}{\sqrt{1 - \tau^2 \sinh^2 t}}, \tag{2–3}$$

once again σ is periodic of period $8\, s_\tau$.

2.2. The singular limit of D_τ as $\tau \to 0$. We now describe some aspects of the surfaces D_τ as $\tau \to 0$. In this limit, D_τ converges to a singular "noded" surface, which is the infinite union of spheres of radius 1 centered at the points $(0,0,2\,k+1)$, $k \in \mathbb{Z}$. From (2–2) or (2–3) we see that

$$s_\tau = -\tfrac{1}{4} \log \tau^2 + \mathcal{O}(1) \qquad \text{as} \quad \tau \to 0.$$

To study this limit more closely, notice that the family of rescaled surfaces $\tau^{-2}\, D_\tau$ converges to a catenoid of revolution around the z-axis. Thus when τ is small, D_τ is well approximated by a sequence of spheres along the z-axis connected by small catenoidal necks. This is central in much of what follows, and so in the remainder of this subsection we make this quantitative by recalling the behavior of the functions σ and κ as $\tau \to 0$. The expansions below are not hard to derive, but we refer to [**8**] for detailed proofs.

DEFINITION 2. *The notation $g = \mathcal{O}_{\mathcal{C}^\infty}(f)$ means that for any $k \geq 0$ there exists $c_k > 0$ such that*

$$|\partial_s^k g| \leq c_k |f|$$

on the domains of definition of these functions.

To obtain the asymptotics of σ as $\tau \to 0$, define

$$r(s) := \frac{\tau}{2} e^{\sigma(s)}.$$

This function satisfies the ODE

$$(\partial_s r)^2 = r^2 - \Big(\frac{\tau^4}{16} \pm \frac{\tau^2}{2} r^2 + r^4\Big),$$

where the plus or minus sign is chosen according to the sign of τ (+ when $\tau > 0$ and $-$ when $\tau < 0$). Because σ attains its minimum when $s = 0$, $r(s) \geq r(0)$ for all s. Hence we can write $r(s) := r(0) \cosh w(s)$ for some function w. This new function satisfies

$$(\partial_s w)^2 = 1 \mp \frac{\tau^2}{2} - r(0)^2 (\cosh^2 w + 1).$$

Now simply from the definition of $\sigma(0)$ we have

$$r(0) = \frac{\tau^2}{4} + \mathcal{O}(\tau^4) \qquad \text{as} \quad \tau \to 0,$$

and therefore

$$w(s) = \Big(1 \mp \frac{\tau^2}{2}\Big)^{1/2} s + \mathcal{O}_{\mathcal{C}^\infty}(\tau^4 \cosh^2 s).$$

Since $s_\tau \sim -\frac{1}{4} \log \tau^2$, $\tau^4 \cosh^2 s \leq c$ when $|s| \leq 4 s_\tau$, and so, in order to evaluate the function σ, this estimate is only interesting when $s \in [-4\,s_\tau, 4\,s_\tau]$. However, by periodicity we obtain an estimate for w for all $s \in \mathbb{R}$. From here it follows that

$$\kappa(s) = \pm \frac{\tau^2}{2} s + \mathcal{O}_{\mathcal{C}^\infty}(\tau^4 \cosh^2 s),$$

according to the sign of τ (with + when $\tau > 0$ and $-$ with $\tau < 0$) but this is only of interest when $s \in [-2\,s_\tau, 2\,s_\tau]$. It is possible to refine this argument to obtain a more precise expansion for κ in the larger interval $[-4\,s_\tau, 4\,s_\tau]$, but we omit this since it will not be needed.

The image of X_τ restricted to $(0, 2\,s_\tau] \times S^1$ may also be written as the graph of functions U_τ, which is defined over an annulus in the xy-plane. (The image of X_τ restricted to $[-2s_\tau, 0) \times S^1$ is the graph of $-U_\tau$.) To analyze U_τ as $\tau \to 0$, we use the function $r = \frac{\tau}{2} e^{\sigma(s)}$, as above and set

$$r_\tau := \frac{\tau}{2} e^{\sigma(-s_\tau)}.$$

For the moment, observe that $r_\tau \sim \tau^{3/2}$ as τ tends to 0. Expanding κ in terms of r near $r = r_\tau$, we find that when $r_\tau/2 \leq r \leq 2r_\tau$,

$$\Big|(r\,\nabla)^k \Big(U_\tau(r,\theta) \pm \frac{\tau^2}{4} \log\Big(\frac{8r}{\tau^2}\Big)\Big)\Big| \leq c_k |\tau|^3 \quad \text{for all } k \geq 0, \tag{2–4}$$

according to the sign of τ (with + when $\tau > 0$ and $-$ with $\tau < 0$). Similar estimates are valid on a larger set, but this behavior near $r = r_\tau$ will be the most crucial later.

2.3. The Jacobi operator on a Delaunay surface. Let Σ be a CMC surface. Any surface which is $\mathcal{C}^2$ close to it may be represented as a normal graph

$$\Sigma_w = \{x + w(x)\nu(x) : x \in \Sigma\},$$

where ν is the unit normal vector field and w is a (small) scalar function. Σ_w is itself CMC provided w satisfies a nonlinear second order elliptic equation. This equation will be discussed in more detail later, but it is well-known that its linearization, $\mathcal{L}$, which is usually called the Jacobi operator, is the sum of the Laplace–Beltrami operator on Σ and the norm squared of the second fundamental form,

$$\mathcal{L} = \Delta_\Sigma + |A_\Sigma|^2.$$

A solution w of the equation $\mathcal{L}w = 0$ is called a Jacobi field on Σ. Ideally, such a function is the tangent vector of a one-parameter family of CMC deformations of Σ. This may not be the case, however, and in general, if w is a Jacobi field, then the elements in any one-parameter family of surfaces $\{\Sigma_{w(\varepsilon)} : |\varepsilon| < \varepsilon_0\}$, where $w'(0) = w$, only satisfy the constant mean curvature equation to second order at $\varepsilon = 0$. Nonetheless, an understanding of the Jacobi fields and mapping properties of the Jacobi operator is fundamental to any account of the deformation theory of Σ.

We denote by $\mathcal{L}_\tau$ the Jacobi operator associated to the Delaunay surface D_τ. In terms of the isothermal parametrization from §2.1,

$$\mathcal{L}_\tau = \frac{4}{\tau^2 e^{2\sigma}}\big(\partial_s^2 + \partial_\theta^2 + \tau^2 \cosh(2\sigma)\big).$$

We analyze instead the simpler operator

$$L_\tau := \partial_s^2 + \partial_\theta^2 + \tau^2 \cosh(2\sigma), \tag{2–5}$$

with the factor $4/(\tau^2 e^{2\sigma})$ removed; it is clear that the mapping properties of one of these operators implies the corresponding properties of the other, and also, their nullspaces are the same. Observe that, since σ is $8\,s_\tau$ periodic and even, the potential in L_τ is $4\,s_\tau$-periodic.

We now undertake a thorough analysis of the operator L_τ.

2.4. Jacobi fields on D_τ. The uniqueness theorem for 2-ended Alexandrov-embedded CMC surfaces [**6**] implies that the only CMC deformations of an entire Delaunay surface are the obvious ones, namely those arising from rigid motions of $\mathbb{R}^3$ and changes in the Delaunay parameter. The corresponding spaces of Jacobi fields are either bounded or linearly growing along the Delaunay axis. All other Jacobi fields grow exponentially in one direction or the other along this axis, and hence do not correspond to actual CMC deformations. In the next subsection we describe this former class of "geometric" Jacobi fields, while in the one following that we discuss some features of the latter class. While these do not lead to global CMC deformations, they do correspond to CMC deformations over half of D_τ, and in any case, it is necessary to understand them in order to describe the mapping properties of $\mathcal{L}_\tau$.

2.4.1. *Geometric Jacobi fields.* As indicated above, there is a special collection of Jacobi fields on D_τ which correspond to explicit geometric deformations of this surface. This family is six-dimensional; there is a three-dimensional space associated to translations in $\mathbb{R}^3$, a two-dimensional space associated to the rotations of the Delaunay axis, and a one-dimensional space associated to changing the Delaunay parameter. We now describe six Jacobi fields, which we denote by $\Phi_\tau^{j,\pm}$, $j = 0, \pm 1$, which form a basis for this space. The motivation for this notation will be made apparent in subsection 2.4.2.

We first treat the case where $\tau \neq 1$.

Let us start with $\Phi_\tau^{0,+}$. This Jacobi field corresponds to an infinitesimal translation of D_τ along its axis, and so is obtained by projecting the constant vector field $(0,0,1)$ (which is the Killing field associated to this family of translations) along the normal vector field N_τ on D_τ. It is geometrically obvious that $\Phi_\tau^{0,+}$ depends only on s, and is periodic in s, hence is bounded as $s \to \pm\infty$.

Next, the two Jacobi fields $\Phi_\tau^{\pm 1,+}$ correspond to translations of D_τ in the two directions orthogonal to its axis. These are calculated by projecting the constant vector fields $(1,0,0)$ and $(0,1,0)$ along N_τ, and hence are once again bounded in s. Moreover,

$$\Phi_\tau^{1,+}(s,\theta) = \phi_\tau^{1,+}(s)\cos\theta, \qquad \text{and} \qquad \Phi_\tau^{-1,+}(s,\theta) = \phi_\tau^{-1,+}(s)\sin\theta.$$

Continuing, the two Jacobi fields

$$\Phi_\tau^{1,-}(s,\theta) = \phi_\tau^{1,-}(s)\cos\theta, \qquad \text{and} \qquad \Phi_\tau^{-1,-}(s,\theta) = \phi_\tau^{-1,-}(s)\sin\theta$$

correspond to infinitesimal rotations of D_τ about its axis. These are given by projecting the Killing fields $(z, 0, -x_1)$ and $(0, z, -x_2)$ (corresponding to rotations in the space $\mathbb{R}^3$ with coordinates (x_1, x_2, z)) along N_τ, and hence grow linearly in s.

Finally, $\Phi_\tau^{0,-}$ corresponds to the derivative of the one-parameter family D_τ obtained by varying the Delaunay parameter τ. Since D_τ are surfaces of revolution, this Jacobi field depends only on s. It is linearly growing when $\tau \neq 1$.

The definitions for $\Phi_\tau^{0,-}$, $\Phi_\tau^{\pm 1,-}$ and $\Phi_\tau^{\pm 1,+}$ when $\tau = 1$ are exactly the same, but since D_1 is a cylinder, the definition for $\Phi_1^{0,+}$ above yields 0. There is an intrinsic way to define the Jacobi fields $\Phi_\tau^{0,\pm}$ for all $\tau \in (0,1]$ by regarding the change of Delaunay parameter and translation along the axis as playing the role of polar coordinates in the space of parameters. More precisely, let (ρ, α) be the ordinary polar coordinates (corresponding to Cartesian coordinates $(y_1, y_2) = (\rho\cos\alpha, \rho\sin\alpha)$). Then there is a smooth map from a small ball in $\mathbb{R}^2$ into the two-dimensional space of Delaunay surfaces sharing the same axis, given by

$$[0,1) \times [0, 2\pi) \ni (\rho, \alpha) \longrightarrow D_{1-\rho} + \left(0,\, 0,\, \frac{\kappa_{1-\rho}(8\, s_{1-\rho})}{4\pi}\, \alpha\right).$$

Then for any value of τ, the Jacobi fields $\Phi_\tau^{0,\pm}$ have the same span as the variations of this family with respect to the Cartesian coordinates.

Further details of these calculations, as well as explicit expressions of these geometric Jacobi fields in terms of σ and κ can be found in [**8**].

2.4.2. *Jacobi fields of exponential type and indicial roots.* We now fit the special family of Jacobi fields discussed above into a broader context.

The operator L_τ is invariant with respect to rotations about the Delaunay axis, and is reduced by the eigendecomposition for the cross-sectional operator ∂_θ^2. This

reduces the analysis of L_τ to that of the countable family of operators

$$L_{\tau,j} := \partial_s^2 + (\tau^2 \cosh(2\sigma) - j^2), \qquad j = 0, 1, 2, \ldots.$$

When $j \neq 0$, the operator $L_{\tau,j}$ occurs with multiplicity 2, corresponding to the two eigenfunctions $e^{\pm ij\theta}$.

Except in certain exceptional cases described below, for each j there exists a complex number $\zeta_{\tau,j}$, with $\operatorname{Re}\zeta_{\tau,j} \geq 0$, and a basis of solutions $\Psi_\tau^{j,\pm}$ of each of these ordinary differential operators such that

$$\Psi_\tau^{j,\pm}(s + 4s_\tau) = e^{\pm 4\,\zeta_{\tau,j}\, s_\tau}\, \Psi_\tau^{j,\pm}(s).$$

Thus when $\zeta_{\tau,j}$ is real, $\Psi_\tau^{j,\pm}$ grows exponentially as $s \to \pm\infty$ and decays exponentially as $s \to \mp\infty$. It happens that either $\zeta_{\tau,j}$ is real or it is pure imaginary. In this case, the corresponding solutions are oscillatory, hence bounded.

The exceptional cases noted above occur when there is one bounded (oscillatory) solution and one solution which grows linearly. For example, we have already seen that this is the case when $j = 0, \pm 1$.

One way to prove these facts is to recall that since the potential in $L_{\tau,j}$ has period $4s_\tau$, there exists a 2×2 matrix $T_{\tau,j}$ such that for any solution w of $L_{\tau,j}w = 0$ on $\mathbb{R}$,

$$\begin{pmatrix} w(s + 4\,s_\tau) \\ \partial_s w(s + 4\,s_\tau) \end{pmatrix} = T_{\tau,j} \begin{pmatrix} w(s) \\ \partial_s w(s) \end{pmatrix}.$$

Using the Wronskian related to the operator $L_{\tau,j}$, it is easy to see that the determinant of this matrix is equal to 1. Since the matrix $T_{\tau,j}$ has real entries, we see that the roots of its characteristic polynomial $\lambda_{\tau,j}^{\pm}$ satisfy $\lambda_{\tau,j}^{+}\lambda_{\tau,j}^{-} = 1$, and $\lambda_{\tau,j}^{+} + \lambda_{\tau,j}^{-} \in \mathbb{R}$. So we can write them as

$$\lambda_{\tau,j}^{\pm} = e^{\pm 4\,\zeta_{\tau,j}\, s_\tau},$$

where $\zeta_{\tau,j}$ is either real or purely imaginary. In the case where $\lambda_{\tau,j}^{\pm} = 1$ or -1 it may happen that the matrix $T_{\tau,j}$ cannot be diagonalized and this corresponds to the exceptional cases we were mentioning above. We leave the details of checking the statements for the last section.

DEFINITION 3. *For each j we define the indicial roots of the operator $L_{\tau,j}$ to be the pair of numbers $\pm\gamma_{\tau,j}$ where*

$$\gamma_{\tau,j} := Re\,\zeta_{\tau,j} \geq 0.$$

Thus

$$\Gamma_\tau := \{\pm\gamma_{\tau,j} : j \in \mathbb{N}\}$$

is the set of all indicial roots of the operator L_τ.

It is not necessary to express these indicial roots exactly (and indeed, it is probably impossible to do so), but the following estimates will suffice for our nefarious purposes.

PROPOSITION 1. *The indicial roots of L_τ satisfy the following properties:*

(i) *For any $\tau \in (-\infty, 0) \cup (0, 1]$, $\gamma_{\tau,0} = \gamma_{\tau,1} = 0$.*
(ii) *When $j \geq 2$ and $\tau \in (-\sqrt{j^2 - 2}, 0) \cup (0, 1]$, $\gamma_{\tau,j} > 0$ and $L_{\tau,j}$ satisfies the maximum principle.*
(iii) *For any $j \geq 2$, $\lim_{\tau\to 0} \gamma_{\tau,j} = j$.*

PROOF. Property (i) follows from the remarks above. Notice that because of the existence of linearly growing solutions, $\zeta_{\tau,j}$, $j = 0, 1$, is identically zero for all τ, instead of just having real part zero.

To prove property (ii), it suffices to show that when τ is in the stated range, the potential in $L_{\tau,j}$ is strictly negative and so this operator satisfies the maximum principle. Hence it cannot have bounded solutions and therefore its indicial roots must be real and nonzero. To show this, first assume that $\tau > 0$. Then

$$\tau^2 \cosh(2\sigma) - j^2 = 2\tau^2 \cosh^2\sigma - \tau^2 - j^2 = 2 - \tau^2 - j^2 - 2(\partial_s \sigma)^2,$$

which is strictly negative when $j \geq 2$. On the other hand, when $\tau < 0$,

$$\tau^2 \cosh(2\sigma) - j^2 = 2\tau^2 \sinh^2\sigma + \tau^2 - j^2 = 2 + \tau^2 - j^2 - 2(\partial_s \sigma)^2,$$

which is strictly negative when $-\sqrt{j^2 - 2} < \tau < 0$.

Finally, Property (iii) reflects the fact that as τ tends to 0, the potential in L_τ is arbitrarily close to 0 on sets which are arbitrary large. To be more precise, according to Proposition 13 of [**8**], for any $\eta > 0$ there exist numbers $s_\eta > 0$ and $\tau_0 > 0$ such that when $|\tau| \in (0, \tau_0)$,

$$\tau^2 \cosh(2\sigma) \leq \eta,$$

when $s \in [s_\eta, 4s_\tau - s_\eta]$. It follows from this that the indicial roots of $L_{\tau,j}$ must converge to the indicial roots of the operator $\partial_s^2 - j^2$ as $\tau \to 0$, see Proposition 20 in [**8**] for details. □

This proposition proves that there exists a number $\tau_* \leq -\sqrt{2}$ such that

$$\text{(2–6)} \qquad \big(j \geq 2 \text{ and } \tau \in (\tau_*, 0) \cup (0, 1]\big) \Longrightarrow \big(\gamma_{\tau,j} > 0, \text{ and } L_{\tau,j} \text{ satisfies the maximum principle.}\big)$$

2.5. Mapping properties of L_τ. We shall call the image by X_τ of $[s_0, \infty) \times S^1$, for any $s_0 \in \mathbb{R}$, a half-Delaunay surface, and sometimes denote it by $D_\tau^+(s_0)$. Our goal in the remainder of this section is to study CMC perturbations of half-Delaunay surfaces with prescribed boundary values. To do this we require rather precise knowledge of the mapping properties of the Jacobi operators, and their inverses, on these surfaces. In fact, we need to understand these mapping properties uniformly as $\tau \to 0$. In this subsection we give careful statements of the results we need. Results of this type (for fixed τ) were originally proved in [**11**] for L_τ acting on weighted Sobolev spaces; these were reformulated for the more convenient family of weighted Hölder spaces in [**8**], and the issue of uniformity in the singular limit was also addressed there.

Let us begin by defining these weighted Hölder spaces on $D_\tau^+(s_0)$.

DEFINITION 4. *Let $r \in \mathbb{N}$, $\alpha \in (0,1)$, and $\mu \in \mathbb{R}$. Then the function space $\mathcal{E}_\mu^{r,\alpha}([s_0, \infty) \times S^1)$ consists of those functions $w \in \mathcal{C}_{loc}^{r,\alpha}([s_0, \infty) \times S^1)$ such that*

$$\|w\|_{\mathcal{E}_\mu^{r,\alpha}} := \sup_{s \geq s_0} e^{-\mu s} \|w\|_{\mathcal{C}^{r,\alpha}([s,s+1] \times S^1)} < \infty.$$

Here $\|\cdot\|_{\mathcal{C}^{r,\alpha}([s,s+1] \times S^1)}$ is the usual Hölder norm on $[s, s+1] \times S^1$. Also,

$$[\mathcal{E}_\mu^{r,\alpha}([s_0, \infty) \times S^1)]_0$$

is the subspace of functions vanishing at $s = s_0$.

Observe that the function $s \mapsto e^{\mu s}$ is in $\mathcal{E}^{r,\alpha}_{\mu}(\mathbb{R}^+ \times S^1)$.

It is clear that

$$L_\tau : [\mathcal{E}^{2,\alpha}_{\mu}([s_0, +\infty) \times S^1)]_0 \longrightarrow \mathcal{E}^{0,\alpha}_{\mu}([s_0, +\infty) \times S^1) \tag{2–7}$$

is bounded for all $\mu \in \mathbb{R}$. However, it is not Fredholm for every weight. In fact, the existence of a solution of $L_\tau w = 0$ which grows or decays exactly like $e^{\pm\gamma_{\tau,j}s}$ can be used to show that this mapping does not have closed range when $\mu = \pm\gamma_{\tau,j}$ for any j. However, it is proved in [**8**] (see also [**11**]) that (2–7) is Fredholm when $\mu \notin \Gamma_\tau$.

We first consider what happens for τ fixed. For simplicity, denote by $L_\tau(\mu)$ the operator in (2–7); hence this notation indicates the weighted space on which we are letting L_τ act. A basic observation is that $L_\tau(\mu)$ and $L_\tau(-\mu)$ are essentially dual to one another. (Of course, since we are working on Hölder spaces, this is not quite true. The analogous statement for weighted L^2 spaces is true, however, and this may be used to deduce the following statements.) An important consequence of this is that when $\mu \notin \Gamma_\tau$, then $L_\tau(\mu)$ is surjective if and only if $L_\tau(-\mu)$ is injective; furthermore, if this is the case, then the dimension of the kernel of $L_\tau(\mu)$ is equal to the dimension of the cokernel of $L_\tau(-\mu)$. Summarizing all of this, the precise result is

PROPOSITION 2. *Assume that $\mu \in (\gamma_{\tau,j}, \gamma_{\tau,j+1})$ for some $j \in \mathbb{N}$, then $L_\tau(\mu)$ is surjective and has a kernel of dimension $2j+1$. Thus when $\mu \in (-\gamma_{\tau,j+1}, -\gamma_{\tau,j})$ for some $j \in \mathbb{N}$, then $L_\tau(\mu)$ is injective, and has a cokernel of dimension $2j+1$.*

Notice that the first of these assertions follows from the discussion of Jacobi fields in the last subsection. In fact, we need all $\Psi^{+,j}_\tau e^{\pm ij\theta}$ and $\Psi^{-,j}_\tau e^{\pm ij\theta}$ to lie in $\mathcal{E}^{r,\alpha}_{\mu}([s_0, \infty))$ in order to take a linear combination of these which vanishes at s_0.

We shall henceforth assume that $\tau \in (\tau_*, 0) \cup (0, 1]$ and $\mu \in (-\gamma_{\tau,2}, -\gamma_{\tau,1}) = (-\gamma_{\tau,2}, 0)$. Although the previous Proposition states that $L_\tau(\mu)$ is not surjective in this case, there is a way to modify this mapping so as to be surjective by augmenting the domain. In fact, from the comments in the last section, it is clear that if we augment the domain of $L_\tau(\mu)$, for μ in this (negative) range, with a 'deficiency space'

$$\mathcal{W}_\tau := \text{Span}\,\{\Psi^{j,\pm}_\tau(s)e^{ij\theta} \quad : \quad j = 0, \pm 1\},$$

which is simply the span of the geometric Jacobi fields, then the nullspace of $L_\tau(-\mu)$ is contained in this extended domain, that is,

$$\ker L_\tau(-\mu) \subset [\mathcal{E}^{2,\alpha}_{\mu}([s_0, \infty) \times S^1) \oplus \mathcal{W}_\tau]_0.$$

The deficiency subspace $\mathcal{W}_\tau$ is 6-dimensional, while from Proposition 2 again, $\dim \ker L_\tau(-\mu) = 3$. Thus projecting this nullspace onto $\mathcal{W}_\tau$ defines a three-dimensional subspace $\mathcal{N}_\tau \subset \mathcal{W}_\tau$. Choosing any complement $\mathcal{K}_\tau$, so that

$$\mathcal{W}_\tau = \mathcal{N}_\tau \oplus \mathcal{K}_\tau,$$

then

$$L_\tau : [\mathcal{E}^{2,\alpha}_{\mu}([s_0, \infty) \times S^1) \oplus \mathcal{K}_\tau]_0 \longrightarrow \mathcal{E}^{0,\alpha}_{\mu}([s_0, \infty) \times S^1) \tag{2–8}$$

is injective. With slightly more work, as explained in [**8**], [**11**], it is possible to show that it is also surjective. Altogether, we have

PROPOSITION 3. *For $\tau \in (\tau_*, 0) \cup (0, 1]$, $\mu \in (-\gamma_{\tau,2}, -\gamma_{\tau,1})$ (which is the same as $(-\gamma_{\tau,2}, 0)$ since $\gamma_{\tau,1} = 0$), and any choice of complement $\mathcal{K}_\tau$ to $\mathcal{N}_\tau$ in $\mathcal{W}_\tau$, then the mapping (2–8) is an isomorphism.*

It is more convenient for us to rephrase this result using the fact that the space of boundary traces $\{w(s_0, \cdot) : w \in \mathcal{K}_\tau\}$ lies in the span of $e^{ij\theta}$, $j = 0, \pm 1$:

PROPOSITION 4. *Fix $\tau \in (\tau_*, 0) \cup (0, 1]$ and $\mu \in (-\gamma_{\tau,2}, -\gamma_{\tau,1}) = (-\gamma_{\tau,2}, 0)$. Then for any $s_0 \in \mathbb{R}$, there exists a bounded mapping*

$$G_{\tau,s_0} : \mathcal{E}^{0,\alpha}_\mu([s_0, \infty) \times S^1) \longrightarrow \mathcal{E}^{2,\alpha}_\mu([s_0, \infty) \times S^1)$$

such that for any $f \in \mathcal{E}^{0,\alpha}_\mu([s_0, \infty) \times S^1)$, the function $w = G_{\tau,s_0}(f)$ satisfies

$$\begin{cases} L_\tau w = f & \text{in } (s_0, \infty) \times S^1, \\ w \in \operatorname{Span}\{e^{-i\theta}, 1, e^{i\theta}\} & \text{on } \{s_0\} \times S^1. \end{cases}$$

Moreover the norm of G_{τ,s_0} is bounded independently of $s_0 \in \mathbb{R}$.

As explained earlier, we must also understand the behavior of this inverse as $\tau \to 0$. To explain the statement of this next result, recall that, for all $j \geq 2$, $\gamma_{\tau,j} \to j$ as $\tau \to 0$, so that any fixed $\mu \in (-2, -1)$ is eventually in $(-\gamma_{\tau,2}, -\gamma_{\tau,1}) = (-\gamma_{\tau,2}, 0)$ provided τ is chosen small enough.

PROPOSITION 5. *Fix any $\mu \in (-2, -1)$. Then, there exists a number $\tau_0 > 0$ such that the norm of G_{τ,s_0} is uniformly bounded, independently of $s_0 \in \mathbb{R}$ and $|\tau| \in (0, \tau_0)$.*

The precise range of μ between -2 and -1 is very important here; this result fails when, for example, $\mu \in (-1, 0)$. This is Proposition 21 in [**8**], and the proof can be found there.

To conclude, we need some facts about the Poisson operator. However, we only require these in the limit as $\tau \to 0$, and so it suffices to study the Poisson operator for the very simple operator

$$\Delta_0 := \partial_s^2 + \partial_\theta^2$$

on $[0, \infty) \times S^1$.

LEMMA 1. *For any $h \in \mathcal{C}^{2,\alpha}(S^1)$ such that*

$$\int_{S^1} h(\theta)\, d\theta = \int_{S^1} h(\theta) e^{\pm i\theta}\, d\theta = 0,$$

(i.e., the function h is orthogonal to 1 and $e^{\pm i\theta}$ in the L^2 sense on S^1) there exists a unique function $w = \mathcal{P}(h) \in \mathcal{E}^{2,\alpha}_{-2}([0, +\infty) \times S^1)$ such that $\Delta_0 w = 0$ on $(0, +\infty) \times S^1$, $w(0, \theta) = h(\theta)$, and moreover, $\|w\|_{\mathcal{E}^{2,\alpha}_{-2}} \leq c \|h\|_{\mathcal{C}^{2,\alpha}}$.

The proof is straightforward.

2.6. CMC surfaces close to a half-Delaunay surface. We shall now study the problem of finding all CMC surfaces near to a given half-Delaunay surface $D^+_\tau(s_0)$. These will be parametrized by their boundary values at $s = s_0$. As before, we require some knowledge of the behavior of this solution as $\tau \to 0$.

2.6.1. *The mean curvature operator on a Delaunay surface.* Using the isothermal parametrization (2–1), the unit normal vector field N_τ on D_τ is given by

$$N_\tau(s,\theta) := (-\tau \cosh\sigma(s) \cos\theta, -\tau \cosh\sigma(s) \sin\theta, \partial_s\sigma(s)), \tag{2–9}$$

when $\tau \in (0,1]$ and is given by

$$N_\tau(s,\theta) := (\tau \sinh\sigma(s) \cos\theta, \tau \sinh\sigma(s) \sin\theta, -\partial_s\sigma(s)), \tag{2–10}$$

when $\tau \in (-\infty, 0)$. Normal graphs over D_τ then admit parametrizations of the form

$$X_w : (s,\theta) \longmapsto X_\tau(s,\theta) + w(s,\theta)\, N_\tau(s,\theta),$$

where w is any function which is suitably small.

The mean curvature of the normal graph of a function w about D_τ is calculated by a fairly complicated nonlinear elliptic expression of w which we shall not write out in full. This surface has mean curvature equal to 1 if and only if w is a solution of the corresponding equation, which we write simply as $\mathcal{M}_\tau(w) = 0$. We need to know a bit about the structure of this operator, which is proved in [**8**]:

PROPOSITION 6. *The equation $\mathcal{M}_\tau(w) = 0$ can be written in the form*

$$L_\tau w = \tau\, e^\sigma\, Q_\tau\left(\frac{w}{\tau\, e^\sigma}\right) \tag{2–11}$$

where L_τ is the (modified) Jacobi operator (2–5), and where Q_τ is a nonlinear second order differential operator which satisfies

$$Q_\tau(0) = 0 \qquad D_w Q_\tau(0) = 0 \qquad \text{and} \qquad D_w^2 Q_\tau(0) = 0.$$

The Taylor expansion of Q_τ (in $w/\tau e^\sigma$) has coefficients that are uniformly bounded in s, along with all their derivatives, independently of $\tau \in (\tau_*, 0) \cup (0, 1]$.

We shall specialize now and suppose that $s_0 = -s_\tau$. Then, from the discussion in §2.1, we see that the surface $D_\tau^+(s_\tau)$ (which we now simply call D_τ^+) is nearly flat close to its boundary, and it will be more convenient computationally to use a slightly different parametrization of nearby surfaces, replacing the unit normal N_τ by a small perturbation of it, $\bar{N}_\tau$, which is constant (and in fact vertical, downward pointing) for all s close to $-s_\tau$.

To define $\bar{N}_\tau$, choose a smooth cutoff function χ_τ which is nonnegative and which satisfies $\chi_\tau = 1$ when $s \geq -s_\tau + 2$ and $\chi_\tau = 0$ when $s \leq -s_\tau + 1$. Now set

$$\bar{N}_\tau := \chi_\tau\, N_\tau - (1 - \chi_\tau)\,(0,0,1).$$

This satisfies

$$\left|\nabla^k \left(\bar{N}_\tau \cdot N_\tau - 1\right)\right| \leq c_k\, |\tau|, \qquad \text{for all} \quad k \geq 0 \quad \text{and} \quad s \in [-s_\tau, -s_\tau + 2]. \tag{2–12}$$

We henceforth use the parametrization

$$\bar{X}_w : (s,\theta) \longmapsto X_\tau(s,\theta) + w(s,\theta)\, \bar{N}_\tau(s,\theta), \qquad (s,\theta) \in [-s_\tau, \infty) \times S^1.$$

Denote by $D_\tau^+(w)$ the surface obtained this way.

It follows from (2–11) and from (2–12) that $D_\tau^+(w)$ has constant mean curvature equal to 1 if and only if w satisfies a nonlinear equation of the form

$$\tilde{\mathcal{M}}_\tau := L_\tau w - \tilde{Q}_\tau(w) = 0,$$

where

$$\tilde{Q}_\tau(w) := \tau\, \bar{L}_\tau w + \tau\, e^\sigma\, \bar{Q}_\tau\left(\frac{w}{\tau\, e^\sigma}\right).$$

Here $\bar{Q}_\tau$ satisfies the same properties as listed for Q_τ in Proposition 6, and in fact these operators agree when $s \geq -s_\tau + 2$. The linear operator $\tau \bar{L}_\tau$ represents the deviation between the linearizations corresponding to the parametrizations X_w using N_τ and $\bar{X}_w$ using $\bar{N}_\tau$. From (2–12), $\bar{L}_\tau$ has coefficients supported in $[-s_\tau, -s_\tau + 2] \times S^1$ which are bounded in any $\mathcal{C}^{k,\alpha}$, uniformly in τ. The details of this change of parametrization are contained in [**10**].

2.6.2. *The nonlinear Poisson problem.* We are now ready to solve the nonlinear boundary problem $\tilde{\mathcal{M}}_\tau(w) = 0$ on D_τ^+ with the value of w at $s = -s_\tau$ (almost) specified.

Fix $\mu \in (-2, -1)$ and $\delta > 0$. Let h be any element of $\mathcal{C}^{2,\alpha}(S^1)$ which is orthogonal to 1 and $e^{\pm i\theta}$, in the L^2 sense, and which satisfies

$$\|h\|_{\mathcal{C}^{2,\alpha}} \leq \delta\, |\tau|^3.$$

Now define the approximate solution

$$w_h := \mathcal{P}(h)(\,\cdot + s_\tau, \cdot),$$

where $\mathcal{P}$ is the Poisson operator for Δ_0 from Lemma 1. It suffices to use this Poisson operator, rather than the one for L_τ, because L_τ is very close to Δ_0 in a long interval around the boundary. Since we are using norms with exponential weight factors, these operators differ by a very small amount in norm. From the bounds in this lemma, the shift by s_τ in the s-variable, and the fact $s_\tau \sim -\frac{1}{4}\log \tau^2$, we have

$$\|w_h\|_{\mathcal{E}^{2,\alpha}_{-2}} \leq c\, |\tau|\, \|h\|_{\mathcal{C}^{2,\alpha}}. \tag{2–13}$$

We shall now search for a solution w, which we write as $w = w_h + v$, of $\tilde{\mathcal{M}}_\tau(w) = 0$. The function v will lie in $\mathcal{E}^{2,\alpha}_\mu([-s_\tau, \infty) \times S^1)$ and should satisfy

$$\begin{cases} L_\tau v = & \tilde{Q}_\tau(w_h + v) - L_\tau w_h \quad \text{in } (-s_\tau, \infty) \times S^1, \\ v(-s_\tau, \cdot) \in & \operatorname{Span}\{1, e^{\pm i\theta}\}. \end{cases}$$

The reason we are only requiring those eigencomponents of $v(s,\theta) = \sum v_j(s) e^{ij\theta}$ with $|j| \geq 2$ to vanish is that we shall be using the inverse $G_{\tau,-s_\tau}$ from Proposition 4, which does not allow these low eigencomponents to be specified.

We solve this equation by finding a fixed point of the mapping

$$\mathcal{N}_\tau(v) := G_{\tau,-s_\tau}\left(\tilde{Q}_\tau(w_h + v) - L_\tau w_h\right).$$

From (2–13), together with the fact that $L_\tau w_h = (L_\tau - \delta_0)(w_h)$, we have

$$\|L_\tau w_h\|_{\mathcal{E}^{0,\alpha}_\mu} \leq c\, |\tau|\, \|h\|_{\mathcal{C}^{2,\alpha}}, \qquad \|\tau\, \bar{L}_\tau w_h\|_{\mathcal{E}^{0,\alpha}_\mu} \leq c\, |\tau|^{1-\mu/2}\, \|h\|_{\mathcal{C}^{2,\alpha}},$$

and

$$\|\tau\, e^\sigma\, \tilde{Q}_\tau\left(\frac{w_h}{\tau\, e^\sigma}\right)\|_{\mathcal{E}^{0,\alpha}_\mu} \leq c\, |\tau|^{-(3+\mu)/2}\, \|h\|^2_{\mathcal{C}^{2,\alpha}}.$$

Only the final estimate uses that the norm of h is small, and in fact, it is only really necessary to assume that $\|h\|_{\mathcal{C}^{2,\alpha}} \leq c_0\, |\tau|^{3/2}$ for $c_0 > 0$ sufficiently small.

At this point it is straightforward, using Proposition 5, to show that there exists $c_* > 0$ and $\tau_0 > 0$, such that when $|\tau| \in (0, \tau_0)$, the nonlinear mapping $\mathcal{N}_\tau$ is a contraction in the ball

$$\mathcal{B} := \left\{v : \|v\|_{\mathcal{E}^{2,\alpha}_\mu} \leq c_*\, |\tau|\, \|h\|_{\mathcal{C}^{2,\alpha}}\right\},$$

and hence $\mathcal{N}_\tau$ has a unique fixed point in this ball. Observe that $\tau_0 > 0$ depends on δ while $c_* > 0$ does not depend on δ.

In summary, we have proved

THEOREM 6. *Fix $\mu \in (-2,-1)$ and $\delta > 0$. There are numbers $\tau_0 > 0$ and $c > 0$ such that for each τ with $0 < |\tau| < \tau_0$ and for every $h \in \mathcal{C}^{2,\alpha}(S^1)$ which is orthogonal to 1 and $e^{\pm i\theta}$ and which satisfies $\|h\|_{2,\alpha} \leq \delta |\tau|^3$, there exists an embedded CMC surface $D_\tau(h)$, parameterized by*

$$\bar{X}_w := X_\tau + w\,\bar{N}_\tau \qquad \text{in} \qquad [-s_\tau, \infty) \times S^1.$$

The function $w = v_h + v$ here lies in $\mathcal{E}^{2,\alpha}_\mu([-s_\tau,\infty)\times S^1)$ and satisfies

$$\|v_h\|_{\mathcal{E}^{2,\alpha}_{-2}} + \|v\|_{\mathcal{E}^{2,\alpha}_\mu} \leq c\,|\tau|\,\|h\|_{\mathcal{C}^{2,\alpha}}$$

and finally, $v(-s_\tau,\cdot) \in \operatorname{Span}\{1, e^{\pm i\theta}\}$.

One of the main reasons for modifying the unit normal vector field N_τ to $\bar{N}_\tau$ is because with this definition, the region near the boundary of $D_\tau(h)$ is a vertical graph over an annulus with outer boundary ∂B_{r_τ} in $\mathbb{R}^2$. Here $r_\tau = \frac{\tau}{2} e^{\sigma(-s_\tau)} \sim |\tau|^{3/2}$ is defined in §2.2.

We conclude this subsection with some estimates on the graph function in this representation. To do this, we first define some function spaces:

DEFINITION 5. *For $r \in \mathbb{N}$, $\alpha \in (0,1)$ and $\mu \in \mathbb{R}$, we let $\mathcal{C}^{r,\alpha}_\mu(\mathbb{R} - \{0\})$ be the space of functions $w \in \mathcal{C}^{r,\alpha}_{loc}(\mathbb{R} - \{0\})$ such that*

$$\|w\|_{\mathcal{C}^{r,\alpha}_\mu} := \sup_{\rho > 0} \rho^{-\mu}\, \|w(\rho\,\cdot)\|_{\mathcal{C}^{r,\alpha}(\overline{B_2} - B_1)} < \infty.$$

If Ω is a closed subset of $\mathbb{R}^2 - \{0\}$, we define the space $\mathcal{C}^{r,\alpha}_\mu(\Omega)$ as the space of restriction of functions of $\mathcal{C}^{r,\alpha}_\mu(\mathbb{R}^2 - \{0\})$ to Ω. This space is naturally endowed with the induced norm.

Notice that the space $\mathcal{C}^{r,\alpha}_\mu(\overline{B_R} - \{0\})$ corresponds precisely to the space

$$\mathcal{E}^{r,\alpha}_\mu([s_0,\infty)\times S^1),$$

where $s_0 = -\log R$, under the elementary change of variables $s = -\log r$. However, we shall not use this obvious change of variables but rather the more complicated change of variables specific to our problem

$$r := \frac{\tau}{2}\, e^{\sigma(s)},$$

for $s \in [-s_\tau, 0)$ and $\theta \in S^1$. Using the estimates of §2.2, we obtain

$$r\,\partial_r = -(1 + \mathcal{O}(\cosh^{-2} s))\,\partial_s, \tag{2–14}$$

for all $s \in [-s_\tau, 0)$.

We finally translate the surface $D_\tau(h)$ by $\pm\frac{\tau^2}{4}\log\left(\frac{8}{\tau^2}\right)$ along the vertical axis. This surface will still be denoted by $D_\tau(h)$. Using the expansion (2–4) together with (2–14), we see that near ∂B_{r_τ}, the surface $D_\tau(h)$ is the graph of the function

$$\overline{B_{r_\tau}} - B_{r_\tau/2} \ni x \longmapsto \mp\frac{\tau^2}{4}\log r - W_h(x) + V_{\tau,h}(x), \tag{2–15}$$

according to the sign of τ (with $-$ when $\tau > 0$ and $+$ with $\tau < 0$). Here W_h denotes the unique harmonic extension of h in the ball B_{r_τ} and $V_{\tau,h}$ is bounded in

$\mathcal{C}^{2,\alpha}_0(\overline{B_{r_\tau}} - B_{r_\tau/2})$, by a constant that does not depend on δ nor on τ, times $|\tau|^3$. This factor of $|\tau|^3$ has its origin in (2–4).

Observe that, reducing τ_0 if this is necessary, we can assume that the mapping $h \to V_{\tau,h}$ is continuous and in fact smooth. With little work we also find that

$$\|V_{\tau,\tilde{h}} - V_{\tau,h}\|_{\mathcal{C}^{2,\alpha}_0} \leq c\,|\tau|^{1+\mu/2}\,\|\tilde{h} - h\|_{\mathcal{C}^{2,\alpha}} \tag{2–16}$$

for some constant $c > 0$ which does not depend on δ, nor on τ.

3. The Geometry and Analysis of k-Ended CMC Surfaces

3.1. Moduli space theory for k-ended CMC surfaces. We now briefly sketch the moduli space theory for k-unduloids as developed in [**7**]. Because it is no harder to do so, we extend this and consider the deformation theory for complete, finite topology CMC surfaces with k ends, each one of which is asymptotic to a Delaunay unduloid or nodoid with Delaunay parameter $\tau > \tau_*$, where τ_* is defined in (2–6). The statements and proofs are identical in this slightly broader context.

DEFINITION 6. *The moduli space $\mathcal{M}^{\tau_*}_{g,k}$ consists of the set of all complete constant mean curvature surfaces of finite topology, with genus g and k ends $E_1, \ldots, E_k$ such that each end E_j is asymptotic to a half Delaunay surface $D^+_{\tau_j}$ with $\tau_j \in (\tau_*, 0) \cup (0, 1]$.*

Now decompose each $\Sigma \in \mathcal{M}^{\tau_*}_{g,k}$ into a union of a compact component K and ends E_ℓ, $\ell = 1, \ldots, k$ which are slightly overlapping (for example, require that each $K \cap E_\ell$ is an annulus). For each ℓ choose standard isothermal coordinates (s, θ) for the model Delaunay end D_{τ_ℓ}, so that E_ℓ is parametrized by

$$Y_\ell := X_\ell + w_\ell\, N_\ell : [0, \infty) \times S^1 \longrightarrow E_\ell.$$

Here X_ℓ is the standard parametrization (2–1) for D_{τ_ℓ}; each function w_ℓ decays exponentially and in fact

$$w_\ell \in \mathcal{E}^{2,\alpha}_{-\gamma_{\tau_\ell,2}}([0, \infty) \times S^1). \tag{3–1}$$

DEFINITION 7. *For $r \in \mathbb{N}$, $\alpha \in (0, 1)$ and $\mu \in \mathbb{R}$, let $\mathcal{D}^{r,\alpha}_\mu(\Sigma)$ be the space of functions $v \in \mathcal{C}^{r,\alpha}(\Sigma)$ for which*

$$\|v\|_{\mathcal{D}^{r,\alpha}_\mu} := \|\, v|_K \,\|_{\mathcal{C}^{r,\alpha}} + \sum_{\ell=1}^{k} \|\, v \circ Y_\ell|_{E_\ell} \,\|_{\mathcal{E}^{r,\alpha}_\mu} < \infty.$$

Let $\mathcal{L}_\Sigma = \Delta_\Sigma + |A_\Sigma|^2$ denote the Jacobi operator Σ. Because of the asymptotic structure of the ends of Σ, the various mapping and regularity properties of this operator may be deduced from the analogous properties for $\mathcal{L}_{\tau_\ell}$. In particular, the set of indicial roots for $\mathcal{L}_\Sigma$, given by

$$\Gamma_\Sigma := \{\pm\,\gamma_{\tau_\ell,j} : j \in \mathbb{N}, \quad \ell = 1, \ldots, k\},$$

determines the weighted spaces on which L_Σ is Fredholm as well as the asymptotic behavior of solutions of the homogeneous equation $\mathcal{L}_\Sigma w = 0$. In particular, it follows from the analysis of §2.5, that when $\mu \notin \Gamma_\Sigma$, then

$$\mathcal{L}_\Sigma : \mathcal{D}^{2,\alpha}_\mu(\Sigma) \longrightarrow \mathcal{D}^{0,\alpha}_\mu(\Sigma)$$

is Fredholm. To keep track of the value of the weight parameter, we write this mapping as $\mathcal{L}_\Sigma(\mu)$. As before, the operators $\mathcal{L}_\Sigma(\mu)$ and $\mathcal{L}_\Sigma(-\mu)$ are essentially dual

to one another which implies that when $\mu \notin \Gamma_\Sigma$, the operator $\mathcal{L}_\Sigma(\mu)$ is surjective if and only if the operator $\mathcal{L}_\Sigma(-\mu)$ is injective, and moreover

$$\dim \ \ker(\mathcal{L}_\Sigma(\mu)) = \dim \ \operatorname{coker}(\mathcal{L}_\Sigma(-\mu)).$$

We now give the precise definition of nondegeneracy.

DEFINITION 8. *The surface $\Sigma \in \mathcal{M}^{\tau_*}_{g,k}$ is nondegenerate if $\mathcal{L}_\Sigma(\mu)$ is surjective for all $\mu > 0$ with $\mu \notin \Gamma_\Sigma$, or equivalently, if $\mathcal{L}_\Sigma(-\mu)$ is injective for all $\mu > 0$.*

Thus, a surface is nondegenerate if it has no decaying Jacobi fields. Note in particular that from the definition of τ_*, the Delaunay surfaces D_τ are nondegenerate when $\tau > \tau_*$.

The basic result from [**7**] is this:

THEOREM 7. *Fix any element $\Sigma \in \mathcal{M}^{\tau_*}_{g,k}$. If Σ is nondegenerate, then some neighborhood of Σ in the moduli space $\mathcal{M}^{\tau_*}_{g,k}$ is a real analytic manifold of dimension $3k$. In general, the moduli space is a locally real analytic variety, i.e., there is a neighborhood of $\Sigma \in \mathcal{M}^{\tau_*}_{g,k}$ which is identified, via a real analytic diffeomorphism, with the zero set of a real analytic function defined in a finite-dimensional Euclidean space.*

We sketch the main ideas in the proof, but only in the nondegenerate case. The general case is somewhat more intricate, but is based on precisely the same ideas.

The first point is that there exist analogues of the six linearly independent geometric Jacobi fields $\Phi^{j,\pm}_{\tau_\ell}$, $j = 0, \pm 1$, on each end E_ℓ, $\ell = 1, \ldots, k$ of Σ. These are denoted $\Phi^{j,\pm}_\ell$.

LEMMA 2. *Let $\mu \in (-\gamma_{\tau_\ell,2}, 0)$. Then these geometric Jacobi fields $\Phi^{j,\pm}_\ell$ can be chosen to satisfy*

$$\Phi^{j,\pm}_\ell \circ Y_\ell - \Phi^{j,\pm}_{\tau_\ell} \in \mathcal{E}^{2,\alpha}_\mu(E_\ell).$$

PROOF. Except for $\Phi^{0,-}_\ell$, these asymptotics may be deduced from the same constructions as in §2.4.1. The existence and asymptotics of the remaining Jacobi fields $\Phi^{0,-}_\ell$ may be deduced by a perturbation argument using Proposition 4 and (3–1). □

The proof of Theorem 7 is based on the implicit function theorem, but to apply this theorem we need to find function spaces on which the nonlinear mean curvature operator acts and on which $\mathcal{L}_\Sigma$ is surjective. Unfortunately, $\mathcal{L}_\Sigma(\mu)$ is never surjective when $\mu < 0$. This may appear discouraging since the nonlinear operator as given only acts on spaces consisting of functions which decay along the ends. To remedy this, we define the $6k$-dimensional deficiency space

$$\mathcal{W}_\Sigma := \oplus_{\ell=1}^k \operatorname{Span}\{\, \chi_\ell \, \Phi^{j,\pm}_\ell : j = -1, 0, 1\},$$

where χ_ℓ is a cutoff function equal to 0 near $E_\ell \cap K$ and equal to 1 on $Y_\ell([c, \infty) \times S^1)$ for some $c > 0$.

PROPOSITION 7. *Assume that Σ is nondegenerate and fix $\mu \in (-\inf_\ell \gamma_{\tau_\ell,2}, 0)$. Then the mapping*

$$\mathcal{L}_\Sigma : \mathcal{D}^{2,\alpha}_\mu(\Sigma) \oplus \mathcal{W}_\Sigma \longrightarrow \mathcal{D}^{0,\alpha}_\mu(\Sigma) \tag{3–2}$$

is surjective and its nullspace $\mathcal{N}_\Sigma$ is $3k$-dimensional. Hence there exists a $3k$-dimensional subspace $\mathcal{K}_\Sigma \subset \mathcal{W}_\Sigma$ such that

$$\mathcal{L}_\Sigma : \mathcal{D}^{2,\alpha}_\mu(\Sigma) \oplus \mathcal{K}_\Sigma \longrightarrow \mathcal{D}^{0,\alpha}_\mu(\Sigma) \tag{3–3}$$

is an isomorphism.

The proof can be found in [**7**].

It is possible to make sense of the mean curvature operator on elements of the domain space in (3–2) using the fact that elements of $\mathcal{W}_\Sigma$ correspond to geometric motions as in §2.4.1. Indeed, decomposing $u \in \mathcal{D}^{2,\alpha}_\mu(\Sigma) \oplus \mathcal{W}_\Sigma$ as

$$u = u' + u'',$$

with $u' \in \mathcal{D}^{2,\alpha}_\mu(\Sigma)$ and $u'' \in \mathcal{W}_\Sigma$, we then let Σ_u denote the surface which is the normal graph by the function u' over the surface obtained by slightly deforming the ends of Σ in the manner prescribed by the components of $u'' \in \mathcal{W}_\Sigma$. (More precisely, one defines a $6k$-dimensional real analytic parameter space $\mathcal{P}$ of geometric deformations of Σ such that the differentials of curves in $\mathcal{P}$ through Σ correspond to the geometric Jacobi fields $\Phi^{j,\pm}_\ell$.) The mean curvature of Σ_u is identified with some function $H(u)$ defined on Σ, and by Proposition 7, when Σ is nondegenerate, the differential of this map with respect to u at $u = 0$ is surjective.

3.2. CMC surfaces close to a k-ended CMC surface.

3.2.1. *Geometric preparations.* Let $\Sigma \in \mathcal{M}^{\tau_*}_{g,k}$ be nondegenerate and fix a point $p_0 \in \Sigma$. Assume by a rigid motion that $p_0 = 0$ and the oriented normal vector to Σ at that point is $(0, 0, -1)$, so that the tangent plane $T_0\Sigma$ is the xy-plane. Then in some neighborhood of 0, Σ can be represented as a vertical graph $z = u_0(x, y)$ for some function u_0 defined on a ball B_ρ. By the assumptions above, $u_0(0) = \nabla u_0(0) = 0$.

Now for any sufficiently small vector $a = (a_{-1}, a_0, a_1) \in \mathbb{R}^3$, we rotate and translate Σ slightly so that this new surface Σ_a is the graph of a function u_a, also defined on B_ρ, such that

$$u_a(0) = a_0, \qquad \partial_x u_a(0) = a_1, \qquad \text{and} \qquad \partial_y u_a(0) = a_{-1}.$$

Also, for $0 < r \leq \rho$, let $\Sigma_{a,r}$ denote the complement in Σ_a of the graph of u_a on B_r. Finally, set

$$p_a := (0, 0, u_a(0)).$$

Next, let us modify the unit normal N_a on Σ_a to a new unit vector field $\bar{N}_a$ on Σ_a such that in $\Sigma_a - \Sigma_{a,\rho/2}$, $\bar{N}_a \equiv (0, 0, -1)$, while on $\Sigma_{a,\rho}$, $\bar{N}_a = N_{\Sigma_a}$. The linearization of the mean curvature operator with respect to this new vector field, $\bar{\mathcal{L}}_a$, is a slight perturbation of $\mathcal{L}_{\Sigma_a}$. Any mapping property for $\mathcal{L}_\Sigma$ immediately transforms to one for $\bar{\mathcal{L}}_a$, hence in particular Proposition 7 holds when $\mathcal{L}_\Sigma$ is replaced by $\bar{\mathcal{L}}_a$, provided a is chosen small enough.

3.2.2. *The mean curvature operator for graphs.* The mean curvature operator for the vertical graph of a function $u : \mathbb{R}^2 \supset B_\rho \to \mathbb{R}$ (with downward pointing normal) is given by

$$H(u) := -\frac{1}{2}\mathrm{div}\left(\frac{\nabla u}{(1 + |\nabla u|^2)^{1/2}}\right).$$

Since $\bar{\mathcal{L}}_a w = D_u H|_{u=u_a}(w)$ in $\Sigma_a - \Sigma_{a,\rho/2}$, we obtain that

$$\bar{\mathcal{L}}_a w = \tfrac{1}{2}\mathrm{div}\,\frac{\nabla w}{(1+|\nabla u_a|^2)^{1/2}} - \tfrac{1}{2}\,\mathrm{div}\left(\frac{\nabla w \cdot \nabla u_a}{(1+|\nabla u_a|^2)^{3/2}}\,\nabla u_a\right)$$

there. The main point is that $\bar{\mathcal{L}}_a$ is close to the standard Laplace operator. More precisely, we have:

LEMMA 3. *Fix any $\nu \in \mathbb{R}$. There exist $a_*, c > 0$ such that when $|a_{\pm 1}| < a_*$, then for $0 < \rho_1 < \rho_2$ and any $w \in \mathcal{C}^{2,\alpha}_{\nu}(\overline{B_{\rho_2}} - B_{\rho_1})$, we have:*

$$\|\bar{\mathcal{L}}_a w - \tfrac{1}{2}\,\Delta w\|_{\mathcal{C}^{0,\alpha}_{\nu-2}(\overline{B_{\rho_2}} - B_{\rho_1})} \leq c\,\left((|a_1|^2 + |a_{-1}|^2) + \rho_2^2\right)\,\|\nabla w\|_{\mathcal{C}^{1,\alpha}_{\nu-1}(\overline{B_{\rho_2}} - B_{\rho_1})}.$$

3.2.3. *Mapping properties of $\bar{\mathcal{L}}_a$.* The proof of the following result can be found in [**2**]:

LEMMA 4. *Fix $-1 < \nu < 0$ and $0 < \rho_1 < \rho_2/2$. Then there exists an operator*

$$G_{\rho_1,\rho_2} : \mathcal{C}^{0,\alpha}_{\nu-2}(\overline{B_{\rho_2}} - B_{\rho_1}) \longrightarrow \mathcal{C}^{2,\alpha}_{\nu}(\overline{B_{\rho_2}} - B_{\rho_1})$$

with the following properties. For any $f \in \mathcal{C}^{0,\alpha}_{\nu-2}(\overline{B_{\rho_2}} - B_{\rho_1})$, the function $w = G_{\rho_1,\rho_2}(f)$ is a solution of the problem

$$\Delta w = f \qquad \text{in} \quad B_{\rho_2} - \overline{B_{\rho_1}}$$

with boundary data $w = 0$ on ∂B_{ρ_2} and w equal to a constant on ∂B_{ρ_1}. In addition,

$$\|G_{\rho_1,\rho_2}(f)\|_{\mathcal{C}^{2,\alpha}_{\nu}} \leq c\,\|f\|_{\mathcal{C}^{0,\alpha}_{\nu-2}}$$

for some constant $c > 0$ independent of ρ_1 and ρ_2.

We now define the function spaces suitable for the analysis of these operators on the punctured surfaces $\Sigma_a - \{p_a\}$. We identify all functions defined near p_a on Σ_a with functions on $B_\rho \subset \mathbb{R}^2$ using the fixed coordinates induced by the functions u_a.

DEFINITION 9. *Let $r \in \mathbb{N}$, $0 < \alpha < 1$ and $\nu, \mu \in \mathbb{R}$. Then $\mathcal{D}^{r,\alpha}_{\mu,\nu}(\Sigma_a - \{p_a\})$ is the space of functions $w \in \mathcal{C}^{r,\alpha}_{\mathrm{loc}}(\Sigma_a - \{p_a\})$ for which*

$$\|w\|_{\mathcal{D}^{r,\alpha}_{\mu,\nu}} := \|w\|_{\mathcal{D}^{r,\alpha}_{\mu}(\Sigma_{a,\rho/2})} + \|w \circ u_a^{-1}\|_{\mathcal{C}^{r,\alpha}_{\nu}(B_\rho - \{0\})} < \infty.$$

Thus μ is the weight on the other ends of Σ_a and ν regulates growth or decay near p_a. As usual, if Ω is any closed subset of $\Sigma_a - \{p_a\}$, then $\mathcal{D}^{r,\alpha}_{\mu,\nu}(\Omega)$ is the space of restriction of functions in $\mathcal{D}^{r,\alpha}_{\mu,\nu}(\Sigma_a - \{p_a\})$ to Ω, endowed with the induced norm.

PROPOSITION 8. *Fix $\mu \in (-\inf_\ell \gamma_{\tau_\ell,2}, 0)$, $-1 < \nu < 0$ and $0 < \alpha < 1$. Recall the number r_τ defined in §2.2. Then, there exist $\tau_0 > 0$ and $\alpha_* > 0$, such that for all $\tau \in (0, \tau_0)$, and for all $|a_{\pm 1}| < a_*$, there exists an operator*

$$\bar{\mathcal{G}}_\tau : \mathcal{D}^{0,\alpha}_{\mu,\nu-2}(\Sigma_{a,r_\tau}) \longrightarrow \mathcal{D}^{2,\alpha}_{\mu,\nu}(\Sigma_{a,r_\tau}) \oplus \mathcal{K}_\Sigma,$$

such that for all $f \in \mathcal{D}^{0,\alpha}_{\mu,\nu-2}(\Sigma_{a,r_\tau})$, the function $w = \bar{\mathcal{G}}_\tau(f)$ is a solution of

$$\bar{\mathcal{L}}_a w = f$$

in Σ_{a,r_τ}, *with* w *constant on* $\partial\Sigma_{a,r_\tau}$. *Furthermore, there exists* $c > 0$ *which is independent of* τ *and* f, *such that*

$$\|w\|_{\mathcal{D}^{2,\alpha}_{\mu,\nu}\oplus\mathcal{K}} \leq c\,\|f\|_{\mathcal{D}^{0,\alpha}_{\mu,\nu-2}}.$$

PROOF. First choose $\rho_* \in (0,\rho]$ and define

$$w := 2\,\chi\,G_{\rho_*,r_\tau}(f),$$

where G_{ρ_*,r_τ} is the operator in Lemma 4 and where χ is a radial cutoff function identically equal to 1 in $B_{\rho_*/2}$ and equal to 0 outside B_{ρ_*}. By construction the support of the function g defined by $g := \frac{1}{2}\,\Delta w - f$ is disjoint from B_{ρ_*}. Now, using Proposition 7, we define

$$w' := \chi'\,\bar{\mathcal{L}}_a^{-1}(g),$$

where χ' is another radial cutoff function equal to 1 outside B_{2r_τ} and vanishing in B_{r_τ}.

It is easy to check that

$$\|f - \bar{\mathcal{L}}_a\,w'\|_{\mathcal{D}^{0,\alpha}_{\mu,\nu-2}} \leq c\,(\rho_*^2 + |a_{\pm1}|^2 + r_\tau^{-\nu})\,\|f\|_{\mathcal{D}^{0,\alpha}_{\mu,\nu-2}},$$

and also that

$$\|w'\|_{\mathcal{D}^{2,\alpha}_{\mu,\nu}} \leq c\,\|f\|_{\mathcal{C}^{0,\alpha}_{\mu,\nu-2}},$$

for some constant $c > 0$ which depends neither on τ nor on ρ_*. The result follows immediately by a simple perturbation argument, provided ρ_* and τ are chosen small enough. □

We conclude this subsection with a simple result regarding the Poisson operator for Δ which is close to Lemma 1 in spirit.

LEMMA 5. *For any* $g \in \mathcal{C}^{2,\alpha}(S^1)$ *such that* $\int_{S^1} g = 0$, *there exists a unique function* $w \in \mathcal{C}^{2,\alpha}_{-1}(\mathbb{R}^2 - B_1)$ *which is a solution of*

$$\begin{cases} \Delta w = 0 & \text{in } \mathbb{R}^2 - B_1, \\ w = g & \text{on } \partial B_1, \end{cases} \tag{3–4}$$

and which satisfies $\|w\|_{\mathcal{C}^{2,\alpha}_{-1}} \leq c\,\|g\|_{\mathcal{C}^{2,\alpha}}$ *for some constant* $c > 0$ *which does not depend on* g.

We write this solution of (3–4) as $P(g)$.

3.2.4. *The nonlinear Poisson problem.* Using Proposition 7, let γ be the solution of

$$\bar{\mathcal{L}}_0\gamma = -2\,\pi\,\delta_p, \qquad \text{in} \qquad \Sigma_0 - \{p\}, \tag{3–5}$$

with $\gamma + \chi\,\log r \in \mathcal{D}^{2,\alpha}_{\mu}(\Sigma_0) \oplus \mathcal{K}_{\Sigma_0}$, where χ is, as usual, a cutoff function equal to 1 in $\Sigma_0 - \Sigma_{0,\rho/2}$ and vanishing in $\Sigma_{0,\rho}$ and $r = |(x,y)|$.

We now observe that there are three global Jacobi fields Φ_j, $j = 0, \pm1$, on Σ_0 such that

$$\Phi_0 = 1 + \mathcal{O}(r), \qquad \Phi_1 = r\,\cos\theta + \mathcal{O}(r^2), \qquad \text{and} \qquad \Phi_{-1} = r\,\sin\theta + \mathcal{O}(r^2)$$

in B_ρ. Indeed, Φ_0 is obtained by projecting the constant Killing field $(0,0,1)$ (corresponding to a vertical translation) onto the normal vector field, while $\Phi_{\pm1}$

are obtained by projecting the Killing fields $(z, 0, -x)$ and $(0, z, -y)$ (corresponding to the two rotations about the horizontal axes). By adding an appropriate linear combination of these three Jacobi fields to γ, we may also assume that $\gamma + \chi \log r$ and its gradient vanishes at 0.

LEMMA 6. *If γ is defined as above, then for all $k \geq 0$, there exists a constant $c_k > 0$ such that*

$$\left|\nabla^k (\gamma + \log r)\right| \leq c_k r^{2-k} |\log r| \tag{3–6}$$

in B_ρ, where $r := |x|$.

The proof is straightforward.

Now fix $\mu \in (-\inf_\ell \gamma_{\tau_\ell,2}, 0)$, $\nu \in (-1/3, 0)$ and $\delta > 0$. Suppose that $a_j \in \mathbb{R}$, $j = 0, \pm 1$, satisfy

$$|a_0| + |\tau|^{3/2} |a_{\pm 1}| \leq \delta |\tau|^3.$$

Finally, for any $g \in \mathcal{C}^{2,\alpha}(S_1)$ such that

$$\int_{S^1} g = 0 \qquad \text{and} \qquad \|g\|_{\mathcal{C}^{2,\alpha}} \leq \delta |\tau|^3,$$

we define

$$w_g := P(g)(r_\tau^{-1} u_a^{-1}(\cdot)).$$

This is the bounded harmonic extension of g to $\mathbb{R}^2 - B_{r_\tau}$. By Lemma 5 there exists a constant $c > 0$ such that

$$\|w_g\|_{\mathcal{C}^{2,\alpha}_{-1}} \leq c\, r_\tau \|g\|_{\mathcal{C}^{2,\alpha}}. \tag{3–7}$$

Now define

$$\tilde{w}_{\tau,g} := \chi\, w_g \mp \frac{\tau^2}{4} \gamma,$$

(with $-$ when $\tau > 0$ and $+$ when $\tau < 0$) where χ is the cutoff function which is defined just after the definition of γ.

We wish to find $v \in \mathcal{D}^{2,\alpha}_{\mu,\nu}(\Sigma_{a,r_\tau}) \oplus \mathcal{K}_{\Sigma_a}$ so that the graph over Σ_a of the function $w := \tilde{w}_{\tau,g} + v$, using the vector field $\bar{N}_a$ is CMC. (Note that, just as in §3.1, we do not actually consider the graph of w over Σ_{a,r_τ}, but rather decompose $w = w' + w''$ and consider the graph of w' over some deformation $\Sigma_{a,r_\tau,w''}$ of Σ_{a,r_τ}.) This is equivalent to finding the solution of some nonlinear elliptic operator which we write formally simply as

$$\bar{\mathcal{L}}_a v = Q(\tilde{w}_{g,\tau} + v) - \bar{\mathcal{L}}_a \tilde{w}_{g,\tau}.$$

As usual, we use Proposition 8 to rephrase this as a fixed point problem for the operator

$$\mathcal{M}_\tau(v) := \bar{\mathcal{G}}_\tau \left(Q(\tilde{w}_{g,\tau} + v) - \bar{\mathcal{L}}_a \tilde{w}_{g,\tau}\right).$$

From the estimates (3–6) and (3–7), we can show with little work that

$$\|\bar{\mathcal{L}}_a \tilde{w}_{g,\tau}\|_{\mathcal{D}^{0,\alpha}_{\mu,\nu-2}} \leq c_\delta |\tau|^{7/2}, \qquad \text{and} \qquad \|Q(\tilde{w}_{g,\tau})\|_{\mathcal{D}^{0,\alpha}_{\mu,\nu-2}} \leq c\, \tau^{3-3\nu/2},$$

where $c > 0$ does not depend on δ, provided τ is small enough. It is only to ensure that the first estimate is much smaller than the final estimate that we need to restrict ν to lie in $(-\frac{1}{3}, 0)$.

It is then routine to show that for any fixed $\delta > 0$, there exist $c_* > 0$ and $\tau_0 > 0$ such that when $\tau \in (0, \tau_0)$, the nonlinear mapping $\mathcal{M}_\tau$ is a contraction mapping in the ball

$$\widehat{\mathcal{B}} := \{v : ||v||_{\mathcal{D}^{2,\alpha}_{\mu,\nu} \oplus \mathcal{K}} \leq c_* \, \tau^{3-3\nu/2}\},$$

and thus has a unique fixed point in this ball. Again, τ_0 depends on δ, while c_* does not.

In summary, we have proved:

THEOREM 8. *Fix constants $\mu \in (-\inf_\ell \gamma_{\tau_\ell,2}, 0)$, $\nu \in (-1/3, 0)$, $\alpha \in (0,1)$ and $\delta > 0$. Suppose τ sufficiently small and $a \in \mathbb{R}^3$ with $|a_0| + |\tau|^{3/2} |a_{\pm 1}| \leq \delta |\tau|^3$. Then for any $g \in \mathcal{C}^{2,\alpha}(S^1)$ with*

$$\int_{S^1} g = 0 \qquad \text{and} \qquad \|g\|_{\mathcal{C}^{2,\alpha}} \leq \delta |\tau|^3,$$

there exists a CMC surface with boundary which is close to Σ_{a,r_τ} and such that a collar neighborhood of its boundary can be parameterized as a vertical graph

$$\text{(3–8)} \quad \overline{B_{2r_\tau}} - B_{r_\tau} \ni x \longrightarrow \mp \frac{\tau^2}{4} \log r + a_0 + (a_1 \cos\theta + a_{-1} \sin\theta)\, r - \widehat{W}_g(x) + \widehat{V}_{\tau,a,g}(x),$$

with $-$ when $\tau > 0$ and $+$ when $\tau < 0$. Here $\widehat{W}_g$ is the unique bounded harmonic extension of g to $\mathbb{R}^2 - B_{r_\tau}$ and $\widehat{V}_{\tau,a,g}$ is a function which is bounded in $\mathcal{C}^{2,\alpha}_0(\overline{B_{2r_\tau}} - B_{r_\tau})$ by $c|\tau|^3$, where $c > 0$ does not depend on τ.

We can also see from this that $\widehat{V}_{\tau,a,g}$ depends smoothly on the parameters (τ, a, g). In fact, with a little work, we find that

(3–9)

$$\begin{aligned} &\|\widehat{V}_{\tilde{\tau},\tilde{a},\tilde{g}} - V_{\tau,a,g}\|_{\mathcal{C}^{2,\alpha}_0} \\ &\leq c\,\tau^{3\nu/2} \left(\tau^{3/2} \|\tilde{g} - g\|_{\mathcal{C}^{2,\alpha}} + \tau^2 \,|(\tilde{a}_{-1}, \tilde{a}_1) - (a_{-1}, a_1)|\right) + c\,\tau^2 \,|\tilde{\tau} - \tau| \end{aligned}$$

for some constant $c > 0$ which does not depend on δ nor on τ.

4. Adding a Delaunay End

We now assemble the results established in the previous sections and prove the main gluing theorem:

THEOREM 9. *Let $\Sigma \in \mathcal{M}^{\tau_*}_{g,k}$ be nondegenerate. Then there exists $\tau_0 > 0$, $\mathcal{U}$ a neighborhood of p in Σ and a smooth gluing map*

$$\mathfrak{G}_\Sigma : \mathcal{U} \times \{\tau : 0 < |\tau| < \tau_0\} \longrightarrow \mathcal{M}^{\tau_*}_{g,k+1}$$

where (p, τ) is mapped to the surface obtained by gluing D^+_τ to Σ at p. The surfaces in the range of this map $\mathfrak{G}$ are all nondegenerate.

We prove this theorem in the next two subsections.

4.1. Step 1: existence. Let $\Sigma \in \mathcal{M}^{\tau_*}_{g,k}$ be nondegenerate and fix a point $p \in \Sigma$ and $\delta > 0$ sufficiently large enough, and apply the results of the previous sections. These give the existence of a $\tau_0 > 0$ such that when $0 < |\tau| < \tau_0$, there are two families of CMC surfaces with boundary: $D_\tau(h)$ and $\Sigma(\tilde\tau, a, g)$. The first of these, defined at the end of section 2.6.2, are the perturbed half-Delaunay surfaces, where the boundary values $h \in \mathcal{C}^{2,\alpha}(S^1)$ satisfy

$$\int_{S^1} h = \int_{S^1} h e^{\pm i\theta} = 0, \qquad \text{and} \qquad \|h\|_{2,\alpha} \leq \kappa\tau^3.$$

The second are the CMC perturbations of $\Sigma_{a,\rho/2}$ (defined at the end of the previous section), where the parameters are defined as follows. First,

$$\tilde\tau := \pm\sqrt{\tau^2 + 4\,t},$$

with $+$ when $\tau > 0$ and $-$ when $\tau < 0$, where $t \in \mathbb{R}$ is any number such that

$$|\log\tau|\,|t| \leq \delta\,|\tau|^3;$$

next, $a \in \mathbb{R}^3$ satisfies

$$|a_0| + |\tau|^{3/2}\,|a_{\pm 1}| \leq \delta\,|\tau|^3;$$

finally, $g \in \mathcal{C}^{2,\alpha}(S^1)$ is such that

$$\int_{S_1} g = 0 \qquad \text{and} \qquad \|g\|_{2,\alpha} \leq \kappa\,|\tau|^3.$$

If we can now find choices of g, h, t and a so that

$$\Sigma(\tilde\tau, a, g) \sqcup D_\tau(h)$$

is $\mathcal{C}^1$ across the joining interface, then the existence will be established. This is because the equation is noncharacteristic at the boundaries where these surfaces intersect, and standard regularity theory for the mean curvature equation shows that this union is then a $\mathcal{C}^\infty$ surface. We denote the CMC surfaces obtained in this way by $\mathfrak{G}_\Sigma(p,\tau)$. These look like Σ with an additional Delaunay end, unduloidal when $\tau > 0$ and nodoidal when $\tau < 0$, attached at the point $p \in \Sigma$.

We now show that we can match the Cauchy data by choosing these parameters correctly. By construction, near each of their respective boundaries, the surfaces $\Sigma(\tilde\tau, a, g)$ and $D_\tau(h)$ are vertical graphs over the xy-plane. Therefore, in view of (2–15) and (3–8), it suffices to solve the equations

(4–1)

$$\begin{cases} \mp t\log r + a_0 + a_1\,r\,\cos\theta + a_{-1}\,r\,\sin\theta - \widehat{W}_g + \widehat{V}_{\tilde\tau,a,g} &= -W_h + V_{\tau,h} \\ \quad \mp t + a_1\,r\,\cos\theta + a_{-1}\,r\,\sin\theta - r\,\partial_r\widehat{W}_g + r\,\partial_r\widehat{V}_{\tilde\tau,a,g} &= -r\,\partial_r W_h + r\,\partial_r V_{\tau,h}; \end{cases}$$

all functions here are evaluated on ∂B_{r_τ}. These identities correspond to the coincidence of the Dirichlet and Neumann data, respectively, of the two graphs.

To solve these, recall that the mapping

$$\mathcal{P} : \mathcal{C}^{2,\alpha}(S^1) \ni h \longrightarrow r_\tau\,\partial_r(W_h - \widehat{W}_h)(r_\tau\,\cdot) \in \mathcal{C}^{1,\alpha}(S^1)$$

is an isomorphism such that both it and its inverse have norm uniformly bounded in τ. To see this, first observe that this mapping does not depend on r_τ, so we may as well assume that $r_\tau \equiv 1$. Next, note that $\mathcal{P}$ is a linear first order elliptic self-adjoint pseudodifferential operator with principal symbol $-2\,|\xi|$. Thus, it is enough to check that it is injective. Now, if $\mathcal{P}(h) = 0$, then the function w which equals $\widehat{W}_h$ in $\mathbb{R}^2 - B_1$ and W_h in B_1 is a global solution of $\Delta w = 0$ on all of $\mathbb{R}^2$; furthermore,

w belongs to $\mathcal{C}^{2,\alpha}(B_1) \cap [\mathcal{C}^{2,\alpha}_{-1}(\mathbb{R}^2 - B_1) \oplus \mathrm{Span}\{\log r\}]$. No such function exists, and so first w and then h must be trivial.

Now define

$$h^- := g \pm t \log r_\tau \qquad \text{and} \qquad h^+ := h + a_0 + a_1 r_\tau \cos\theta + a_{-1} r_\tau \sin\theta.$$

It is easy to see that (4–1) reduces to a fixed point problem

$$(h^+, h^-) = \mathbf{C}_\tau(h^+, h^-)$$

in the space $\mathcal{E} := (\mathcal{C}^{2,\alpha}(S^1))^2$. But (2–15) and (3–9) imply that $\mathbf{C}_\tau : \mathcal{E} \longrightarrow \mathcal{E}$ is a contraction mapping defined in the ball of radius $\delta |\tau|^3$ in $\mathcal{E}$ into itself, provided δ is sufficiently large and $|\tau|$ is sufficiently small. This gives the fixed point and completes the proof of the existence of the gluing map.

4.2. Step 2: nondegeneracy. We now prove that the surfaces $\mathfrak{G}_\Sigma(p, \tau)$ constructed above are nondegenerate when τ is small enough. The proof is by contradiction. Assume that this is not the case, so that there exists a sequence $\tau_n \to 0$, a sequence $p_n \in \Sigma$ tending to p and a sequence of surfaces

$$\Sigma_n := \mathfrak{G}_\Sigma(p_n, \tau_n),$$

for which the operators $\mathcal{L}_{\Sigma_n}$ are not injective on $\mathcal{D}^{2,\alpha}_{\mu_n}(\Sigma_n)$, for some $\mu_n < 0$. In other words, for each n there exists a nontrivial function $w_n \in \mathcal{D}^{2,\alpha}_{\mu_n}(\Sigma_n)$ such that $\mathcal{L}_{\Sigma_n} w_n = 0$. Without loss of generality, we can assume that $p = 0$ and that the tangent plane of Σ at 0 is the xy-plane.

The set indicial roots of $\mathcal{L}_{\Sigma_n}$ decomposes into two groups:

(i) those indicial roots associated to the ends of Σ_n which converge to the ends of Σ;
(ii) those indicial roots associated to the end of Σ_n which is a perturbation of D_{τ_n}.

As $n \to \infty$, elements of the first subset converge to the corresponding indicial roots of Σ, while elements of the second subset converge to elements in $\mathbb{Z} - \{\pm 1\}$. Hence there exist $\mu' \in (-\inf_\ell \gamma_{\tau_\ell,2}, 0)$ and $\tilde{\mu}' \in (-2, -1)$ such that when n is large, w_n is bounded by a multiple of $e^{\mu' s}$ on each end of the "original" ends of Σ_n and by a multiple of $e^{\tilde{\mu}' s}$ on the new end of Σ_n.

Choose $\mu \in (\mu', 0)$ and $\tilde{\mu} \in (\tilde{\mu}', -1)$. By construction, Σ_n decomposes into the union of surfaces, one of which we denote by $\tilde{\Sigma}_n$ and has boundary $\partial\tilde{\Sigma}_n$ which is a normal graph over a compact portion of Σ. The other, which we denote by $\tilde{D}_n$, is a normal graph over the Delaunay surface $D^+_{\tau_n}$. Now define a weight function $\zeta_n > 0$ on Σ_n, such that

- $\zeta_n \sim e^{\mu s}$ on each end of $\tilde{\Sigma}_n$,
- $\zeta_n \sim r^{-\tilde{\mu}}$ near $\partial\tilde{\Sigma}_n$ (for r larger than some fixed constant),
- $\zeta_n \sim |\tau|^{-2\tilde{\mu}} e^{\tilde{\mu} s}$ on $\tilde{D}_n$ (for $s + \log \tau^2$ larger than some fixed constant).

Here $f \sim g$ means that $1/2 \leq f/g \leq 2$. We use the usual cylindrical coordinates to parametrize the various ends of $\tilde{\Sigma}_n$, while in a small annulus near $\partial\tilde{\Sigma}_n$ we use polar coordinates.

Normalize the sequence w_n so that

$$\sup_{\Sigma_n} \zeta_n^{-1} w_n = 1.$$

By the choices of μ and $\tilde{\mu}$, these suprema are achieved at some point $q_n \in \Sigma_n$. We distinguish a few cases according to the behavior of the sequence of points q_n. Note that as $n \to \infty$, the surfaces Σ_n converge to the union of the original surface Σ and an infinite union of spheres of radius 1 centered at the points $(0, 0, 2j+1)$, for $j \in \mathbb{N}$.

Case 1: Suppose that (some subsequence of) the points q_n remain in a fixed end of $\tilde{\Sigma}_n$ and tend to infinity. Then $q_n = (s_n, \theta_n) \in [0, \infty) \times S^1$, where $s_n \to \infty$. Possibly extracting another subsequence, we may assume that

$$\tilde{w}_n := e^{-\mu s_n} w_n(s_n + \cdot, \cdot)$$

converges uniformly on any compact subset of $\mathbb{R} \times S^1$ to a limiting function w_∞ which is a solution of

$$\mathcal{L}_{\tau_\ell} w_\infty = 0$$

on the full Delaunay surface D_{τ_ℓ}. Here τ_ℓ is the Delaunay parameter for that end in the surface Σ. In addition, $|w_\infty| \leq c\, e^{\mu s}$. To see this is impossible, decompose $w_\infty = \sum_j w_j\, e^{ij\theta}$. By the choice of μ we have $w_0 = w_{\pm 1} = 0$ since the nontrivial solutions of $\mathcal{L}_{\tau_\ell} w = 0$ in these eigenspaces are either bounded or blow up linearly at both ends of D_{τ_ℓ}. On the other hand, the restriction of $\mathcal{L}_{\tau_\ell}$ to the eigenspaces with $|j| \geq 2$ satisfies the maximum principle. This implies that $w_j = 0$ for all other values of j. This is a contradiction.

Case 2: Next, suppose that the sequence q_n converges to a point $q_\infty \in \Sigma - \{0\}$. We may clearly assume that w_n converges uniformly on any compact of $\Sigma - \{0\}$ to a solution of

$$\mathcal{L}_\Sigma w_\infty = 0$$

defined on $\Sigma - \{0\}$. But $|w_\infty| \leq c r^{-\tilde{\mu}}$ as $r \to 0$ and $|w_\infty| \leq c\, e^{\mu s}$ at all other ends of Σ. Hence it is a Jacobi field which decays exponentially at all ends, and so must vanish by nondegeneracy.

Case 3: Finally, suppose that p_n tends to a point on the union of spheres centered at the points $(0, 0, 2j+1)$, for $j \in \mathbb{N}$. But p_n corresponds to (s_n, θ_n) in the parameterization of the Delaunay end of parameter τ_n. Both $\tilde{D}_n$ and $\tilde{\Sigma}_n - \tilde{\Sigma}_{n,\rho}$ are normal graphs over D_{τ_n} when ρ is small enough. Thus we may define the rescaled sequence

$$\tilde{w}_n := e^{-\tilde{\mu}\, s_n} w_n(\cdot + s_n, \cdot).$$

It is proved in [**8**] that, as $\tau_n \to 0$, the term of order 0 in $\mathcal{L}_{\tau_n}$ converges either to 0 or to $2 \cosh^{-2} s$ on compact subsets. It is then easy to see that, some subsequence of $(\tilde{w}_n)_n$ converges to w_∞, a nontrivial solution of one of the equations

$$\Delta_0 w_\infty = 0$$

or

$$\Delta_0 w_\infty + \frac{2}{\cosh^2(s + \bar{s})} w_\infty = 0, \qquad \text{for some} \quad \bar{s} \in \mathbb{R}$$

on $\mathbb{R} \times S^1$. In addition, $|w| \leq c\, e^{\tilde{\mu} s}$. This is once again impossible. To see this, decompose $w_\infty = \sum_j w_j(s)\, e^{ij\theta}$. By the choice of $\tilde{\mu}$ we get $w_0 = w_{\pm 1} = 0$ since nontrivial solutions of the homogeneous problems on the eigenspaces $j = 0, \pm 1$ decay at most like $\cosh^{-1} s$ at ∞, whereas $\tilde{\mu} \in (-2, -1)$. On the other hand, the restrictions of these two operators to the eigenspaces with $|j| \geq 2$ satisfy the maximum principle, and so all the remaining components $w_j = 0$. This is again a contradiction.

We have now ruled out all possibilities, and so the surfaces $\mathfrak{G}_\Sigma(p,\tau)$ are non-degenerate when τ is sufficiently small.

5. Analysis of the Forgetful Map

In the remainder of this paper we apply this gluing construction to study some aspects of the topology of $\mathcal{M}_{g,k}$; this will be accomplished using a natural mapping from this space into the more familiar and better understood Teichmüller space of closed Riemann surfaces of genus g with an ordered k-tuple of points deleted. We let $\mathcal{T}_{g,k}$ denote this latter space; any element has the form $\overline{\Sigma} - \{p_1, \ldots, p_k\}$, where $\overline{\Sigma}$ is a compact Riemann surface and the p_j are distinct points on it.

Any CMC surface $\Sigma \in \mathcal{M}_{g,k}$ is conformally equivalent to some element of this Teichmüller space. This allows us to define a forgetful map

$$\mathcal{F}_{g,k} : \mathcal{M}_{g,k} \longrightarrow \mathcal{T}_{g,k}, \tag{5–1}$$

which is given by forgetting the geometric (i.e., CMC) structure of a surface and remembering only its conformal structure. For convenience, we shall usually drop the subscripts (g,k) from the notation for this map.

Our basic goal is to understand the image $\mathcal{I}_{g,k} = \mathcal{I}$ of this mapping. We first show that $\mathcal{F}$ is real analytic. A recent result of Kusner [**5**] states that (a slight modification of) $\mathcal{F}$ is proper. Together, these results show that $\mathcal{I}$ is a closed, subanalytic set. As such, it is stratified by real analytic submanifolds, and so we may define the codimension of $\mathcal{I}$ as the codimension of its stratum of maximal dimension. Using the preceding gluing construction, we prove that $\mathcal{F}$ is surjective when $g = 0$ while for each $g > 0$ its image has codimension which is uniformly bounded (depending on g) as $k \to \infty$. Examined more carefully, this argument also shows that $\mathcal{M}_{g,k}$ detects much of the topology of this Teichmüller space. We conclude by summarizing the ramifications of these results for the differential topological structure of the -MC moduli space.

We stress that the basic ideas here are very intuitive, granting the main gluing theorem. Roughly speaking, the fact that $\mathcal{F}_{0,k}$ is surjective when $g = 0$ follows inductively from the fact that we can glue a half-Delaunay surface at any point p of any fixed nondegenerate $\Sigma \in \mathcal{M}_{0,k'}$, $3 \leq k' \leq k$, without changing the conformal structure much away from p. The result is complicated when $g > 0$ by the fact that there may be constraints on the image of $\mathcal{F}$ which we do not see directly; this is where the real analyticity enters, for it implies that the image does lie in a well-behaved set. The uniform boundedness of the codimension of the image means essentially that the only serious constraints on the image occur when k is small, and that when k is large enough, the (conformal) location of the ends may be chosen freely.

5.1. Analyticity of $\mathcal{F}$. We show now that $\mathcal{F}$ is analytic, not only when near smooth points of $\mathcal{M}_{g,k}$, but even as a map of (possibly singular) real analytic spaces.

To state this result precisely, fix $\Sigma \in \mathcal{M}_{g,k}$ and let g denote its induced metric. We parametrize a neighborhood of Σ in the space of all nearby surfaces in the usual way, using a finite-dimensional family of deformations of Σ which preserve the CMC structure of the ends and then taking normal perturbations of these. Part of the main theorem in [**7**] is that for any $\Sigma \in \mathcal{M}_{g,k}$, there is a neighborhood in the CMC moduli space which lies in some open finite-dimensional analytic submanifold $\mathcal{Q}$ of this infinite-dimensional space. When Σ is nondegenerate, then we may assume

that $\mathcal{Q}$ is a real analytic coordinate chart in $\mathcal{M}_{g,k}$, but in general, the CMC moduli space is the zero set of a real analytic function on $\mathcal{Q}$. To be definite, we regard each element of $\mathcal{Q}$ as an embedding of a fixed surface Σ_0 into $\mathbb{R}^3$; we let q_0 denote the base embedding, corresponding to the original CMC surface Σ. From that construction, we may even assume that every $q \in \mathcal{Q}$ is an analytic embedding, but this is not so important here.

PROPOSITION 9. *Suppose $\Sigma \in \mathcal{M}_{g,k}$, and let the finite dimensional analytic manifold $\mathcal{Q}$ be chosen as above. Then the natural extension of the forgetful map $\mathcal{F}$ which assigns to any $q \in \mathcal{Q}$ the element of $\mathcal{T}_{g,k}$ determined by $q(\Sigma_0)$ is a real analytic mapping.*

Since the CMC moduli space (locally) lies in $\mathcal{Q}$, this result gives what is perhaps the most natural meaning to the statement that $\mathcal{F}$ is real analytic on $\mathcal{M}_{g,k}$ near singular points of this moduli space.

PROOF. Let g_0 denote the base metric on Σ_0, i.e., $g_0 = q_0^*(\delta)$, where δ is the Euclidean metric. Similarly, for $q \in \mathcal{Q}$ we let $g_q = q^*(\delta)$. For each one of these metrics, there is a uniquely determined conformal factor $e^{2\phi_q}$ such that $h_q = e^{2\phi_q} g_q$ is a hyperbolic metric of finite area. These hyperbolic metrics parametrize the Teichmüller space $\mathcal{T}_{g,k}$, and so the theorem will be proved if we show that the map $q \mapsto \phi_q$ is real analytic.

For each end E_j of Σ_0, fix isothermal coordinates (s_j, θ_j). The model Delaunay surface D_{τ_j} for this end has the metric

$$g_{\tau_j} = \frac{1}{4}\tau_j^2\, e^{2\sigma_{\tau_j}}\, (ds_j^2 + d\theta_j^2),$$

and hence

$$g_q = \frac{1}{4}\tau_j^2\, e^{2\sigma_{\tau_j}}\, \left(ds_j^2 + d\theta_j^2 + \mathcal{O}(e^{-\alpha s_j})\right)$$

there, for some $\alpha > 0$. The Delaunay parameters τ_j and the functions σ_{τ_j} depend real analytically on q. We can even modify this conformal factor further and thereby find a function μ_q which depends real analytically on q such that

$$g_{c,q} := e^{2\mu_q} g_q = ds_j^2 + d\theta_j^2 \qquad \text{on} \quad E_j.$$

This is a metric with cylindrical ends. Accordingly, decompose the sought-after conformal factor ϕ_q as $\mu_q + \psi_q$. We must prove that ψ_q depends analytically on q.

Letting Δ_q and K_q denote the Laplace operator and Gauss curvature function for $g_{c,q}$, then ψ_q is the unique solution of the PDE

$$\Delta_q \psi_q - e^{2\psi_q} = K_q.$$

Write ψ_0 for the unique solution when $q = q_0$. Note that $K_q = 0$ on all ends, and so $-\log s_j$ is a solution along each E_j.

We shall define natural function spaces X and Y below (these will be certain weighted Sobolev spaces), on which the mapping

$$\mathcal{N} : \mathcal{Q} \times X \longrightarrow Y,$$

defined by

$$(q, \psi) \quad \longmapsto \quad \Delta_q \psi - e^{2\psi} - K_q$$

is locally surjective near (q_0, ψ_0). To this end, observe that the differential in the second factor is

$$Lw := D_2 \mathcal{N}_{q_0,\psi_0}(w) = \Delta_0 w - 2e^{2\psi_0} w.$$

In terms of the coordinates (s_j, θ_j) on E_j,

$$L|_{E_j} = \partial_{s_j}^2 + \partial_{\theta_j}^2 - 2/{s_j}^2.$$

It remains to choose the function spaces X and Y which have the correct properties, in particular that $L : X \mapsto Y$ is surjective. While this is not too difficult, and equivalent theorems surely exist elsewhere in the literature, we sketch the proof for completeness.

As a first attempt, we can try to use a procedure similar to the one outlined in §3.1; namely, if we define the weighted Sobolev spaces $e^{ms} H^t(\Sigma)$ (so u is in this space if $u \in H^t_{\mathrm{loc}}$ and on each end E_j, $u = e^{ms_j} v$ where v is in the ordinary Sobolev space $H^t(E_j)$), then it is straightforward to show that

$$L : e^{ms} H^{t+2}(\Sigma) \longrightarrow e^{ms} H^t(\Sigma) \tag{5–2}$$

is Fredholm whenever $m \notin \mathbb{Z}$. Furthermore, using the maximum principle, which applies since the term of order zero in L is strictly negative, the mapping (5–2) is injective when $m < 0$; by duality and elliptic regularity, it is then surjective when $m > 0$, $m \neq 1, 2, \ldots$. Again as in §3.1, there is a deficiency space consisting of the linear span of cutoffs of a special set of temperate solutions to $Lu = 0$ along each end. We can determine these temperate solutions using separation of variables and see that they are linear combinations of the functions s^2 and s^{-1}. (For convenience, in many places in the remainder of this proof we drop references to the various ends E_j and also omit the subscript j.) Therefore, if $-1 < m < 0$, then for any $f_1 \in e^{ms} H^t(\Sigma)$ it is possible to find a function

$$u_1 = as^2 + bs^{-1} + \tilde{u}, \qquad \tilde{u} \in e^{ms} H^{t+2}(\Sigma),$$

where a and b are constants, and such that $Lu_1 = f_1$ on Σ. The deficiency space W consists of all linear combinations of these solutions s^2 and s^{-1} on all ends, and thus is $2k$-dimensional. We have shown that

$$L : e^{ms} H^{t+2}(\Sigma) \oplus W \longrightarrow e^{ms} H^t(\Sigma) \tag{5–3}$$

is surjective. A relative index calculation (see [**7**]) shows that the nullspace $\mathcal{B}$ of (5–3) is k-dimensional.

Unfortunately, this is not the end of the story because the nonlinear operator $\mathcal{N}$ does not carry this domain space into $e^{ms} H^t(\Sigma)$; in fact, there is no evident way to use the geometric context to regularize this mapping. Therefore we proceed further.

We first require Sobolev spaces with polynomial rather than exponential weights. Thus for $t \in \mathbb{N}$ and $\nu \in \mathbb{R}$, let $H^t_\nu(\mathbb{R}^+ \times S^1)$ be the space of functions in H^t_{loc} such that $s^{-1/2+k-\nu} \nabla^k u \in L^2(\mathbb{R}^+ \times S^1)$, $k = 0, \ldots, t$. We also define $H^t_\nu(\Sigma)$ as the space of $H^t_{loc}(\Sigma)$ functions which lie in $H^t_\nu(\mathbb{R}^+ \times S^1)$ on each end.

By separation of variables, it is easy to produce a map

$$G_0 : H^t_{\nu-2}(\mathbb{R}^+ \times S^1) \longrightarrow H^{t+2}_\nu(\mathbb{R}^+ \times S^1)$$

for any $\nu \in \mathbb{R}$, $\nu \neq -1, 2$, such that $(\partial_s^2 + \partial_\theta^2 - 2/s^2)\, G_0\, f_0 = f_0$. In other words, G_0 is a right inverse for L. Observe that we do not impose any boundary data, in particular G_0 is not unique. Using G_0 and cutoff functions χ_j on each ends, we can produce an operator

$$\tilde{G}_0 : H^t_{\nu-2}(\Sigma) \longrightarrow H^{t+2}_\nu(\Sigma)$$

such that $u_0 := \tilde{G}_0 f_0$ has the property that Lu_0 vanishes outside a compact set of Σ.

Now, if $f \in H^t_{\nu-2}(\Sigma)$ has compact support and if $t \geq 2$, then f is continuous. Since the constant function 1 is a subsolution for the operator L — in fact $L(1) = -2/s^2$ on each end — we can solve a sequence of equations $Lu_j = f$ in Σ_j with $u_j = 0$ on the boundaries $\partial\Sigma_j$, where Σ_j is a smooth exhaustion of Σ by compact sets; the limit of this sequence is a *bounded* solution u of $Lu = f$. By earlier remarks, this is the only bounded solution of this equation. Since $f = 0$ on an end, then we know that u must be a linear combination of s^2 and $1/s$ and a term which exponentially decreases on that end, and so by boundedness, $u = u_1 + v$ where $u_1 = a/s$ and $v \in e^{-s} H^{t+2}$.

Finally, let $\nu \in (-1, 0]$, $t \geq 2$ and define

$$X = H^{t+2}_{\nu}(\Sigma), \qquad \text{and} \qquad Y = H^t_{\nu-2}(\Sigma).$$

We claim that $L : X \to Y$ is an isomorphism and $\mathcal{N} : \mathcal{Q} \times X \to Y$ is smooth. To prove these, first suppose that $f \in Y$. Let $u_0 := \tilde{G}_0 f$, then $L(u - u_0) = \tilde{f}$ has compact support. Next, there is also a unique bounded solution $\bar{u}$ of $L\bar{u} = \tilde{f}$. As explained above, separating variables on the end, we see that $\tilde{v} := u_0 + \bar{u}$ is the sum of some multiple of $1/s$ and a function in $H^{t+2}_{\nu}(\Sigma)$. Tracing through this procedure, we have found a bounded map $G : Y \to X$ such that $LG = I$. Since L does not have any nullspace in X, this map is an isomorphism. So far, we have only used the fact that $t \geq 2$ and $\nu \in (-1, 2)$.

We also note that $\mathcal{N}$ is a real analytic mapping from a neighborhood of 0 in X to Y. This follows from the considerations above concerning the linear part L of $\mathcal{N}$, and from inspection of the nonlinear error term, which has the form $s^{-2}(e^{2w} - 1 - 2w)$ on each end. This is where the restriction that $\nu \leq 0$ is required.

In any case, we may now apply the real analytic version of the implicit function theorem to get the existence of a real analytic mapping $\Psi : \mathcal{Q} \to X$ such that $\mathcal{N}(q, \Psi(q)) \equiv 0$. Since $\psi_q = \Psi(q)$, we have proved the theorem. □

5.2. The image of $\mathcal{F}$. Recall that a connected component of $\mathcal{M}_{g,k}$ is said to be nondegenerate if it contains an element Σ which is (analytically) nondegenerate in the sense of Definition 1. The principal stratum in a nondegenerate component of $\mathcal{M}_{g,k}$ has dimension $3k$. On the other hand, $\mathcal{T}_{g,k}$ is a real analytic manifold of dimension $6g-6+2k$. Therefore one might hope that $\mathcal{F}$ is surjective, at least when k is sufficiently large.

We now give some results concerning the nature and size of the image $\mathcal{I}$. These require some preliminary definitions.

There is a tautological bundle $\mathbb{V}_{g,k}$ over $\mathcal{T}_{g,k}$ defined by

$$\mathbb{V}_{g,k} = \left\{([(\overline{\Sigma}, p_1, \ldots, p_k)], p) : [(\overline{\Sigma}, p_1, \ldots, p_k)] \in \mathcal{T}_{g,k},\ p \in \overline{\Sigma} - \{p_1, \ldots, p_k\}\right\}.$$

This is the domain of a natural augmentation map

$$\begin{aligned} \mathcal{A} : \mathbb{V}_{g,k} &\longrightarrow \mathcal{T}_{g,k+1}, \\ ([(\overline{\Sigma}, p_1, \ldots, p_k)], p) &\longmapsto [(\overline{\Sigma}, p_1, \ldots, p_k, p)]. \end{aligned}$$

It is clear that $\mathcal{A}$ is an isomorphism.

Next, there is also a tautological bundle over $\mathcal{M}_{g,k}$,

$$\mathbb{U}_{g,k} = \{(\Sigma, p) : \Sigma \in \mathcal{M}_{g,k}, p \in \Sigma\}.$$

For any $(\Sigma, p) \in \mathbb{U}_{g,k}$, let $(0, \tau^*(\Sigma, p))$ denote the largest open interval such that if $0 < \tau < \tau^*(\Sigma, p)$, then the gluing map which attaches a half-Delaunay end with Delaunay parameter τ to the point p exists. It follows from the gluing construction that $\tau^*(\Sigma, p)$ is bounded away from zero on compact sets of Σ, provided Σ lies in a compact set in a nondegenerate stratum of $\mathcal{M}_{g,k}$. Let

$$\mathcal{W}_{g,k} = \{(\Sigma, p, \tau) : (\Sigma, p) \in \mathbb{U}_{g,k}, 0 < \tau < \tau^*(\Sigma, p)\}.$$

This is the natural domain of the gluing map

$$\mathfrak{G} : \mathcal{W}_{g,k} \longrightarrow \mathcal{M}_{g,k+1}.$$

This map is continuous, and in fact, real analytic. We shall often omit the subscripts (g, k) from these bundles when the meaning is clear. Also, if C is any subset either of $\mathcal{M}_{g,k}$ or of $\mathbb{U}_{g,k}$, then we let $\mathcal{W}(C)$ denote the portion of $\mathcal{W}$ lying over C; in particular $\mathcal{W}(\Sigma)$ denotes the natural domain of the gluing map over a fixed surface Σ. We write the CMC surface $\mathfrak{G}_\Sigma(p, \tau)$ as $\Sigma_{p,\tau}$. Finally, note that

$$\lim_{\tau \to 0} \mathcal{F}(\Sigma_{p,\tau}) = \mathcal{A}(\mathcal{F}(\Sigma), p). \tag{5–4}$$

THEOREM 10. *Suppose that $g = 0$ and $k \geq 3$. Then there is a nondegenerate component of $\mathcal{M}_{0,k}$ on which $\mathcal{F}$ is surjective.*

PROOF. We prove this by induction on k. When $k = 3$, then according to [**3**], $\mathcal{M}_{0,3}$ is homeomorphic to a 3-ball, hence in particular is connected; by [**8**] it contains a nondegenerate element (one can also use the result of Theorem 1 to produce a 3-ended nondegenerate CMC surface by adding an end to a Delaunay surface). In other words, $\mathcal{M}_{0,3}$ contains a single component, and this component is nondegenerate. On the other hand, $\mathcal{T}_{0,3}$ consists of a single point. Hence $\mathcal{F}_{0,3}$ is obviously surjective.

Now suppose that $k \geq 3$, and C^0_k is a nondegenerate stratum in some component $C_k \subset \mathcal{M}_{0,k}$ such that $\mathcal{F} : C^0_k \to \mathcal{T}_{0,k}$ is surjective. Choose any point $[(S^2, p_1, \ldots, p_k, p_{k+1})] \in \mathcal{T}_{0,k+1}$ (note that $\overline{\Sigma}$ must be S^2 when $g = 0$), and let $([(S^2, p_1, \ldots, p_k)], p_{k+1})$ be the lift of this point to $\mathbb{V}_{0,k}$. Write p for p_{k+1} for simplicity.

By assumption, there is an element $\Sigma \in C^0_k$ such that $\mathcal{F}(\Sigma) = [(S^2, p_1, \ldots, p_k)]$. Let $\mathcal{B}$ be some neighborhood of (Σ, p) in $\mathbb{U}_{0,k}$. Then for any $(\Sigma', q, \tau) \in \mathcal{B} \times (0, \eta)$, we obtain elements $\Sigma'_{q,\tau} \in \mathcal{M}_{0,k+1}$ and $(\mathcal{F}(\Sigma'), q) \in \mathbb{V}_{0,k}$. The theorem will be proved if we show that

$$\mathcal{A}(\mathcal{F}(\Sigma), p) \in (\mathcal{F} \circ \mathfrak{G})(\mathcal{B}, \eta),$$

when η is small enough. But this is clear from (5–4) using a straightforward degree theory argument.

This proof also shows that the nondegenerate stratum C^0_{k+1} is obtained inductively by gluing half-Delaunay surfaces with very small necks located at arbitrary points $p \in \Sigma$ for all surfaces $\Sigma \in C^0_k$. □

As remarked above, when $g > 0$ we no longer expect $\mathcal{F}_{g,k}$ to be surjective. We shall show instead that its codimension does not become unbounded as $k \to \infty$.

THEOREM 11. *Suppose that $g \geq 1$ and that C_{k_0} is a nondegenerate component of $\mathcal{M}_{g,k_0}$, with nondegenerate stratum $C^0_{k_0}$. Then for each $k = k_0, k_0 + 1, \ldots$ there is a nondegenerate component $C_k \subset \mathcal{M}_{g,k}$ such that codimension of the image $\mathcal{I}_k = \mathcal{F}(C_k)$ is bounded as $k \to \infty$.*

PROOF. This is also proved by an inductive procedure. Let us suppose that for some $k \geq k_0$, there is a nondegenerate element $\Sigma \in \mathcal{M}_{g,k}$; suppose also that rank $D\mathcal{F}|_\Sigma \equiv r$ is maximal amongst all such elements. Let C_k be the component of $\mathcal{M}_{g,k}$ containing Σ and $\mathcal{I}_k$ its image by $\mathcal{F}$ in $\mathcal{T}_{g,k}$. The dimension of the stratum of $\mathcal{I}_k$ through $\mathcal{F}(\Sigma)$ is r, and so $d_k = \operatorname{codim}(\mathcal{I}_k) \leq 6g - 6 + 2k - r$. Let us furthermore choose an r-dimensional analytic submanifold $\mathcal{S}_k$ through Σ such that the restriction of $\mathcal{F}$ to it is an analytic diffeomorphism onto its image. Let $\mathbb{U}(\mathcal{S}_k)$ be the portion of the bundle $\mathcal{U}$ lying over $\mathcal{S}_k$.

We now consider, for some small $\eta > 0$, the restriction of the gluing map

$$\mathfrak{G}_\eta : \mathbb{U}(\mathcal{S}_k) \longrightarrow \mathcal{M}_{g,k+1}.$$

We first claim that the dimension of the image of $\mathfrak{G}_\eta$ is $r+2$. This is straightforward, since the differential of $\mathfrak{G}_\eta$ in the directions of the fibers of $\mathbb{U}(\mathcal{S}_k)$, i.e., letting p vary and $\Sigma' \in \mathcal{S}$ remain fixed, is injective. This produces an $(r+2)$-dimensional submanifold $\mathcal{S}_{k+1}$ in $\mathcal{M}_{g,k+1}$, consisting entirely of nondegenerate points. Using (5–4) again, when η is small enough, $\mathcal{F}(\mathcal{S}_{k+1})$ is $(r+2)$-dimensional. However, since $\dim(\mathcal{T}_{g,k+1}) - \dim(\mathcal{T}_{g,k}) = 2$, we see that the codimension of $\mathcal{F}(\mathcal{S}_{k+1})$ is again $6g - 6 + 2k - r$, and so the codimension of the image of the component of $\mathcal{M}_{g,k+1}$ containing $\mathcal{S}_{k+1}$ is bounded by this same number. This proves the theorem. □

Regarding the global structure of the image $\mathcal{I}_{g,k}$ of $\mathcal{M}_{g,k}$ by $\mathcal{F}$, we quote a recent nice result of Kusner. To state it, recall that the necksize parameter τ of an end E of $\Sigma \in \mathcal{M}_{g,k}$ is the Delaunay parameter of the Delaunay surface to which this end is asymptotic. This value may be determined using the force integral.

PROPOSITION 10. *(Kusner* [**5**]*) Let Σ_ℓ be a sequence of elements in $\mathcal{M}_{g,k}$ and suppose that the necksize parameters $\tau_j^{(\ell)}$ of the ends E_j of Σ_ℓ are all bounded below by some $\eta > 0$. Then either the conformal structures $\mathcal{F}(\Sigma_\ell)$ diverge in $\mathcal{T}_{g,k}$ or else the surfaces Σ_ℓ converge, up to rigid motion, to some limiting surface $\Sigma_\infty \in \mathcal{M}_{g,k}$.*

Kusner's result is somewhat more general, and he phrases it in terms of properness of $\mathcal{F}$.

Combining Propositions 9 and 10, we obtain:

PROPOSITION 11. *The image $\mathcal{I}_{g,k}$ of the forgetful map $\mathcal{F}$, restricted to any (not necessarily nondegenerate) component in $\mathcal{T}_{g,k}$ is a closed subanalytic set.*

We conclude this subsection with a discussion of the fundamental group of $\mathcal{M}_{g,k}$. While we do not determine this group precisely, we examine the homomorphism

$$\mathcal{F}_* : \pi_1(\mathcal{M}_{g,k}) \longrightarrow \pi_1(\mathcal{T}_{g,k}).$$

The group on the right here is well understood and rather complicated. We show that the image of $\mathcal{F}_*$ is a fairly large subgroup.

We first review some facts about the group $\pi_1(\mathcal{T}_{g,k})$, referring to [**1**] for more details. There is a subsidiary forgetful map

$$F' : \mathcal{T}_{g,k} \longrightarrow \mathcal{C}(g,k).$$

The spaces on the right here are the Teichmüller space of conformal structures on a compact surface of genus g (the classical Teichmüller space) and the configuration space of k points on a compact surface of genus g. To define it, recall that we may

identify an element of $\mathcal{T}_{g,k}$ with a hyperbolic metric on the compact surface $\overline{\Sigma}$ along with an ordered k-tuple of distinct points $(p_1, \ldots, p_k)$ on $\overline{\Sigma}$ (rather than finite area complete hyperbolic metrics on $\overline{\Sigma} - \{p_1, \ldots, p_k\}$). Then F' is defined by forgetting the conformal structure. It is well known that

$$F'_* : \pi_1(\mathcal{T}_{g,k}) \longrightarrow \pi_1(\mathcal{C}(g,k))$$

is an isomorphism. We write $\mathcal{F}' = F' \circ \mathcal{F}$.

The space $\mathcal{C}(g,k)$ has a rather interesting topology. Its fundamental group is known as the pure braid group of the surface of genus g on k braids, and we denote it by $B(g,k)$. It is a finitely generated group; the loops Γ_{ij}, $1 \leq i,j \leq k$, $i \neq j$, corresponding to the point p_j traversing a small loop winding around the point p_i once, with all other p_ℓ fixed, comprise a generating set.

THEOREM 12. *When $k \geq 3$, the map*

$$\mathcal{F}'_* : \pi_1(\mathcal{M}_{0,k}) \longrightarrow \pi_1(\mathcal{C}(0,k))$$

is an epimorphism.

THEOREM 13. *Suppose $g \geq 1$ and $\mathcal{M}_{g,k_0}$ contains a nondegenerate component. Then for any $k \geq k_0$, $\mathcal{M}_{g,k}$ contains a nondegenerate component C_k such that the image of the homomorphism*

$$\mathcal{F}'_* : \pi_1(C_k) \longrightarrow \pi_1(\mathcal{C}(g,k))$$

contains the subgroup of $B(g,k)$ generated by the collection of loops Γ_{ij}, $j > k_0$.

The proofs of these two theorems are nearly identical. They rely only on the simple observations that any of the loops Γ_{ij} are in the image of $\mathcal{F}'_*$ when $g = 0$, while when $g > 0$, at least those loops with $j > k_0$ are in the image.

5.3. The structure of the CMC moduli space. We conclude this paper by describing informally what we know at this point about the CMC moduli spaces $\mathcal{M}_{g,k}$.

As before, we let $\mathcal{I}_{g,k}$ denote the image of $\mathcal{M}_{g,k}$ under the forgetful map $\mathcal{F}$. Then it is tempting to think of

$$\mathcal{F} : \mathcal{M}_{g,k} \longrightarrow \mathcal{I}_{g,k} \subset \mathcal{T}_{g,k}$$

as a sort of singular fibration. We have shown that all spaces here are real analytic or subanalytic, hence stratified, and $\mathcal{F}$ is a real analytic mapping. The image $\mathcal{I}_{g,k}$ detects at least some fairly large portion of the fundamental group of $\mathcal{T}_{g,k}$ when $g \geq 1$ and k is large; it detects all of it when $g = 0$ and $k \geq 3$. If $\mathcal{M}_{g,k}(\eta)$ denotes the subset of surfaces $\Sigma \in \mathcal{M}_{g,k}$ with necksizes of all ends of Σ no smaller than η, then by Kusner's theorem, the restriction of $\mathcal{F}$ to this subset is proper. As already noted, one way to interpret this is that if Σ_j is a divergent sequence of surfaces in $\mathcal{M}_{g,k}$ with no end necksizes tending to zero, then necessarily the conformal structures $\mathcal{F}(\Sigma_j)$ must be degenerating. This behavior does indeed occur; for example, the connected sum construction of [**10**], shows that it is possible to construct sequences of surfaces with no end necksizes tending to zero, but with some interior necks pinching off. This corresponds to degeneration in $\mathcal{T}_{g,k}$, and these examples exist even when $g = 0$.

We conclude with a number of open questions:

- Is $\mathcal{M}_{g,k}$ connected? The only case where this is understood (and known to be true) is when $(g,k) = (0,3)$ by [**3**].
- Do the fibers of $\mathcal{F}$ ever have nontrivial topology; for example, do they ever contain homotopically nontrivial loops? This does not occur in $\mathcal{M}_{0,3}$, and if the fibers are always contractible, then the image $\mathcal{I}_{g,k}$ of any component of $\mathcal{M}_{g,k}$ would be a retract of that component.
- Is $\mathcal{F}_{g,k}$ ever surjective when $g > 0$? A heuristic argument against this might be made by considering configurations $(\overline{\Sigma}, p_1, \ldots, p_k)$ where all of the points p_j are contained in a small neighborhood (for example, a small ball relative to the conformally equivalent flat or hyperbolic metric on $\overline{\Sigma}$); it seems likely that the CMC balancing formulae would rule out CMC realizations of such configurations.
- Construct, or prove the existence of, a degenerate CMC surface Σ in some $\mathcal{M}_{g,k}$. Although the possibility of their existence adds significant complications throughout the theory, to date none are known to exist.
- In a related direction, it seems quite likely that every element in $\mathcal{M}_{0,3}$ is nondegenerate, and it would be very useful to know whether this is true. One motivation is that the (otherwise) very explicit geometric knowledge about these surfaces makes them ideally suited as building blocks in gluing constructions, but to use them in this way requires knowing that they are nondegenerate.
- We have not discussed the geometric structure on these CMC moduli spaces. Along these lines, it is proved in [**7**] that on the infinitesimal level, each $\mathcal{M}_{g,k}$ has the structure of a Lagrangian submanifold of a larger symplectic submanifold. More precisely, the tangent space of $\mathcal{M}_{g,k}$ at any nondegenerate point is a Lagrangian subspace of a natural symplectic vector space. It is not too difficult to make this global picture more precise, but the more compelling question is: what can be done with it? Some simple examples and other evidence point to the possibility that there is a tautological one-form on some large subset of this ambient symplectic manifold which becomes singular on the subvariety of surfaces with at least one (asymptotically) cylindrical end. The periods of this one-form appear to have direct geometric meaning. Is there anything else which can be done with this Lagrangian structure?

References

[1] J. Birman, *Braids, links and mapping class groups*, Annals of Mathematical Studies **82**, Princeton University Press, Princeton (1975).

[2] S. Fakhi and F. Pacard, *Existence result for minimal hypersurfaces with a prescribed finite number of planar ends*, Manuscripta Math. **103**:4 (2000), 465–512.

[3] K. Große-Brauckmann, R. Kusner and J. Sullivan, *Triunduloids: embedded constant mean curvatures surfaces with three ends and genus zero*, preprint, math.DG/0102183.

[4] N. Kapouleas, *Complete constant mean curvature surfaces in Euclidean three space*, Ann. of Math. (2) **131** (1990), 239–330.

[5] R. Kusner, *Conformal structure of embedded CMC surfaces*, pp. 585–596 in this volume.

[6] N. Korevaar, R. Kusner and B. Solomon, *The structure of complete embedded surfaces with constant mean curvature*, J. Differential Geometry 30 (1989), 465–503.

[7] R. Kusner, R. Mazzeo and D. Pollack, *The moduli space of complete embedded constant mean curvature surfaces*, Geom. Funct. Anal. **6** (1996), 120–137.

[8] R. Mazzeo and F. Pacard, *Constant mean curvature surfaces with Delaunay ends*, Comm. Anal. Geom. **9**:1 (2001) 169–237.

[9] R. Mazzeo and F. Pacard, *Bifurcating nodoids*, pp. 169–186 in Topology and Geometry: Commemorating SISTAG, Contemporary Mathematics 314, American Mathematical Society, Providence (2002).

[10] R. Mazzeo, F. Pacard and D. Pollack, *Connected sums of constant mean curvature surfaces in Euclidean 3 space*, J. Reine Angew. Math. **536** (2001), 115–165.

[11] R. Mazzeo, D. Pollack and K. Uhlenbeck, *Moduli spaces of singular Yamabe metrics*, J. Amer. Math. Soc. **9**:2 (1996), 303–344.

[12] J. Ratzkin, *An end-to-end gluing construction for surfaces of constant mean curvature*, PhD Thesis, University of Washington (2001).

[13] W. Meeks III, *The topology and geometry of embedded surfaces of constant mean curvature*, J. Differential Geom. **27**:3 (1988), 539–552.

Department of Mathematics, Stanford University, Stanford, CA 94305
E-mail address: mazzeo@math.stanford.edu

Centre de Mathématiques, Université Paris XII, 61, avenue du Général de Gaulle, 94010 Creteil, France
E-mail address: pacard@univ-paris12.fr

University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350
E-mail address: pollack@math.washington.edu

Clay Mathematics Proceedings
Volume 2, 2004

Constructing Mean Curvature 1 Surfaces in H^3 with Irregular Ends

Wayne Rossman, Masaaki Umehara, and Kotaro Yamada

Abstract. With the developments of the last decade on complete constant mean curvature 1 (CMC-1) surfaces in the hyperbolic 3-space H^3, many examples of such surfaces are now known. However, most of the known examples have regular ends. (An end is regular if the hyperbolic Gauss map of the surface has no essential singularity there.) There are some known surfaces with irregular ends, but they are all either reducible or of infinite total curvature. (The surface is reducible if and only if the monodromy of the secondary Gauss map can be simultaneously diagonalized.) Up to now there have been no known complete irreducible CMC-1 surfaces in H^3 with finite total curvature and irregular ends.

The purpose of this paper is to construct countably many 1-parameter families of genus zero CMC-1 surfaces with irregular ends and finite total curvature, which have either dihedral or Platonic symmetries. For all the examples we produce, we show that they have finite total curvature and irregular ends. For the examples with dihedral symmetry and the simplest example with tetrahedral symmetry, we show irreducibility. Moreover, we construct a genus one CMC-1 surface with four irregular ends, which is the first known example with positive genus whose ends are all irregular.

Introduction

Let H^3 denote the unique simply connected complete 3-dimensional Riemannian manifold with constant sectional curvature -1, which we call the hyperbolic 3-space. Associated to a complete finite-total-curvature CMC-1 (constant mean curvature one) conformal immersion $f\colon M \to H^3$ of a Riemann surface M are two meromorphic maps called the hyperbolic Gauss map and the secondary Gauss map, which we denote by G and g respectively (to be defined in the next section). Using these two Gauss maps, we can define two characteristics of the surface f:

(1) It is known that M is biholomorphic to a compact Riemann surface with a finite number of points removed, and hence each end is conformally a punctured disk. Therefore we may consider the order of the hyperbolic Gauss map G at each end, and an end is called *regular* if G has at most a pole singularity at this end. If G has an essential singularity, the end is called *irregular*.

©2004 Clay Mathematics Institute

(2) Although G is single-valued on M, the secondary Gauss map g might be multi-valued on M, so we can have a nontrivial monodromy representation defined on the first fundamental group of M. This monodromy group is a subgroup of SU(2), and if all members of this group can be diagonalized by the same conjugation, we say that the surface f is *reducible*. Otherwise, we say that f is *irreducible*. (Irreducibility depends on a global behavior of the surface but not on individual ends.)

If a CMC-1 immersion is reducible, the surface can be deformed preserving its hyperbolic Gauss map G and Hopf differential ($(2,0)$-part of the second fundamental form, see Section 1). On the other hand, an irreducible surface is the only surface with given hyperbolic Gauss map and Hopf differential.

Recent progress in the theory of CMC-1 surfaces in H^3 has led to the discovery of many new examples of these surfaces. Many examples with regular ends are now known, and various properties of these surfaces are understood. Bryant [**Bry**] found a local description for these surfaces in terms of holomorphic data that initiated this recent progress. The last two authors [**UY1**]–[**UY7**] developed the theory using Bryant's description to find many examples and properties, and work in this direction has been continued by Small [**Sm**], the authors [**RUY1**]–[**RUY5**], Costa–Sousa Neto [**CN**], Earp–Toubiana [**ET1**]–[**ET2**], Yu [**Yu1**]–[**Yu3**], Levi–Rossman [**LR**], Barbosa–Berard [**BB**], do Carmo–Gomes–Lawson–Thorbergsson–Silveira [**CGT, CL, CS**], and others. Regarding properties of the ends of embedded examples, Collin, Hauswirth and Rosenberg [**CHR1**] have recently shown that any embedded CMC-1 surface of finite total curvature is either a horosphere or all of its ends are asymptotic to catenoid cousin ends. In [**CHR1, Yu3**] it is further shown that any irregular end cannot be embedded, and the limit points of such an end are dense at infinity. Recently, Pacard and Pimentel [**PP**] established a method for attaching small handles between tangent horospheres and deforming to produce CMC-1 surfaces, and this construction produces many embedded CMC-1 surfaces of any genus. Also, Karcher [**Kar**] has recently constructed periodic CMC-1 surfaces with fundamental domains in several different types of compact quotients of H^3.

A typical example of an irregular end is the end of the Enneper cousin, a surface first constructed by Bryant [**Bry**]. After that the last two authors [**UY2**] constructed examples of genus zero and two irregular ends, and also many reducible CMC-1 surfaces of genus zero whose ends are all irregular, using deformations from minimal surfaces. (The conclusion of Remark 4.4 in [**UY2**] contains an error. The number of ends should be $ml+2$, and hence the genus of $\hat{M}_0^*$ is zero.) Recently, Daniel [**D**] has investigated irregular ends from the viewpoint of Nevanlinna theory.

After [**UY2**], no further surfaces with irregular ends and finite total curvature had been constructed. (However, such an example with infinite total curvature can be found in [**RUY3**].) In particular, until now no irreducible CMC-1 surfaces with irregular ends and finite total curvature had been known.

The purpose of the paper is to construct countably many 1-parameter families of genus zero CMC-1 surfaces with irregular ends and finite total curvature, which have either dihedral or Platonic symmetries. We further show that examples with dihedral symmetries, and the simplest example with tetrahedral symmetry, must be irreducible. All of our examples have irregular ends of finite type in the sense of Daniel [**D**].

To do the construction, we start with the meromorphic data for the genus zero irreducible CMC-1 surfaces with regular ends found in [**UY3**] and [**RUY1**] and modify this data to make surfaces with irregular ends. The spirit of the construction is similar to the construction of trinoids in [**UY3**], where CMC-1 surfaces with prescribed Gauss maps are constructed by reflecting spherical triangles, and we use monodromy killing arguments as in [**RUY1**] and [**UY6**], but the techniques are brought to bear more intricately here.

In Section 1 we give necessary preliminaries. As our construction is done by reflecting abstract spherical triangles, we discuss this in Section 2, and introduce a method to construct CMC-1 surfaces with irregular ends (Theorem 2.3), which is proved in Section 4. As an application of the theorem, we construct examples of genus zero with either dihedral or Platonic symmetries in Section 3. Finally, in Section 5, we construct a CMC-1 surface of genus 1 with four irregular ends, which is the first known example with positive genus whose ends are all irregular.

1. Preliminaries

Null meromorphic curves. Here we recall from [**UY1, UY7**] some fundamental properties of null meromorphic curves in $\mathrm{SL}(2,\mathbb{C})$.

DEFINITION 1.1. Let $F\colon M \to \mathrm{SL}(2,\mathbb{C})$ be a meromorphic map defined on a Riemann surface M with a local complex coordinate z. Then F is called *null* if $\det(F_z) \equiv 0$. (This condition does not depend on the choice of coordinate z.)

Let $F\colon M \to \mathrm{SL}(2,\mathbb{C})$ be a null meromorphic map. We define a matrix α by

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} := F^{-1}dF,$$

and set

$$g := \alpha_{11}/\alpha_{21}, \qquad \omega := \alpha_{21}. \tag{1.1}$$

Then the pair (g,ω) is a meromorphic function g and a holomorphic 1-form ω on M satisfying

$$F^{-1}dF = \begin{pmatrix} g & -g^2 \\ 1 & -g \end{pmatrix} \omega. \tag{1.2}$$

Conversely, let g be a meromorphic function and ω a meromorphic 1-form on M. Then the ordinary differential equation (1.2) is integrable and the solution F is a null map into $\mathrm{SL}(2,\mathbb{C})$ (since we will always choose the initial condition to be in $\mathrm{SL}(2,\mathbb{C})$) defined on the universal covering of $M \setminus \{\text{poles of } \alpha\}$. In general, F might not be single-valued on M itself, and F may have essential singularities at poles of α. We call the pair (g,ω) the *Weierstrass data* of F.

DEFINITION 1.2. Let

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}$$

be a null meromorphic map of M into $\mathrm{SL}(2,\mathbb{C})$ with Weierstrass data (g,ω). We call

$$G := \frac{dF_{11}}{dF_{21}} = \frac{dF_{12}}{dF_{22}}$$

the *hyperbolic Gauss map* of F. Furthermore, we call g in (1.1) the *secondary Gauss map* and $Q = \omega \cdot dg$ the *Hopf differential* of F.

We remark that the secondary Gauss map g satisfies

$$g = -\frac{dF_{22}}{dF_{21}} = -\frac{dF_{12}}{dF_{11}}.$$

Let $F\colon M \to \mathrm{SL}(2,\mathbb{C})$ be a null meromorphic map. Then for $a, b \in \mathrm{SL}(2,\mathbb{C})$, $\widetilde{F} = aFb^{-1}$ is also a null meromorphic map. Then the associated two Gauss maps $\widetilde{G}$, $\widetilde{g}$, and the Hopf differential $\widetilde{Q}$ of $\widetilde{F}$ are

$$\widetilde{G} = a \star G, \quad \widetilde{g} = b \star g, \quad \text{and} \quad \widetilde{Q} = Q, \tag{1.3}$$

where, for any matrix $a = (a_{ij}) \in \mathrm{SL}(2,\mathbb{C})$ and any function G, $a \star G$ is the Möbius transformation of G by a:

$$a \star G = \frac{a_{11}G + a_{12}}{a_{21}G + a_{22}}. \tag{1.4}$$

Let z be a complex coordinate on a neighborhood U of M. Now we consider the Schwarzian derivatives $S(G)$ and $S(g)$ on U of G and g, where

$$S(G) = \left(\left(\frac{G''}{G'}\right)' - \frac{1}{2}\left(\frac{G''}{G'}\right)^2 \right) dz^2, \qquad \text{with } {}' = \frac{d}{dz}. \tag{1.5}$$

The description of the Schwarzian derivative depends on the choice of complex coordinate z. However, any difference of two Schwarzian derivatives, as a meromorphic 2-differential, is independent of the choice of complex coordinate. Note that the Schwarzian derivative is invariant under Möbius transformations:

$$S(G) = S(a \star G) \qquad \big(a \in \mathrm{SL}(2,\mathbb{C})\big). \tag{1.6}$$

The following identity can be checked:

$$S(g) - S(G) = 2Q. \tag{1.7}$$

Conversely, the following lemma holds:

LEMMA 1.3 ([**Sm, UY3, KUY**]). *Let (G, g) be a pair of meromorphic functions on M such that $S(g) - S(G)$ is not identically zero. Then there exists a unique (up to sign) null meromorphic map $F\colon M \to \mathrm{SL}(2,\mathbb{C})$ such that G and g are the hyperbolic Gauss map and the secondary Gauss map of F.*

Now, for later use, we point out the following elementary fact from linear algebra:

LEMMA 1.4 ([**RUY1**]). *A matrix $a \in \mathrm{SL}(2,\mathbb{C})$ satisfies $a\bar{a} = \mathrm{id}$ if and only if it is of the form*

$$a = \begin{pmatrix} p & i\gamma_1 \\ i\gamma_2 & \bar{p} \end{pmatrix} \qquad \textit{with} \quad \gamma_1, \gamma_2 \in \mathbb{R}, \quad p\bar{p} + \gamma_1\gamma_2 = 1, \quad i = \sqrt{-1}.$$

CMC-1 surfaces in H^3. We identify the Minkowski 4-space L^4, which has the canonical Lorentzian metric $(\,\cdot\,,\,\cdot\,)$ of signature $(-,+,+,+)$, with the space of 2×2 hermitian matrices $\mathrm{Herm}(2)$. More explicitly, $(t, x_1, x_2, x_3) \in L^4$ is identified with the matrix

$$\begin{pmatrix} t + x_3 & x_1 + ix_2 \\ x_1 - ix_2 & t - x_3 \end{pmatrix} \in \mathrm{Herm}(2).$$

The hyperbolic 3-space can be defined as the upper component

$$H^3 = \{\xi = (t, x_1, x_2, x_3) \in L^4 \,|\, (\xi, \xi) = -1\,,\ t > 0\,\}$$

of the hyperboloid in L^4 with the induced metric. In Herm(2) this is represented as

$$H^3 = \{\, X \in \mathrm{Herm}(2)\,;\, \det X = 1\,,\ \mathrm{trace}\, X > 0\,\} = \{aa^*\,;\, a \in \mathrm{SL}(2,\mathbb{C})\},$$

where $a^* = {}^t\bar{a}$. The complex Lie group $\mathrm{SL}(2,\mathbb{C})$ acts isometrically on H^3 by $\rho(a)x = axa^*$, where $a \in \mathrm{SL}(2,\mathbb{C})$ and $x \in H^3$.

Let M be a Riemann surface and $F\colon M \to \mathrm{SL}(2,\mathbb{C})$ a null holomorphic immersion. Then $f = FF^*\colon M \to H^3$ is a conformal CMC-1 immersion. Conversely, for any conformal CMC-1 immersion $f\colon M \to H^3$, there exists a null holomorphic immersion $F\colon \widetilde{M} \to \mathrm{SL}(2,\mathbb{C})$ such that $f = FF^*$, where $\widetilde{M}$ denotes the universal covering of M. We call F a *lift* of the conformal CMC-1 immersion f. Let $\widetilde{F}$ be another lift of f. Then, we have the expression $\widetilde{F} = Fb^{-1}$ for some matrix $b \in \mathrm{SU}(2)$. Let (g,ω) be the Weierstrass data of the null map of the lift F. Then the first fundamental form ds^2 and the second fundamental form $I\!I$ are given by (see [**UY1**], for example)

$$\begin{aligned} ds^2 &= (1+|g|^2)^2\omega\cdot\bar{\omega}, \\ I\!I &= -Q - \overline{Q} + ds^2, \end{aligned} \tag{1.8}$$

where $Q = \omega \cdot dg$ is the Hopf differential of F.

Let $f = FF^*\colon M \to H^3$ be a complete conformal CMC-1 immersion whose total Gaussian curvature is finite. Since the Gaussian curvature K of CMC-1 surface is nonnegative, finiteness of the total Gaussian curvature is equivalent to

$$\int_M (-K)\, dA < \infty,$$

where dA is the area element with respect to the first fundamental form. Then there is a compact Riemann surface $\overline{M}$ and a finite number of points $\{p_1, \ldots, p_n\} \in \overline{M}$ such that $M = \overline{M} \setminus \{p_1, \ldots, p_n\}$. We call each p_j an *end* of f. The hyperbolic Gauss map G on M does not necessarily extend meromorphically on $\overline{M}$. The end p_j is called a *regular* end if p_j is at most a pole singularity of G, and otherwise is called an *irregular* end. Namely, an irregular end is an end at which the hyperbolic Gauss map has an essential singularity.

On the other hand, the Hopf differential Q can always be extended as a meromorphic 2-differential on $\overline{M}$. We denote by $\mathrm{ord}_p Q$ the order of the first nonvanishing term of the Laurent expansion of the Hopf differential Q at $p \in \overline{M}$. (By definition, $\mathrm{ord}_p Q > 0$ at zeros of Q and $\mathrm{ord}_p Q < 0$ at poles of Q.) The following lemma is well-known (cf. [**Bry**], Lemma 2.3 of [**UY1**]).

LEMMA 1.5. *An end p_j is regular if and only if* $\mathrm{ord}_{p_j} Q \geq -2$.

Now we set $d\sigma_f = (-K)ds^2$. Then $d\sigma_f^2$ is a pseudometric of constant curvature 1 with conical singularities (see the appendix of this paper, and also Proposition 4 of [**Bry**]). It follows from (1.8) and the Gauss equation that

$$d\sigma_f^2 = \frac{4\, dg\cdot d\bar{g}}{(1+|g|^2)^2}. \tag{1.9}$$

Hence $d\sigma_f^2$ is the pull-back of the Fubini–Study metric $d\sigma_0^2$ on $\mathbb{CP}^1$ induced by the secondary Gauss map $g\colon \widetilde{M}^2 \to \mathbb{C} \cup \{\infty\} = \mathbb{CP}^1$. By (1.8) and (1.9) we have

$$ds^2 \cdot d\sigma_f^2 = 4\, Q \cdot \overline{Q}. \tag{1.10}$$

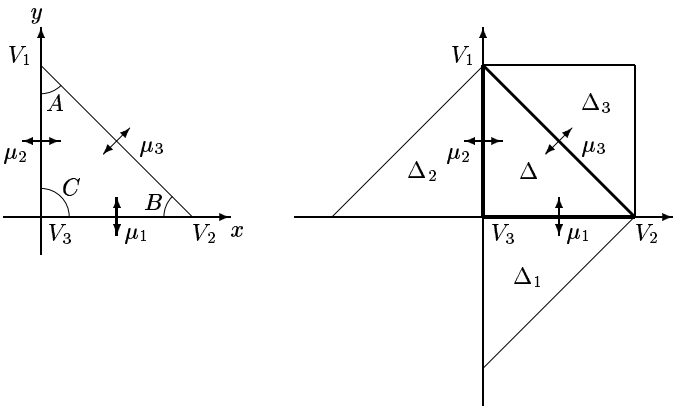


FIGURE 1. The triangle Δ and its reflections.

In addition to having conical singularities at the ends p_j, the pseudometric $d\sigma_f^2$ also has a conical singularity at each umbilic point $q \in M$ of f. The conical order of $d\sigma_f^2$ at each point is defined in the appendix of this paper. Since ds^2 is positive definite at q, we have the following:

LEMMA 1.6. *At umbilic points, the conical order of $d\sigma_f^2$ equals the order of Q.*

2. Reflections of an Abstract Spherical Triangle

In this section, we introduce a method to construct CMC-1 surfaces with irregular ends. In [**RUY1**], examples with regular ends are constructed from the holomorphic data G (the hyperbolic Gauss map) and Q (the Hopf differential). However, since the hyperbolic Gauss map has essential singularities at irregular ends, it is hard to find an explicit expression of G in our case. Thus, our construction is based on the secondary Gauss map g and the Hopf differential Q. Though g is not a well-defined meromorphic function on the surface, the pseudometric $d\sigma_f^2$ as in (1.9) is a spherical metric with conical singularities which is well-defined on the surface. So, we start by constructing a spherical metric $d\sigma_f^2$ with conical singularities, using reflections of spherical triangles (see [**UY3**]).

Abstract spherical triangles. In this section, we consider abstract spherical triangles and their extensions by reflection. First, we shall define abstract triangles. We set

$$\Delta := \{x + iy \in \mathbb{C}\,;\, x \geq 0,\, y \geq 0,\, x + y \leq 1\},$$

and label each vertex V_1, V_2, V_3 of this closed triangular region Δ as in Figure 1. An *abstract spherical triangle* is a pair $(\Delta, d\sigma^2)$, where $d\sigma^2$ is a Riemannian metric defined on Δ with constant curvature 1 such that the three edges forming the boundary $\partial\Delta$ are geodesics. Let A, B and C be the interior angles of Δ at the vertices V_1, V_2, and V_3 with respect to the metric $d\sigma^2$ respectively. The angles A, B and C are positive, but may take any positive values, including those greater than or equal to π. Moreover, if $A, B, C \notin \pi\mathbb{Z}$, then we call the abstract spherical triangle $(\Delta, d\sigma^2)$ *nondegenerate*. The following fact is known:

LEMMA 2.1 ([**UY7**]). *Let $(\Delta, d\sigma^2)$ be a nondegenerate abstract spherical triangle, then the three angles A, B, C satisfy the inequality*

$$\cos^2 A + \cos^2 B + \cos^2 C + 2\cos A \cos B \cos C < 1. \tag{2.1}$$

Conversely, if a triple of positive real numbers (A, B, C) satisfies (2.1), then there exists a unique nondegenerate abstract spherical triangle such that the angles at V_1, V_2 and V_3 are A, B and C, respectively.

PROOF. We take a double (two identical copies) of $(\Delta, d\sigma^2)$ and glue them along their corresponding vertices and edges. Then we get a conformal pseudometric on S^2 with three conical singularities with conical angles $2A$, $2B$, $2C$. Consequently, the conical orders are

$$\frac{A-\pi}{\pi}, \quad \frac{B-\pi}{\pi}, \quad \frac{C-\pi}{\pi},$$

respectively. If $(\Delta, d\sigma^2)$ is nondegenerate, then $A, B, C \notin \pi\mathbb{Z}$. By Corollary 3.2 of [**UY7**], the metric on S^2 is irreducible. Then (2.1) follows from (2.19) of [**UY7**].

Conversely, suppose that (2.1) holds. By Theorem 2.4 of [**UY7**], there exists a unique conformal pseudometric on S^2 with three conical singularities with conical angles $2A, 2B, 2C \notin 2\pi\mathbb{Z}$. The uniqueness of this pseudometric implies that it has a symmetry and can be considered as a gluing of two identical nondegenerate spherical triangles. □

(Note: Recently, an alternative proof and a geometric explanation of this lemma were given in [**FH**] and [**F**].)

The preceding lemma implies that a nondegenerate abstract spherical triangle is uniquely determined by its angles A, B, C. So we denote it by

$$\mathcal{T}(A, B, C) := (\Delta, d\sigma^2).$$

Now we fix a nondegenerate abstract spherical triangle $\mathcal{T}(A, B, C)$. Since Δ is simply connected, there exists a meromorphic function $g\colon \Delta \to \mathbb{C} \cup \{\infty\}$ such that the pull-back of the Fubini–Study metric on $\mathbb{C} \cup \{\infty\} = \mathbb{CP}^1$ by g is $d\sigma^2$. However, such a choice of the developing map has an ambiguity up to an $\mathrm{SU}(2)$-matrix action $g \mapsto a \star g$ $(a \in \mathrm{SU}(2))$. We shall now remove this ambiguity, using a normalization: There exists a unique (up to sign) developing map

$$g = g_{A,B,C} : \Delta \longrightarrow \mathbb{C} \cup \{\infty\}$$

of $d\sigma^2$ satisfying (see Figure 1)

$$e^{-iC} g(V_1) \in \mathbb{R} \cup \{\infty\}, \quad g(V_2) \in \mathbb{R} \cup \{\infty\} \quad \text{and} \quad g(V_3) = 0. \tag{2.2}$$

We call the developing map $g = g_{A,B,C}$ the *normalized developing map* of the triangle $\mathcal{T}(A, B, C)$.

Let μ_j $(j = 1, 2, 3)$ be the reflections of the triangle $\mathcal{T}(A, B, C)$ across the three edges, as in Figure 1. Let Δ_j be the closed domain obtained by reflecting Δ with respect to μ_j (see Figure 1). Then each reflection μ_j can be regarded as an involution on the domain $\Delta \cup \Delta_j$.

LEMMA 2.2 (Monodromy principle). *Let $\mathcal{T}(A, B, C)$ be a nondegenerate abstract spherical triangle and $g_{A,B,C} : \Delta \to \mathbb{C} \cup \{\infty\}$ $(j = 1, 2, 3)$ be the normalized*

developing map of $\mathcal{T}(A,B,C)$. *Then the following identities hold*:

$$\overline{g_{A,B,C}\circ\mu_1} = g_{A,B,C},$$
$$\overline{g_{A,B,C}\circ\mu_2} = e^{-2iC} g_{A,B,C},$$
$$\overline{g_{A,B,C}\circ\mu_3} = \begin{pmatrix} q & i\delta \\ i\delta & \bar{q} \end{pmatrix} \star g_{A,B,C} \qquad (\delta\in\mathbb{R}, \quad q\bar{q}+\delta^2 = 1),$$

where

$$q = \frac{i}{\sin C}(\cos A + e^{iC}\cos B). \tag{2.3}$$

PROOF. To simplify the notation, we set $g = g_{A,B,C}$. Let γ_1, γ_2 and γ_3 be the three edges of Δ which are stabilized by the reflections μ_1, μ_2 and μ_3, respectively. Since the edge γ_1 is a geodesic, (2.2) implies that $g(\gamma_1)$ lies on the real axis. Then by the reflection principle, $\overline{g\circ\mu_1} = g$ holds. Similarly, by (2.2), $e^{-iC}g(\gamma_2)$ is real. Hence

$$\overline{e^{-iC}g\circ\mu_2} = e^{-iC}g$$

holds. Then we have the second assertion.

There exists a rotation a centered at $g(V_2)$ of the unit 2-sphere $S^2(=\mathbb{C}\cup\{\infty\})$ such that the image of $g(V_1)$ is real. Such an isometry a of S^2 can be represented as a matrix $a\in \mathrm{SU}(2)$. Then we have $a\star g(\gamma_3)$ lies on the real axis. Hence by the reflection principle, $\overline{a\star g\circ\mu_3} = a\star g$ holds, and then we have

$$\overline{g\circ\mu_3} = (\bar{a}^{-1}a)\star g.$$

In particular, we have

$$\overline{g\circ\mu_1} = g, \quad \overline{g\circ\mu_2} = e^{-2iC}g, \quad \overline{g\circ\mu_3} = (\bar{a}^{-1}a)\star g. \tag{2.4}$$

Now we glue $\mathcal{T}(A,B,C)$ to a double of $\mathcal{T}(A,B,C)$ along corresponding edges and vertices, giving us a metric constant curvature one with three conical singularities on S^2 with conical angles $2A$, $2B$, and $2C$, just as in the proof of Lemma 2.1. The open domains Δ_1, Δ_2 and Δ_3 can be regarded as the interior of the second triangle in the double of $\mathcal{T}(A,B,C)$. The monodromy of reflections for such metrics on S^2 are determined in [**UY7**]. Then, as shown at the bottom of page 82 of [**UY7**], we have

$$\bar{a}^{-1}a = \begin{pmatrix} q & i\delta \\ i\delta & \bar{q} \end{pmatrix}, \qquad (\delta\in\mathbb{R},\ q\bar{q}+\delta^2 = 1)$$

with q as in (2.3). □

Closed Riemann surfaces generated by three reflections. Let $\overline{M}$ be a closed Riemann surface and $D(\subset \overline{M})$ a simply connected closed domain bounded by three real analytic curves γ_1, γ_2 and γ_3. We label the vertices V_1, V_2, V_3 of this closed triangular region D such that γ_1, γ_2, and γ_3 correspond to the three edges V_2V_3, V_3V_1 and V_1V_2, respectively. The Riemann surface $\overline{M}$ is generated by D if there are three anti-holomorphic reflections μ_1, μ_2, μ_3 of $\overline{M}$ stabilizing the three edges γ_1, γ_2, and γ_3 of D such that any point of $\overline{M}$ can be contained in the image of D by a suitable finite composition of these three reflections (see Figure 2). In this case, D is called a *fundamental domain* of $\overline{M}$. A meromorphic 2-differential Q on $\overline{M}$ is said to be symmetric with respect to D if

$$\overline{Q\circ\mu_j} = Q \qquad (j = 1,2,3)$$

holds.

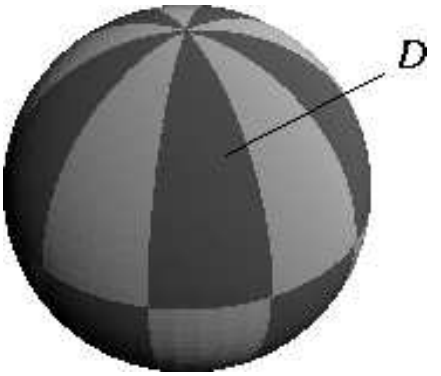


FIGURE 2. A Riemann surface generated by reflections: Here the Riemann surface is the sphere S^2, and is obtained from the triangle D by reflections.

We let $\mathrm{Met}_1(\overline{M})$ denote the set of conformal pseudometrics with conical singularities on $\overline{M}$. A metric $d\sigma^2 \in \mathrm{Met}_1(\overline{M})$ is called *symmetric* with respect to the fundamental domain D if it is invariant under the three reflections μ_1, μ_2 and μ_3. Moreover, if the restriction $(D, d\sigma^2|_D)$ is isometric to $\mathcal{T}(A,B,C)$, we denote the metric by

$$d\sigma^2 = d\sigma^2_{A,B,C}.$$

A meromorphic function g on $\overline{M}$ is called SU(2)-*symmetric* with respect to D if

$$d\sigma_g^2 := \frac{4\,dg \cdot d\bar{g}}{(1+|g|^2)^2} \tag{2.5}$$

is symmetric with respect to D.

The following is the main theorem in this paper:

THEOREM 2.3. *Let $\overline{M}$ be a closed Riemann surface which is generated by a triangular fundamental domain $D \subset \overline{M}$ by reflections, and label the vertices of D as V_1, V_2 and V_3. Let g be an* SU(2)*-symmetric meromorphic function on $\overline{M}$ with respect to D, and let Q be a symmetric meromorphic 2-differential on $\overline{M}$ with respect to D. Suppose that:*

(1) *There exist $A, B_0 \in \mathbb{R}^+ \setminus \pi\mathbb{Z}$ such that $d\sigma_g^2 = d\sigma^2_{A,B_0,\pi/2}$, with $d\sigma_g^2$ as in (2.5).*
(2) *Q is holomorphic on $D \setminus \{V_2\}$ and has a pole at V_2 with $\mathrm{ord}_{V_2} Q \le -3$.*
(3) *The branch point set of g outside the poles of Q equals the zero set of Q, and at each zero of Q, the order of Q is equal to the conical order of $d\sigma_g^2$.*

Let $p_1, \ldots, p_n \in \overline{M}$ be the set of poles of Q. Then, for some $\varepsilon > 0$, there exist a smooth function $B(t)\colon (-\varepsilon, \varepsilon) \to \mathbb{R}$ satisfying $B(0) = B_0$ and a 1-parameter family of conformal CMC-1 *immersions $f_t : \overline{M} \setminus \{p_1, \ldots, p_n\} \to H^3$ for $t \in (-\varepsilon, \varepsilon)$ with the following properties:*

(i) *The Hopf differential of f_t is tQ and $d\sigma^2_{f_t} = d\sigma_{A,B(t),\pi/2}$.*
(ii) *f_t has irregular ends at $\{p_1, \ldots, p_n\}$.*
(iii) *f_t is symmetric with respect to D. That is, the image of f_t is generated by the reflections across the edges of $f_t(D)$.*

REMARK 2.4. The construction method for proving Theorem 2.3 will still work if $\operatorname{ord}_{V_2} Q = -2$. The stronger assumption $\operatorname{ord}_{V_2} Q \leq -3$ in (2) is required only to make the ends irregular (see Lemma 1.5).

The next theorem gives conditions which imply the surfaces f_t in Theorem 2.3 are irreducible.

THEOREM 2.5. *Under the assumptions in Theorem* 2.3, *if* $A \not\equiv \pi/2 \pmod{\pi}$ *and*

$$\oint_\tau \begin{pmatrix} g & -g^2 \\ 1 & -g \end{pmatrix} \frac{Q}{dg} \neq 0 \tag{2.6}$$

for a local loop τ *about* V_2, *then* f_t *is irreducible for* t *sufficiently close to zero.*

The proofs of these theorems are given in Section 4.

3. CMC-1 Surfaces with Dihedral and Platonic Symmetries

In this section, we construct examples of finite total curvature CMC-1 surfaces with irregular ends and either dihedral or Platonic symmetries. Examples with dihedral symmetries and the simplest example with tetrahedral symmetry (the case $m = 1$ in (3.1) and (3.2)) are irreducible. (Though we expect all the other examples to be irreducible, we have not checked them yet.)

CMC-1 surfaces with dihedral symmetries. Let $n \geq 3$ be an integer and

$$M_n := \mathbb{C} \cup \{\infty\} \setminus \{1, \zeta, \dots, \zeta^{n-1}\} \qquad \left(\zeta = \exp\frac{2\pi}{n} i\right).$$

The Jorge–Meeks' n-noid is the complete minimal immersion $f_{n,0}\colon M_n \to \mathbb{R}^3$ given by the Weierstrass representation as

$$f_{n,0} := \operatorname{Re} \int \left((1 - g_{n,0}^2), i(1 + g_{n,0}^2), 2g_{n,0}\right) \frac{Q_{n,0}}{dg_{n,0}},$$

$$\text{where} \qquad g_{n,0} = z^{n-1}, \quad Q_{n,0} = \frac{z^{n-2}}{(z^n - 1)^2}\, dz^2.$$

The $\mathbb{Z}_2$ extension $D_n \times \mathbb{Z}_2$ of the dihedral group D_n acts isometrically on the image of $f_{n,0}$.

There exists a one-parameter family of corresponding CMC-1 immersions of M_n to H^3 such that the hyperbolic Gauss map is $g_{n,0}$ and the Hopf differential is proportional to $Q_{n,0}$ (see [**UY3, RUY1**], and Figure 4 for the $n = 3$ case). Since $\operatorname{ord}_{\zeta^j} Q_{n,0} = -2$ $(j = 0, \dots, n-1)$, the ends of these corresponding surfaces are regular.

However, as we wish to produce CMC-1 surfaces in H^3 whose ends are *not* regular, we now modify our choices for Q and g to accomplish this: Let $m \geq 1$ be an integer and set

$$Q_{n,m} := \frac{z^{n(m+1)-2}}{(z^n - 1)^{2(m+1)}}\, dz^2, \qquad g_{n,m} := z^{n(m+1)-1}.$$

Then $g = g_{n,m}$ and $Q = Q_{n,m}$ are a meromorphic function and a meromorphic 2-differential on $\overline{M} = \mathbb{C} \cup \{\infty\}$ respectively, which are symmetric with respect to the fundamental domain Ω_n as in Figure 3. Moreover, (g, Q) satisfies the assumptions (1)–(3) of Theorem 2.3 for $A = \pi(m+1) - \pi/n$ and $B_0 = \pi/2$. Then, for each n

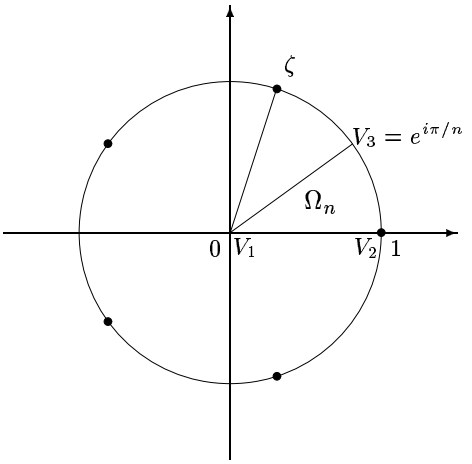


FIGURE 3. The fundamental domain of surfaces with dihedral symmetry.

and m, there exists a one-parameter family of CMC-1 immersions $f_{n,m,t}\colon M_n \to H^3$ $(0 < |t| < \varepsilon)$ with symmetry group $D_n \times \mathbb{Z}_2$ whose Hopf differential is $tQ_{n,m}$ (see Figure 4). The total Gaussian curvature of these surfaces will be approximately $4\pi(n(m+1)-1)$.

Let τ be a loop surrounding the end $V_2 = 1$. Since

$$\frac{Q_{n,m}}{dg_{n,m}} = \left(\frac{1}{n(m+1)-1}\right)\frac{dz}{(z^n-1)^{2(m+1)}},$$

we have

$$\oint_\tau \frac{Q_{n,m}}{dg_{n,m}} = \left(\frac{2\pi i}{n(m+1)-1}\right)\operatorname*{Res}_{z=1}\frac{1}{(z^n-1)^{2(m+1)}} \neq 0.$$

Thus by Theorem 2.5, the surfaces are irreducible for sufficiently small t.

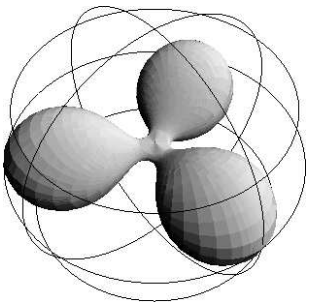
CMC-1 surfaces with tetrahedral symmetries. It is well known that there exists a minimal immersion

$$f_0\colon M := \mathbb{C}\cup\{\infty\}\setminus\{p_1,\dots,p_4\} \longrightarrow \mathbb{R}^3$$

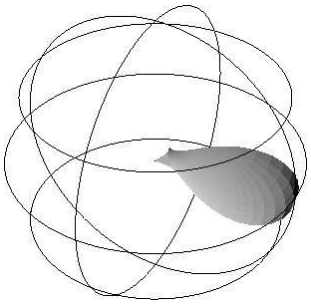
with four catenoid ends and tetrahedral symmetry [**Kat, Xu, BR, UY3**] and corresponding CMC-1 surfaces in H^3 [**UY3, RUY1**] with regular ends.

We denote by Q_0 and g_0 the Hopf differential and the Gauss map of f_0 respectively. Since each end is asymptotic to a catenoid, $\operatorname{ord}_{p_j} Q_0 = -2$ $(j = 1,\dots,4)$. Then there exists four umbilic points (zeros of Q_0) $q_1,\dots,q_4$ such that $\operatorname{ord}_{q_j} Q_0 = 1$ $(j = 1,\dots,4)$. The Gauss map g_0 is a meromorphic function on $\mathbb{C}\cup\{\infty\}$ whose branch points are $\{q_1,\dots,q_4\}$ each with branch order 1. Moreover, M is obtained by reflections across the edges of the fundamental domain D, which is a triangle with vertices $V_1 = q_1$, $V_2 = p_1$, V_3. (See Figure 5. See also the construction in [**UY3**].) The Hopf differential Q_0 and the Gauss map g_0 are symmetric with respect to the fundamental domain D.

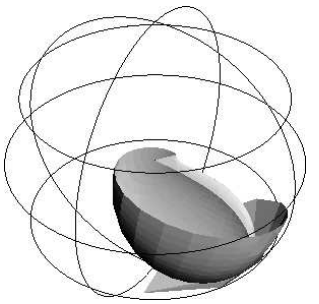
Consider the Schwarzian derivative $S(g_0)$ of g_0, as in (1.5), where z is the usual complex coordinate of $\mathbb{C}\cup\{\infty\}$. Then $S(g_0)$ is a meromorphic 2-differential on $\mathbb{C}\cup\{\infty\}$. Moreover, since g_0 is symmetric, $S(g_0)$ is invariant under reflections about the edges of D.



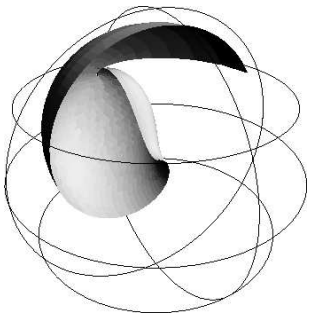
(a) $f_{3,0,t}$



(b) Fundamental piece of $f_{3,0,t}$



(c) Fundamental piece of $f_{3,1,t}$



(d) Enneper's cousin

FIGURE 4. Surfaces with dihedral symmetry: Figure (a) shows a CMC-1 surface with dihedral symmetry and three regular ends (the surface corresponding to a Jorge–Meeks surface) in the Poincaré model of H^3, and figure (b) is the fundamental region of the surface in figure (a). Figure (c) shows the fundamental piece of $f_{3,1,t}$, a surface of dihedral symmetry with three irregular ends. The central part of $f_{3,1,t}$ is similar to that of the hyperbolic correspondence of a Jorge–Meeks surface. The end of $f_{3,1,t}$ seen here is similar to the end of an Enneper cousin [**Bry**], which is shown in figure (d).

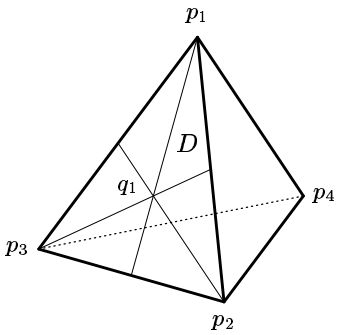


FIGURE 5. The fundamental domain of surfaces with tetrahedral symmetry.

The branch points of g_0 are poles of $S(g_0)$, and each pole of $S(g_0)$ has order -2. So $S(g_0)$ has four poles of order 2 at the q_j $(j = 1, \dots, 4)$ and is holomorphic on $\mathbb{C} \cup \{\infty\} \setminus \{q_1, \dots, q_4\}$. Since the total order of a meromorphic 2-differential on $\mathbb{C} \cup \{\infty\}$ is -4, $S(g_0)$ has four zeros (counting multiplicity). If there exists a zero of $S(g_0)$ on the interior of D, $S(g_0)$ has at least 24 zeros because $\mathbb{C} \cup \{\infty\}$ consists of 24 copies of the fundamental region D, which is impossible. Similarly, if a zero of $S(g_0)$ lies on the interior of a edge of D, $S(g_0)$ has at least 12 zeros, which is also impossible. If the vertex V_3 of D is a zero of $S(g_0)$, there exist six zeros, which is again impossible. Since $V_1 = q_1$ is a pole of $S(g_0)$, the set of zeros of $S(g_0)$ must be $\{p_1, \dots, p_4\}$, and $\operatorname{ord}_{p_j} S(g_0) = 1$ for $j = 1, \dots, 4$.

For an integer $m \geq 1$, we define a meromorphic 2-differential Q_m as

$$Q_m := \frac{{Q_0}^{m+1}}{S(g_0)^m}, \tag{3.1}$$

where ${Q_0}^{m+1}$ (resp. $S(g_0)^m$) is the symmetric product of $m+1$ copies of Q_0 (resp. m-copies of $S(g_0)$). Since the poles of Q_0 and the zeros of $S(g_0)$ are $\{p_1, \dots, p_4\}$, Q_m has the poles $\{p_1, \dots, p_4\}$ and is holomorphic on M. More precisely,

$$\operatorname{ord}_{p_j} Q_m = -3m - 2 \qquad \text{and} \qquad \operatorname{ord}_{q_j} Q_m = 3m + 1 \qquad (j = 1, \dots, 4)$$

hold. Since Q_0 and $S(g_0)$ are invariant under the reflections, so is Q_m.

Consider an abstract spherical triangle $\mathcal{T}(A, B_0, C)$ with

$$A = \pi m + \frac{2}{3}\pi, \quad B_0 = \frac{\pi}{3}, \quad C = \frac{\pi}{2}, \tag{3.2}$$

and identify it with the fundamental domain D. Then we have a pseudometric $d\sigma^2_{A,B_0,C} \in \operatorname{Met}_1(\mathbb{C} \cup \{\infty\})$. Since neighborhoods of V_1, V_2 and V_3 are generated by six, six and four copies of the fundamental domain D, respectively, $d\sigma^2_{A,B_0,C}$ is a pseudometric whose conical orders are all integers. Hence by Proposition A.1 in Appendix, the developing map g_m of $d\sigma^2_{A,B_0,C}$ is meromorphic on $\mathbb{C} \cup \{\infty\}$.

Then one can easily check that (g_m, Q_m) satisfies the assumptions (1)–(3) of Theorem 2.3. Hence for each m, there exists a one-parameter family of CMC-1 immersions $\{f_{m,t}\,;\, 0 < |t| < \varepsilon\}$ of M into H^3 with irregular ends and tetrahedral symmetry.

Finally, we check irreducibility for $m = 1$. We may set

$$M = \mathbb{C} \cup \{\infty\} \setminus \{1, \zeta, \zeta^2, \infty\}, \qquad \text{where} \quad \zeta = \exp\frac{2}{3}\pi i,$$

and

$$g_0 = \frac{1}{3\sqrt{2}}\left(z - \frac{4}{z^3}\right), \qquad Q_0 = \frac{z(z^3+8)}{(z^3-1)^2}\,dz^2$$

(see page 221 of [**UY3**]). Hence the umbilic points are $\{0, -2, -2\zeta, -2\zeta^2\}$. By direct calculation, we have

$$Q_1 = \frac{{Q_0}^2}{S(g_0)} = \frac{1}{96}\frac{z^4(z^3+8)^4}{(z^3-1)^5}\,dz^2.$$

On the other hand, g_1 is a meromorphic function which branches at the umbilic points with branch order 4. Then we have $\deg g_1 = 9$ by the Riemann–Hurwitz formula. Choose a rotation $a \in \mathrm{SU}(2)$ such that $a \star g_1(q_1) = \infty$, where we set $q_1 = 0$. Then q_j $(j = 2, 3, 4)$ are not poles of $g := a \star g_1$, because the multiplicity of g at q_j is 5 for each j and $\deg g = 9$. Moreover, $d\sigma^2_{g_1} = d\sigma^2_g$ has a conical singularity

Symmetry	$\#\{p_j\}$	$\#\{q_j\}$	$\mathrm{ord}_{q_j} Q_0$	$\mathrm{ord}_{p_j} Q_m$	$\mathrm{ord}_{q_j} Q_m$	A	B_0
Dihedral	n	2	$n-2$	$2(m+1)$	$n(m+1)-2$	$\pi\left(m+1-\frac{1}{n}\right)$	$\frac{\pi}{2}$
Tetrahedral	4	4	1	$-3m-2$	$3m+1$	$\pi\left(m+\frac{2}{3}\right)$	$\frac{\pi}{3}$
Octahedral	8	6	2	$-3m-2$	$4m+2$	$\pi\left(m+\frac{4}{3}\right)$	$\frac{\pi}{3}$
Octahedral	6	8	1	$-4m-2$	$3m+1$	$\pi\left(m+\frac{2}{3}\right)$	$\frac{\pi}{4}$
Icosahedral	20	12	3	$-3m-2$	$5m+3$	$\pi\left(m+\frac{4}{5}\right)$	$\frac{\pi}{3}$
Icosahedral	12	20	1	$-5m-2$	$3m+1$	$\pi\left(m+\frac{2}{3}\right)$	$\frac{\pi}{5}$

TABLE 1. Data for CMC-1 surfaces with Platonic symmetries.

at 0 with conical order 4. Hence, by symmetry, we have $g_1(\zeta z) = \zeta g_1(z)$, and we may write

$$dg = \beta \frac{(z^3+8)^4}{z^6(z^3-a^3)^2}\, dz$$

for some nonzero constants a and β. Such a function g exists if and only if all residues of the right-hand side vanish, which is equivalent to $a^3 = 16$. Then one can check irreducibility by direct calculation and Theorem 2.5.

CMC-1 surfaces with Platonic symmetries. There are genus zero minimal surfaces in $\mathbb{R}^3$ with catenoidal ends and the symmetry of any Platonic solid ([**Kat, Xu, BR, UY3**]). By similar arguments to the tetrahedral case above, one can obtain CMC-1 immersions with irregular ends and any Platonic symmetry. Table 1 shows the data for such surfaces.

4. Proof of the Main Theorem

PROOF OF THEOREM 2.3. *Step 1.* Take a real number $B \notin \pi\mathbb{Z}$ and let

$$\begin{aligned} &\rho_1 = \mathrm{id}, \qquad \rho_2 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \\ &\rho_3 = \rho_3(B) = \begin{pmatrix} q(B) & i\delta(B) \\ i\delta(B) & \bar{q}(B) \end{pmatrix} \qquad (\delta \in \mathbb{R},\ q\bar{q} + \delta^2 = 1), \end{aligned} \tag{4.1}$$

where

$$q(B) = i\cos A - \cos B. \tag{4.2}$$

Then we have

$$\bar{\rho}_j \rho_j = \mathrm{id} \qquad (j = 1, 2, 3).$$

Since $\mathcal{T}(A, B_0, \pi/2)$ is nondegenerate (by the assumption (1)), A, B_0 and $C = \pi/2$ satisfy the relation (2.1). Then for each B sufficiently close to B_0, there exists an abstract spherical triangle with angles A, B and $\pi/2$. We identify the domain $D \subset \overline{M}$ with this triangle. Then, by reflecting the metric, we have $d\sigma^2_{A,B,\pi/2} \in \mathrm{Met}_1(\overline{M})$. Let $M := \overline{M} \setminus \{p_1, \ldots, p_n\}$ and $\pi\colon \tilde{M} \to M$ be the universal covering.

By Proposition A.1 in the appendix of this paper, the developing map $\hat{g}_{A,B,\pi/2}$ of $d\sigma^2_{A,B,\pi/2}$ is defined on $\widetilde{M}$. To simplify the notation, we set

$$\hat{g}_B := \hat{g}_{A,B,\pi/2} \colon \widetilde{M} \longrightarrow \mathbb{C} \cup \{\infty\}.$$

Then by the monodromy principle (Lemma 2.2), we have

$$\overline{\hat{g}_B \circ \mu_j} = \rho_j \star \hat{g}_B \qquad (j = 1, 2, 3).$$

Step 2. One may regard the triangle $D \setminus \{V_2\}$ as generating $\widetilde{M}$ by the three reflections across its edges. We denote these reflections by $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$; that is, each $\hat{\mu}_j$ $(j = 1, 2, 3)$ is an antiholomorphic transformation on $\widetilde{M}$ which preserves the j'th edge of the triangle $D \setminus \{V_2\}$. We set

$$\hat{Q} := Q \circ \pi.$$

Let $F = F_{t,B}$ be a solution of the following ordinary differential equation on $\widetilde{M}$:

$$F^{-1} dF = t \begin{pmatrix} \hat{g}_B & -\hat{g}_B^2 \\ 1 & -\hat{g}_B \end{pmatrix} \frac{\hat{Q}}{d\hat{g}_B}, \qquad F(V_3) = \mathrm{id}. \tag{4.3}$$

Such a solution $F_{t,B}$ is uniquely determined on $\widetilde{M}$. By (4.3), the right-hand side of the ordinary differential equation is traceless and so $F_{t,B}$ takes values in $\mathrm{SL}(2, \mathbb{C})$.

Then $\overline{F_{t,B} \circ \hat{\mu}_j}$ $(j = 1, 2, 3)$ has the Hopf differential $\hat{Q} = \overline{\hat{Q} \circ \hat{\mu}_j}$ and the secondary Gauss map satisfies $\overline{\hat{g}_B \circ \hat{\mu}_j} = \rho_j \star \hat{g}_B$. However, $F_{t,B}\, \rho_j^{-1}$ also has the Hopf differential $\hat{Q}$ and secondary Gauss map $\rho_j \star \hat{g}_B$, by (1.3). Thus, by (4.3), we have

$$\left(\overline{F_{t,B} \circ \hat{\mu}_j}\right)^{-1} d\left(\overline{F_{t,B} \circ \hat{\mu}_j}\right) = \left(F_{t,B}\, \rho_j^{-1}\right)^{-1} d\left(F_{t,B}\, \rho_j^{-1}\right) = \rho_j \begin{pmatrix} \hat{g}_B & -\hat{g}_B^2 \\ 1 & -\hat{g}_B \end{pmatrix} \frac{\hat{Q}}{d\hat{g}_B}\, \rho_j^{-1}.$$

This implies that $\overline{F_{t,B} \circ \hat{\mu}_j}$ and $F_{t,B}\, \rho_j^{-1}$ are both solutions of the same ordinary differential equation, and thus they differ only by the choice of initial values at the base point V_3. So there exists a matrix $\sigma_j(t, B) \in \mathrm{SL}(2, \mathbb{C})$ such that

$$\overline{F_{t,B} \circ \hat{\mu}_j} = \sigma_j(t, B)\, F_{t,B}\, \rho_j^{-1} \qquad (j = 1, 2, 3). \tag{4.4}$$

Since $\overline{F_{t,B} \circ \hat{\mu}_j(V_3)} = F(V_3) = \mathrm{id}$ for $j = 1, 2$, we have

$$\mathrm{id} = \sigma_j(t, B)\, \rho_j^{-1} \qquad (j = 1, 2).$$

Thus

$$\sigma_1(t, B) = \rho_1 = \mathrm{id}, \qquad \sigma_2(t, B) = \rho_2 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \tag{4.5}$$

In particular, the matrices $\sigma_1(t, B)$ and $\sigma_2(t, B)$ do not depend on t nor the angle B. Since $F_{0,B} = \mathrm{id}$, (4.4) implies that

$$\sigma_3(0, B) = \rho_3(B) = \begin{pmatrix} q(B) & i\delta(B) \\ i\delta(B) & \overline{q(B)} \end{pmatrix}. \tag{4.6}$$

Step 3. Next we describe the matrix $\sigma_3(t, B)$. We have

$$\begin{aligned} F_{t,B} &= F_{t,B} \circ \hat{\mu}_3 \circ \hat{\mu}_3 = \overline{\overline{F_{t,B} \circ \hat{\mu}_3 \circ \hat{\mu}_3}} \\ &= \overline{\sigma_3(t, B)\left(F_{t,B} \circ \hat{\mu}_3\right)\rho_3(B)^{-1}} = \overline{\sigma_3(t, B)}\, \sigma_3(t, B)\, F_{t,B}\, \rho_3(B)^{-1} \overline{\rho_3(B)^{-1}} \\ &= \overline{\sigma_3(t, B)}\, \sigma_3(t, B)\, F_{t,B}, \end{aligned}$$

where we used the fact $\overline{\rho_3(B)}\rho_3(B)$ is the identity. Thus we have

$$\overline{\sigma_3(t,B)}\sigma_3(t,B) = \mathrm{id}\,.$$

By Lemma 1.4, the matrix $\sigma_3(t,B)$ can be written in the form

$$\sigma_3(t,B) = \begin{pmatrix} p(t,B) & i\nu_1(t,B) \\ i\nu_2(t,B) & \overline{p(t,B)} \end{pmatrix} \qquad \text{with } \nu_1,\nu_2 \in \mathbb{R}, \quad p\bar{p} + \nu_1\nu_2 = 1. \tag{4.7}$$

We also have

$$\begin{aligned} F_{t,B}\circ\hat{\mu}_2\circ\hat{\mu}_3 &= \overline{\overline{F_{t,B}\circ\hat{\mu}_2\circ\hat{\mu}_3}} = \overline{\sigma_3(t,B)\big(F_{t,B}\circ\hat{\mu}_2\big)\rho_3(B)^{-1}} \\ &= \overline{\sigma_3(t,B)}\,\sigma_2(t,B)\,F_{t,B}\,\rho_2(B)^{-1}\,\overline{\rho_3(B)^{-1}}. \end{aligned}$$

We may assume that the small disk centered at V_1 consists of $2l$-copies of D. Let b be the branching order of g at V_1. By the condition (3), we have

$$A = \pi\frac{b+1}{l}. \tag{4.8}$$

Since $\hat{\mu}_2\circ\hat{\mu}_3$ is the rotation of angle $2A$ at V_1, we have $(\hat{\mu}_2\circ\hat{\mu}_3)^l = \mathrm{id}$:

$$F_{t,B} = F_{t,B}\circ(\hat{\mu}_2\circ\hat{\mu}_3)^l = \big(\overline{\sigma_3(t,B)}\,\sigma_2(t,B)\big)^l F_{t,B}\big(\overline{\rho_3(B)}\,\rho_2(B)\big)^{-l}.$$

On the other hand, one can easily check that the eigenvalues of $\overline{\rho_3(B)}\rho_2(B)$ are $\{-e^{iA}, -e^{-iA}\}$. Then by (4.8), the eigenvalues of $\big(\overline{\rho_3(B)}\rho_2(B)\big)^l$ are $\{\pm 1, \pm 1\}$, that is

$$\big(\overline{\rho_3(B)}\rho_2(B)\big)^l = \pm\,\mathrm{id}\,.$$

So we have

$$F_{t,B} = \pm\big(\overline{\sigma_3(t,B)}\sigma_2(t,B)\big)^l F_{t,B}$$

which implies that

$$\big(\overline{\sigma_3(t,B)}\sigma_2(t,B)\big)^l = \pm\,\mathrm{id}\,.$$

This implies that the eigenvalues of $\overline{\sigma_3(t,B)}\sigma_2(t,B)$ are of the form $\{e^{\pi iN/l}, e^{-\pi iN/l}\}$ for some integer N. Since $\overline{\sigma_3(t,B)}\sigma_2(t,B)$ is continuous with respect to the parameter t, we can conclude that the eigenvalues of $\overline{\sigma_3(t,B)}\sigma_2(t,B)$ do not change by t. In particular,

$$\operatorname{trace}\overline{\sigma_3(t,B)}\sigma_2(t,B) = \operatorname{trace}\overline{\sigma_3(0,B)}\sigma_2(0,B). \tag{4.9}$$

On the other hand, since $F_{t,B}\circ\hat{\mu}_2\circ\hat{\mu}_3 = \overline{\sigma_3(t,B)}\sigma_2(t,B)\,F_{t,B}\,(\overline{\rho_3(B)}\rho_2(B))^{-1}$ and $F_{0,B} = \mathrm{id}$, we have

$$\overline{\sigma_3(0,B)}\sigma_2(0,B) = \overline{\rho_3(B)}\rho_2(B). \tag{4.10}$$

By (4.9), (4.10) and (4.7), we have

$$2\operatorname{Im}p(t,B) = \operatorname{trace}\overline{\sigma_3(t,B)}\,\sigma_2(t,B) = \operatorname{trace}\overline{\rho_3(B)}\,\rho_2(B) = 2\cos A.$$

Step 4. Since $\sigma_3(0,B) = \rho_3(B)$, we have

$$\operatorname{Re}p(0,B_0) = -\cos B_0.$$

Now we would like to find a real-valued smooth function $B(t)$ such that

$$\operatorname{Re}p(t,B(t)) = -\cos B_0 \qquad (B(0) = B_0).$$

By the implicit function theorem, a sufficient condition for the existence of such a $B(t)$ is

$$\left.\frac{\partial \operatorname{Re} p(t,B)}{\partial B}\right|_{(t,B)=(0,B_0)} \neq 0,$$

and by (4.7), (4.6), (4.2) and the assumption (1), we have

$$\begin{aligned}\left.\frac{\partial \operatorname{Re} p(t,B)}{\partial B}\right|_{(t,B)=(0,B_0)} &= \left.\frac{\partial \operatorname{Re} p(0,B)}{\partial B}\right|_{B=B_0} = \left.\frac{\partial \operatorname{Re} q(B)}{\partial B}\right|_{B=B_0} \\ &= -\left.\frac{\partial \cos B}{\partial B}\right|_{B=B_0} = \sin B_0 \neq 0.\end{aligned}$$

This proves the existence of such a $B(t)$ ($|t| < \varepsilon$) for a sufficiently small $\varepsilon > 0$.

Step 5. When $t = 0$, it holds that $\sigma_3(t,B) = \rho_3(B)$, so $\nu_1\nu_2 > 0$ for sufficiently small t ($|t| < \varepsilon$), by continuity. Now we set

$$F_t := \begin{pmatrix} u(t) & 0 \\ 0 & u(t)^{-1} \end{pmatrix} F_{t,B(t)}, \qquad u(t) = \sqrt[4]{\frac{\nu_2(t,B(t))}{\nu_1(t,B(t))}}.$$

Obviously, F_t satisfies the ordinary differential equation (4.3) for $B = B(t)$. In particular, F_t has the Hopf differential $t\hat{Q}$ and the secondary Gauss map $\hat{g}_{B(t)}$. Then by (4.5) and (4.7), we get the relations

$$\overline{F_t \circ \mu_j} = \sigma_j(t)\, F_t\, \rho_j(B(t)) \qquad (j = 1,2,3),$$

where

$$\sigma_1(t) = \mathrm{id}, \qquad \sigma_2(t) = \rho_2 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$$

and

$$\sigma_3(t) = \begin{pmatrix} p & i\sqrt{\nu_1\nu_2} \\ i\sqrt{\nu_1\nu_2} & \bar{p} \end{pmatrix} \qquad \big(p = p(t,B(t)),\ \nu_j = \nu_j(t,B(t)),\ j = 1,2\big).$$

Since $\operatorname{Im} p(t,B(t)) = \cos A$ and $\operatorname{Re} p(t,B(t)) = -\cos B_0$, we have

$$p(t,B(t)) = i\cos A - \cos B_0 = q(B_0).$$

Thus we have $\sigma_j(t) = \rho_j(B_0)$ for $j = 1,2,3$. We now set

$$f_t := F_t F_t^* \colon \widetilde{M} \longrightarrow H^3.$$

By (1.8), the first fundamental form of f_t is given by

$$ds^2 := \big(1 + |\hat{g}_{B(t)}|^2\big)^2 \left|\frac{\hat{Q}}{d\hat{g}_{B(t)}}\right|^2.$$

By the condition (3) of the theorem, it is positive definite, and thus f_t is a conformal CMC-1 immersion whose Hopf differential is $t\hat{Q}$ and the secondary Gauss map $\hat{g}_{B(t)}$.

Step 6. We shall now prove that the conformal CMC-1 immersion f_t is single-valued on $M = \overline{M} \setminus \{p_1, \ldots, p_n\}$. Let s be a nonnegative integer and $\hat{\mu}_{i_1}, \hat{\mu}_{i_2}, \ldots, \hat{\mu}_{i_{2s}}$ be sequences of three reflections $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ on M such that

$$\pi \circ \hat{\mu}_{i_1} \circ \hat{\mu}_{i_2} \circ \cdots \circ \hat{\mu}_{i_{2s}} = \pi,$$

where $\pi\colon \widetilde{M} \to M$ be the universal covering. To show the single-valued property of f_t, it is sufficient to show that $f_t = f_t \circ \hat{\mu}_{i_1} \circ \hat{\mu}_{i_2} \circ \cdots \circ \hat{\mu}_{i_{2s}}$. In fact, we have

$$F_t \circ \hat{\mu}_{i_1} \circ \hat{\mu}_{i_2} \circ \cdots \circ \hat{\mu}_{i_{2s}} = \left(\overline{\rho_{i_1}(B_0)}\rho_{i_2}(B_0) \dots \overline{\rho_{i_{2s-1}}(B_0)}\rho_{i_{2s}}(B_0)\right) F_t \left(\overline{\rho_{i_1}(B(t))}\rho_{i_2}(B(t)) \dots \overline{\rho_{i_{2s-1}}(B(t))}\rho_{i_{2s}}(B(t))\right).$$

Since g is single-valued on M, $\hat{g} := g \circ \pi$ satisfies

$$\hat{g} = \hat{g} \circ \hat{\mu}_{i_1} \circ \cdots \circ \hat{\mu}_{i_{2s}}.$$

On the other hand, since $\hat{g}$ is the secondary Gauss map of F_{t,B_0}, we have by (1.3) that

$$\hat{g} = \hat{g} \circ \hat{\mu}_{i_1} \circ \cdots \circ \hat{\mu}_{i_{2s}} = \left(\overline{\rho_{i_1}(B_0)}\,\rho_{i_2}(B_0) \dots \overline{\rho_{i_{2s-1}}(B_0)}\,\rho_{i_{2s}}(B_0)\right) \star g.$$

Thus we can conclude that

$$\overline{\rho_{i_1}(B_0)}\,\rho_{i_2}(B_0) \dots \overline{\rho_{i_{2s-1}}(B_0)}\,\rho_{i_{2s}}(B_0) = \pm\,\mathrm{id}\,.$$

Since $\rho_{i_j}(B(t)) \in \mathrm{SU}(2)$ $(j = 1, \dots, 2s)$, this implies that $f_t = F_t F_t^*$ is single-valued on M.

Also, by the assumption (2) and Lemma 1.5, each end is irregular. The Hopf differential of f_t is Q_t, and the pseudometric $d\sigma^2_{f_t}$ defined in (1.9) is $d\sigma^2_{A,B(t),\pi/2}$. Since they are symmetric with respect to D, (1.10) and (1.8) imply that the first and second fundamental forms of f_t are invariant under the reflections μ_j $(j = 1, 2, 3)$. Then, by the fundamental theorem of surfaces, f_t is symmetric with respect to D. □

PROOF OF THEOREM 2.5. Let τ be a loop surrounding the point V_2 with the base point V_3 and T the covering transformation of $\widetilde{M}$ corresponding to τ. Suppose that a neighborhood of V_2 is generated by $2k$-copies of D. Then

$$T := (\hat{\mu}_3 \circ \hat{\mu}_1)^k,$$

so

$$F_{t,B} \circ \tau = F_{t,B} \circ (\hat{\mu}_3 \circ \hat{\mu}_1)^k = \left(\overline{\sigma_3(t,B)}\,\sigma_1\right)^k F_{t,B} \left(\rho_1^{-1}\overline{\rho_3(B)^{-1}}\right)^k = \overline{\sigma_3(t,B)}^k F_{t,B}\,\overline{\rho_3(B)}^{-k}.$$

Here, by the argument of Step 6 of the proof of Theorem 2.3, we have $\rho_3(B)^k = \pm\,\mathrm{id}$. Hence at the base point V_3,

$$\left.\frac{\partial F_{t,B} \circ \tau}{\partial t}\right|_{(t,B)=(0,B_0)} = \left.\frac{\partial}{\partial t}\overline{\sigma_3(t,B)}^k F_{t,B}\,\overline{\rho_3(B)}^{-k}\right|_{(t,B)=(0,B_0)} = \pm \left.\frac{\partial}{\partial t}\overline{\sigma_3(t,B)}^k\right|_{(t,B)=(0,B_0)} \tag{4.11}$$

since $F_{t,B}(V_3) = \mathrm{id}$.

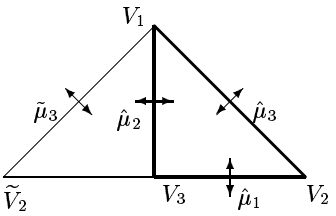


FIGURE 6. $\widetilde{V}_2$ and $\tilde{\mu}_3$.

Since $F_{t,B}$ is a solution of (4.3), we have

$$\begin{aligned}\frac{\partial}{\partial z}\left(\left.\frac{\partial F_{t,B}}{\partial t}\right|_{(t,B)=(0,B_0)}\right)dz &= d\left(\left.\frac{\partial F_{t,B}}{\partial t}\right|_{(t,B)=(0,B_0)}\right)\\ &= \left.\frac{\partial}{\partial t}dF_{t,B}\right|_{(t,B)=(0,B_0)}\\ &= \left.\frac{\partial}{\partial t}\left(tF_{t,B}\begin{pmatrix}\hat{g}_B & -\hat{g}_B^2\\ 1 & -\hat{g}_B\end{pmatrix}\frac{\hat{Q}}{d\hat{g}_B}\right)\right|_{(t,B)=(0,B_0)}\\ &= \begin{pmatrix}\hat{g} & -\hat{g}^2\\ 1 & -\hat{g}\end{pmatrix}\frac{\hat{Q}}{d\hat{g}},\end{aligned}$$

because $\hat{g}_{B_0} = \hat{g} = g \circ \pi$. Then by the assumption and (4.11), we have

$$\left.\frac{\partial}{\partial t}\overline{\sigma_3(t,B)}^k\right|_{(t,B)=(0,B_0)} = \pm\oint_\tau \begin{pmatrix}g & -g^2\\ 1 & -g\end{pmatrix}\frac{Q}{dg} \neq 0.$$

This implies that $\sigma_3(t) = \sigma_3(t, B(t))$ is not constant on $\{t\,;\,|t|<\varepsilon\}$ for sufficiently small $\varepsilon > 0$, and hence we have

$$B(t) \neq B_0 \qquad \text{for} \quad 0 < |t| < \varepsilon. \tag{4.12}$$

Hence the eigenvalues of $\rho_3(t) := \rho_3(B(t))$ are different from those of $\rho_3(B_0)$.

The secondary Gauss map $\hat{g}_t$ of F_t changes by the covering transformation T as

$$\hat{g}_t \circ T = \hat{g}_t \circ (\hat{\mu}_3 \circ \hat{\mu}_1)^k = \overline{\rho_3(t)}^k \star \hat{g}_t.$$

Now, let $\widetilde{V}_2 = \hat{\mu}_2(V_2)$ (see Figure 6). We denote by $\tilde{\mu}_3$ the reflection about the edge $V_1\widetilde{V}_2$. Then we have

$$\tilde{\mu}_3 = \hat{\mu}_2 \circ \hat{\mu}_3 \circ \hat{\mu}_2.$$

Let $\tilde{\tau}$ be a loop surrounding $\widetilde{V}_2$ with base point V_3, and let $\widetilde{T}$ be the covering transformation corresponding to $\tilde{\tau}$. Then we have

$$\widetilde{T} = (\hat{\mu}_1 \circ \tilde{\mu}_3)^k = (\hat{\mu}_1 \circ \hat{\mu}_2 \circ \hat{\mu}_3 \circ \hat{\mu}_2)^k,$$

and

$$\hat{g}_t \circ \widetilde{T} = \left(\overline{\rho_1}\,\rho_2\,\overline{\rho_3(t)}\,\rho_2\right)^k \star \hat{g}_t = (-1)^{k-1}\rho_2\left(\overline{\rho_3(t)}^k\right)\rho_2 \star \hat{g}_t.$$

So, to show irreducibility, it is sufficient to show that the matrices

$$a := \overline{\rho_3(t)}^k \qquad \text{and} \qquad b := \rho_2\overline{\rho_3(t)}^k\rho_2 = \rho_2\, a\, \rho_2$$

do not commute. By (4.1), $\overline{\rho_3(t)^k}\rho_3(t)^k = \mathrm{id}$ holds. Then by Lemma 1.4, the off-diagonal components of a are coincide and pure imaginary. Set

$$a = \overline{\rho_3(t)}^k = \begin{pmatrix} r & i\beta \\ i\beta & \bar{r} \end{pmatrix} \qquad (\beta \in \mathbb{R},\ r\bar{r} + \beta^2 = 1).$$

Then we have

$$b = \rho_2\, a\, \rho_2 = \begin{pmatrix} -r & i\beta \\ i\beta & -\bar{r} \end{pmatrix},$$

and

$$[a, b] = ab - ba = \begin{pmatrix} 0 & -2\beta \operatorname{Im} r \\ 2\beta \operatorname{Im} r & 0 \end{pmatrix},$$

that is, a and b commute if and only if $\beta = 0$ or r is a real number.

First, we consider the case $\beta = 0$. Then $a = \overline{\rho_3(t)}^k$ is a diagonal matrix whose eigenvalues are different from ± 1 for sufficiently small t, because $B(t)$ is not constant. In particular, the two eigenvalues of a are distinct. This implies that the eigenspaces of a coincide of those of $\overline{\rho_3(t)}$. Since a is diagonal, this implies that $\rho_3(t)$ is also a diagonal matrix, a contradiction.

Next, assume r is real. Then there exists a real number θ such that

$$a = \overline{\rho_3(t)}^k = \begin{pmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{pmatrix} = P \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} P^{-1},$$

$$\text{where } P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad \theta \in \mathbb{R} \setminus \pi\mathbb{Z}.$$

In this case, $(P^{-1}\overline{\rho_3(t)}P)^k$ is a diagonal matrix whose eigenvalues differ from ± 1, for sufficiently small $t \neq 0$. Then, by a similar argument to the previous case, we have $P^{-1}\overline{\rho_3}(t)P$ is diagonal. If $A \not\equiv \pi/2 \pmod{\pi}$, this contradicts (4.1) and (4.2). Hence a and b do not commute. □

5. An Example of Genus One

In the final section, we construct an example of a CMC-1 surface of genus one with four irregular ends.

Let $\Gamma = \mathbb{Z} \oplus i\mathbb{Z}$ be the lattice of Gaussian integers of $\mathbb{C}$ and let

$$\overline{M} := \mathbb{C}/2\Gamma.$$

We consider $\overline{M}$ as the torus obtained from the square $[-\frac{1}{2}, \frac{3}{2}] \times [-\frac{1}{2}, \frac{3}{2}]$ in $\mathbb{C} = \mathbb{R}^2$ by identified opposite edges. Take a triangle D on $\overline{M}$ as in Figure 7. Then $\overline{M}$ is obtained from D by reflecting across the edges of D.

Using the Weierstrass $\wp$ function with respect to Γ (not with respect to 2Γ), we set

$$Q = \left(\wp'(z)\right)^2 dz^2.$$

Then Q has poles at $\{p_1, p_2, p_3, p_4\} = \{0, 1, 1+i, i\}$, each with order 6. The $\wp$-function with respect to the square lattice has the properties

$$\overline{\wp(\bar{z})} = \wp(z), \qquad \overline{\wp(-\bar{z})} = \wp(z), \qquad \overline{\wp(i\bar{z})} = -\wp(z).$$

Hence Q is symmetric with respect to D.

Consider an abstract spherical triangle $\mathcal{T}(A, B_0, C)$ with

$$A = \frac{3}{4}\pi, \qquad B_0 = \frac{\pi}{2}, \qquad C = \frac{3}{2}\pi,$$

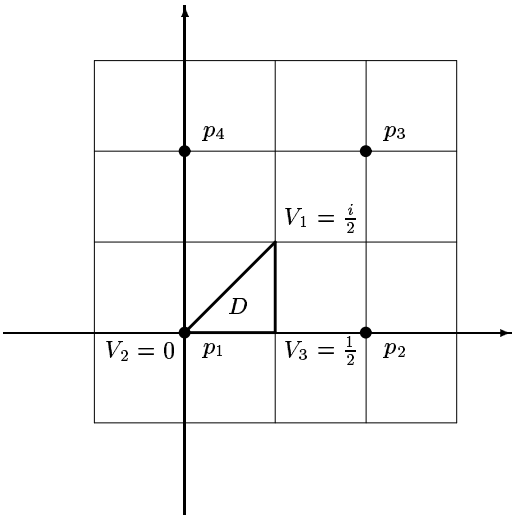


FIGURE 7. Fundamental domain of the torus.

and identify the triangle with the fundamental region D. Then the metric of $\mathcal{T}(A, B_0, C)$ can be extended to $d\sigma^2_{A,B_0,C} \in \mathrm{Met}_1(\overline{M})$ by reflections. Since A, B_0 and C satisfy (2.1), $d\sigma^2_{A,B_0,C}$ is nondegenerate. Let g be the developing map of $d\sigma^2_{A,B_0,C}$. Since the conical orders of $d\sigma^2_{A,B_0,C}$ are integers, g is well-defined on $\mathbb{C}$.

Now, we prove that g is well-defined on $\overline{M}$. By the monodromy principle (Lemma 2.2), one can choose g such that $\overline{g \circ \hat{\mu}_j} = \rho_j \star g$ $(j = 1, 2, 3)$, where

$$\rho_1 := \mathrm{id}, \qquad \rho_2 := \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad \rho_3 := \frac{1}{\sqrt{2}} \begin{pmatrix} i & \pm i \\ \pm i & -i \end{pmatrix}$$

and $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$ are reflections along the edges V_3V_1, V_1V_2 and V_2V_3, respectively. We denote by τ_1 and τ_2 the translations $z \mapsto z + 1$ and $z \mapsto z + i$, respectively. Then

$$\tau_1 = \hat{\mu}_2 \circ \hat{\mu}_3 \circ \hat{\mu}_1 \circ \hat{\mu}_3, \qquad \tau_2 = \hat{\mu}_3 \circ \hat{\mu}_2 \circ \hat{\mu}_3 \circ \hat{\mu}_1$$

holds. So we have

$$g \circ \tau_1 = \overline{\rho_2}\, \rho_3\, \overline{\rho_1}\, \rho_3 \star g = \rho_2 \star g,$$
$$g \circ \tau_2 = \overline{\rho_3}\, \rho_2\, \overline{\rho_3}\, \rho_1 \star g = \pm \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \star g.$$

Thus

$$g(z + 2) = g \circ \tau_1 \circ \tau_1(z) = g(z), \qquad g(z + 2i) = g \circ \tau_2 \circ \tau_2(z) = g(z)$$

hold. This shows that g is invariant under the action of the lattice 2Γ. Hence g is a meromorphic function on $\overline{M}$.

One can easily see that the same result as Theorem 2.3 holds when $C = 3\pi/2$, instead of $\pi/2$. Hence we have a one-parameter family $\{f_t\}$ of CMC-1 immersions of $\overline{M} \setminus \{p_1, p_2, p_3, p_4\}$ into H^3 with irregular ends.

Appendix A

For a compact Riemann surface $\overline{M}$ and points $p_1, \dots, p_n \in \overline{M}$, a conformal metric $d\sigma^2$ of constant curvature 1 on $M := \overline{M} \setminus \{p_1, \dots, p_n\}$ is an element of $\mathrm{Met}_1(\overline{M})$ if there exist real numbers $\beta_1, \dots, \beta_n > -1$ so that each p_j is a conical singularity of conical order β_j, that is, if $d\sigma^2$ is asymptotic to $c_j|z-p_j|^{2\beta_j}\, dz \cdot d\bar{z}$ at p_j, for $c_j \neq 0$ and z a local complex coordinate around p_j. We call the formal sum

$$D := \sum_{j=1}^{n} \beta_j\, p_j \tag{A.1}$$

the *divisor* corresponding to $d\sigma^2$. For a pseudometric $d\sigma^2 \in \mathrm{Met}_1(\overline{M})$ with divisor D, there is a holomorphic map $g\colon \widetilde{M} \to \mathbb{CP}^1$ such that $d\sigma^2$ is the pull-back of the Fubini–Study metric of $\mathbb{CP}^1$. This map, called the *developing map* of $d\sigma^2$, is uniquely determined up to Möbius transformations $g \mapsto a \star g$ for $a \in \mathrm{SU}(2)$. We have the expression

$$\pi^* d\sigma^2 = \frac{4\, dg \cdot d\bar{g}}{(1+|g|^2)^2},$$

where $\pi : \widetilde{M} \to M$ is a covering projection.

Consider $d\sigma^2 \in \mathrm{Met}_1(\overline{M})$ with divisor D as in (A.1) and with the developing map g. Since the pull-back of the Fubini–Study metric of $\mathbb{CP}^1$ by g is invariant under the deck transformation group $\pi_1(M)$ of $M := \overline{M} \setminus \{p_1, \dots, p_n\}$, there is a representation

$$\rho_g\colon \pi_1(M) \longrightarrow \mathrm{SU}(2)$$

such that

$$g \circ T^{-1} = \rho_g(T) \star g \qquad (T \in \pi_1(M)).$$

The metric $d\sigma^2$ is called *reducible* if the image of ρ_g is a commutative subgroup in $\mathrm{SU}(2)$, and is called *irreducible* otherwise. Since the maximal abelian subgroup of $\mathrm{SU}(2)$ is $\mathrm{U}(1)$, the image of ρ_g for a reducible $d\sigma^2$ lies in a subgroup conjugate to $\mathrm{U}(1)$, and this image might be simply the identity. We call a reducible metric $d\sigma^2$ $\mathcal{H}^3$*-reducible* if the image of ρ_g is the identity, and $\mathcal{H}^1$*-reducible* otherwise (for more on this, see [**RUY1**, Section 3]).

The following assertion was needed in Section 2:

PROPOSITION A.1. *Let $d\sigma^2$ be a metric of constant curvature 1 defined on $M = \overline{M} \setminus \{p_1, \dots, p_n\}$ whose conical order at each p_j is an integer. Then the developing map g of $d\sigma^2$ is single-valued on the universal covering of M.*

Let $p_1, \dots, p_{n-1}$ be distinct points in $\mathbb{C}$ and $p_n = \infty$. We set

$$M_{p_1,\dots,p_n} := \mathbb{C} \cup \{\infty\} \setminus \{p_1, p_2, \dots, p_n\} \qquad (p_n = \infty)$$

and $\widetilde{M}_{p_1,\dots,p_n}$ its universal covering.

COROLLARY A.2. *Let $d\sigma^2$ be a metric of constant curvature 1 defined on $M_{p_1,\dots,p_n}$ $(p_n = \infty)$ whose conical order at each p_j is an integer. Then the developing map g of $d\sigma^2$ is single-valued on $M_{p_1,\dots,p_n}$ and extends as a meromorphic function on $\mathbb{C} \cup \{\infty\}$. Moreover, the divisor of $d\sigma^2$ coincides with the ramification divisor of the meromorphic function g.*

References

[BB] J. L. Barbosa and P. Berard, *Eigenvalue and "twisted" eigenvalue problems, applications to CMC surfaces*, J. Math. Pures Appl. **79** (2000), 427–450.

[BR] J. Berglund and W. Rossman, *Minimal surfaces with catenoid ends*, Pacific J. Math. **171** (1995), 353–371.

[Bry] R. Bryant, *Surfaces of mean curvature one in hyperbolic space*, Astérisque **154–155** (1987), 321–347.

[CGT] M. P. do Carmo, J. de M. Gomes and G. Thorbergsson, *The influence of the boundary behaviour on hypersurfaces with constant mean curvature in* H^{n+1}, Comment. Math. Helvetici **61** (1986), 429–441.

[CL] M. P. do Carmo and H. B. Lawson, *On Alexandrov-Berstein theorems in hyperbolic space*, Duke Math. J. **50(4)** (1983), 995–1003.

[CHR1] P. Collin, L. Hauswirth and H. Rosenberg, *The geometry of finite topology Bryant surfaces*, Ann. of Math. **153** (2001), 623–659.

[CHR2] ———, *The gaussian image of mean curvature one surfaces in* H^3 *of finite total curvature*, Advanced Studies in Pure Mathematics 29, Minimal Surfaces, Geometric Analysis and Symplectic Geometry 2000.

[CN] C. J. Costa and V. F. de Sousa Neto, *Mean curvature* 1 *surfaces of Costa type in hyperbolic 3-space*, Tôhoku Math. J. **53** (2001), 617–628.

[CS] M. P. do Carmo and A. M. da Silveira, *Index and total curvature of surfaces with constant mean curvature*, Proc. Amer. Math. Soc. **110(4)** (1990), 1009–1015.

[D] B. Daniel, *Surfaces de Bryant dans* H^3 *de type fini*, Bull. Sci. Math. **126** (2002), 581–594.

[ET1] R. Sa Earp and E. Toubiana, *On the geometry of constant mean curvature one surfaces in hyperbolic space*, Illinois J. Math. **45** (2001), 371–401.

[ET2] ———, *Meromorphic data for mean curvature one surfaces in hyperbolic space*, to appear in Tôhoku Math. J.

[F] M. Furuta, *Gluing spherical triangles* (Kyûmen sankaku kei wo hariawaseru to), in Japanese, Sûgaku Tsûshin **6** (2001), 4–27.

[FH] M. Furuta and Y. Hattori, *Two-dimensional spherical space forms*, preprint.

[GGN] C. C. Góes, M. E. E. L. Galvão and B. Nelli, *A type Weierstrass representation for surfaces in hyperbolic space with mean curvature one*, An. Acad. Brasil Cienc. **70**:1 (1998), 1–6.

[Kar] H. Karcher, *Hyperbolic constant mean curvature one surfaces with compact fundamental domains*, pp. 311–323 in this volume.

[Kat] S. Kato, *Construction of n-end catenoids with prescribed flux*, Kodai Math. J **18** (1995), 86–98.

[KUY] M. Kokubu, M. Umehara and K. Yamada, *An elementary proof of Small's formula for null curves in* $\mathrm{PSL}(2,\mathbb{C})$ *and an analogue for Legendrian curves in* $\mathrm{PSL}(2,\mathbb{C})$, to appear in Osaka J. Math., math.DG/0209258.

[LR] L. L. de Lima and W. Rossman, *On the index of mean curvature* 1 *surfaces in* H^3, Indiana Univ. Math. J. **47(2)** (1998), 685–723.

[MU] C. McCune and M. Umehara, *An analogue of the UP-iteration for constant mean curvature one surfaces in hyperbolic 3-space*, to appear in Differential Geometry and Its Applications.

[PP] F. Pacard and F. A. Pimentel, *Attaching handles to Bryant surfaces*, preprint, math.DG/0112224.

[Ros] H. Rosenberg, *Bryant surfaces*, The global theory of minimal surfaces in flat spaces (Martina Franca, 1999), pp. 67–111, Lecture Notes in Math., Vol. 1775, Springer-Verlag, 2002.

[RR] L. Rodriguez and H. Rosenberg, *Half-space theorems for mean curvature one surfaces in hyperbolic space*, Proc. Amer. Math. Soc. **126** (1998), 2763–2771.

[RS] W. Rossman and K. Sato, *Constant mean curvature surfaces with two ends in hyperbolic space*, Experimental Math. **7(2)** (1998), 101–119.

[RUY1] W. Rossman, M. Umehara and K. Yamada, *Irreducible constant mean curvature 1 surfaces in hyperbolic space with positive genus*, Tôhoku Math. J. **49** (1997), 449–484.

[RUY2] ———, *A new flux for mean curvature* 1 *surfaces in hyperbolic* 3*-space, and applications*, Proc. Amer. Math. Soc. **127** (1999), 2147–2154.

[RUY3] ———, *Mean curvature* 1 *surfaces with low total curvature in hyperbolic* 3*-space I*, preprint, math.DG/0008015.

[RUY4] ———, *Mean curvature* 1 *surfaces with low total curvature in hyperbolic* 3*-space II*, preprint, math.DG/0102035, to appear in Tôhoku J. Math.

[RUY5] ———, *Period problems for mean curvature one surfaces in* H^3 *(with application to surfaces of low total curvature)*, math.DG/0102185, to appear in the proceedings of the MSJIRI meeting of the Mathematical Society of Japan, Tokyo 2000.

[Sm] A. J. Small, *Surfaces of constant mean curvature* 1 *in* H^3 *and algebraic curves on a quadric*, Proc. Amer. Math. Soc. **122** (1994), 1211–1220.

[Tro1] M. Troyanov, *Metric of constant curvature on a sphere with two conical singularities*, in "Differential Geometry", Lect. Notes in Math. vol. 1410, Springer-Verlag, (1989), 296–306.

[UY1] M. Umehara and K. Yamada, *Complete surfaces of constant mean curvature-*1 *in the hyperbolic* 3*-space*, Ann. of Math. **137** (1993), 611–638.

[UY2] ———, *A parameterization of Weierstrass formulae and perturbation of some complete minimal surfaces of* $\mathbb{R}^3$ *into the hyperbolic* 3*-space*, J. reine u. angew. Math. **432** (1992), 93–116.

[UY3] ———, *Surfaces of constant mean curvature-c in* $H^3(-c^2)$ *with prescribed hyperbolic Gauss map*, Math. Ann. **304** (1996), 203–224.

[UY4] ———, *Another construction of a CMC* 1 *surface in* H^3, Kyungpook Math. J. **35** (1996), 831–849.

[UY5] ———, *A duality on CMC* 1 *surfaces in hyperbolic* 3*-space and a hyperbolic analogue of the Osserman Inequality*, Tsukuba J. Math. **21** (1997), 229-237.

[UY6] ———, *Geometry of surfaces of constant mean curvature* 1 *in the hyperbolic* 3*-space*, Suugaku Expositions **10(1)** (1997), 41–55.

[UY7] ———, *Metrics of constant curvature* 1 *with three conical singularities on the* 2*-sphere*, Illinois J. Math. **44(1)** (2000), 72–94.

[Xu] Y. Xu, *Symmetric minimal surface in* $\mathbb{R}^3$, Pacific J. Math. **171** (1995), 203–224.

[Yu1] Z. Yu, *Value distribution of hyperbolic Gauss maps*, Proc. Amer. Math. Soc. **125** (1997), 2997–3001.

[Yu2] ———, *The inverse surface and the Osserman Inequality*, Tsukuba J. Math. **22** (1998), 575–588

[Yu3] ———, *Surfaces of constant mean curvature one in the hyperbolic three-space with irregular ends*, Tôhoku Math. J. **53** (2001) 305–318.

Department of Mathematics, Faculty of Science, Kobe University, Rokko, Kobe 657-8501, Japan

E-mail address: wayne@math.kobe-u.ac.jp

Department of Mathematics, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan

E-mail address: umehara@math.wani.osaka-u.ac.jp

Faculty of Mathematics, Kyushu University 36, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

E-mail address: kotaro@math.kyushu-u.ac.jp

Clay Mathematics Proceedings
Volume 2, 2004

Conformal Structures and Necksizes of Embedded Constant Mean Curvature Surfaces

Rob Kusner

Dedicated to my father and mentor, David B. Kusner, on the occasion of his seventieth birthday

Abstract. Let $\mathcal{M} = \mathcal{M}_{g,k}$ denote the space of properly embedded (or Alexandrov immersed) constant mean curvature (CMC) surfaces of genus g with k (labeled) ends, modulo rigid motions, endowed with the real analytic structure described in [**20**]. Let $\mathcal{P} = \mathcal{P}_{g,k} = \mathcal{R}_{g,k} \times \mathbb{R}^k_+$ be the space of parabolic structures over Riemann surfaces of genus g with k (marked) punctures, the real analytic structure coming from the $3g - 3 + k$ local complex analytic coordinates on the Riemann moduli space $\mathcal{R}_{g,k}$. Then the *parabolic classifying map*, $\Phi : \mathcal{M} \to \mathcal{P}$, which assigns to a CMC surface its induced conformal structure and asymptotic necksizes, is a proper, real analytic map. It follows that Φ is closed and in particular has closed image. For genus $g = 0$, this can be used to show that every conformal type of multiply punctured Riemann sphere occurs as a CMC surface, and — under a nondegeneracy hypothesis — that Φ has a well-defined (mod 2) degree. This degree vanishes, so generically an *even* number of CMC surfaces realize any given conformal structure and asymptotic necksizes (compare [**11, 12**] for the case $k = 3$).

Introduction

Besides their beauty and variety, perhaps the most important reason minimal surfaces have been so thoroughly investigated is because they can be conformally parametrized via holomorphic functions of a complex variable, a long-familiar tool to many mathematicians. This representation theory was worked out by Enneper and Weierstrass in the middle of the 19th century, leading to important local results and interesting periodic examples. But not until the 1960s did Osserman [**27**] prepare the way for a global theory by showing that a complete minimal surface with finite total curvature is conformally parametrized by meromorphic data on a finitely punctured, finite genus Riemann surface. The global Enneper–Weierstrass representation enables one to employ algebro-geometric methods from the theory of complex curves to study such minimal surfaces. Over the past two decades it has been used in conjunction with analytic methods by a number of authors — many represented in this volume — to prove deep and striking results about properly embedded minimal surfaces of finite total curvature or finite topology (finite genus and a finite number of ends).

©2004 Clay Mathematics Institute

The theory of complete, properly embedded surfaces with *nonzero* constant mean curvature (CMC) has developed more recently and without the benefit of these holomorphic methods. While more than a century elapsed between Delaunay's classification [**7**] of the CMC rotation surfaces (two ends, genus zero) and Alexandrov's [**2**] proof that the only compact (zero ends) CMC surface is the round sphere, only in the past decade have interesting noncompact examples of finite topology CMC surfaces (with three or more ends) been constructed by Kapouleas [**16**], using difficult analytic methods. At about the same time, the first steps toward a global theory of these surfaces were taken by Meeks and by Korevaar, Kusner and Solomon: they found obstructions to the existence of CMC surfaces with finite topology — none have one end [**24**], two-ended surfaces are Delaunay unduloids [**19**] — and developed an asymptotic theory for these surfaces. In fact the fundamental result of [**19**] is that each (annular) end of a CMC surface has an unduloid asymptote. An immediate consequence is that a finite topology CMC surface can be conformally parametrized by a finitely punctured Riemann surface of finite genus.

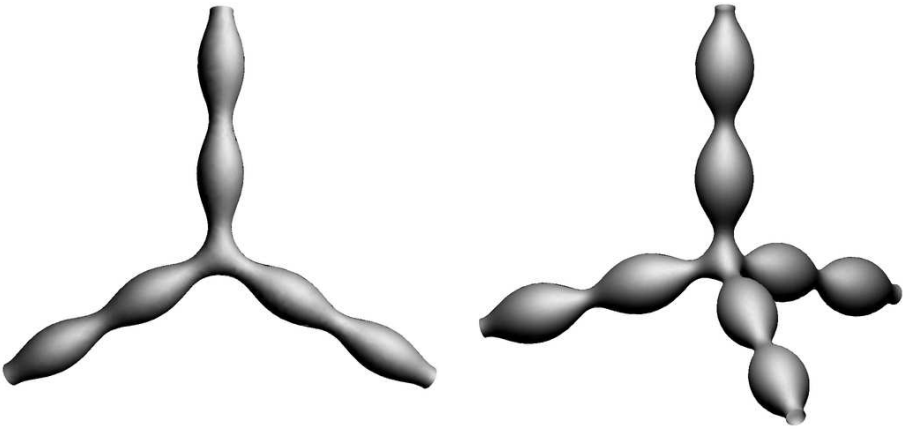


FIGURE 1. Equilateral triunduloid and tetrunduloid (images by Nick Schmitt, GANG).

Can such a global conformal parametrization be made explicit — enough to prove something new about CMC surfaces, or at least to experiment with them using a computer? Using a loop-group analogue of the Enneper–Weierstrass representation — dubbed the DPW representation after its discoverers, Dorfmeister, Pedit and Wu [**8**] — Schmitt has recently taken up this question at GANG in Amherst (see Figure 1, and `www.gang.umass.edu`). He adapted the DPW representation to construct, in terms of explicit meromorphic data on a punctured Riemann surface, conformal immersions of finite topology constant mean curvature surfaces whose ends are asymptotically Delaunay unduloids (or nodoids) [**17, 30**]. This approach is still under development, and little is known about how to prove that the examples it constructs have the symmetries or embeddedness properties required. Nevertheless, it appears to recapture the earlier results of Große-Brauckmann, Kusner and Sullivan [**11, 12**] about three-ended CMC surfaces of genus zero, the *triunduloids*.

Our investigations [**11, 12**] made use of a special classifying map from the space of all triunduloids to an open three-ball. We relied heavily on the real analytic variety structure of CMC moduli space developed by Kusner, Mazzeo, and Pollack [**20**],

as well as the compactness results of Korevaar and Kusner [**18**], to prove this map is a homeomorphism. The present paper also uses these transcendental methods, and was motivated by a desire to weave these together with the holomorphic methods arising from the DPW representation in order to better comprehend the fabric of CMC moduli spaces.

The main result of this paper is a properness theorem for a general classifying map which assigns a CMC surface its induced conformal structure and asymptotic necksizes. As noted in the abstract, consequences of this theorem include information about which conformal structures and necksizes — input data for the DPW representation — can arise for CMC surfaces, as well as about how many surfaces might share given data. In particular, we show that these surfaces generally occur in pairs, a phenomenon first observed by Karcher and Große-Brauckmann [**10**] for surfaces with special symmetry. Several outstanding open problems about the moduli spaces of CMC surfaces are also discussed.

There is some interplay between our results and those of Mazzeo, Pacard and Pollack, so here an effort is made to focus on what is complementary; the reader is encouraged to refer to their nice paper [**23**], as well as to [**12**], for additional background and related material. The author wishes to thank Rafe Mazzeo, Frank Pacard, Dan Pollack, Jesse Ratzkin and Rick Schoen for stimulating discussions, and Karsten Große-Brauckmann for much thoughtful advice. Thanks are also due to Nicos Kapouleas, Hermann Karcher, Nick Korevaar, Remi Langevin, Franz Pedit, Wayne Rossman, Nick Schmitt, John Sullivan and Mike Wolf for their interest and comments, and especially to David Hoffman and Harold Rosenberg for their invitation to present this at the 2001 Clay/MSRI summer workshop. The research reported here was supported in part by NSF grant DMS-0076085.

1. Background on the CMC Moduli Space

Let $\mathcal{M} = \mathcal{M}_{g,k}$ be the moduli space of all CMC surfaces of genus g with k (labeled) ends, modulo rigid motions. Here, as usual, we scale the mean curvature to be 1, and include in $\mathcal{M}$ not only the embedded surfaces, but also those which are immersed in the sense of Alexandrov: these are boundaries of properly immersed 3-dimensional manifolds (compare [**3, 12, 19**]). Some of the results mentioned in the introduction can then be expressed in terms of $\mathcal{M}$ as follows:

- $\mathcal{M}_{0,0}$ is a point, represented by any unit sphere, and $\mathcal{M}_{g,0}$ is empty for all $g > 0$ [**2**];
- $\mathcal{M}_{g,1}$ is empty for every g [**24**];
- $\mathcal{M}_{0,2}$ consists of the unduloids (up to rigid motion), but $\mathcal{M}_{g,2}$ is empty for all $g > 0$ [**19**];
- $\mathcal{M}_{g,k}$ is nonempty for every $g \geq 0$, $k > 2$ [**16**] — for example, $\mathcal{M}_{0,3}$ is a 3-ball [**11, 12**];
- $\mathcal{M}_{g,k}$ is a real analytic variety of (virtual) dimension $3k - 6$, for $k > 2$ [**20**].

To understand this real analytic structure on $\mathcal{M}$, it will be useful to introduce the *pre-moduli* space $\widetilde{\mathcal{M}}$, that is, the space of CMC surfaces *before* dividing by rigid motions. We also let $\mathcal{U} = \widetilde{\mathcal{M}}_{0,2}$ denote the pre-moduli space of all unduloids. The discussion below will describe the real analytic variety structure on $\widetilde{\mathcal{M}}$ in terms

of an explicit real analytic manifold structure on $\mathcal{U}$, in order to demonstrate the following (almost tautological) result:

PROPOSITION 1.1. *The map $\widetilde{\mathcal{M}}_{g,k} \to \mathcal{U}^k$ which assigns a* CMC *surface its k unduloid asymptotes is real analytic, and is equivariant with respect to the actions of rigid motions. Moreover, there is a real analytic necksize function $s : \mathcal{U} \to \mathbb{R}_+$, taking values in $(0, 1]$. It is invariant under rigid motions, and so gives rise to a real analytic asymptotic necksize map $(s_1, \ldots, s_k) : \mathcal{M}_{g,k} \to \mathbb{R}^k_+$, taking values in $(0, 1]^k$.*

PROOF. The equivariance is clear when we take the diagonal action by rigid motions on the k-fold product $\mathcal{U}^k$: as a CMC surface undergoes a rigid motion, each unduloid asymptote undergoes the same motion. Furthermore, once s is defined, it should be clear that $(s_1, \ldots, s_k)$ is defined by composition with the asymptotes map and passage to the quotient by rigid motions.

Recall that an unduloid [**19**] is a periodic CMC surface of revolution determined by its axis, its neckphase — where along the axis it is nearest the axis — and its necksize. (Of course, modulo rigid motions, an unduloid is specified by its necksize alone.) So the space $\mathcal{U}$ is naturally a fiber bundle over the space $T\mathbb{S}^2$ of (oriented) lines in $\mathbb{R}^3$. The base space $T\mathbb{S}^2$ has a natural analytic structure as a homogeneous space for the group of rigid motions. The total space $\mathcal{U}$ will then be expressed as a 6-dimensional real analytic manifold provided an analytic manifold structure can be exhibited on the typical fiber $\mathcal{D}$, consisting of all unduloids with a fixed, oriented axis (the two ends of the unduloid are labeled according to the axis orientation).

Observe that $\mathcal{D}$ is topologically an open disk, with necksize and neckphase as polar coordinates. The cylinder, whose necksize is maximal and whose neckphase is not well defined, is at the center of $\mathcal{D}$, fixed under rotation (neckphase-shift). The boundary of $\mathcal{D}$ corresponds to the zero necksize limits of unduloids — chains of spheres along the given axis. The natural analytic structure on $\mathcal{D}$ comes from its (locally homogeneous) realization as an open subset of the unit sphere $\mathbb{S}^2$, indeed as $\mathbb{S}^2 - \{p\}$ for a single point p.

This spherical picture of $\mathcal{D}$ also played a key role in [**12**], leading to a natural interpretation of necksize — there referring to the length of the shortest nontrivial loop on an unduloid — as the spherical distance n between p and another point $q \in \mathcal{D} = \mathbb{S}^2 - \{p\}$. Unfortunately, n is not an analytic coordinate on $\mathcal{D}$: at the cylinder (corresponding to q antipodal to p) the function n is only Lipshitz. This will be remedied by reinterpreting *necksize* here to mean $s = \frac{1}{2}(1 - \cos n)$.

Although s may seem geometrically less natural than n, it turns out to be more natural in other ways. For example, being a height function on $\mathbb{S}^2$ in the direction p, it is symplectically dual to neckphase (angle around p) for the homogeneous symplectic structure on $\mathbb{S}^2$. For the present paper, however, the important point is that s is analytic with respect to the natural (locally homogeneous) analytic structure on $\mathcal{U}$, and has the invariance properties and range indicated in the statement of the proposition. In fact, we could have equally well used the *weight* or *force* modulus [**19**] of an unduloid, $f = n(2\pi - n)$, which is also an analytic coordinate on $\mathcal{U}$, even at the cylinder ($n = \pi$).

The tangent space to $\mathcal{U}$ at an unduloid can be identified with the 6-dimensional linear space $\mathcal{V}$ of *geometric* Jacobi fields (see [**19, 20**]). Thus $\mathcal{V}$ gives analytic local coordinates on $\mathcal{U}$ in a neighborhood of this unduloid. We shall pretend that $\mathcal{V}$ defines global analytic coordinates on $\mathcal{U}$ in what follows, but since checking

analyticity is a local condition, this pretense only amounts to a mild abuse of notation.

Using the asymptotics result [**19**], one can construct [**20**] a $6k$-dimensional linear space $\mathcal{W}$ of functions on any surface $\Sigma \in \widetilde{\mathcal{M}}_{g,k}$ which grow at each end of Σ like those in $\mathcal{V}$. There are various ways to realize such a $\mathcal{W}$ — either as an orthogonal complement (in an appropriate Hilbert space, compare [**20**]) to the functions which decay on every end of Σ, or as a quotient space (compare [**22**]) by such functions — but the salient point is that there is a linear isomorphism from $\mathcal{W}$ to $\mathcal{V}^k$, and thus a real analytic map from $\mathcal{W}$ to $\mathcal{U}^k$.

So in order to establish the proposition above, it would suffice to analytically embed $\widetilde{\mathcal{M}}$ into $\mathcal{W}$, at least locally, in the natural way: at a CMC surface Σ, represent any nearby CMC surface in $\widetilde{\mathcal{M}}$ via its asymptotic behavior, which is encoded by a point of $\mathcal{W}$. One of the main results in [**20**] is that this is possible when Σ is *nondegenerate*, that is, provided Σ supports no nontrivial L^2-Jacobi fields. In this case all nearby CMC surfaces form an analytic manifold of dimension $3k$ — half that of $\mathcal{W}$ — and $\widetilde{\mathcal{M}}_{g,k}$ is (locally) a Lagrangian submanifold with respect to a natural symplectic structure defined [**20**] on $\mathcal{W}$ (or on $\mathcal{V}^k$, or even directly on $\mathcal{U}^k$).

For the general case, a neighborhood of Σ in $\widetilde{\mathcal{M}}$ may no longer analytically embed in $\mathcal{W}$, but it can still be embedded as a variety in a larger linear space $\mathcal{W} \times \mathcal{K}$ where $\mathcal{K}$ is a finite-dimensional linear space accounting for the L^2-Jacobi fields on Σ (compare [**20**]). Clearly the composition of analytic maps — inclusion of this variety in $\mathcal{W} \times \mathcal{K}$, followed by projection to $\mathcal{W}$, then the linear isomorphism to $\mathcal{V}^k$, and finally, the analytic local coordinates map into $\mathcal{U}^k$ — is the asymptotes map, which is thus analytic. □

2. A Proper Classifying Map and Its Consequences

One can investigate the topology of $\mathcal{M}$ by studying maps with special properties from $\mathcal{M}$ to other known spaces. As noted in the introduction, this is the method used to classify the triunduloids: there is a homeomorphism Ψ from $\mathcal{M}_{0,3}$ to an open 3-ball, realized by the space of ordered triples of points on $\mathbb{S}^2$ (up to rotations): Ψ identifies the asymptotic necksizes of the triunduloid with the spherical distances for the triple [**11, 12**]. Unfortunately, the triunduloid classifying map Ψ requires the existence of a mirror symmetry which is not necessarily present for CMC surfaces with four or more ends, and so we will introduce a more general classifying map here to investigate them.

We have already observed that the *necksize map* $(s_1, \ldots, s_k) : \mathcal{M}_{g,k} \to \mathbb{R}^k_+$, is real analytic. Now consider the *forgetful map* $\phi : \mathcal{M}_{g,k} \to \mathcal{R}_{g,k}$ which assigns the CMC surface Σ the conformal class $[\Sigma]$ of its induced metric from $\mathbb{R}^3$. In other words, ϕ "forgets" everything but the underlying punctured Riemann surface. It is well known (compare [**1, 15, 32, 31**]) that the Riemann moduli space $\mathcal{R}_{g,k}$ can be given a real analytic structure in two (equivalent) ways, either in terms of holomorphic quadratic differentials with prescribed poles at the punctures ($3g-3+k$ complex analytic local coordinates coming from the cotangent bundle to $\mathcal{R}_{g,k}$) or in terms of finite-area hyperbolic metrics ($6g-6+2k$ real analytic Fenchel–Nielsen coordinates). One could exploit the Hopf differential for the Gauss map of a constant mean curvature surface to derive the following result via holomorphic quadratic differentials, but fortunately Mazzeo, Pacard and Pollack have already derived this [**23**] by proving a nontrivial result in semilinear elliptic partial differential equations;

they show the unique hyperbolic metric on Σ — conformal to its induced metric as a CMC surface — depends analytically on Σ, and thus [**23**]:

PROPOSITION 2.1. *The forgetful map ϕ is real analytic.*

Although it is tempting to work with the forgetful map ϕ alone, there are at least two reasons why another map will be more suitable. First, in order to find homotopy invariants for maps between noncompact spaces, the maps are generally required to be proper. Unfortunately ϕ itself is not proper: compare $\mathcal{R}_{0,3}$, which is a single point (and thus compact), to its preimage $\mathcal{M}_{0,3}$ under ϕ, which is an open (and thus noncompact) 3-ball. In general ϕ fails to be proper because a sequence of CMC surfaces diverges in $\mathcal{M}$ by having asymptotic necksizes tend to zero, while remaining in a compact family of conformal types. Second, the meromorphic data for the DPW representation of a CMC surface with asymptotically unduloidal ends can be geometrically interpreted in terms of flat connections on a rank-2 complex vector bundle over a punctured Riemann surface, with parabolic holonomy determined by a positive number (related to the asymptotic necksize) at each of the punctures (compare [**5, 30**]).

These considerations suggest extending the Riemann moduli space to the *parabolic moduli space* $\mathcal{P}_{g,k} = \mathcal{R}_{g,k} \times \mathbb{R}^k_+$ which assigns a positive real number to each puncture of $[\Sigma]$, and defining the *parabolic classifying map*

$$\Phi : \mathcal{M}_{g,k} \to \mathcal{P}_{g,k} = \mathcal{R}_{g,k} \times \mathbb{R}^k_+$$

as the product of the forgetful map ϕ with the asymptotic necksize map $(s_1, \ldots, s_k)$. The main result of this paper is the following:

THEOREM 2.2. *The parabolic classifying map $\Phi : \mathcal{M} \to \mathcal{P}$ is real analytic and proper.*

Note that the real analyticity of Φ follows immediately from Propositions 1.1 and 2.1. Before giving a proof of properness (see Section 3), we first point out several consequences:

COROLLARY 2.3. *The parabolic classifying map $\Phi : \mathcal{M} \to \mathcal{P}$ is closed. In fact, its image is a closed real analytic subvariety, which represents a (mod 2) homology class in $\mathcal{P}$. The forgetful map ϕ also has closed image in $\mathcal{R}$.*

PROOF. Since $\mathcal{P}$ is a compactly generated Hausdorff space, we can apply a lemma from general topology (see [**12**], Section 5) to show that any proper map to $\mathcal{P}$ is closed; in particular, Φ has closed image $\Phi(\mathcal{M}) \subset \mathcal{P}$. The homology class in $\mathcal{P}$ is carried by the image of Φ, or more precisely, by the proper (mod 2) cycle which is the push-forward via Φ of the fundamental (mod 2) cycle of the real analytic variety $\mathcal{M}$ (compare [**6, 33**]).

The forgetful map ϕ is the composition of Φ with the projection from $\mathcal{P}$ to $\mathcal{R}$. Unfortunately, projection is generally not a closed map, so the proof that the image of ϕ is closed does not follow immediately from the fact that the image of Φ is closed. Nevertheless, the analyticity of the forgetful map ϕ (Proposition 2.1) suffices to show it has closed image (compare [**23**]). □

Next, observe that any component of $\mathcal{M}_{g,k}$ containing a nondegenerate CMC surface is a real analytic variety of dimension $3k - 6$ [**20**], and this coincides with $6g-6+3k$ — the dimension of $\mathcal{P}_{g,k}$ — precisely when $g = 0$. This is a third important

reason to introduce $\mathcal{P}$, since it leads to a degree theory, at least in case of genus zero surfaces (we shall assume $g = 0$ for the remainder of this section, with the exception of the final remark):

THEOREM 2.4. *For any component $\mathcal{N}$ of $\mathcal{M}_{0,k}$ containing a nondegenerate* CMC *surface, the restriction of Φ to $\mathcal{N}$ has a well-defined degree (mod 2). Moreover, this degree is zero.*

PROOF. Because the domain $\mathcal{N}$ and the range $\mathcal{P}$ have the same dimension, the ideas in [**6, 12, 33**] let us define this degree as the number of pre-images (mod 2) for any regular value of Φ. The connectedness of $\mathcal{P}$ and properness of Φ imply that the pre-images of any two regular values can be joined by a compact, 1-dimensional (semi-analytic) variety in $\mathcal{N}$, which necessarily has an even number of endpoints, so the numbers of pre-images agree (mod 2). Then the obvious upper bound on necksize by that of a cylinder shows that this (mod 2)-degree must vanish. □

In particular, it follows that genus zero CMC surfaces generically occur in pairs:

COROLLARY 2.5. *For any nondegenerate surface $\Sigma \in \mathcal{M}_{0,k}$ which is a regular point of Φ, there is a corresponding surface $\Sigma' \in \mathcal{M}_{0,k}$ with the same conformal structure and necksizes.*

Of course, there is no reason to expect there are exactly *two* such surfaces with the same data, only an *even* number (see Section 5 for further discussion of this phenomenon).

Finally, we observe that every conformal type of punctured Riemann sphere is realized by a CMC surface. Although we make no use of this below, this observation is useful from the perspective of the DPW representation, since it guarantees the search for DPW data on a particular punctured Riemann sphere will (in principle) always succeed. Although this is also proven via a rather different argument — using end attachment at (almost) any point of a (nondegenerate) CMC surface, and induction on the number of ends — in [**23**], it is interesting to present this direct argument, based on ideas coming from Kapouleas' original construction of CMC surfaces [**16**]:

THEOREM 2.6. *The forgetful map $\phi : \mathcal{M}_{0,k} \to \mathcal{R}_{0,k}$ is surjective.*

PROOF. From the second part of Corollary 2.3, it suffices to show that a dense set of conformal types is realized. Using a little linear algebra, it is not hard to see that, up to conformal transformations of $\mathbb{S}^2$, any configuration of $k > 1$ points $\xi_1, \ldots, \xi_k$ on $\mathbb{S}^2$ can be balanced with *positive* weights: that is, regarding the ξ_j as unit vectors in $\mathbb{R}^3$, there exists a positive solution $f_1, \ldots, f_k$ to the linear relation (force balancing)

$$f_1\xi_1 + \cdots + f_k\xi_k = 0.$$

For sufficiently small forces f_j, the Kapouleas construction [**16**] then gives a CMC surface Σ of genus zero with k ends asymptotic to unduloids whose jth axis is approximately in the direction ξ_j and whose corresponding necksize is approximately f_j. Because Σ is obtained by quasiconformally attaching punctured disks (corresponding to the unduloid ends) at the boundaries of arbitrarily small disks about each ξ_j, its conformal structure $[\Sigma]$ can be made to lie within any prescribed neighborhood of $[\mathbb{S}^2 - \{\xi_1, \ldots, \xi_k\}]$ in $\mathcal{R}_{0,k}$. □

It is quite easy to see that ϕ cannot be surjective for $g > 0$. For instance, when $k = 3$ a CMC surface Σ must have a plane of reflection symmetry, so the Riemann surface $[\Sigma] = \phi(\Sigma)$ must have a certain anti-conformal involution with more than g component curves in the fixed point set; this necessary condition, for example, allows only rectangular tori in $\mathcal{M}_{1,3}$.

There is an interesting generalization of Φ and $\mathcal{M}_{g,k}$ which admits results analogous to Theorems 2.2 and 2.4 (and their Corollaries) in case $g > 0$. To understand this, recall that any $\Sigma \in \mathcal{M}_{g,k}$ bounds an immersed genus-g handlebody Ω; because Σ closes when we follow a loop in Ω means we have the *trivial* representation of $\pi_1(\Omega)$ into the group of rigid motions. The new idea is to consider *all* such representations, where Σ may no longer have closed periods around the g generators of $\pi_1(\Omega)$; the corresponding space $\mathcal{M}^*_{g,k}$ of *half-periodic* CMC surfaces forms a variety whose (virtual) dimension is $6g$ more than that of $\mathcal{M}_{g,k}$. A fundamental domain for a half-periodic surface still defines a punctured Riemann surface, and it still has k asymptotic necksizes; thus we get an *extended parabolic classifying map* Φ^* in the obvious way:

REMARK 2.7. *The extended parabolic classifying map $\Phi^* : \mathcal{M}^*_{g,k} \to \mathcal{P}_{g,k}$ is real analytic and proper. Any component $\mathcal{N}^*$ of $\mathcal{M}^*$ containing a nondegenerate half-periodic* CMC *surface is $(3k - 6 + 6g)$-dimensional (the same as for $\mathcal{P}$), and thus the restriction of Φ^* to $\mathcal{N}^*$ has a well-defined (mod 2) degree, which must vanish.*

We plan to explore the consequences of these observations about Φ^*, along with the problem of closing periods of $\Sigma \in \mathcal{M}^*_{g,k}$, in a future paper.

3. Proof of Properness

The proof that the parabolic classifying map $\Phi : \mathcal{M} \to \mathcal{P}$ is proper relies on the *a priori* estimates for CMC surfaces developed in [**18**], namely:

LEMMA 3.1. *Any $\Sigma \in \mathcal{M}_{g,k}$ lies in a uniform tubular neighborhood of a piecewise linear 1-complex with k rays and at most $k + 3g - 3$ finite segments. Moreover, a sequence $\Sigma(i)$ of* CMC *surfaces, all of which lie in a compact family of uniform tubular neighborhoods, diverges in $\mathcal{M}_{g,k}$ if and only if the length $\ell(i)$ of the shortest nontrivial loop on $\Sigma(i)$ tends to zero (compare* [**18**]*).*

A compact family of uniform tubular neighborhoods is characterized by requiring the lengths of all segments to be uniformly bounded above, although these lengths may go to zero, as may the angles between the rays or edges. For sequences of CMC surfaces $\Sigma(i)$ lying in such a compact family of tubular neighborhoods, uniform linear area and total curvature estimates [**18**] still hold, and uniform pointwise curvature estimates also hold under the hypothesis that the $\ell(i)$ are bounded away from zero (using blow-up arguments of the kind we outline below).

The proof also depends on a description of divergent sequences of punctured Riemann surfaces in terms of the conformal moduli of certain nontrivial annuli. We say that an embedded annulus A on a punctured Riemann surface is *essential* (or nonperipheral) provided it is not homotopic to a single point of the surface, nor to a single puncture. Such an annulus then has a (necessarily finite) conformal modulus $[A] = m \in (1, \infty)$ defined by conformally mapping A into $\mathbb{C}$ so that one boundary component goes to the unit circle, and the other goes to the circle of radius m. The basic fact which we will use is the following:

LEMMA 3.2. *A sequence of punctured Riemann surfaces* $[\Sigma(i)]$ *diverges in* $\mathcal{R}_{g,k}$ *if and only if there exist essential annuli* $A(i) \subset \Sigma(i)$ *whose moduli* $[A(i)] = m(i) \to \infty$ *as* $i \to \infty$.

PROOF. This is a standard result in Teichmüller theory, using extremal length (compare, for instance, [**32**]) or Fenchel–Nielsen coordinates (compare [**1, 15, 31**]), although it seems difficult to find the exact statement in the literature, particularly for the case of a punctured Riemann surface. In effect, the existence of a family of essential annuli $A(i) \subset \Sigma(i)$ whose moduli diverges to infinity is equivalent to having a family of geodesics (in the homotopy class of $A(i)$) whose length in the conformal hyperbolic metric on $[\Sigma(i)]$ tends to zero. The Mumford–Bers compactness theorem (compare [**26, 4**]) implies that only these $[\Sigma(i)]$ diverge in $\mathcal{R}_{g,k}$. □

Proof of the properness theorem. To show that the classifying map Φ is proper it will suffice to prove the contrapositive, namely, that a divergent sequence $\Sigma(i)$ in $\mathcal{M}_{g,k}$, whose asymptotic necksizes $s_1(i), \ldots, s_k(i)$ are uniformly bounded away from zero, must contain divergent essential annuli as above, implying $\Phi(\Sigma(i))$ is divergent in $\mathcal{P}_{g,k}$.

So suppose a sequence $\Sigma(i)$ diverges in $\mathcal{M}$, with all asymptotic necksizes $s_j(i) \geq \ell > 0$. Then we have two (non-exclusive) alternatives:

- $L(i) \to \infty$, where $L(i)$ is the length of the longest finite segment in the 1-complex whose uniform tubular neighborhood contains $\Sigma(i)$; or
- $\ell(i) \to 0$, where $\ell(i)$ is the length of some nontrivial closed curve, which is not homotopic to an end of $\Sigma(i)$.

In the first case, the Alexandrov symmetrization method [**18, 19**] shows there is a sequence of essential annuli $A(i) \subset \Sigma(i)$ which are approximately unduloidal on greater and greater lengths (comparable to $L(i) \to \infty$, though the constants are not explicitly computable). Since any unduloid is periodic, the essential annuli approximating an unduloid have moduli $[A(i)]$ which grow exponentially with length, and thus diverge.

In the second case, we make a blow-up argument (compare [**12, 18, 19**]) to get a non-flat, finite total curvature minimal surface M bounding an immersed 3-manifold. There are several subcases to consider.

First, suppose M is embedded, so its top and bottom ends are catenoidal. Let A be a punctured-disk representative for such an end on M. Thus as $i \to \infty$, the end A is approximated by rescaled copies of annuli $A(i) \subset \Sigma(i)$ with moduli $[A(i)] \to [A] = \infty$. Since A is catenoidal, it has nonzero force (see [**12, 18, 19**]), as must the $A(i)$, which implies they are homologically nontrivial. And although the forces (and thus the necksizes) of the $A(i)$ tend to zero, because we are assuming the necksizes of the ends of $\Sigma(i)$ are bounded away from zero, it follows that the $A(i)$ are not homotopic to any end of $\Sigma(i)$, and so must be essential.

Next, suppose the blow-up minimal surface M is only immersed. If each end of M is simple, then each must be either catenoidal or planar. An application of the strong half-space theorem [**14**], using the immersed 3-manifold in place of $\mathbb{R}^3$, shows that not all ends of M can be planar; so we get a catenoidal end with nonzero force, and a punctured-disk representative A, to use just as in the argument above.

Finally, in case an end of the blow-up minimal surface M is non-simple, we use a "slicing trick" (compare with the instructive example in Section 4): we find a smoothly embedded plane P in $\mathbb{R}^3$, which meets a representative annulus A for

this end in a curve which cannot bound a compact immersed surface on P; it will then follow that A must have been approximated by rescaled essential annuli $A(i)$ on $\Sigma(i)$. Such a P can be constructed by taking a large radius cylinder with axis parallel to the asymptotic unit normal of A, and capping off by a hemisphere very high up on this cylinder (choose this normal to be on the side which bounds the immersed 3-manifold and think of this side as up). Its curve of intersection with the rescaled approximating $A(i)$ has turning number in P equal to the asymptotic winding number of A, which is at least 2. Endow P with an intrinsically flat metric and apply the Gauss–Bonnet formula to see this curve cannot bound a compact immersed surface in P: its winding number would equal the Euler number of a connected surface with boundary, which is at most 1. This shows that this curve of intersection cannot bound in the 3-manifold itself. It follows that the annuli $A(i) \subset \Sigma(i)$ approximating A must have been homotopically nontrivial in the immersed 3-manifold bounded by $\Sigma(i)$ as well, and thus essential in $\Sigma(i)$.

This completes the proof of our main result.

4. An Instructive Example: the Enneper Surface

The reader may find it instructive to apply this slicing trick to decide whether a particular immersed surface can actually be extended to a properly immersed 3-manifold. Take, for example, the Enneper surface, a minimally immersed $\mathbb{R}^2$ in $\mathbb{R}^3$; it has one end, of winding number 3. Note that, in general, the Frankel–Lawson [**9, 21**] argument shows the inclusion of the bounding minimal surface into the 3-manifold induces an epimorphism on fundamental groups. So if Enneper were to bound, it would bound an immersed 3-ball (with one boundary puncture, corresponding to the one end of Enneper). But the slicing trick implies that the intersection curve of Enneper with a slicing plane P, constructed as above, has turning number 3, and therefore the curve would be homotopically nontrivial in the 3-ball, which is absurd. Thus:

COROLLARY 4.1. *The Enneper surface is not immersed in the sense of Alexandrov.*

5. Some Open Problems

To conclude, it may be interesting to contemplate the following (open) problems which are related to the results above:

• What is the image of the parabolic classifying map Φ in $\mathcal{P}$, or its intersection with a slice of fixed conformal type? In other words, how does the set of allowable necksizes depend on the underlying punctured Riemann surface? For $g = 0$ and $k = 3$, there is only one conformal type, so these questions coincide: the image of Φ is known to be the simplex determined by the spherical triangle inequalities for the necksizes of a triunduloid [**12**]. The general case seems closely related to the Biswas inequalities [**5, 30**] which arise in the study of flat connections on parabolic vector bundles.

• Is the cardinality of $\Phi^{-1}([\Sigma])$ finite, and is $2^{|\chi(\Sigma)|}$ an upper bound? The absolute value of the Euler characteristic $|\chi(\Sigma)| = 2g - 2 + k$ is the number of trousers in a decomposition of Σ. Thus the possible choices of "innie" or "outie" configurations at each trousers nexus saturates this bound. (This whimsical terminology — suggested by the proximity of each trousers nexus to the umbilic points on the surface —

describes either a neck-like or bubble-like geometry near the trousers nexus.) Again, for $g = 0$ and $k = 3$ we have $|\chi(\Sigma)| = 1$, and this bound is sharp [**12**]: Φ is two-to-one, except at the maximal necksize "fold" where it is one-to-one.

• Is $\mathcal{M}_{g,k}$ connected? And — at least on some connected component of $\mathcal{M}$ — how far from an isomorphism is the homomorphism on fundamental groups induced by Φ — or, equivalently, by ϕ? In case $g = 0$ it follows from [**12**] and [**29**] (or more directly from [**23**]) that, on what is presumably the only component — the one containing the *coplanar* CMC surfaces [**13**] — of $\mathcal{M}_{0,k}$, this is an epimorphism for any k.

• Does $\mathcal{M}_{g,k}$ carry a complete metric of nonpositive curvature, in some suitable sense, and is it thus an Eilenberg–MacLane $K(\pi, 1)$-space? The identification of $\mathcal{M}_{0,3}$ with the (hyperbolic) 3-ball in [**12**] is very suggestive, but it is not clear where to find a metric (hyperbolic or otherwise) directly from the CMC geometry, even in this special case. Note that $\mathcal{P}_{g,k}$ does carry such a metric, and is a $K(\pi, 1)$ with $\pi = \pi_1(\mathcal{R}_{g,k})$, which when $g = 0$ is a (colored) braid group.

• Can one construct a properly (Alexandrov immersed or) embedded *minimal* surface of finite topology which is degenerate (in the sense of [**22, 25**]), either explicitly, or by global methods on minimal surface moduli space (for instance, might one exhibit a classifying map whose critical points must be degenerate minimal surfaces, and which is forced to have a critical point via Morse theory)? And if one works L^2-orthogonal to the extra Jacobi fields, can one glue unduloids to the ends of the minimal surface to get a degenerate CMC surface in $\mathcal{M}$? There exist minimal surfaces with all ends planar which *are* degenerate, but these are not immersed in the sense of Alexandrov; similar comments apply to the immersed constant mean curvature tori, first constructed by Wente [**34**], and later classified by Pinkall and Sterling [**28**].

• Is there a computable way to detect when an immersed surface in $\mathbb{R}^3$ is immersed in the sense of Alexandrov? As we have noted above, the trick of slicing with a smoothly embedded plane, and then deciding by winding number considerations that the intersection curves cannot bound an immersed domain in this plane, provides a *necessary* condition, but it would be interesting to have a general obstruction theory. The corresponding *sufficient* condition is more difficult to formulate: for example, if all such planes slice the surface in curves that bound immersed surfaces in this plane, it is tempting to think that ideas from ambient Morse theory could be used to construct the immersed 3-manifold by carefully stacking such slices. Of course, one problem is the potential non-uniqueness of the immersed surface which extends the immersed curve in each slice.

References

[1] W. Abikoff, The real analytic theory of Teichmüller space, Springer LNM 820 (1980).
[2] A. D. Alexandrov, Uniqueness theorems for surfaces in the large V, Vestnik Leningrad Univ. 19:13 (1958), 5–8; Amer. Math. Soc. Transl. (Ser. 2) 21 (1962), 412–416.
[3] A. D. Alexandrov, A characteristic property of spheres, Ann. Mat. Pura Appl. 58 (1962), 303–315.
[4] L. Bers, A remark on Mumford's compactness theorem, Israel J. Math. 12 (1972), 400–407.
[5] I. Biswas, On the existence of unitary flat connections over the punctured sphere with given local monodromy around the punctures, Asian J. Math. 3 (1999), 333–344.
[6] A. Borel and A. Haefliger, La classe d'homologie fondamentale d'un espace analytique, Bull. Soc. Math. France 89 (1961), 461–513.

[7] C. Delaunay, Sur la surface de révolution, dont la courbure moyenne est constante, Journal de mathématiques 6 (1841), 309–320.
[8] J. Dorfmeister, F. Pedit, and H. Wu, Weierstrass type representation of harmonic maps into symmetric spaces, Comm. Anal. Geom. 6 (1998), 633–667.
[9] T. Frankel, On the fundamental group of a compact minimal submanifold, Annals of Math. 83 (1966), 68–73.
[10] K. Große-Brauckmann, New surfaces of constant mean curvature, Math. Zeit. 214 (1993), 527–565.
[11] K. Große-Brauckmann, R. Kusner, and J. M. Sullivan, Constant mean curvature surfaces with three ends, Proc. Natl. Acad. Sci. USA 97 (2000), 14067–14068.
[12] K. Große-Brauckmann, R. Kusner, and J. M. Sullivan, Triunduloids: Embedded constant mean curvature surfaces with three ends and genus zero, J. Reine Angew. Math. 564 (2003), 35–61.
[13] K. Große-Brauckmann, R. Kusner, and J. M. Sullivan, Coplanar constant mean curvature surfaces (preprint, 2003).
[14] D. Hoffman and W. H. Meeks, The strong halfspace theorem for minimal surfaces, Invent. Math. 101 (1990), 373–377.
[15] Y. Imayoshi and M. Taniguchi, An introduction to Teichmüller spaces, Springer-Verlag, Tokyo, 1992.
[16] N. Kapouleas, Complete constant mean curvature surfaces in Euclidean three-space, Annals of Math. 131 (1990), 239–330.
[17] M. Kilian, I. McIntosh, and N. Schmitt, New constant mean curvature surfaces, Experiment. Math. 9 (2000), 595–611.
[18] N. Korevaar and R. Kusner, The global structure of constant mean curvature surfaces, Invent. Math. 114 (1993), 311–332.
[19] N. Korevaar, R. Kusner, and B. Solomon, The structure of complete embedded surfaces with constant mean curvature, J. Diff. Geom. 30 (1989), 465–503.
[20] R. Kusner, R. Mazzeo, and D. Pollack, The moduli space of complete embedded constant mean curvature surfaces, Geom. Funct. Anal. 6 (1996), 120–137.
[21] H. B. Lawson, The unknottedness of minimal embeddings, Invent. Math. 11 (1970), 183–187.
[22] R. Mazzeo and F. Pacard, Constant mean curvature surfaces with Delaunay ends, Comm. Anal. Geom. 9 (2001), 169–237.
[23] R. Mazzeo, F. Pacard, and D. Pollack, The conformal theory of Alexandrov embedded constant mean curvature surfaces, pp. 525–559 in this volume.
[24] W. H. Meeks, The topology and geometry of embedded surfaces of constant mean curvature, J. Diff. Geom. 27 (1988), 539–552.
[25] S. Montiel and A. Ros, Schrödinger operators associated to a holomorphic map, Global differential geometry and global analysis, Springer LNM 1481 (1991), 148–174.
[26] D. Mumford, A remark on Mahler's compactness theorem, Proc. Amer. Math. Soc. 28 (1971), 289–294.
[27] R. Osserman, A survey of minimal surfaces, Van Nostrand Reinhold, New York, 1969.
[28] U. Pinkall and I. Sterling, On the classification of constant mean curvature tori, Annals of Math. 130 (1989), 407–451.
[29] J. Ratzkin, An end-to-end gluing construction for surfaces of constant mean curvature, PhD dissertation, U. Washington, 2001.
[30] N. Schmitt, Constant mean curvature trinoids (in preparation).
[31] M. Seppälä and T. Sorvali, Geometry of Riemann surfaces and Teichmüller spaces, North-Holland Mathematics Studies 169, North-Holland Publishing, Amsterdam, 1992.
[32] K. Strebel, Quadratic Differentials, Ergibnisse der Mathematik 3.5, Springer-Verlag, Berlin, 1984.
[33] D. Sullivan, Combinatorial invariants of analytic spaces, Proc. of Liverpool Singularities Symp., Springer LNM 192 (1971), 165–169.
[34] H. Wente, Counterexample to a conjecture of H. Hopf, J. M Pacificath. 121 (1986), 193–243.

Center for Geometry, Analysis, Numerics & Graphics (GANG), Department of Mathematics, University of Massachusetts, Amherst, MA 01003
E-mail address: kusner@math.umass.edu

Clay Mathematics Proceedings
Volume 2, 2004

Uniqueness of the Riemann Minimal Surfaces

Joaquín Pérez
(Joint work with William H. Meeks III
and Antonio Ros)

Abstract. The Riemann minimal examples constitute the only known properly embedded genus zero minimal surfaces in $\mathbb{R}^3$ with infinitely many ends. They have been studied over the years, giving rise to many statements directed to conclude their uniqueness. Here we revise some of them, and we point out some new results in that direction.

1. Posing the Problem. The Solution in the Periodic Setting

The surfaces we will handle appeared for the first time in a posthumous paper by Riemann [**23**], where he classified all minimal surfaces in $\mathbb{R}^3$ foliated by circles and straight lines in parallel planes. These surfaces are the plane, the helicoid, the catenoid and a one-parameter family $\{R_t\}_{t>0}$ of surfaces with infinitely many ends asymptotic to parallel planes uniformly distributed. Since then, they have been known as the *Riemann minimal surfaces.* Each R_t is a properly embedded minimal surface in $\mathbb{R}^3$ which intersects horizontal planes at integer heights in parallel straight lines $\{l_k\}_{k\in\mathbb{Z}}$ orthogonal to the plane $\{x_2 = 0\}$; the remaining horizontal sections are circles of radius varying from $+\infty$ at the lines $\{l_k\}_k$ to a minimum radius exactly at heights in $1/2 + \mathbb{Z}$. The symmetry group of R_t consists of a reflection symmetry with respect to the plane $\{x_2 = 0\}$, 180^o rotations about the lines l_k, a translation by vector $T_t = (t, 0, 2)$ and 180^o rotations about lines $l_k + \frac{1}{4}T_t$ that cut the surface orthogonally at antipodal points of the circles of minimum radius; see Figure 1. Each R_t is conformally a vertical cylinder punctured at integer heights; such punctures correspond to the ends. In particular, R_t has genus zero (i.e., R_t is a *planar domain*) and two limit ends. Moreover, for any positive integer n the quotient surface R_t/nT_t has genus one in $\mathbb{R}^3/nT_t$ and $2n$ planar ends.

The main open question about these surfaces is this:

Question 1. *Are $\{R_t\}_{t>0}$ the unique properly embedded minimal planar domains in $\mathbb{R}^3$ with two limit ends?*

The main difficulty of Question 1 comes from the fact that there are no assumptions on the size of the symmetry group of the surfaces. This problem has been attacked by many mathematicians assuming different symmetry properties; see for instance [**4, 6, 9, 10, 19, 20, 21, 24**]. Among this type of results, perhaps the strongest one is in Meeks, Pérez and Ros [**12**]:

©2004 Clay Mathematics Institute

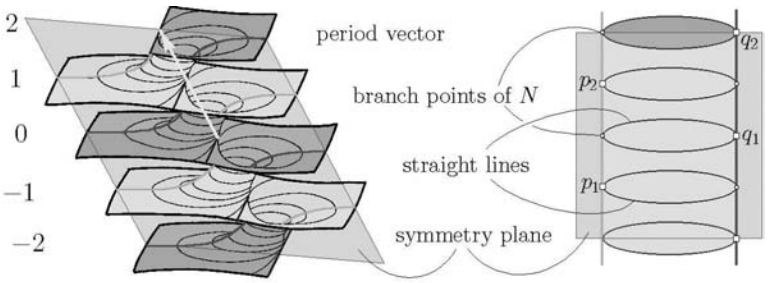


FIGURE 1. Left: One of Riemann's examples. Right: A conformal representation.

THEOREM 1. *Let $M \subset \mathbb{R}^3$ be a properly embedded periodic minimal planar domain with two limit ends. Then, M is one of the Riemann minimal examples.*

PROOF. As M has more than one end, it admits a tangent plane at infinity[1], which, in the sequel, will be supposed horizontal. In this setting, Frohman and Meeks [**5**] showed that the ends are ordered by heights. By periodicity, M has a top and a bottom limit end, thus all middle ends are simple and annular. As the symmetry group of M is infinite, the structure theorem of Callahan, Hoffman and Meeks [**1**] guarantees the existence of either a nontrivial screw motion or a translation that preserves M, and in both cases the quotient surface of M modulo this symmetry has genus one and a finite number of planar ends. Using properties of the flux, Pérez and Ros [**21**] excluded the screw motion symmetry in this setting, thus Theorem 1 follows directly from the next result:

THEOREM 2 (Meeks, Pérez and Ros [**12**]). *Let $M \subset \mathbb{R}^3/T$ be a properly embedded minimal surface with genus one and a finite number of planar ends, T being a nontrivial translation. Then, M is a quotient of a Riemann minimal surface.*

SKETCH OF PROOF. Fix a positive integer n. Let $\mathcal{S}$ be the space of all properly embedded minimal tori in a quotient of $\mathbb{R}^3$ by a translation T (depending on the surface) with $2n$ horizontal planar ends. Denote by $\mathcal{R} \subset \mathcal{S}$ the family of Riemann minimal examples, then the statement reduces to prove that $\mathcal{R} = \mathcal{S}$.

The general description of an element $M \in \mathcal{S}$ is as follows: The $2n$ planar ends of M are separated by a positive distance (maximum principle at infinity [**8**] or [**16**]). Because of embeddedness, consecutive ends of M have opposite (vertical) limit normal vectors, so we can label $(p_1, q_1, \ldots, p_n, q_n)$ the ends of M by increasing order of the heights with $N(p_i) = (0, 0, -1)$, $N(q_i) = (0, 0, 1)$, $1 \leq i \leq n$; see Figure 2.

The stereographic projection of the Gauss map of M, say g, extends as a meromorphic map of degree $2n$ to the torus $\overline{M}$ obtained by attaching to M its planar ends. Thus $g(p_i) = 0^2$, $g(q_i) = \infty^2$, M has no points with vertical normal vector and each compact horizontal section γ is a Jordan curve. As the flux around a planar end is zero, the flux of M along γ, denoted by $\mathrm{Flux}(M, \gamma)$, does not

[1]Callahan, Hoffman and Meeks [**1**] proved that if $M \subset \mathbb{R}^3$ is a properly embedded minimal surface with more than one end, then there exists a complete, embedded, minimal, finite total curvature end E contained in one of the regions of $\mathbb{R}^3 - M$. The plane through the origin orthogonal to the limit normal vector to E is called the *limit tangent plane at infinity* to M (it only depends on M).

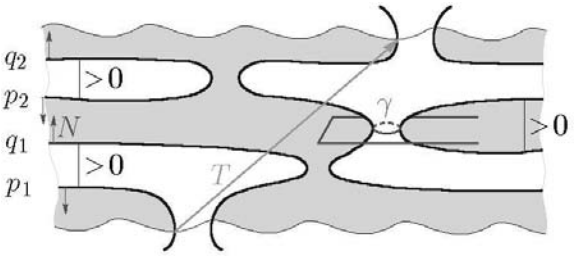


FIGURE 2. A generic $M \in \mathcal{S}$.

depend on the height of γ. By transversality, $\mathrm{Flux}(M, \gamma)$ cannot be horizontal. These facts suggest introducing in $\mathcal{S}$ the following normalization: We rescale and orient all surfaces $M \in \mathcal{S}$ so that $\mathrm{Flux}(M, \gamma)$ has third coordinate equal to 1. Thus, $\mathrm{Flux}(M, \gamma)$ is completely determined by the so-called *Flux map*

$$F : \mathcal{S} \to \mathbb{R}^2 - \{(0,0)\} \ / \ F(M) = (a, b) \text{ if } \mathrm{Flux}(M, \gamma) = (a, b, 1).$$

The fact that F omits the value $(0, 0)$ follows from application of the López–Ros deformation; see [**21**]. This Flux map has three important properties. Two of them are of topological nature, when we endow $\mathcal{S}$ with the uniform topology on compact sets:

- F is a proper map.
- F is an open map.
- There exists $\varepsilon > 0$ such that if $|F(M)| < \varepsilon$, then $M \in \mathcal{R}$.

The proofs of the above three properties are delicate and we will not give them here. They strongly depend on curvature estimates for sequences $\{M_i\} \subset \mathcal{S}$ such that $\{F(M_i)\}$ is bounded. These curvature estimates together with an analysis of the possible limits of sequences $\{M_i\} \subset \mathcal{S}$ give the properness of F. The openness follows from extending F to a holomorphic map P from the complex n-dimensional manifold $\mathcal{W}$ of all allowed Weierstrass data (without killing the periods) to $\mathbb{C}^n$ (P essentially expresses the conditions for a generic point in $\mathcal{W}$ to produce a minimal immersion, recall that the number of ends is $2n$), and applying to P the local open mapping theorem for holomorphic finite[2] maps of several variables. The fact that P is a finite map around any $M \in \mathcal{S} \subset \mathcal{W}$ depends on the properness of F. Finally, the third property, the uniqueness result for almost vertical flux, is proved by extending P holomorphically to a suitable point in the closure of the moduli space $\mathcal{W}$ (more precisely, to the limit point of any sequence $\{M_i\} \subset \mathcal{S}$ with $\{F(M_i)\} \to 0$), and using the inverse mapping theorem to conclude bijectivity of P in a neighborhood of that point.

On the other hand, the topological subspace $\mathcal{R} \subset \mathcal{S}$ of Riemann surfaces has two key properties:

- $\mathcal{R}$ is closed in $\mathcal{S}$.
- $\mathcal{R}$ is open in $\mathcal{S}$.

[2]A holomorphic map between complex manifolds of the same dimension is said to be *finite* if the inverse image of any point is finite.

Closedness of $\mathcal{R}$ in $\mathcal{S}$: any Riemann example is characterized by being foliated by circles and lines in parallel planes, a condition preserved by limits in the uniform topology on compact sets. Openness of $\mathcal{R}$ in $\mathcal{S}$: Pérez [**20**] characterized $\mathcal{R}$ as the only *nondegenerate* minimal surfaces in $\mathcal{S}$, this nondegeneracy being defined in terms of the spaces of Jacobi functions on such surfaces. As the condition to be nondegenerate is open in the generic setting (see [**20, 22**]), $\mathcal{R}$ must be open in $\mathcal{S}$.

With these ingredients in mind, the proof of Theorem 2 is easy. By contradiction, if $\mathcal{R} \neq \mathcal{S}$ then $\mathcal{S} - \mathcal{R}$ is open and closed in $\mathcal{S}$, thus the restriction $F : \mathcal{S} - \mathcal{R} \to \mathbb{R}^2 - \{(0,0)\}$ must be an open and proper map, in particular it is surjective. This contradicts the third property of F and finishes the proof of Theorem 2.

2. Progress in the Nonperiodic Setting

In the sequel we will give some ideas that seem to be useful to find a solution of Question 1, when no periodicity is assumed. All that follows is part of work in progress [**13**].

Let $\mathcal{S}$ be the space of all properly embedded minimal surfaces with genus zero and two limit ends. The description of any $M \in \mathcal{S}$ is as follows. As M has more than one end, it admits a tangent plane at infinity, which we will assume to be horizontal. By the work of Frohman and Meeks [**5**], the ends of M are ordered by heights. Thanks to a recent paper of Collin, Kusner, Meeks and Rosenberg [**3**] all middle ends are simple, thus the top and bottom ends are limit ends, and the remaining middle ends can be labeled E_i, $i \in \mathbb{Z}$ (E_i is above E_j whenever $i > j$). Topologically, M is an infinite vertical cylinder $\mathcal{C}$ punctured in a discrete set whose height diverges to $+\infty$ and $-\infty$.

The following easy argument shows that M has no catenoidal ends: suppose that M contains a catenoidal end E_0, say with positive logarithmic growth $c_0 > 0$. As M is embedded, all ends E_i with $i > 0$ must be catenoidal with logarithmic growth $c_i \geq c_0$. Now, take a sequence of closed curves $\{\gamma_i \mid i \geq 0\} \subset M$, all homologous to the generator of $H^1(\mathcal{C}, \mathbb{Z})$, and such that each γ_i divides M in two noncompact components Ω_i^-, Ω_i^+, with $\Omega_0^+ \supset \Omega_1^+ \supset \Omega_2^+ \supset \cdots$ and $E_i \subset \Omega_i^+$ for all $i \geq 0$; see Figure 3. We can also assume that the Gauss map is never vertical along the γ_i-curves.

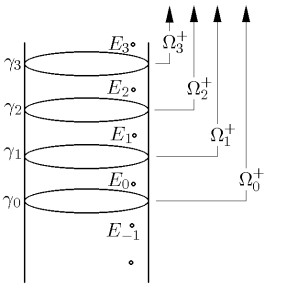


FIGURE 3. A topological picture of M.

For each $i > 0$, the domain $\Omega_0^+ - \Omega_i^+$ has boundary $\gamma_0 \cup \gamma_i$ and contains i catenoidal ends, all with positive third coordinate of flux. A direct application of

the divergence theorem shows that the total flux of $\Omega_0^+ - \Omega_i^-$ can vanish only if the sum of the logarithmic growths of ends inside $\Omega_0^+ - \Omega_i^-$ converges to some finite value, in contradiction with $c_i \geq c_0 > 0$ for all i. Thus, all middle ends of M must be planar. More generally, Collin, Kusner, Meeks and Rosenberg [**3**] have proved that for a properly embedded minimal surface $M \subset \mathbb{R}^3$ with two limit ends, every middle end of M can be represented by a proper subdomain with compact boundary contained in a horizontal slab, from where the nonexistence of catenoidal ends in our settings follows directly.

To discover the conformal structure of M, divide the surface by cutting with a transversal horizontal plane $\{x_3 = c\}$ between two consecutive ends. Define $M^+ = M \cap \{x_3 \geq c\}$ and $M^- = M \cap \{x_3 \leq c\}$, so that $M = M^+ \cup M^-$. By a theorem of Collin, Kusner, Meeks and Rosenberg [**3**], M^+ is parabolic in the sense that bounded harmonic functions on M^+ are determined by their boundary values. After attaching the simple ends inside M^+, we will obtain a new parabolic surface $\overline{M^+}$ that is topologically a punctured disk, thus $\overline{M^+}$ is also conformally a punctured disk and the same holds for M^-. Now we glue M^+ and M^- along the compact section at height c, so we get $M \cup \{\text{planar ends}\} = \mathbb{C} - \{0\}$ conformally. In particular, the conformal structure of M is $\mathbb{C} - [\{0\} \cup \{p_i\}]$, where $\{p_i\} \subset \mathbb{C} - \{0\}$ is a sequence that accumulates at 0 and ∞, each p_i corresponding to a planar end E_i.

Furthermore, the third coordinate function x_3 extends to a proper harmonic function on the punctured disk $\overline{M^+} = \{0 < |z| \leq 1\}$ with constant boundary values, thus $x_3(z) = \lambda \log |z|$ for a certain $\lambda < 0$ and consequently, g misses $0, \infty$ in M^+ (analogously for M^-). Now it is not difficult to obtain (up to a rescaling of M so that the third coordinate of the flux along any compact horizontal section is 1) the Weierstrass data $(g, \phi_3 = \frac{\partial x_3}{dz} dz = \frac{dz}{z})$, $z \in \mathbb{C} - \big(\{0\} \cup \{p_i\}\big)$. Finally, all the poles and zeros of the Gauss map at the ends are double, because otherwise the gradient of x_3 would vanish at those points.

For any $M \in \mathcal{S}$ we can now consider the *Shiffman function* of M, which is a Jacobi function[3] defined as

$$u = \Lambda \frac{\partial \kappa}{\partial y} = \operatorname{Im}\left(\frac{3}{2} \Big(\frac{g'}{g} \Big)^2 - \frac{g''}{g} - \frac{1}{|g|^2 + 1} \Big(\frac{g'}{g} \Big)^2 \right),$$

where $(g(\xi), \phi_3 = d\xi)$, $\xi = x + iy \in \mathbb{C}$ is the Weierstrass data of M in the ξ-plane (the relation between ξ and the above z is $z = e^{\xi}$), $ds^2 = \Lambda^2 |d\xi|^2$ is the induced metric, κ is the curvature of the planar section $\{x = \text{constant}\}$ and the prime stands for derivation with respect to the ξ-parameter. As Λ is a positive function, the zeros of u coincide with the zeros of the derivative of κ with respect to y (or to any other parameter in the planar section $\{x = \text{constant}\}$). As a consequence, u vanishes identically if and only if M is foliated by circles or lines in horizontal planes, that is, $M \in \mathcal{R}$.

The following statement shows that the Riemann minimal examples admit another characterization in terms of the Shiffman function.

PROPOSITION 1. *Let $M \in \mathcal{S}$. If the Shiffman function of M is a linear function of the Gauss map, then $M \in \mathcal{R}$.*

---

[3]A *Jacobi function* is a solution of the *Jacobi equation* $\Delta u + \|\nabla N\|^2 u = 0$ on M, where M is the Gauss map of the surface. The *Jacobi operator* is defined as $L = \Delta + \|\nabla N\|^2$.

PROOF. Let

$$f = \frac{3}{2}\Big(\frac{g'}{g}\Big)^2 - \frac{g''}{g} - \frac{1}{|g|^2+1}\Big(\frac{g'}{g}\Big)^2,$$

so $u = \mathrm{Im}(f)$ is the Shiffman function of M. We claim that if $u = \langle N, a\rangle$ where N is the Gauss map of M and $a \in \mathbb{R}^3$, then $f = \langle N, z_0\rangle$ for certain $z_0 \in \mathbb{C}$.

Recall the well-known correspondence between the linear space $\mathcal{J}(M)$ of all Jacobi functions on a given minimal surface M and the linear space of branched minimal immersions with the same Gauss map N as M: each $v \in \mathcal{J}(M)$ defines a conformal harmonic map (thus a branched minimal immersion) $X_v = \nabla v + vN : M \to \mathbb{R}^3$ where ∇v stands for the spherical gradient of v through the Gauss map. The Gauss map of X_v is N and its support function is $\langle X_v, N\rangle = v$. Furthermore, $v \in \mathcal{J}(M)$ is linear[4] if and only if X_v is constant, and if $v^* \in \mathcal{J}(M)$ is a Jacobi-conjugate[5] of $v \in \mathcal{J}(M)$, then X_v, X_{v^*} are conjugate branched minimal immersions. For details, see Montiel and Ros [**18**]. Coming back to our setting, if the Shiffman function u is linear on M, then the corresponding branched minimal immersion X_u is constant, and thus its minimal conjugate $(X_u)^* = X_{u^*}$ is also constant, which implies that $u^* = -\mathrm{Re}(f)$ is also linear. This proves our claim.

Once we know that $f = \langle N, z_0\rangle$, a simple calculation (using the expression of f and the relationship between N and g) yields an equation of the type

$$(g')^2 = g(\alpha g^2 + \beta g + \delta) \tag{2–1}$$

for certain $\alpha, \beta, \delta \in \mathbb{C}$. Then, (g, g') is a branched holomorphic covering from the cylinder $\mathcal{C} = M \cup \{\text{planar ends}\}$ onto the compact Riemann surface

$$\Sigma = \{(z, w) \in \overline{\mathbb{C}}^2 \mid w^2 = z(\alpha z^2 + \beta z + \delta)\}.$$

On the other hand, by the uniqueness of solutions of the holomorphic first order differential equation (2–1) with a given initial value, if a, b are distinct points in $\mathcal{C}$ with $g(a) = g(b) \neq 0, \infty$, then $g(z + b - a) = g(z)$ for all z. As we can always find distinct points $a, b \in \mathcal{C}$ with this property (take points close enough to different ends with the same normal limit vector), we conclude that $b - a$ is a period of g and also of g'. This implies that the pair (g, g') is well defined on the torus $\mathcal{C}/(b-a)$, and gives rise to a holomorphic map $\pi : \mathcal{C}/(b-a) \to \Sigma$.

If Σ is a sphere, then the total ramification number of π is $B(\pi) = 2\mathrm{degree}(\pi)$. But the pullback by π of the meromorphic differential $\frac{dz}{w}$ on Σ is $\frac{dg}{g'} = d\xi$, which has no poles nor zeros on $\mathcal{C}$. In particular, the branch points of π occur exactly at $\pi^{-1}(\text{poles of } \frac{dz}{w})$ and thus $B(\pi) = \mathrm{degree}(\pi)$, a contradiction. Therefore, Σ is also a torus and $\pi : \mathcal{C}/(b-a) \to \Sigma$ is unbranched. Finally, the Weierstrass data $(g, d\xi)$ of M projects on Σ as $\big(z, c\frac{dz}{w}\big)$, $c \in \mathbb{C} - \{0\}$, producing a properly embedded minimal torus with two planar ends in certain $\mathbb{R}^3/T$, which has to be a Riemann example by Theorem 1. Now the proposition is proved. □

DEFINITION 1. *A minimal planar domain $M \in \mathcal{S}$ is called quasiperiodic if there exist sequences $\{p(j)^+\}, \{p(j)^-\} \subset M$ with $\{x_3(p(j))\} \to \pm\infty$, and surfaces $M_\infty^+, M_\infty^- \in \mathcal{S}$ such that the sequence of translated surfaces $\{M - p(j)^\pm\}$ converges to $M_\infty^\pm$ on compact subsets of $\mathbb{R}^3$.*

[4]Given $a \in \mathbb{R}^3$, the function $\langle N, a\rangle$ is always Jacobi. Such functions will be called *linear*.

[5]Two Jacobi functions $v, v^* \in \mathcal{J}(M)$ are called *Jacobi-conjugate* if there exists a globally defined complex solution h of the Jacobi equation on M such that $v = \mathrm{Re}(h)$ and $v^* = \mathrm{Im}(h)$.

Obviously quasiperiodicity generalizes periodicity, so a remarkable advance in the solution of Question 1 would be characterizing the Riemann minimal examples among all quasiperiodic minimal planar domains. On the other hand, being quasiperiodic is closely related to having global curvature estimates, in the following manner.

LEMMA 1. *If the absolute Gaussian curvature of $M \in \mathcal{S}$ is bounded above, then M is quasiperiodic.*

SKETCH OF PROOF. M cuts transversely every horizontal plane at a height different from the end heights and the vertical part of the flux is normalized to be 1, hence Lemma 2 in [**12**] ensures that for any divergent sequence $\{p(i)\} \subset M$ with $N(p(i))$ horizontal (here N is the Gauss map of M), the minimum of the absolute Gaussian curvature $|K|$ in disks $D(i) \subset M$ centered at $p(i)$ with certain positive radius independent of i is bounded away from zero. As $|K|$ is globally bounded above, a subsequence of $\{M - p(i)\}$ must converge to a properly embedded, nonflat minimal surface M_∞ in $\mathbb{R}^3$ with genus zero and bounded curvature. We can find points where N is horizontal between any two consecutive planar ends of M, thus in order to finish the proof it suffices to check that M_∞ has two limit ends.

By contradiction, first suppose that M_∞ has only one end, in particular it is simply connected. A very recent theorem of Meeks and Rosenberg [**17**] ensures that M_∞ must be a Helicoid. As the Gauss map of M_∞ omits the vertical direction, this Helicoid must be vertical. Now the contradiction comes from the fact that the third component of the flux of $M - p(i)$ equals one while the same component in a vertical helicoid is infinite.

Now assume that the number of ends of M_∞ is finite. By Collin [**2**], M_∞ will have finite total curvature, thus it is a catenoid, López–Ros [**11**]. The above remark on the Gauss map shows that this catenoid is vertical, hence its flux is also vertical. This is a contradiction, because the flux of $M - p(i)$ coincides with the flux of M, which is not vertical (this follows from the application of the López–Ros deformation to pieces of M obtained by intersection with horizontal slabs, bounded by two compact convex planar curves $\gamma(i), \gamma(i+1)$ in horizontal planes, where $\gamma(i)$ converges to the neck of M_∞).

Finally, assume that M_∞ has only one limit end. Let a be a unit vector such that $\Pi = \{a\}^\perp$ is the tangent plane at infinity of M_∞. If a is vertical, then M_∞ has either a highest or a lowest end, which must be catenoidal by the Strong Halfspace Theorem [**7**]. Now it is not difficult to deduce that M_∞ has vertical flux, which is impossible by a suitable modification of the argument in the last paragraph. If a is not vertical, then the embeddedness of M_∞ guarantees that the limit normal vector of all ends is $\pm a$. With a little work one can show that the Gauss map N_∞ of M_∞ takes a vertical value, so the open mapping property of N_∞ ensures that the same holds for $M - p(i)$ with i large, a contradiction. This completes the proof. □

REMARK 1. *Considerable progress has been achieved since this article was written. In a recent paper* [**15**]*, Meeks, Pérez and Ros proved that a properly embedded minimal surface in $\mathbb{R}^3$ with finite genus and exactly one limit end cannot exist. This avoids the argument in the last paragraph of the proof of Lemma 1. As an important consequence of this nonexistence theorem and of results in* [**14**]*, the same authors have given a proof of the validity of Conjecture 1 below.*

The converse of Lemma 1 was unknown before the time this meeting was held. During the weeks here at MSRI, the authors were able to conclude its validity. In fact, they proved a stronger result.

THEOREM 3. *Let $\{M_i\} \subset \mathcal{S}$ be a sequence of surfaces with $|F(M_i)|$ bounded. Then, the Gaussian curvature of the M_i is uniformly bounded.*

The proof of this result is based on the argument used by the authors in order to obtain curvature estimates in the periodic setting [**12**], together with an appropriate use of recent results by Colding and Minicozzi about minimal laminations and curvature estimates for embedded minimal disks (see [**17**] and references therein).

Theorem 3 immediately gives that any surface in $\mathcal{S}$ has bounded Gaussian curvature, thus Lemma 1 shows that quasiperiodicity is shared by any surface in $\mathcal{S}$. The boundedness of the curvature for any surface in $\mathcal{S}$ should in fact hold in much more generality. Let us state it as a conjecture.

CONJECTURE 1. *Every properly embedded minimal surface in $\mathbb{R}^3$ with finite genus has bounded Gaussian curvature.*

To finish these notes, we present a situation in which several of the tools explained above can be applied to give a uniqueness result for the Riemann minimal examples.

THEOREM 4. *There exists $\varepsilon > 0$ such that if $|F(M)| < \varepsilon$ for $M \in \mathcal{S}$, then M is a Riemann minimal example.*

REMARK 2. *Note that Theorem 4 is nothing but the generalization to the non-periodic setting of the third property of the periodic Flux map (see page 599). At this point, the reader could ask if the other properties of F and the ones of $\mathcal{R}$ in $\mathcal{S}$ remain valid, in order to give an affirmative answer to Question 1: the properness of F follows directly from Theorem 3, while the proof of the closedness of $\mathcal{R}$ in $\mathcal{S}$ remains valid now. The authors have proved partial results concerning the openness of $\mathcal{R}$ in $\mathcal{S}$ with an argument similar in nature to the proof that we present below. All these facts reduce the affirmative answer to Question 1 to prove the openness of the Flux map in the quasiperiodic setting. This is still work in progress.*

SKETCH OF PROOF. The first step is proving that given any sequence $\{M_i\} \subset \mathcal{S}$ with $F(M_i) \to 0$ and any choice of points $p_i \in M_i$ where the Gauss map of M_i is horizontal, the sequence $\{M_i - p_i\}$ has a subsequence that converges to a vertical catenoid with flux $(0,0,1)$ on compact subsets of $\mathbb{R}^3$. This follows from Theorem 3 plus an appropriate analysis of the possible limits of $\{M_i\}$. Then, for i large enough, the surface M_i is arbitrarily close to an infinite stack of vertical catenoids (all with neck lengths equal to 2π), joined by almost flat horizontal graphs, each containing one end and one finite branch point where the Gauss map takes an almost vertical value. Hence Theorem 4 reduces to the following result.

THEOREM 5. *If $M \in \mathcal{S}$ is a quasiperiodic planar domain and M is close enough to an infinite discrete collection of vertical catenoids joined by almost flat graphs containing the ends, then $M \in \mathcal{R}$.*

SKETCH OF PROOF. Given $M \in \mathcal{S}$ with Gauss map g, we will call a *nodal domain* of $\{|g| = 1\}$ any component Ω of $[M \cup \{\text{planar ends}\}] - \{|g| = 1\}$. Nodal domains will play a key role in the proof, so we need a detailed description of their

geometry. Fix such a nodal domain $\Omega \subset M$. As M is extremely close to a stack of vertical catenoids with vertical flux 1 joined by almost flat graphs, Ω must be bounded by two almost round circles extremely close to the necks of consecutive vertical catenoids, and contains exactly one planar end of M, where g has a double zero or pole. $g|_\Omega$ is a holomorphic branched two-sheeted covering of a half-sphere with two branch points (both of the type z^2), one at the end contained in Ω and the other one at a nonvertical value of g, extremely close to the vertical; see Figure 4.

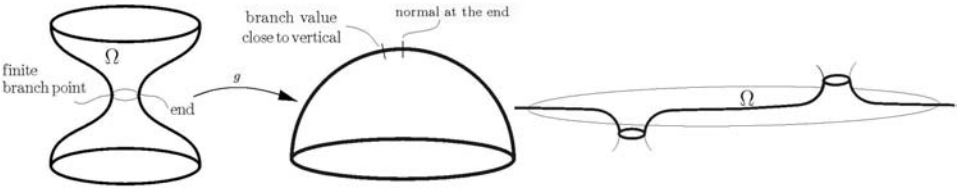


FIGURE 4. A nodal domain Ω of $\{|g| = 1\}$.

The argument for the Proof of Theorem 5 divides into three steps.

STEP 1. *Let $M \in \mathcal{S}$ be quasiperiodic, with Gauss map g and Shiffman function u. If M is close enough to an infinite discrete collection of vertical catenoids joined by almost flat graphs containing the ends, then for any nodal domain Ω of $\{|g| = 1\}$ one has:*

$$Q_\Omega(u,u) \geq 0, \qquad \textit{with equality if and only if } \ M \in \mathcal{R},$$

where Q_Ω denotes the index form associated to the Jacobi operator:

$$Q_\Omega(v,v) = \int_\Omega (\|\nabla v\|^2 - \|\nabla N\|^2 v^2) ds^2$$

for v in the Sobolev space of Ω.

SKETCH OF PROOF OF STEP 1. The statement is equivalent to proving that if $\{M_i\} \subset \mathcal{S} - \mathcal{R}$ is a sequence of quasiperiodic minimal planar domains (none being a Riemann minimal example) that converges to an infinite stack of vertical catenoids, then for i large enough $Q_{\Omega_i}(u_i, u_i) > 0$ holds, where u_i is the Shiffman function of M_i, $\Omega_i \subset M_i$ is any nodal domain of $\{|g_i| = 1\}$ (g_i is the Gauss map of M_i) and Q_{Ω_i} is the index form for the Jacobi operator on Ω_i. Arguing by contradiction, assume that $\{M_i\} \subset \mathcal{S} - \mathcal{R}$ converges to an infinite stack of vertical catenoids and for each i we find a nodal domain Ω_i of $\{|g_i| = 1\}$ such that $Q_{\Omega_i}(u_i, u_i) \leq 0$.

As $i \to \infty$, Ω_i degenerates into the union of two opposite half-spheres Ω^+, Ω^- joined by a common point e (the south pole in Ω^+, the north pole in Ω^-) where the end and the finite branch point of each Ω_i collapse. Note that u_i is not identically zero. We will rescale u_i in order to have a good limit (up to a subsequence) on $(\Omega^+ \cup \Omega^-) - \{e\}$: changing the induced metric ds_i^2 with the spherical metric on Ω_i, the index form $Q_{\Omega_i}(u_i, u_i)$ can be expressed as

$$0 \geq \int_{\Omega_i} (\|\nabla_i u_i\|_i^2 - \|\nabla_i N_i\|_i^2 u_i^2) ds_i^2 = \int_{\Omega_i} (\|\nabla u_i\|^2 - 2u_i^2) dA, \tag{2–2}$$

where N_i is the Gauss map of M_i, all the objects in the first integral are computed with respect to the induced metric ds_i^2 and the objects in the second integral refer

to the spherical metric. Note that by analyticity of the Jacobi functions, u_i cannot be identically zero on Ω_i and it makes sense to define

$$v_i = \frac{1}{\|u_i\|_{L^2(\Omega_i, dA)}} u_i$$

on Ω_i. Now, v_i is a Jacobi function with L^2-norm 1 in the spherical metric, and by (2–2) $\|\nabla v_i\|^2_{L^2(\Omega_i)} \leq 2$. Standard compactness and regularity theorems for elliptic equations ensure that there exists a function $v \in C^\infty(\Omega^+ \cup \Omega^-)$ such that, after extracting a subsequence, $\{v_i\}$ converges to v smoothly on compact subsets of $(\Omega^+ \cup \Omega^-) - \{e\}$, and v satisfies $\Delta v + 2v = 0$ in $\Omega^+ \cup \Omega^-$. Set $v^+ = v|_{\Omega^+}$, $v^- = v|_{\Omega^-}$. From the convergence of v_i to v one deduces $\|v\|_{L^2(\Omega^+ \cup \Omega^-)} = 1$ and $\|\nabla v\|^2_{L^2(\Omega^+ \cup \Omega^-)} \leq 2$, which imply

$$Q^{\Delta+2}_{\Omega^+}(v^+, v^+) + Q^{\Delta+2}_{\Omega^-}(v^-, v^-) \leq 0, \tag{2–3}$$

where $Q^{\Delta+2}_{\Omega^\pm}$ stands for the index form associated to the operator $\Delta + 2$ in the corresponding round half-sphere. In particular, one of the two terms in the left-hand-side of (2–3) must be nonpositive, say $Q^{\Delta+2}_{\Omega^+}(v^+, v^+) \leq 0$. Let $\varphi = v^+ - cf_3$, where $f_3(x) = \langle x, e_3 \rangle$ on Ω^+ (here $e_3 = (0,0,1)$) and $c = \left(\int_{\Omega^+} f_3 \, dA\right)^{-1} \left(\int_{\Omega^+} v^+ \, dA\right) \in \mathbb{R}$. Thus,

$$Q^{\Delta+2}_{\Omega^+}(\varphi, \varphi) = Q^{\Delta+2}_{\Omega^+}(v^+, v^+) - 2cQ^{\Delta+2}_{\Omega^+}(v^+, f_3) + Q^{\Delta+2}_{\Omega^+}(f_3, f_3).$$

But

$$Q^{\Delta+2}_{\Omega^+}(f_3, f_3) = \int_{\Omega^+} (\|\nabla f_3\|^2 - 2f_3^2) dA = \int_{\Omega^+} [\operatorname{div}(f_3 \nabla f_3) - f_3(\Delta f_3 + 2f_3)] \, dA = 0,$$

the last equality being true because f_3 vanishes at $\partial\Omega^+$ and because $\Delta f_3 + 2f_3 = 0$ in the sphere. A similar argument shows that $Q^{\Delta+2}_{\Omega^+}(v^+, f_3) = 0$, hence

$$Q^{\Delta+2}_{\Omega^+}(\varphi, \varphi) = Q^{\Delta+2}_{\Omega^+}(v^+, v^+) \leq 0.$$

As $\int_{\Omega^+} \varphi \, dA = 0$ and the second eigenvalue for the Neumann problem of the Laplacian in the round half-sphere $\mathbb{S}^2(1) \cap \{x_3 \leq 0\}$ is 2, with eigenfunctions $\langle x, a \rangle$ (a horizontal), one gets that $Q^{\Delta+2}_{\Omega^+}(\varphi, \varphi) = 0$ thus $\varphi(x) = \langle x, a \rangle$ for certain $a \perp e_3$. This shows that v^+ must be linear on Ω^+. Plugging this information into (2–3), it follows that $Q^{\Delta+2}_{\Omega^-}(v^-, v^-) = 0$, thus v^- must also be linear on Ω^-.

The following argument shows that both v^+, v^- are identically zero. By the four vertex theorem, the Shiffman function u_i has at least four zeros on each compact horizontal section of Ω_i, so the same holds for v_i. Passing to the limit of v_i and of the compact horizontal sections on Ω_i, one deduces that v^+ must have at least one zero on each horizontal circle in Ω^+ (with at most one exception). By contradiction, assume that the linear function v^+ is not identically zero. The above distribution of zeros forces v^+ to be of the form $v^+ = \langle x, a \rangle$ with $a \in \mathbb{S}^2(1)$ horizontal. Such v^+ has exactly two zeros on each horizontal circle $\Gamma^+ \subset \Omega^+$, which obliges the (at least four) zeros of v_i on the compact horizontal section Γ_i^+ converging to Γ^+, to collapse into the two zeros of v^+ in Γ^+. In particular, the gradient of v^+ must vanish at $\Gamma^+ \cap \{v^+ = 0\} = \Gamma^+ \cap \{a\}^\perp$, a contradiction. This proves that $v^+ = 0$ in Ω^+, and similarly $v^- = 0$ in Ω^-.

Now denote by p_i a point of maximum of $|u_i|$ in Ω_i (this maximum exists because the Shiffman function extends smoothly through the ends of M). Clearly, p_i is also a point of maximum for $|v_i|$ in Ω_i. Assume for the moment that $p_i \in \Omega_i$

does not collapse into e when $i \to \infty$. As $\{v_i\}$ converges smoothly to $v = 0$ on compact subsets of $(\Omega^+ \cup \Omega^-) - \{e\}$, it follows that

$$1 = \|v_i\|_{L^2(\Omega_i)} \leq |v_i(p_i)| \int_{\Omega_i} dA = 4\pi |v_i(p_i)| \to 0,$$

a contradiction. Thus, we have only to prove that p_i does not collapse into e as $i \to \infty$. Ω_i is conformally equivalent to a compact cylinder $[a_i^-, a_i^+] \times \mathbb{S}^1$, where $a_i^- < 0 < a_i^+$ and $a_i^\pm \to \pm\infty$ as $i \to \infty$. By contradiction, if $\{p_i\} \to e$ as $i \to \infty$, then we can place p_i in the model of $\Omega_i = [a_i^-, a_i^+] \times \mathbb{S}^1$ at a fixed point $p_0 \in \{0\} \times \mathbb{S}^1$ for all i. Define $w_i = u_i / |u_i(p_i)|$ (note that $|u_i(p_i)| \neq 0$ because $M_i \in \mathcal{S} - \mathcal{R}$). Then w_i is a solution of the Jacobi equation in $\Omega_i = [a_i^-, a_i^+] \times \mathbb{S}^1$, which in the flat metric is written as $\Delta_0 w_i + \|\nabla_0 N_i\|_0^2 w_i = 0$. Using that $\{\|\nabla_0 N_i\|_0^2\}$ converges to zero on compact subsets of $\mathbb{R} \times \mathbb{S}^1$ (this follows because any compact subset of $\mathbb{R} \times \mathbb{S}^1$, when viewed in the spherical covering model of Ω_i, must collapse entirely into e, thus the spherical area of its image by N_i goes to zero) and that $|w_i|$ is globally bounded by 1 in $[a_i^-, a_i^+] \times \mathbb{S}^1$, we deduce that a subsequence of $\{w_i\}$ converges smoothly on compact subsets of $\mathbb{R} \times \mathbb{S}^1$ to a harmonic function $w_\infty : \mathbb{R} \times \mathbb{S}^1 \to [-1, 1]$. As $\mathbb{R} \times \mathbb{S}^1$ is parabolic, w_∞ must be constant. As $|w_i(p_0)| = 1$, we get $w_\infty = \pm 1$ on $\mathbb{R} \times \mathbb{S}^1$. Finally, the contradiction follows from a counting argument of the zeros of w_i in each compact horizontal section, similar to the one we did above. Now Step 1 is proved.

STEP 2. *Let $M \in \mathcal{S}$ be any quasiperiodic minimal planar domain for which Step 1 holds. By quasiperiodicity, there exist sequences of points $p(j)^+, p(j)^- \in M$ with $\{x_3(p(j)^\pm)\} \to \pm\infty$, such that $\{M - p(j)^\pm\}$ converges on compact subsets of $\mathbb{R}^3$ to minimal surfaces $M_\infty^+, M_\infty^- \in \mathcal{S}$. Then, M_∞^+ and M_∞^- are the same Riemann minimal example.*

PROOF OF STEP 2. For $j \in \mathbb{N}$ fixed, denote by u_j^+ the Shiffman function of $M - p(j)^+$, which is $p \in M - p(j)^+ \mapsto u_j^+(p) = u(p + p(j)^+)$, u being the Shiffman function of M. The smooth convergence of $\{M - p(j)^+\}$ to M_∞^+ implies that $\{u_j^+\}$ converges smoothly on compact subsets of M_∞^+ to the Shiffman function u_∞^+ of M_∞^+.

Note that, as Step 1 holds for M, it also holds for any translation of M and after taking smooth limits, it holds for M_∞^+. Fix a nodal domain $\Omega_\infty \subset M_\infty \cup \{\text{planar ends}\}$ of $\{|g_\infty| = 1\}$, where g_∞ represents the Gauss map of M_∞^+. Let γ_∞ be one of the boundary curves of Ω_∞. As γ_∞ is compact, we find a sequence $\gamma(j)$ of curves inside M such that $\{\gamma(j) - p(j)^+\}$ converges smoothly to γ_∞. Now call $\Omega(j)$ to the domain in M bounded by $\gamma(j) \cup \gamma(j+1)$, see Figure 5.

Clearly, we can choose the curves $\gamma(j)$ $(j \in \mathbb{N})$ so that $\Omega(j)$ is a (finite) union of nodal domains of $\{|g| = 1\}$, $\Omega(j) = \Omega_{1,j} \cup \cdots \cup \Omega_{k(j),j}$. Using Step 1 in each nodal domain and integrating by parts:

$$0 \leq \sum_{i=1}^{k(j)} Q_{\Omega_{i,j}}(u, u) = \sum_{i=1}^{k(j)} \int_{\partial \Omega_{i,j}} u \frac{\partial u}{\partial \eta} ds = \int_{\partial \Omega(j)} u \frac{\partial u}{\partial \eta} ds,$$

where η is the unit conormal vector exterior to $\Omega_{i,j}$ along its boundary, and the second equality follows from the observation that if two nodal domains among $\Omega_{1,j}, \ldots, \Omega_{k(j),j}$ share a common part of their boundaries, then the corresponding conormal vectors along such common part are opposite. Furthermore, the last

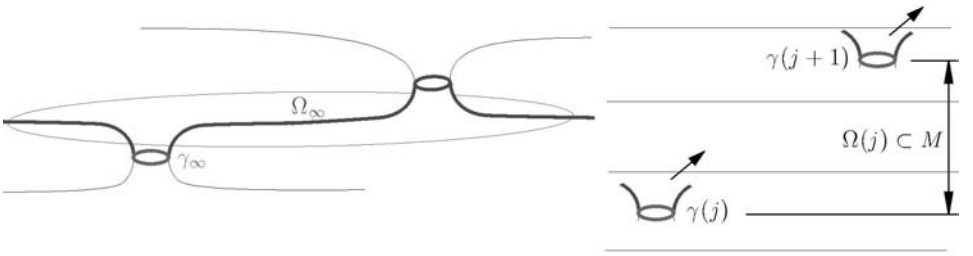


FIGURE 5. Left: The nodal domain $\Omega_\infty \subset M^+_\infty$. Right: The domain $\Omega(j) \subset M$.

integral equals

$$\int_{\gamma(j+1)} u\frac{\partial u}{\partial \eta}ds - \int_{\gamma(j)} u\frac{\partial u}{\partial \eta}ds = \int_{\gamma(j+1)-p(j+1)^+} u^+_{j+1}\frac{\partial u^+_{j+1}}{\partial \eta}ds - \int_{\gamma(j)-p(j)^+} u^+_j\frac{\partial u^+_j}{\partial \eta}ds.$$

Taking limits,

$$0 \leq \lim_{j\to\infty}\sum_{i=1}^{k(j)} Q_{\Omega_{i,j}}(u,u) = \int_{\gamma_\infty} u^+_\infty\frac{\partial u^+_\infty}{\partial \eta}ds - \int_{\gamma_\infty} u^+_\infty\frac{\partial u^+_\infty}{\partial \eta}ds = 0.$$

In particular, $Q_{\Omega_{i,j}}(u,u)$ can be made arbitrarily small for j large and for all $i = 1,\ldots,k(j)$. Translating by $p(j)$ and taking limits on the nodal domains inside $M - p(j)^+$ that converge to the nodal domain $\Omega_\infty \subset M^+_\infty$, we obtain $Q_{\Omega_\infty}(u^+_\infty, u^+_\infty) = 0$. Using Step 1 again for M^+_∞, it follows that $M^+_\infty \in \mathcal{R}$. Similarly we deduce $M^-_\infty \in \mathcal{R}$. Finally, that $M^+_\infty = M^-_\infty$ comes from the fact that both surfaces are limits of translations of the same surface M, thus they share the same flux as M, and the flux characterizes elements in $\mathcal{R}$.

STEP 3. *If Step 1 holds for $M \in \mathcal{S}$ quasiperiodic, then $M \in \mathcal{R}$.*

PROOF OF STEP 3. By Step 2, there exist sequences of points $p(j)^+$, $p(j)^- \in M$, $\{x_3(p(j)^\pm)\} \to \pm\infty$, such that $\{M - p(j)^\pm\}$ converges smoothly to the same Riemann example $M^+_\infty = M^-_\infty = R \in \mathcal{R}$. As before, let $\Omega_\infty \subset R \cup \{\text{planar ends}\}$ be a nodal domain of $\{|g_\infty| = 1\}$ and let $\gamma_\infty \subset \partial\Omega_\infty$ be one of the boundary curves of Ω_∞. As $M^+_\infty = M^-_\infty$, we can find curves $\gamma(j), \Gamma(j) \subset M$ such that both sequences $\{\gamma(j) - p(j)^+\}$ and $\{\Gamma(j) - p(j)^-\}$ converge smoothly to γ_∞. Denote by $\Omega'(j) \subset M$ the domain bounded by $\gamma(j) \cup \Gamma(j)$; see Figure 6.

Again, we can choose the curves $\gamma(j), \Gamma(j)$ so that $\Omega'(j)$ is a finite union of nodal domains of $\{|g| = 1\}$. By the same argument as in the proof of Step 2, we have

$$0 \leq Q_{\Omega'(j)}(u,u) = \int_{\partial\Omega'(j)} u\frac{\partial u}{\partial \eta}ds = \int_{\gamma(j)} u\frac{\partial u}{\partial \eta}ds - \int_{\Gamma(j)} u\frac{\partial u}{\partial \eta}ds.$$

The first integral in the right-hand-side is

$$\int_{\gamma(j)} u\frac{\partial u}{\partial \eta}ds = \int_{\gamma(j)-p(j)^+} u^+_j\frac{\partial u^+_j}{\partial \eta}ds \overset{(j\to\infty)}{\longrightarrow} \int_{\gamma_\infty} u^+_\infty\frac{\partial u^+_\infty}{\partial \eta}ds,$$

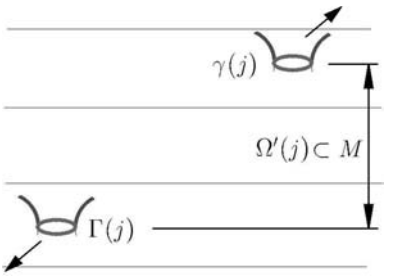


FIGURE 6. The domain $\Omega'(j) \subset M$.

which is zero because u_∞^+ is identically zero ($M_\infty^+ \in \mathcal{R}$). Analogously,

$$\lim_{j\to\infty} \int_{\Gamma(j)} u \frac{\partial u}{\partial \eta} ds = 0,$$

and hence $\lim_{j\to\infty} Q_{\Omega'(j)}(u,u) = 0$. Finally, consider a nodal domain $\Omega \subset \Omega'(j)$ of $\{|g| = 1\}$. By Step 1, one has $0 \leq Q_\Omega(u,u) \leq Q_{\Omega'(j)}(u,u)$. Taking limits, $Q_\Omega(u,u) = 0$ and again by Step 1, $M \in \mathcal{R}$. Now the theorem is proved.

References

[1] M. Callahan, D. Hoffman and W. H. Meeks III, *The structure of singly periodic minimal surfaces*, Invent. Math. **99** (1990), 455–481.

[2] P. Collin, *Topologie et courbure des surfaces minimales proprement plongées de* $\mathbb{R}^3$, Ann. of Math. 2nd Series **145(1)** (1997), 1–31.

[3] P. Collin, R. Kusner, W. H. Meeks III and H. Rosenberg, *The topology, geometry and conformal structure of properly embedded minimal surfaces*, preprint.

[4] Y. Fang and F. Wei, *On Uniqueness of Riemann's examples*, Proc. Amer. Math. Soc. **126(5)** (1998), 1531–1539.

[5] C. Frohman and W. H. Meeks III, *The ordering theorem for the ends of properly embedded minimal surfaces*, Topology **36(3)** (1997), 605–617.

[6] D. Hoffman, H. Karcher and H. Rosenberg, *Embedded minimal annuli in* $\mathbb{R}^3$ *bounded by a pair of straight lines*, Comment. Math. Helv. **66(4)** (1991), 599–617.

[7] D. Hoffman and W. H. Meeks III, *The strong half-space theorem for minimal surfaces*, Invent. Math. **101** (1990), 373–377.

[8] R. Langevin and H. Rosenberg, *A maximum principle at infinity for minimal surfaces and applications*, Duke Math. J. **57** (1988), 819–828.

[9] F. J. López, M. Ritoré and F. Wei, *A characterization of Riemann's minimal surfaces*, J. Diff. Geom. **47(2)** (1997), 376–397.

[10] F. J. López and D. Rodríguez, *Properly immersed singly periodic minimal cylinders in* $\mathbb{R}^3$, Michigan Math. J. **45(3)** (1998), 507–528.

[11] F. J. López and A. Ros, *On embedded complete minimal surfaces of genus zero*, J. Diff. Geom. **33** (1991), 293–300.

[12] W. H. Meeks III, J. Pérez and A. Ros, *Uniqueness of the Riemann minimal examples*, Invent. Math. **131** (1998), 107–132.

[13] W. H. Meeks III, J. Pérez and A. Ros, *Minimal planar domains in* $\mathbb{R}^3$, in preparation.

[14] W. H. Meeks III, J. Pérez and A. Ros, *The geometry of minimal surfaces of finite genus I: curvature estimates and quasiperiodicity*, preprint.

[15] W. H. Meeks III, J. Pérez and A. Ros, *The geometry of minimal surfaces of finite genus II: nonexistence of one limit end examples*, preprint.

[16] W. H. Meeks III and H. Rosenberg, *The maximum principle at infinity for minimal surfaces in flat three manifolds*, Comm. Math. Helv. **65**, 2 (1990), 255–270.

[17] W. H. Meeks III and H. Rosenberg, *The uniqueness of the helicoid and the asymptotic geometry of properly embedded minimal surfaces with finite topology*, preprint.
[18] S. Montiel and A. Ros, *Schrödinger operators associated to a holomorphic map*, Proceedings Conference on Global Differential Geometry and Global Analysis, Berlin, 1990, Lecture Notes in Math. **1481**, 147–174.
[19] J. Pérez, *Riemann bilinear relations on minimal surfaces*, Math. Ann. **310(2)** (1998), 307–332.
[20] J. Pérez, *On singly-periodic minimal surfaces with planar ends*, Trans. of the A.M.S. **349(6)** (1997), 2371–2389.
[21] J. Pérez and A. Ros, *Some uniqueness and nonexistence theorems for embedded minimal surfaces*, Math. Ann. **295(3)** (1993), 513–525.
[22] J. Pérez and A. Ros, *The space of properly embedded minimal surfaces with finite total curvature*, Indiana Univ. Math. J. **45** (1996), 177–204.
[23] B. Riemann, *Über die Fläche vom kleinsten Inhalt bei gegebener Begrenzung*, Abh. Königl. d. Wiss. Göttingen, Mathem. Cl., **13** (1867), 3–52.
[24] E. Toubiana, *On the minimal surfaces of Riemann*, Comment. Math. Helv. **67(4)** (1992), 546–570.

Departamento de Geometria y Topologia, Universidad de Granada, Fuentenueva s/n, 18071, Granada, Spain
E-mail address: jperez@ugr.es

Clay Mathematics Proceedings
Volume 2, 2004

The Mathematical Protein Folding Problem

Yi Fang

Abstract. We propose a geometric variational problem, the mathematical protein folding problem. Let $U \subset \mathbb{R}^3$ be a solid body which can change shapes (preserving the volume $V(U) > 0$) within a certain admissible structural space $\mathcal{P}(U)$. Each $P \in \mathcal{P}(U)$ indexes a special structure (shape) U_P of U. In other words, we may imagine U as a transformer. The problem then is to seek a special structure U_P, with $P \in \mathcal{P}(U)$, and a Lipschitz surface enclosing U_P that minimizes a structural function of three variables: the volume enclosed by the surface, the area of the surface, and the area of a specifically defined subset of the surface. A special case of this problem is the isoperimetric problem, which corresponds to the admissible structural space being all possible volume-preserving shape changes, maybe confined in certain subsets of $\mathbb{R}^3$, and a specific structural function, namely the area alone.

This problem has its origin in computational biology, it comes from a new hypothesis on the principle of protein folding, the geometric equilibrium principle.

1. The Problem

Transformers. Let $U \subset \mathbb{R}^3$ be a solid body, i.e., a closed submanifold bounded by an embedded closed Lipschitz surface ∂U. Let $Y \subset U$ be a closed subset consisting of a collection of solid bodies such that $0 \leq V(Y) < V(U)$, where $V(U)$ is the volume of U. The subset Y could be disconnected and could be empty. Due to the biological origin, we will call Y the *hydrophobic part* of U.

Suppose that U can change its shape in $\mathbb{R}^3$. Each shape of U is a *structure* of U. Furthermore, we assume that these changes of shape preserve the volumes $V(U)$ and $V(Y)$. An example of this kind of changing shape solid bodies is a colored transformer, where $Y \subset U$ could be, say, the white colored part. We will call such an U a *transformer* occasionally.

Admissible structural spaces. In general, we only consider a subset of the all possible structures of U. Hence we assume that there is a criterion to determine which shapes are eligible (under consideration). For example, we may want a transformer to become a space shuttle but not a dinosaur, therefore, a dinosaur shape is not eligible. Eligible shapes will be called *admissible structures*, the set of all admissible structures is called the *admissible structural space*, and is denoted by $\mathcal{P} = \mathcal{P}(U)$. For each $P \in \mathcal{P}(U)$, denote U_P for the specific structure of U

©2004 Clay Mathematics Institute

corresponding to P, and Y_P the corresponding shape of Y under U_P. The shape change is volume-preserving then means that $V(U_P) = V(U)$ and $V(Y_P) = V(Y)$.

REMARK 1. In many cases, $\mathcal{P}(U)$ has a topology under which $P \to U_P$ and $P \to Y_P$ changes continuously for some topology for the subsets of $\mathbb{R}^3$, say the Hausdorff distance.

REMARK 2. In the protein folding case of Appendices A and B, the shape changes are actually homeomorphisms between the interiors of various U_P's, $P \in \mathcal{P}(U)$. There might be no more generality when we only require volume preserving in the mathematical protein folding problem.

Enclosing domains and surfaces. A bounded domain $\Omega \subset \mathbb{R}^3$ is Lipschitz if its boundary $S = \partial\Omega$ is an embedded closed Lipschitz surface. Let $\overline{\Omega}$ be the closure of Ω in $\mathbb{R}^3$. For a $P \in \mathcal{P}$, if $U_P \subset \overline{\Omega}$ and Ω is Lipschitz, then we call Ω an *enclosing domain* of U_P, and call $S = \partial\Omega$ an *enclosing surface* of U_P. Denote by $\mathcal{D}_P$ the set of all enclosing domains of U_P, $\mathcal{S}_P$ the set of all enclosing surfaces of U_P. Note that $S \in \mathcal{S}$ if and only if $S = \partial\Omega$, $\Omega \in \mathcal{D}_P$.

We work with congruence classes, so that if we say that $U_{P_1} = U_{P_2}$, we mean that they are in the same congruence class, etc. Note that if $U_{P_1} \neq U_{P_2}$, i.e., they are not congruent, then $\mathcal{S}_{P_1} \neq \mathcal{S}_{P_2}$ and $\mathcal{D}_{P_1} \neq \mathcal{D}_{P_2}$, in the sense of congruency.

Also in the sense of congruency, the mapping $P \to U_P$ is not necessarily one to one.

Hydrophobic surfaces. For each $P \in \mathcal{P}$ and each $S \in \mathcal{S}_P$, we define

(1–1)
$$S_{Y_P} := \{q \in S : \operatorname{dist}(q, \partial^* Y_P) < \operatorname{dist}(q, \partial U_P - \partial^* Y_P)\}, \quad \partial^* Y_P = \partial Y_P \cap \partial U_P,$$

where dis is the distance. Due to its biological origin, we will call $S_{Y_P} \subset S$ the *hydrophobic surface* of the enclosing surface S.

Structural functions. Let $\mathbb{R}_+ = [0, \infty)$ and $E : \mathbb{R}^3_+ \to \mathbb{R}_+$ be a continuous function. We assume that E is nondecreasing, i.e., for any two fixed variables, E is a nondecreasing function of the remaining variable. To avoid difficulties, we also assume that $\lim_{|(x,y,z)|\to\infty} E(x, y, z) = \infty$. Also due to the biological origin, we call such a function E a *structural function.*

The mathematical protein folding problem. Let $V(\Omega)$, $A(S)$, and $A(S_{Y_P})$ be the volume and areas of Ω, S, and S_{Y_P}, respectively. Our geometric variational problem then is: find a $P_0 \in \mathcal{P}$ and an $\Omega_0 \in \mathcal{D}_{P_0}$, $\partial\Omega_0 = S_0 \in \mathcal{S}_{P_0}$, such that

(1–2)
$$E(V(\Omega_0), A(S_0), A((S_0)_{Y_{P_0}})) = \inf_{P\in\mathcal{P},\, \Omega\in\mathcal{D}_P} E(V(\Omega), A(S), A(S_{Y_P})).$$

Such a triple (U_{P_0}, Ω_0, S_0), if it exists, will be called a minimizer. The U_{P_0}, Ω_0, and S_0 will be called the *solution structure*, the *solution domain*, and the *solution surface*, respectively.

We will denote the problem (1–2) as $\mathcal{X}(U, \mathcal{P}, E)$ and call it the *mathematical protein folding problem* due to its biological origin. We can also specify the topology of the enclosing surfaces and then seek solution surfaces among that topological class. Then we denote the problem (1–2) by $\mathcal{X}(U, \mathcal{P}, E)_q$, where the non-negative integer q is the genus of the required enclosing and solution surfaces. When considering $\mathcal{X}(U, \mathcal{P}, E)_q$, we use $\mathcal{D}^q_P$ and $\mathcal{S}^q_P$ to indicate the topological requirement for the enclosing domains and surfaces.

The real interest in biology is $\mathcal{X}(U, \mathcal{P}, E)_0$, where U is the peptide amino acid sequence of a globular protein, please see Appendices A and B.

REMARK 3. When U is a solid body with $V(U) > 0$, $Y = \varnothing$, $\mathcal{P}$ comprises all possible volume-preserving shape changes in $\mathbb{R}^3$ and the structural function E is the surface area ($E(x, y, z) = y$), then the problem $\mathcal{X}(U, \mathcal{P}, E)$ reduces to the isoperimetric problem in $\mathbb{R}^3$, and the solution surface is a round sphere. Further, we can restrict the admissible structural space of the isoperimetric problem by requiring that each U_P must be contained in a specific region of $\mathbb{R}^3$, say a box B. In case that $V(B) > V(U)$ but a ball of volume $V(U)$ cannot be contained in B, we get a different admissible structural space $\mathcal{P}'$. Take the structural function E as the surface area again, then the problem $\mathcal{X}(U, \mathcal{P}', E)$ is still called an isoperimetric problem and it still has a solution surface, but obviously the solution surface is no longer a round sphere. Please see [**13**] for the isoperimetric problem.

2. Some Geometric Features of Solution Surfaces

To show some flavor of the geometric properties of a solution surface S to the mathematical protein folding problem $\mathcal{X}(U, \mathcal{P}, E)$, we will deduce the first variation formula and then derive the Euler–Lagrange equation in a very special case. Even by this elementary simple treatment we obtain an interesting property of a solution surface.

Let U and $\mathcal{P}(U)$ be as in the last section. For any $P \in \mathcal{P}$, we define a Lipschitz function $f_P : \mathbb{R}^3 \to \mathbb{R}$,

$$f_P(p) = \operatorname{dist}(p, \partial U_P - \partial^* Y_P) - \operatorname{dist}(p, \partial^* Y_P), \quad \partial^* Y_P = \partial Y_P \cap \partial U_P, \quad p \in \mathbb{R}^3.$$

Then $\mathbb{R}^3 = Z_+ \cup Z_0 \cup Z_-$, where

$$Z_\pm = \{p \in \mathbb{R}^3;\ f_P(p) > (<)\, 0\}, \quad Z_0 = \{p \in \mathbb{R}^3;\ f_P(p) = 0\}.$$

Let $S \in \mathcal{S}_P$. By the definition of the hydrophobic surface, $S_{Y_P} = S \cap Z_+$ and $\partial S_{Y_P} \subset S \cap Z_0$.

First variation. Let (U_P, Ω, S) be a minimizer and assume that $\mathcal{P}(U)$ has a smooth structure. Let $\{P_t\}_{-\epsilon<t<\epsilon} \subset \mathcal{P}$, $\epsilon > 0$, be a smooth path in $\mathcal{P}$ such that $P_0 = P$. Define f_{P_t} and $Z_\pm^t$, Z_0^t similarly as above. Let $O \subset \mathbb{R}^3$ be open and $\overline{\Omega} \subset O$. Let $\{\phi_t\}_{-\epsilon<t<\epsilon}$ be a 1-parameter family of diffeomorphisms $O \to O$ such that $\phi(0, x) = \phi_0(x) = x$, for $x \in O$, and out of a compact set $K \subset O$, $\phi(t, x) = \phi_t(x) = x$, $-\epsilon < t < \epsilon$. For such $\{\phi_t\}$ we will say that it is supported in K. Then the variational vector field X is

$$X(x) = \left.\frac{d\phi(t, x)}{dt}\right|_{t=0}, \quad x \in O.$$

Since $\{\phi_t\}$ is supported in K, on $O - K$, $X \equiv (0, 0, 0)$. We will assume that X is C^1 and take $O = \mathbb{R}^3$ in this article. Then we consider $\Omega_t = \phi_t(\Omega)$, $S_t = \phi_t(S)$, etc. We say that $\{\phi_t\}$ is an *eligible variation* with respect to $\{P_t\}$ if $U_{P_t} \subset \phi_t(\overline{\Omega})$ for $-\epsilon < t < \epsilon$. If it is only true that $U_{P_t} \subset \phi_t(\overline{\Omega})$ for $0 \le t < \epsilon$, then we say that $\{\phi_t\}$ is *one-side eligible*. For example, if S is C^2 and N is its inward unit normal, then $\phi_t(x) = x + tX(x)$ is always at least one-side eligible with respect to the constant path $\{P_t \equiv P\}$, as long as X satisfies

$$X|_S = \xi N, \quad \xi \le 0 \text{ is a compactly supported } \mathrm{C}^1 \text{ function.}$$

We may treat Ω and $S = \partial\Omega$ as rectifiable varifolds and consider the first variation of $E(V(\Omega_t), A(S_t), A((S_t)_{Y_{P_t}}))$ at $t = 0$, where $(S_t)_{Y_{P_t}} = S_t \cap Z_+^t$. From now on, we will assume that E is C^1. Since (U_P, Ω, S) is a minimizer, for eligible $\{\phi_t\}$ with respect to $\{P_t\}$, we obtain

$$\frac{d}{dt}\Big|_{t=0} E(V(\Omega_t), A(S_t), A((S_t)_{Y_{P_t}})) = 0. \tag{2–1}$$

For one-side eligible $\{\phi_t\}$ with respect to $\{P_t\}$, we have that for $t \geq 0$,

$$\frac{d}{dt}\Big|_{t=0} E(V(\Omega_t), A(S_t), A((S_t)_{Y_{P_t}})) \geq 0. \tag{2–2}$$

Special considerations. We only consider the simple case that E is C^1 and that $P_t \equiv P$ (so we do not need the smooth structure of $\mathcal{P}$), then

$$\begin{aligned}\frac{d}{dt}\Big|_{t=0} & E(V(\Omega_t), A(S_t), A((S_t)_P)) \\ &= E_x(V(\Omega), A(S), A(S_P))\frac{d}{dt}\Big|_{t=0} V(\Omega_t) + E_y(V(\Omega), A(S), A(S_P))\frac{d}{dt}\Big|_{t=0} A(S_t) \\ &\qquad + E_z(V(\Omega), A(S), A(S_P))\frac{d}{dt}\Big|_{t=0} (A((S_t)_P)).\end{aligned}$$

Since S is Lipschitz and orientable, it has an inward unit normal vector N for $\mathcal{H}^2$ (Hausdorff measure) almost everywhere. So

$$\frac{d}{dt}\Big|_{t=0} V(\Omega_t) = \int_\Omega \operatorname{div} X \, d\mathcal{H}^3 = -\int_S X{\cdot}N \, d\mathcal{H}^2 \tag{2–3}$$

(see [**14**, p. 51], for example), where $\cdot$ denotes the inner product in $\mathbb{R}^3$. From now on we will further assume that S has generalized mean curvature vectors $\mathbf{H}$. Then for $\mathcal{H}^2$ almost everywhere, $\mathbf{H} = 2HN$ and $H \in L^1(\mathcal{H}^2 \llcorner S)$, see page 81 of [**14**], and

$$\frac{d}{dt}\Big|_{t=0} A(S_t) = \int_S \operatorname{div}_S X \, d\mathcal{H}^2 = -2\int_S H(X{\cdot}N) \, d\mathcal{H}^2. \tag{2–4}$$

Again we only consider a simple case that $K \subset Z_+ \cup Z_-$. Then for each $p \in K$, if $|t|$ is small enough, then $p \in Z_\pm$ if and only if $\phi(t,p) \in Z_\pm$. Because K is compact, there is an $\eta > 0$, such that whenever $|t| < \eta$, p and $\phi_t(p)$ are both in Z_+ or both in Z_-, for every $p \in K$. That is, $\phi_t(S) \cap Z_+ = \phi_t(S \cap Z_+)$ and $\phi_t(S) \cap Z_- = \phi_t(S \cap Z_-)$ for $|t| < \eta$. Thus in the case that $K \subset Z_+ \cup Z_-$ and $P_t \equiv P$ hence $Z_+^t \equiv Z_+$, we have

$$(S_t)_{Y_{P_t}} = S_t \cap Z_+^t = \phi_t(S) \cap Z_+ = \phi_t(S \cap Z_+) = \phi_t(S_{Y_P}), \quad |t| < \eta.$$

Now since $K \cap Z_0 = \varnothing$, X vanishes on $\partial S_{Y_P} \subset S \cap Z_0$, we have

$$\begin{aligned}\frac{d}{dt}\Big|_{t=0} A((S_t)_{Y_P}) &= \frac{d}{dt}\Big|_{t=0} A(\phi_t(S_{Y_P})) \\ &= \int_{S_{Y_P}} \operatorname{div}_S X \, d\mathcal{H}^2 = -2\int_S H\chi_{S_{Y_P}}(X{\cdot}N) \, d\mathcal{H}^2,\end{aligned} \tag{2–5}$$

where $\chi_{S_{Y_P}}$ is the characteristic function of S_{Y_P}.

We now assume that E is C^1 and S has generalized mean curvature vectors $\mathbf{H}$. Then for eligible first variation $\{\phi_t\}$ with respect to the constant path $\{P_t \equiv P\}$

with variational vector field X compactly supported in $K \subset Z_+ \cup Z_-$, by (2–1), (2–3), (2–4), and (2–5), we have

$$\int_{S\cap(Z_+\cup Z_-)} [(E_x + 2H(E_y + E_z\chi_{S_{Y_P}})](X\cdot N)\, d\mathcal{H}^2 = 0, \tag{2–6}$$

and under the same assumptions, if $\{\phi_t\}$ is only one-side eligible with respect to $\{P_t \equiv P\}$, then

$$\int_{S\cap(Z_+\cup Z_-)} [(E_x + 2H(E_y + E_z\chi_{S_{Y_P}})](X\cdot N)\, d\mathcal{H}^2 \leq 0. \tag{2–7}$$

The Euler–Lagrange equation. We further assume that

$$\frac{\partial E}{\partial x_i} > 0, \quad i = 1, 2, 3. \tag{2–8}$$

Let $\Omega_1 = \mathbb{R}^3 - U_P$, then it is an non-empty domain. It is easy to see that if $K \subset \Omega_1$ is compact, then any $\{\phi_t\}$ which is supported in K is eligible with respect to the constant path $\{P_t \equiv P\}$, hence (2–6) is true for any C^1 vector field supported in $K \subset \Omega_1 \cap (Z_+ \cup Z_-)$.

Consider the surface $S_Q - \partial S_{Y_P} = S \cap \Omega_1 \cap (Z_+ \cup Z_-)$, where $S_Q = S \cap \Omega_1$. Using parameter representation for the surface $S_Q - \partial S_{Y_P}$, we see that locally it is a domain in $\mathbb{R}^2$ with $L^\infty(\mathcal{H}^2)$ metrics. Hence the use of a mollifier to $L^\infty(\mathcal{H}^2 \llcorner S)$ vector field compactly supported in $S_Q - \partial S_{Y_P}$ is justified. Therefore, (2–6) is true for any $L^\infty(\mathcal{H}^2 \llcorner S)$ vector field X compactly supported in $S_Q - \partial S_{Y_P}$. Thus we can take $X = \xi N$, where ξ is an $L^\infty(\mathcal{H}^2 \llcorner S)$ function compactly supported in $S_Q - \partial S_{Y_P}$, and obtain

$$\int_{S_Q - \partial S_{Y_P}} [(E_x + 2H(E_y + E_z\chi_{S_{Y_P}})]\xi\, d\mathcal{H}^2 = 0.$$

Since ξ was arbitrary in $L_c^\infty(S_Q - \partial S_{Y_P}, \mathcal{H}^2 \llcorner S)$ and $H \in L^1(\mathcal{H}^2 \llcorner S)$, we see that

$$E_x + 2H(E_y + E_z\chi_{S_{Y_P}}) = 0, \quad \mathcal{H}^2 \text{ a.e. on } S_Q - \partial S_{Y_P}.$$

By the assumption (2–8) about the structural function E,

$$H = \frac{-E_x}{2(E_y + E_z)} < 0, \quad \mathcal{H}^2 \text{ a.e. on } S_Q \cap S_{Y_P},$$

and

$$H = \frac{-E_x}{2E_y} < 0, \quad \mathcal{H}^2 \text{ a.e. on } S_Q - \overline{S_{Y_P}}.$$

Note that E_x, E_y, and E_z take values at $(V(\Omega), A(S), A(S_{Y_P}))$, hence they are constants. Thus $S_Q - \partial S_{Y_P}$ consists of constant mean curvature pieces.

Since N is the inward normal, we will say that $S_Q - \partial S_{Y_P}$ is *inward mean concave* ($H < 0$). If ∂U_P and $S_Q - \partial S_{Y_P}$ are both piecewise C^2 and ∂U_P is *inward mean convex* ($H \geq 0$) on its smooth part of ∂U_P, denoted by $(\partial U_P)_M$ (the model for protein meets this criterion), then at the intersection of $\overline{S_Q - \partial S_{Y_P}} \cap (\partial U_P)_M$, S will not be C^2. In fact, otherwise the mean curvature should vanish at the intersection, but that is impossible since the number of the negative values of the mean curvature is finite. Also for the same reason, S will not be C^2 along ∂S_{Y_P}.

Piecewise constant mean curvature. We can summarize the above results as follows:

THEOREM 1. *Let (U_P, Ω, S) be a minimizer of a mathematical protein folding problem $\mathcal{X}(U, \mathcal{P}, E)$. Suppose that $E \in C^1$ and satisfies (2–8). Assume that S has generalized mean curvature vector $\mathbf{H} = 2HN$, $H \in L^1(\mathcal{H}^2 \llcorner S)$. Let $\Omega_1 = \mathbb{R}^3 - U_P$ and $S_Q = S \cap \Omega_1$, then $S_Q \cap \partial U_P = \varnothing$ and on $S_Q \cap S_{Y_P}$,*

$$H \equiv \frac{-E_x(V(\Omega), A(S), A(S_{Y_P}))}{2(E_y(V(\Omega), A(S), A(S_{Y_P})) + E_z(V(\Omega), A(S), A(S_{Y_P}))} < 0, \tag{2–9}$$

and on $S_Q - \overline{S_{Y_P}}$,

$$H \equiv \frac{-E_x(V(\Omega), A(S), A(S_{Y_P}))}{2E_y(V(\Omega), A(S), A(S_{Y_P}))} < 0. \tag{2–10}$$

In particular, S_Q has piecewise constant mean curvature. Especially, if ∂U_P is piecewise smooth and is inward mean convex on $\partial U_P \cap S \neq \varnothing$, then S is at most $C^{1,1}$.

This is only a glimpse of the characteristics of a solution surface to a mathematical protein folding problem $\mathcal{X}(U, \mathcal{P}, E)$ under some minor assumption about the regularity of S and E. We will study the general geometric properties, as well as the existence, uniqueness, and regularity theories for the mathematical protein folding problem elsewhere.

3. Main Questions

Existence. For general C^1 structural function E, it turns out that even the admissible structural space $\mathcal{P}(U)$ has a compact topology, due to the regularity problem, the existence of a minimizer of $\mathcal{X}(U, \mathcal{P}, E)$ is not easy to prove. It is even harder for the problem $\mathcal{X}(U, \mathcal{P}, E)_q$. The real protein folding problem is a problem of $\mathcal{X}(U, \mathcal{P}, E)_0$.

One interesting question is that for a fixed non-negative integer q, what kind of structural functions E can guarantee that $\mathcal{X}(U, \mathcal{P}, E)_q$ has a minimizer?

Uniqueness. Another interesting question is the uniqueness of minimizers. Again, for some structural functions, there may be more than one minimizer. Both for $\mathcal{X}(U, \mathcal{P}, E)$ and $\mathcal{X}(U, \mathcal{P}, E)_q$, what kinds of properties of the structural functions E can guarantee that the solution is unique?

Regularity and geometric properties. One can ask a very general question: what are the general geometric features of the solution surfaces to $\mathcal{X}(U, \mathcal{P}, E)$? We have shown that if $S_Q - \partial S_P$ is C^2, then it is rather well behaved as long as E satisfies (2–8). Indeed the key arguments in the mathematical protein folding problem will be the proofs of various regularity properties of the solution surfaces. As we have shown, in some case some singular set is bound to exist. In particular, the real protein folding problem is such a case.

Generalizations. Finally, one can also consider the same problem in general Riemannian manifolds, with or without boundary; or in higher dimensions; or relax the solid body assumption about U; or consider certain prescribed boundary conditions, say, the Plateau condition; etc.

4. The Origin

The problem $\mathcal{X}(U, \mathcal{P}, E)_0$ comes from trying to solve the protein folding problem in computational biology.

Proteins. A protein is a macromolecule consisting of a sequence of amino acids connected by peptide linkages. There are 20 different amino acids, hence the number of amino acid sequences with the same length N is 20^N.

Some amino acids do not like water; they are called *hydrophobic side chains*; others do not care and are called *hydrophilic side chains.*

A globular protein in its native cellular environment folds to a specific shape, its *native structure.* This native structure greatly affects the protein's biological function.

The protein folding problem. The problem of predicting the native structure of a protein from the knowledge of its amino acid sequence is called the *protein folding problem.* Predicting the native structure of a protein guided by a (hypothesized) general principle, and using *only* the knowledge of the protein's amino acid sequence, is called an *ab initio*, or *first principle*, prediction. So far, all *ab initio* principle predictions are guided by the *thermodynamic principle* that hypothesizes that the native structure should have minimal free energy.

The grand challenge. Since it is important and very difficult, the protein folding problem has been called a grand challenge in computational molecular biology. See for example, [**8**, pp. 19–20].

The thermodynamic principle predictions have not been successful for over 40 years of research. Besides the intrinsic complexity of the proteins, they are huge macromolecules, the difficulties in calculating energy functions perhaps are the main reasons for the failure of the predictions so far.

A mathematical point of view. The structures of proteins are geometric shapes, hence the protein folding problem can be seen as a geometric problem and might be solved by geometric methods. A principle of geometric flavor (a consequence of the thermodynamic principle?) may give us an alternative to the difficult calculations of the free energy.

The geometric features of the native structures. In order to work in geometry, we need to know what are the geometric characteristics of the native structures of globular proteins. The following facts are well known:

(1) Globular proteins tend to fold into fairly compact units and that's how they got their name [**16**, p. 524]. Except in a few cases, the core of a globular protein has no holes (one or more water molecules occupy a hole) [**1**, p. 14]. Hence a globular protein has a simply connected enveloping surface.

(2) The packing density of a globular protein is very high (around 0.75); see [**11**].

(3) The globular protein in natural state attempts to bury its hydrophobic side chains inside its core as much as possible; see [**1**, p. 90], [**3**], [**9**], and [**10**].

(4) The simply connected enveloping (accessible) surface tends to have smaller surface area in the native shape than in an unnatural shape; see [**2**], [**6**], [**7**], [**9**], [**10**], and [**12**].
(5) For long amino acid sequences, the fold may divide into several domains, see [**1**], page 29.

Coincidences or nature's law? It seems that these geometric features indicate that nature gets protein structures via a minimization of some geometric quantities. For philosophy of nature's employing free physical and mathematical laws in biology, please see [**5**] and [**15**].

We take the view that the above geometric features may give a good characterization of the native structures of globular proteins. We will suggest a new hypothesis of the protein folding principle as a geometric minimization process as below.

Seeing proteins as transformers. The amino acid sequence of a protein can be taken as a closed solid body U (a gigantic transformer). We can use the union of the hydrophobic side chains as Y. This transformer U can change shapes (structures) while preserving the volumes $V(U)$ and $V(Y)$. We can also define an admissible structural space $\mathcal{P} = \mathcal{P}(U)$ according to the physicochemical properties of the protein (see Appendices A and B, especially the condition (A–2)).

Formulating a geometric variation. Taking U as a transformer, then the packing density can be defined by

$$\text{density} = \frac{V(U)}{V_O}, \tag{4–1}$$

where $V_O \geq V(U)$ is the total volume that the protein really occupies.

To have variations, we consider V_O as the volume $V(\Omega)$ of any enclosing domain $\Omega \in \mathcal{D}_P^0$ of U_P, where $P \in \mathcal{P}(U)$. Hence to get higher density allowed in nature, we should reduce the volume $V_O = V(\Omega)$ in (4–1).

Consider an enclosing surface $S \in \mathcal{S}_P^0$, $P \in \mathcal{P}$. We want $A(S)$ to be small. Furthermore, we can define the hydrophobic surface S_{Y_P} as in (1–1). Then the fact that hydrophobic side chains tend to be buried inside as much as possible is equivalent to the fact that the hydrophobic surface area is as small as possible.

Finally, to fit the well-known geometric features, we should simultaneously minimize the volume, the area of the enclosing surface, and the area of the hydrophobic surface, in a coherent way. To reflect the coherent simultaneous minimization, we need a structural function E as defined in Section 1. Since for amino acid sequences of different lengths the folding may obey different rules, E could vary with the length n of the sequence. We write E_n to emphasize this dependence on n.

A new hypothesis. Now, according to these well-known geometric features, we are ready to formulate a hypothesis. In [**4**] the author hypothesizes:

HYPOTHESIS 1 (Geometric Equilibrium Principle). *Let U be a protein sequence of length n. Think of U as a transformer and let $\mathcal{P}(U)$ be a fixed admissible structural space for all allowed space shapes of U (under the restrictions (A–2) in Appendix A). The native structure of U is U_{P_0}, where (U_{P_0}, Ω_0, S_0) is a minimizer of the problem $\mathcal{X}(U, \mathcal{P}, E_n)_0$ for some structural function E_n. That is,*

$$E_n(V(\Omega_0), A(S_0), A((S_0)_{Y_{P_0}})) = \inf_{P \in \mathcal{P},\, \Omega \in \mathcal{D}_P^0} E_n(V(\Omega), A(S), A(S_{Y_P})). \tag{4–1}$$

This is then the origin of the variational problem $\mathcal{X}(U, \mathcal{P}, E)$.

Acknowledgment

I sincerely thank Jiri Novotny, who introduced me to the protein folding problem. The conversations and correspondences with him were important for me to develop the geometric equilibrium principle hypothesis for protein folding. Jenn-Fang Hwang has made invaluable suggestions and criticisms on various drafts of this article.

Appendix A. A Mathematical Model of Proteins

Any mathematical model of the protein folding problem under the geometric equilibrium principle is a specific identification of the admissible structural space $\mathcal{P}(U)$ for a given protein U. Here we present one possible mathematical model, it is based on the space filling model used in biochemistry. Anybody who has watched a child playing with a "transformer" may think of this model as a transformer whose basic units are rather simple — they are balls from which are deleted a number of solid spherical segments. The complexity of a protein comes from the fact that the number of balls is huge (from several thousand to several million). The change of shape is also like playing with a transformer, rotating one part while fixing another part.

Axes and planes of bonds. We will use round balls to represent atoms, different kinds of atoms have different radii. A plane cuts a ball into two solid spherical segments, the section of the cutting is a disk. Now suppose that we have two solid spherical segments and their cutting section disks have the same radius. Gluing the two spherical segments along their cutting section disks, we have a chemical bond connecting the two atoms. Let B_1 and B_2 be atoms connected by a bond $D(x)$, where $D(x)$ is a disk centered at x. Let $\vec{l}$ be a unit vector perpendicular to D and issued from x. We will call $\vec{l}$ the *axis* of the bond and the plane P containing D the *plane* of the bond. The *dihedral angle between two bonds* is the dihedral angle between their planes.

Basic convex bodies. An atom may have several bonds connecting to other atoms or atom groups. We define the *basic convex body* as a ball cutting out a number of (disjoint) spherical segments. These basic convex bodies are *building bricks* of proteins, i.e., a protein U is constructed by many basic convex bodies gluing along cutting section disks (pairs with the same radius) such that ∂U is an embedded closed piecewise spherical surface.

Disconnecting bonds. The topological space of the protein U has the feature that if one cuts U along a bond, then either the cutting disconnects U into two connected components U_1 and U_2, or U remains connected after the cutting.

If cutting along a bond of U disconnects U into two components, this bond is a *disconnecting bond.*

Effective bonds. We define a bond D to be *effective* if and only if it is a disconnecting bond and the two components of cutting U along D are both not rotationally symmetric around the bond axis $\vec{l}_D$.

Relative rotations. Let D_j, $j = 1, \cdots, k(U)$, be all effective bonds in U. Let $\vec{l}_j$ be the axis of the effective bond D_j, P_j be the plane of the effective bond D_j. Now we consider relatively rotating U around an axis $\vec{l}_j$.

Fix a basic convex body in U, denote it by o. Since there are two connected components U_1 and U_2 after cutting along bond D_j, we can fix either of them and rotate the other around $\vec{l}_j$. To be specific, we will fix the part which contains the basic convex body o, and rotate the other part, the *rotating part*. We can measure the rotation by the rotating degree θ around $\vec{l}_j$, where $\theta \in [-\pi, \pi]$ and the counter-clockwise around positive direction of $\vec{l}_j$ is positive (right-hand principle).

Rule of changing shapes (how to play with this transformer). The above relative rotations in effective bonds are the only allowed ways to change the shapes of U. But in the real protein case, some effective bonds are not allowed to rotate, for example, the peptide bonds, that makes the real protein folding problems easier.

Restrictions of rotations. All the rotations are subject to the condition (A–2) below. We allow several rotations to take part simultaneously, as long as (A–2) is obeyed. Hence a permitted rotation can be identified as a point (or vector) as

$$\text{(A–1)} \qquad \Theta := (\theta_1, \theta_2, \ldots, \theta_{k(U)}) \in [-\pi, \pi] \times [-\pi, \pi] \times \cdots \times [-\pi, \pi] = \prod_{j=1}^{k(U)} I_j,$$

where $I_j = [-\pi, \pi]$, $j = 1, \ldots, k(U)$. For example, the initial configuration corresponds to the point $\Theta_0 := (0, 0, \ldots, 0) \in \prod_{j=1}^{k(U)} I_j$.

Physicochemical considerations. To represent the physicochemical properties of the protein, we have to restrict further the rotations allowed. For example, without restriction, any two parts in the amino acid sequence may come close when embedded in space, even touch each other. Let A be the set of the 20 amino acids, or atoms, or atom groups, then there are constant $\epsilon(X, Y) \geq 0$ for X, $Y \in A$, such that in certain situations (such as in the cell environment) X and Y keep distance $\epsilon(X, Y)$ from each other. Let C_i, $i = 1, \ldots, n$, be the amino acids (or atoms, or atom groups) in the sequence, then there are constants $\epsilon_{i,j} \geq 0$ such that it must be satisfied in any possible shape (structure):

$$\text{(A–2)} \qquad \operatorname{dist}(C_i, C_j) \geq \epsilon_{i,j} = \epsilon(X, Y), \quad \text{if } X = C_i,\ Y = C_j; \quad 1 \leq i, j \leq n.$$

Hence condition (A–2) helps to decide the admissible structural space $\mathcal{P}(U)$.

Defining the admissible structural space. We also consider the initial positions of the centers of the basic convex bodies $\vec{R} := (\vec{r}_1, \ldots, \vec{r}_{I(U)})$, where $\vec{r}_i \in B_0^i(\rho_i) \subset \mathbb{R}^3$ and $B_0^i(\rho_i)$ is the closed ball of radius $\rho_i > 0$ and centered at $(0, 0, 0)$. Let $\vec{R}_0$ be the most possible initial set of coordinates of the ball centers by physicochemical knowledge. Let

$$S \subset \prod_{i=1}^{k(U)} I_i \times \prod_{j=1}^{I(U)} B_0^j(\rho_j)$$

be the set of all eligible rotations and initial center coordinates satisfying (A–2). Let $\mathcal{P} = \mathcal{P}(U)$ be a connected component of S containing $P_0 = (\Theta_0, \vec{R}_0)$. Then we

can set $P = (\Theta, \vec{R}) \in \mathcal{P}$ as the admissible structural space in (1), and try to get a minimizer.

Topologies. We give $\mathcal{P}$ the subspace topology of $\prod_{i=1}^{k(U)} J = I_i \times \prod_{j=1}^{k(U)} B_0^j(\rho_j)$, then since (A–2) is a closed condition, $\mathcal{P}$ is compact. Moreover, the interior of $\mathcal{P}$ has a natural differential structure.

Let B_i, $i = 1, \ldots, I(U)$ be the list of basic convex bodies of U. Let $W_i(P)$ be their (ball) centers under U_P, $P \in \mathcal{P}(U)$. We will fix $W_1(P) \equiv (0,0,0)$. Then (A–2) can be translated to the distances between $W_i(P)$, according to their radii. Furthermore, we can define distance between U_{P_1} and U_{P_2}, P_1, $P_2 \in \mathcal{P}(U)$, as follows:

$$d(U_{P_1}, U_{P_2})^2 = \min_{S \in \mathrm{SO}^+(3)} \sum_{i=1}^{N} |W_i(P_1) - S(W_i(P_2))|^2,$$

where $\mathrm{SO}^+(3)$ is the 3-dimensional rotation group.

Thus $P \to U_P$ is a continuous map.

Appendix B. Implementation in Computational Biology

The real goal in biology. Although in mathematics the solution surface S_0 is more interesting, the real goal here is to find the structure U_{P_0}. The special properties of solution surfaces can help us in searching for the structure U_{P_0}.

Existence. An important issue is that, even if we can give a pure existence proof for the existence of a minimizer to the protein folding problem (1), it does not really help us to implement an algorithm to calculate a solution by a computer. A constructive proof of existence is the key to practically solving the real protein folding problem in biology.

Selecting structural functions. For the first step of serious investigation, one may try to study these questions for linear structural function,

$$E(x,y,z) = \mu_1 x + \mu_2 y + \mu_3 z, \quad \sum \mu_i = 1, \quad \mu_i \geq 0.$$

Or consider a homothetically homogeneous version,

$$E(x,y,z) = \mu_1 x^{2/3} + \mu_2 y + \mu_3 z, \quad \sum \mu_i = 1, \quad \mu_i \geq 0.$$

The biological significance is that in writing $E(x^{\frac{2}{3}}, y, z)$ and requiring that E be a homogeneous function, we obtain the invariance of the problem $\mathcal{X}(U, \mathcal{P}, E)_0$ against homotheties in $\mathbb{R}^3$. Consider the real sizes of proteins (they are tiny), the structural function E being invariant under homotheties is very important.

For references on protein structures and protein folding phenomena we recommend the book [**1**] and the survey article [**3**] and [**16**].

References

[1] C. Branden and J. Tooze, *Introduction to Protein Structure*, 2nd edition, Garland, 1998.
[2] C. Chothia, The nature of the accessible and buried surfaces in proteins, *J. Mol. Biol* 105 (1976), 1–12.
[3] K. A. Dill, Dominant forces in protein folding, *Biochemistry* 29 (1990), 7133–7155.
[4] Y. Fang, Geometric equilibrium principle: a hypothesis for protein folding, preprint, 2002.
[5] B. Goodwin, *How the leopard changed its spots*, Scribner, 1994.
[6] J. Janin, Surface area of globular proteins, *J. Mol. Biol* 105 (1976), 13–4.
[7] J. Janin, Surface and inside volumes in globular proteins, *Nature* 277 (1979), 491–492.

[8] M. Kanehisa, *Post-genome informatics*, Oxford University Press, 2000.
[9] J. Novotny, R. Bruccoleri, and M. Karplus, An analysis of incorrectly folded protein models: Implications for structure predictions, *J. Mol. Biol.* 177 (1984), 787–818.
[10] J. Novotny, A. A. Rashin, and R. Bruccoleri, Criteria that discriminate between native proteins and incorrectly folded models, *Proteins* 4 (1988), 19–30.
[11] F. M. Richards, The interpretation of protein structures: Total volume, group volume distributions and packing density, *J. Mol. Biol.* 82 (1974), 1–14.
[12] F. M. Richards, Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng.* 6 (1977), 151–176.
[13] A. Ros, Isoperimetric surfaces, pp. 175–209 in this volume.
[14] L. Simon, *Lectures on geometric measure theory*, Proc. of the Centre for Math. Analysis 3, Australian National University, 1983.
[15] I. Stewart, *Life's other secret: the new mathematics of the living world*, Penguin, 1999.
[16] W. R. Taylor, A. C. W. May, N. P.Brown, and A. Aszódi, Protein structure: geometry, topology and classification, *Rep. Prog. Phys.*, 64 (2001), 517–590.

Centre for Bioinformation Science, School of Mathematical Sciences & John Curtin School of Medical Research, Australian National University, Canberra ACT 0200, Australia

E-mail address: yi@maths.anu.edu.au

Clay Mathematics Proceedings
Volume 2, 2004

Minimal and CMC Surfaces Obtained by Ribaucour Transformations

Keti Tenenblat

Introduction

This article describes recent results obtained in my joint work with A. Corro and W. Ferreira, contained in [CFT1] and [CFT2], on Ribaucour transformations as a method of obtaining minimal, constant mean curvature (cmc) and linear Weingarten surfaces from a given such surface.

In the last two decades, a great deal of research activity has been devoted to constructing new complete minimal and cmc surfaces in $\mathbb{R}^3$ (see for example [Co], [HM], [JM]). The main tool in such constructions has been the Weierstrass representation for minimal surfaces, while cmc surfaces have been obtained by using different methods, after the first example of a non-totally umbilical compact immersed cmc surface was found in [W1]. These methods include a perturbation approach [K1; K2], integrable systems [PS] and the conjugate cousin method first introduced by B. Lawson [L] and later used by Karcher [Ka].

In our work we considered Ribaucour transformations to construct not only minimal and cmc surfaces, but also linear Weingarten surfaces. Ribaucour transformations for constant Gaussian curvature and constant mean curvature surfaces, including minimal surfaces, were considered at the beginning of the last century [Bi]. However, the first families of complete minimal surfaces based upon the use of the Ribaucour transformation were obtained recently in [CFT1]. Motivated by this work, we proved in [CFT2] that the classical theory of these transformations could be extended to linear Weingarten surfaces and therefore provided a unified version of the classical theory (see [RS] and [BS] for extensions of other results to special elliptic linear Weingarten surfaces).

We observe that linear Weingarten surfaces are locally parallel to surfaces of constant Gaussian curvature or to minimal surfaces. However, the well-known Ribaucour transformations for these surfaces cannot be applied to produce complete linear Weingarten surfaces, since these parallel constructions in general produce curves of singularities.

In what follows, we start recalling briefly the classical theory of Ribaucour transformation for surfaces. Imposing a one-parameter algebraic condition on a Ribaucour transformation, one has a correspondence between linear Weingarten

Partially supported by CNPq and PRONEX..

©2004 Clay Mathematics Institute

surfaces. Starting with such a surface, the system of equations is integrable and provides a family of new Weingarten surfaces. As a consequence of this theory one has the corresponding results for minimal and cmc surfaces. In this report, we state the main results and we include some applications of the theory, which provide interesting families of complete minimal, cmc and linear Weingarten surfaces, associated to the catenoid and to the cylinder. The reader is referred to [CFT1] and [CFT2] for proofs and other applications.

There is a two-parameter family of complete minimal surfaces associated to the catenoid. For generic values of the parameter c of the Ribaucour transformation, the minimal surfaces are not periodic and have an infinite number of embedded planar ends. However, special values for c, related to irreducible rational numbers, produce 1-periodic surfaces with any finite number of embedded planar ends and two nonplanar ends.

Similarly, associated to the cylinder there is a two-parameter family of complete linear Weingarten, immersed surfaces, explicitly given. These are 1-periodic n-bubble surfaces for special values of c and nonperiodic (with an infinite number of bubbles) otherwise. We point out that the family associated to the cylinder shows the existence of infinitely many complete hyperbolic, linear Weingarten surfaces in $\mathbb{R}^3$. The existence of such surfaces is unexpected, since these surfaces, as well as the surfaces of constant negative curvature, are in correspondence with solutions of the sine-Gordon equation, and Hilbert's theorem proves that there are no complete surfaces of constant negative curvature in $\mathbb{R}^3$.

1. Ribaucour Transformations for Linear Weingarten Surfaces

Let M and $\tilde{M}$ be orientable surfaces of $\mathbb{R}^3$ without umbilic points. We denote by N and $\tilde{N}$ their Gauss map. We say that M and $\tilde{M}$ are *associated by a Ribaucour transformation* if and only if there exists a differentiable function h defined on M and a diffeomorphism $\psi : M \to \tilde{M}$ such that:

(a) at corresponding points, the normal lines intersect at an equidistant point i.e., $p + h(p)N(p) = \psi(p) + h(p)\tilde{N}(\psi(p))$, for all $p \in M$;
(b) the center manifold is two-dimensional i.e., the subset $p + h(p)N(p)$, $p \in M$, is a surface;
(c) ψ preserves lines of curvature.

We say that M and $\tilde{M}$ are *locally associated by a Ribaucour transformation* if for all $p \in M$ there exists a neighborhood of p in M which is associated by a Ribaucour transformation to an open subset of $\tilde{M}$. Similarly, one may consider the notion of *parametrized surfaces locally associated by a Ribaucour transformation.*

The following results give a characterization of Ribaucour transformations.

THEOREM 1.1. *Let M be an orientable surface of $\mathbb{R}^3$ and N its Gauss map. Let e_i, $1 \le i \le 2$ be orthonormal principal directions, λ^i the corresponding principal curvatures, i.e., $dN(e_i) = \lambda_i e_i$. A surface $\tilde{M}$ is associated to M by a Ribaucour transformation if and only if the function $h : M \to R$ is a generic solution of*

$$dZ^j(e_i) + Z^i\omega_{ij}(e_i) - Z^iZ^j\lambda^i = 0, \quad 1 \le i \ne j \le 2 \tag{1–1}$$

where the ω_{ij} are the connection forms of the frame e_i and $Z^i = dh(e_i)/(1 + h\lambda^i)$.

Let ω_i denote the dual frame of e_i. If h is a nonvanishing function which satisfies equation (1–1), $1/h\sum_{i=1}^{2} Z^i\omega_i$ is a closed 1-form and hence there exists a

nonvanishing function Ω, defined on a simply connected domain, such that $d\Omega(e_i) = \Omega Z^i/h$. We define $\Omega_i = d\Omega(e_i)$ and $W = \Omega/h$. With this notation, equation (1–1) is equivalent to a linear system. In fact, a function h is a solution of (1–1) defined on a simply connected domain, if and only if $h = \Omega/W$, where Ω and a nonvanishing function W satisfy

$$d\Omega_i(e_j) = \Omega_j \omega_{ij}(e_j), \quad \text{for } i \neq j, \tag{1–2}$$

$$d\Omega = \sum_{i=1}^{2} \Omega_i \omega_i, \tag{1–3}$$

$$dW = -\sum_{i=1}^{2} \Omega_i \lambda^i \omega_i. \tag{1–4}$$

It is easy to see that equation (1–2) is the integrability condition of the system of equations (1–3), (1–4) for Ω and W. Moreover, if a surface in $\mathbb{R}^3$ is parametrized by $X : U \subset \mathbb{R}^2 \to M$, one can show that a surface $\tilde{M}$ is locally associated to M, by a Ribaucour transformation, if and only if, there exist differentiable functions $W, \Omega, \Omega_i : V \subset U \to R$, which satisfy (1–2)–(1–4) and $\tilde{X} : V \subset \mathbb{R}^2 \to \tilde{M}$, is a parametrization of $\tilde{M}$ given by

$$\tilde{X} = X - \frac{2\Omega}{S}\left(\sum_i \Omega_i e_i - WN\right), \tag{1–5}$$

where

$$S = \sum_i (\Omega_i)^2 + W^2. \tag{1–6}$$

From now on, whenever we say that two surfaces are locally associated by a Ribaucour transformation we are assuming that there is a local solution of the system (1–2)-(1–4), that the surfaces have no umbilic points and they are locally related as in (1–5).

A *linear Weingarten surface* of $\mathbb{R}^3$ is a surface whose Gaussian and mean curvature K and H satisfy a linear relation $\alpha + \beta H + \gamma K = 0$, where $\alpha, \beta, \gamma \in R$. A sufficient condition for a Ribaucour transformation to transform a linear Weingarten surface into another such surface was given in [CFT2]. This result, which is our next theorem, provides a unified version of the classical results for constant mean and Gaussian curvatures and extends these results to linear Weingarten surfaces.

THEOREM 1.2. *Let M and $\tilde{M}$ be regular surfaces of $\mathbb{R}^3$, which are associated by a Ribaucour transformation, such that the normal lines intersect at a distance function h. Assume that $h = \Omega/W$ is not constant along the lines of curvature and the functions Ω_i, Ω and W satisfy the additional relation*

$$S = 2c(\alpha\Omega^2 + \beta\Omega W + \gamma W^2), \tag{1–7}$$

where S is defined by (1–6), $c \neq 0$ and α, β, γ are real constants. Then $\tilde{M}$ is a linear Weingarten surface satisfying $\alpha+\beta\tilde{H}+\gamma\tilde{K} = 0$, if and only if $\alpha+\beta H+\gamma K = 0$ holds for the surface M, where K, H and $\tilde{K}$, $\tilde{H}$ are the Gaussian and mean curvatures of M and $\tilde{M}$, respectively.

One may ask if the system (1–2)–(1–4) with the additional condition (1–7) is integrable, whenever we start with a linear Weingarten surface. The answer to this question is affirmative.

THEOREM 1.3 (Integrability). *Let M be a linear Weingarten surface of $\mathbb{R}^3$ satisfying $\alpha + \beta H + \gamma K = 0$. Then the system of equations (1–2)–(1–4) and (1–7) is integrable and the solution is uniquely determined on a simply connected domain U, by any given initial condition satisfying (1–7). Moreover, whenever $\alpha \neq 0$, any solution of the system defined on U is either identically zero (and hence annihilates S) or else the function S does not vanish on U.*

As a consequence, one can prove that if M is a linear Weingarten surface locally parametrized by $X : U \subset \mathbb{R}^2 \to M \subset \mathbb{R}^3$, then any linear Weingarten parametrized surface $\tilde{X}$, locally associated to X by a Ribaucour transformation as above, is given by (1–5), where e_i are orthogonal principal directions, Ω, Ω_i, W are solutions of (1–2)–(1–4) and (1–7), and $\tilde{X}$ is an immersed regular surface defined on

$$\tilde{U} = \left\{ (u_1, u_2) \in U : T^2 + 2TQH + Q^2K \neq 0 \right\},$$

where $T = \alpha\Omega^2 - \gamma W^2$ and $Q = 2\gamma\Omega W + \beta\Omega^2$. In general, the surface $\tilde{X}$ depends on four parameters. However, in some cases the number of parameters may reduce to one (the parameter c), if one considers surfaces which are congruent by rigid motions of $\mathbb{R}^3$. Special cases of the above results include the minimal surfaces and the cmc surfaces.

The cmc surfaces are obtained by considering $\alpha = -H \neq 0$, $\beta = 1$, $\gamma = 0$ and hence the algebraic condition (1–7) reduces to $S = 2c\Omega(-H\Omega + W)$, where c satisfies the relation $c(c - 2H) > 0$. For any nontrivial solution of the system (1–2)–(1–4) and (1–7), defined on a simply connected domain U, the function S does not vanish. Hence, if $X : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ is a cmc surface, then a cmc surface $\tilde{X}$ associated to X by a Ribaucour transformation is regular on open subsets of U, where X has no umbilic points.

The case of the minimal surfaces is obtained by considering $\alpha = 0$, $\beta = 1$, and $\gamma = 0$ and the algebraic condition reduces to $S = 2c\Omega W$. One can show that the Ribaucour transformations for minimal surfaces are related to producing embedded planar ends for the new associated minimal surfaces. In fact such ends are produced by the isolated zeros of S, where Ω does not vanish.

The reader is referred to [CFT1] for proofs and details in the case of the minimal surfaces and to [CFT2] for the linear Weingarten and cmc surfaces.

2. Minimal Surfaces Associated to the Catenoid

The first families of complete minimal surfaces obtained by Ribaucour transformations were given in [CFT1], by applying the method to Enneper's surface and to the catenoid. The family of minimal surfaces associated to Enneper's surface is explicitly given and depends on three parameters. Each surface of this family has infinite total curvature and corresponds to a complete immersion of a sphere punctured at an infinite number of points, which are contained on a circle and accumulate at the pole. All except one of the infinite number of ends are embedded planar ends, whose positions are determined by the parameters.

The family associated to the catenoid has more interesting features. Depending on the value of the parameter c of the Ribaucour transformation, the associated surface may have infinitely many embedded planar ends (see [CHM] for minimal surfaces with an infinite number of annular ends) or any finite number of embedded planar ends (see [JM] for minimal surfaces with any finite number of catenoid ends). Moreover, each surface has one or two nonplanar ends. We point out that the family

of minimal surfaces associated to the catenoid are of genus zero and contain a special class of 1-periodic surfaces. In the following results we give a brief description of the family associated to the catenoid.

PROPOSITION 2.1. *Consider the catenoid parametrized by*

$$X(u_1,u_2) = (\cos u_2 \cosh u_1,\ \sin u_2 \cosh u_1,\ u_1)\,.$$

Excluding the catenoid and up to rigid motions of $\mathbb{R}^3$, *a parametrized surface* $\tilde{X}_c(u_1,u_2)$ *is a minimal surface locally associated to* X *by a Ribaucour transformation as above, if and only if,*

$$\tilde{X}_c = X - \frac{\cosh u_1}{c}(\cos u_2, \sin u_2, 0) + \frac{1}{c(f+g)}(f' X_{u_1} - g' X_{u_2}), \tag{2–1}$$

where $c \neq 0$, $f(u_1)$ *and* $g(u_2)$ *are given as follows:*

(a) *if* $c = 1/2$, *then*

$$f = \frac{(c_1 u_1 + b_1)^2}{2c_1}, \quad g = \frac{c_1 u_2^2}{2},$$

where $c_1 \neq 0, b_1 \in R$, *and the function* $\tilde{X}_{1/2}$ *is defined on* $\mathbb{R}^2 \setminus \{p_1\}$ *with*

$$p_1 = -\frac{1}{c_1}(b_1,\ 0);$$

(b) *if* $2c-1 > 0$, *then* $f = \sin(A + \sqrt{2c-1}\,u_1)$, $\quad g = \pm\cosh(\sqrt{2c-1}\,u_2)$, $A \in R$ *and the function* $\tilde{X}_c$ *is defined on* $\mathbb{R}^2 \setminus \{p_k : k \in Z\}$, *where*

$$p_k = \frac{1}{\sqrt{2c-1}}\,(\mp\pi/2 - A + 2k\pi,\ 0)\,.$$

(c) *if* $1-2c > 0$, *then* $f = \pm\cosh(A + \sqrt{1-2c}\,u_1)$, $\quad g = \sin(\sqrt{1-2c}\,u_2)$, $A \in R$ *and the function* $\tilde{X}_c$ *is defined on* $\mathbb{R}^2 \setminus \{p_k : k \in Z\}$, *where*

$$p_k = \frac{1}{\sqrt{1-2c}}\,(-A,\ \mp\pi/2 + 2k\pi)\,.$$

One can prove:

PROPOSITION 2.2. *Any minimal surface locally associated to the catenoid by a Ribaucour transformation given by Proposition 2.1. is complete.*

As one may expect from Proposition 2.1, the geometric properties of the minimal surfaces associated to the catenoid are quite distinct, depending on the value of the parameter c. Our next two results describe these geometric properties. In particular, for special values of c, namely when $\sqrt{1-2c} = n/m$ is a rational number, the associated surface, which will be denoted by $\tilde{X}_{(n,m)}$, is 1-periodic in the variable u_2. Figure 1 contains several examples of such surfaces.

PROPOSITION 2.3. *Let* c *be a real number such that* $0 \neq c < 1/2$ *and* $\sqrt{1-2c} = n/m$ *is an irreducible rational number, with* $n \neq m$. *Consider* $\tilde{X}_{(n,m)}$ *the family of minimal surfaces associated to the catenoid as in Proposition 2.1.c). Let* $\mathcal{F}^{\pm}$ *be the two ends of any surface of* $\tilde{X}_{(n,m)}$ *corresponding to* $u_1 \to \pm\infty$. *Then*

(a) *Any surface of* $\tilde{X}_{(n,m)}$ *is a complete minimal surface corresponding to an immersion of a sphere punctured at* $n+2$ *points: the two poles and* n *points contained on a circle.*

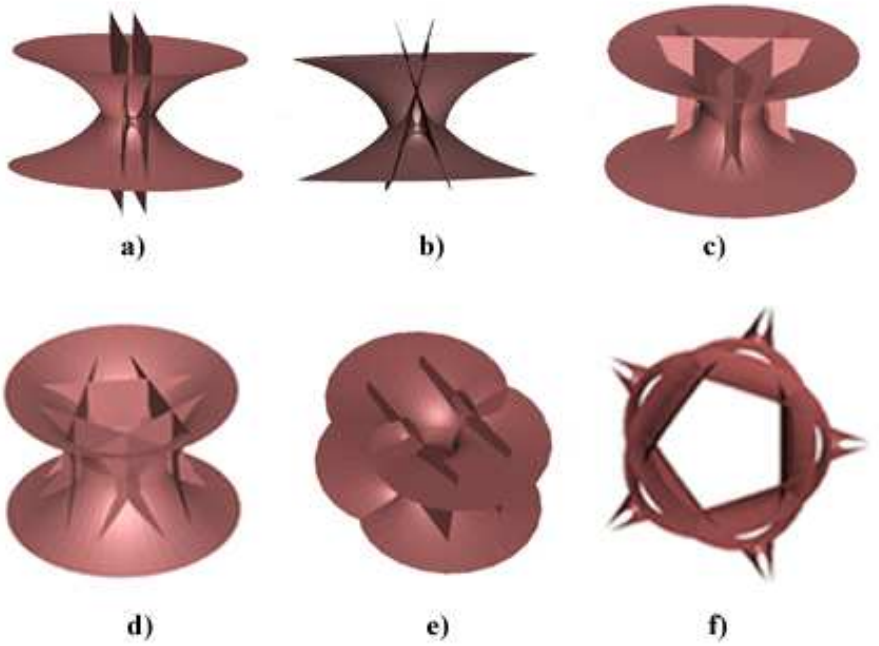


FIGURE 1. Complete, minimal, 1-periodic surfaces $\tilde{X}_{(n,m)}$ associated to the catenoid by Ribaucour transformations: (a) $\tilde{X}_{(2,1)}$ with $A = 0$. (b) $\tilde{X}_{(2,1)}$ with $A = 0.5$. (c) $\tilde{X}_{(3,1)}$ with $A = 0$. (d) $\tilde{X}_{(5,1)}$ with $A = 0$. (e) $\tilde{X}_{(2,3)}$ with $A = 0$. (f) $\tilde{X}_{(5,2)}$ with $A = 0$ an upper view.

(b) *Its Gauss map $\tilde{N}$ satisfies*

$$\lim_{u_1 \to \pm\infty} \tilde{N}(u_1, u_2) = (0, 0, \pm 1), \qquad \lim_{(u_1,u_2)\to p_k} \tilde{N}(u_1, u_2) = N(p_k),$$

where N is the normal map of the catenoid and the p_k, for $k = 1, \ldots, n$, are defined by Proposition 2.1(c).

(c) *Any surface of $\tilde{X}_{(n,m)}$ has n embedded planar ends and two ends $\mathcal{F}^{\pm}$ of geometric index m, that grow asymptotically as the ends of the catenoid. In particular, $\mathcal{F}^{\pm}$ are embedded if and only if $m = 1$; in this case $\mathcal{F}^{\pm}$ are catenoid ends.*

(d) *The total curvature of any surface $\tilde{X}_{(n,m)}$ is equal to $-4\pi(n + m)$.*

It follows from the above result that for any integer $j \geq 3$, there exists a finite number of two-parameter families of immersed complete minimal surfaces $\tilde{X}_{(n,m)}$ with $j = n + m$, whose total curvature is $-4\pi j$. Any such surface has at least one and at most $j - 1$ embedded planar ends and two nonplanar ends. In particular, any surface of $\tilde{X}_{(j-1,1)}$ has $j - 1$ planar ends and two catenoid ends.

Whenever c does not satisfy the conditions of Proposition 2.3, the surface $\tilde{X}$ is not periodic in any variable. These surfaces are treated in the following result.

PROPOSITION 2.4. *Let $\tilde{X}_c$ be the family of minimal surfaces locally associated to the catenoid by a Ribaucour transformation given in Proposition 2.1. Assume $2c-1 \geq 0$ or $2c-1 < 0$ and $\sqrt{1-2c}$ is not a rational number. Then*

(a) *Any surface of the family $\tilde{X}_c$ is a complete minimal surface corresponding to a sphere punctured at two points if $2c-1=0$, and punctured at an infinite number of points contained on a circle otherwise.*
(b) *Its normal map $\tilde{N}$ satisfies $\lim_{(u_1,u_2)\to p_k} \tilde{N}(u_1,u_2) = N(p_k)$, where N is the normal map of the catenoid and p_k are given by Proposition 2.1(a)–(c) respectively.*
(c) *Except for one, all ends are planar.*
(d) *Any surface of $\tilde{X}_c$ has infinite total curvature.*

We observe that whenever $c = 1/2$, any surface of $\tilde{X}_{1/2}$ is an immersed minimal surface with two ends. One of them, corresponding to p_1, is planar and the other one, corresponding to the pole, is not embedded. Moreover, as a consequence of Propositions 2.3 and 2.4, any minimal surface of the family $\tilde{X}_c$ given by (2–1) such that $0 \neq c < 1/2$, with infinite total curvature, is the limit of a sequence of minimal surfaces of the family $\tilde{X}_{(n,m)}$, whose total curvature is $-4\pi(n+m)$.

We conclude this section observing that the families of minimal surfaces, associated to the catenoid, depend on two parameters. The geometric properties of the surfaces are determined by c while the position of the planar ends is determined by the other parameter.

3. Linear Weingarten and CMC Surfaces Associated to the Cylinder

In this section, we describe a two-parameter family of linear Weingarten surfaces and a one-parameter family of cmc surfaces associated to the cylinder by Ribaucour transformations. Depending on the value of the parameter c of the transformation, the associated surfaces are periodic in one variable or not. In the periodic case, the total absolute curvature of the surfaces is a multiple of 8π and the surfaces have a finite number of bubbles, characterized by the number of points of maximum and minimum for the Gaussian curvature. In the nonperiodic case, the total absolute curvature is infinite and there are infinitely many bubbles.

PROPOSITION 3.1. *Consider the cylinder parametrized by*

$$X(u_1,u_2) = (\cos(u_2), \sin(u_2), u_1) \qquad (u_1,u_2) \in \mathbb{R}^2 \tag{3–1}$$

as a linear Weingarten surface satisfying $-1/2 + H + \gamma K = 0$. A parametrized surface is a linear Weingarten surface locally associated to X by a Ribaucour transformation as in Section 1 if and only if it is given by

$$\tilde{X}_{c\gamma} = X - \frac{2(f+g)}{c[(2\gamma+1)g^2 - f^2]}(f' X_{u_1} + g' X_{u_2} - gN), \tag{3–2}$$

where N is the inner unit normal vector field of the cylinder, $c \neq 0$ and γ are real constants, such that

$$\xi(c,\gamma) = 1 - c(2\gamma+1) \tag{3–3}$$

and c are not simultaneously positive, and $f(u_1)$, $g(u_2)$ are solutions of the equations

$$f'' + cf = 0, \qquad g'' + \xi g = 0, \tag{3–4}$$

with initial conditions satisfying

$$\left((f')^2+(g')^2+\xi g^2+cf^2\right)(u_1^0,u_2^0)=0.$$

Moreover, $\tilde{X}_{c\gamma}$ is a regular surface defined on the subset of $U\subset\mathbb{R}^2$ where

$$\left((f+g)^2+2\gamma g^2\right)\left(f^2+2(2\gamma+1)fg+(2\gamma+1)g^2\right)\neq 0.$$

One can show that the linear Weingarten surfaces $\tilde{X}_{c\gamma}$ associated to the cylinder and parametrized by (3–2) (excluding the cylinder), have curves of singularity if $c\xi\geq 0$. Moreover, if $c\xi<0$, then, up to rigid motions of $\mathbb{R}^3$, the surface $\tilde{X}_{c\gamma}$ is determined by the functions

$$f=\varepsilon_1\sqrt{|\xi|}\sin(\sqrt{c}\,u_1),\qquad g=\varepsilon_2\sqrt{c}\cosh(\sqrt{|\xi|}u_2)\qquad \text{if } c>0,\xi<0$$
$$f=\varepsilon_1\sqrt{\xi}\cosh(\sqrt{|c|}u_1),\qquad g=\varepsilon_2\sqrt{|c|}\sin(\sqrt{\xi}\,u_2)\qquad \text{if } c<0,\xi>0$$

where $\varepsilon_i=\pm 1$, $c\neq 0$ and γ are real numbers and $\xi(c,\gamma)$ is defined by (3–3). In order to characterize when the surface $\tilde{X}_{c\gamma}$ is complete, we consider the two connected components of $\mathbb{R}^2$ described in Figure 2. The functions that determine the regions of Figure 2 are defined by

$$h_1(c,\gamma)=2c(2\gamma+1)\left(\sqrt{2\gamma(2\gamma+1)}-2\gamma\right)-1,$$
$$h_2(c,\gamma)=2c\left(\sqrt{2|\gamma|}+2\gamma\right)-1,$$
$$h_3(c,\gamma)=-2c(2\gamma+1)\left(\sqrt{2\gamma(2\gamma+1)}+2\gamma\right)-1,$$
$$h_4(c,\gamma)=2c\left(\sqrt{2|\gamma|}-2\gamma\right)+1,$$
$$h_5(c,\gamma)=-2c(2\gamma+1)\left(\sqrt{2\gamma(2\gamma+1)}-2\gamma\right)+1.$$

One can prove that the connected region in Figure 2, where $c<0$, contains an infinite number of curves such that the corresponding surfaces are complete 1-periodic linear Weingarten surfaces, which have finite total absolute curvature. For pairs (c,γ) not on these curves, the associated surfaces are not periodic in any variable. More precisely:

PROPOSITION 3.2. *Any linear Weingarten surface $\tilde{X}_{c\gamma}$, given by Proposition 3.1 is complete if and only if $c\xi(c,\gamma)<0$ and the pair (c,γ) belongs to one of the connected components described in Figure 2.*

(a) *If $c<0$ and $\sqrt{\xi(c,\gamma)}=n/m$ is an irreducible rational number, then $\tilde{X}_{c\gamma}$ is an immersion of a cylinder into $\mathbb{R}^3$, with two ends of geometric index m and n isolated points of maximum (respectively minimum) for the Gaussian curvature. Moreover, the total curvature of $\tilde{X}_{c\gamma}$ is zero, while its total absolute curvature is $8\pi n$. The ends are embedded if and only if $m=1$; in this case they are cylindrical ends.*

(b) *If $c>0$ or $c<0$ and $\sqrt{\xi}$ is not a rational number, then $\tilde{X}_{c\gamma}$ is an immersion of $\mathbb{R}^2$ into $\mathbb{R}^3$ with an infinite number of isolated critical points of its Gaussian curvature.*

One can also show that the complete linear Weingarten surfaces $\tilde{X}_{c\gamma}$ considered in Proposition 3.2. are asymptotically close to the cylinder. The symmetries of the surfaces are quite distinct for (c,γ) in each connected component of Figure 2 and they are given explicitly (see [CFT2]) in terms of rotations, translations and reflections with respect to certain planes. The lines of curvature $\tilde{X}_{c\gamma}(u_1,u_2^0)$ are

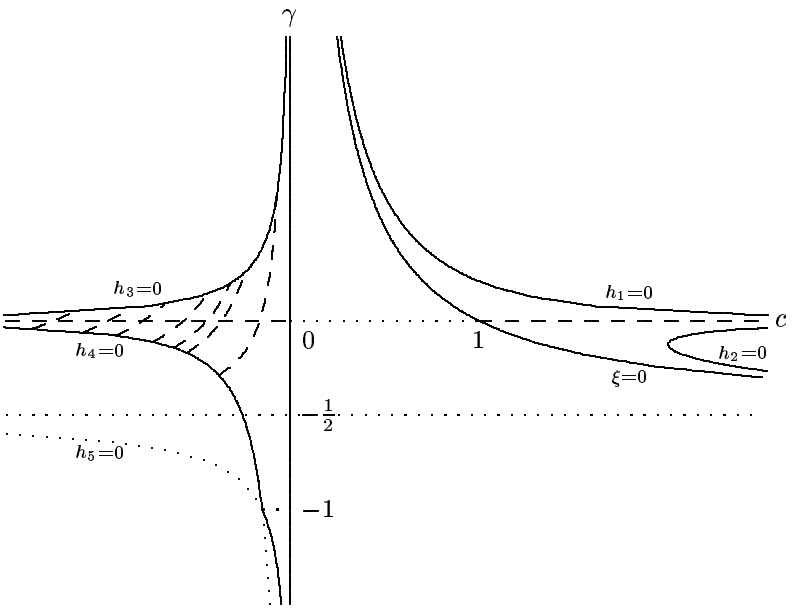


FIGURE 2. Any pair (c, γ), in each of the two connected components, generates a complete linear Weingarten surface, which satisfies the relation $-1/2 + H - \gamma K = 0$ and it is cmc when $\gamma = 0$. The dashed curves in the left region, given by $1 - c(2\gamma + 1) = n^2/m^2$, generate 1-periodic n-bubble surfaces with two ends of geometric index m (see Figure 3). For other values of (c, γ) the surfaces are not periodic.

planar, while the curves $\tilde{X}_{c\gamma}(u_1^0, u_2)$, when $f'(u_1^0) \neq 0$, are contained on a sphere centered at $(0, 0, f/f'(u_1^0))$ with radius $\sqrt{1 + (f/f')^2}$. Whenever $f'(u_1^0) = 0$ (it can only occur for $c > 0$), then $\tilde{X}_{c\gamma}(u_1^0, u_2)$ are planar lines of curvature. In Figure 3, one can visualize some of the surfaces given by $\tilde{X}_{c\gamma}$.

We observe that the linear Weingarten surfaces given by $\tilde{X}_{c\gamma}$ are tubular surfaces when $\gamma = -1/2$, since they satisfy $\Delta = \beta^2 - 4\alpha\gamma = 0$, and they provide examples of complete surfaces with $\Delta < 0$ and $\Delta > 0$. We point out that the linear Weingarten surfaces with no umbilic points for which $\Delta < 0$ (respectively $\Delta > 0$) correspond (see for example [T]) to solutions of a hyperbolic (respectively elliptic) differential equation, namely the sine-Gordon equation (respectively elliptic sinh-Gordon or cosh-Gordon equations). Hilbert's theorem asserts that there are no complete surfaces of constant negative Gaussian curvature immersed in $\mathbb{R}^3$ and it is well known that such surfaces correspond to solutions of the sine-Gordon equation. However, the examples $\tilde{X}_{c\gamma}$, where $c < 0$ and $\gamma < -1/2$, show that there are complete linear Weingarten surfaces immersed in $\mathbb{R}^3$ for which $\Delta < 0$.

In what follows, we describe a one-parameter family of $1/2$-cmc surfaces obtained from the cylinder by Ribaucour transformations. The family is contained in the class of linear Weingarten surfaces described in Proposition 3.1. and are explicitly obtained by considering $\gamma = 0$. These surfaces could also be obtained directly

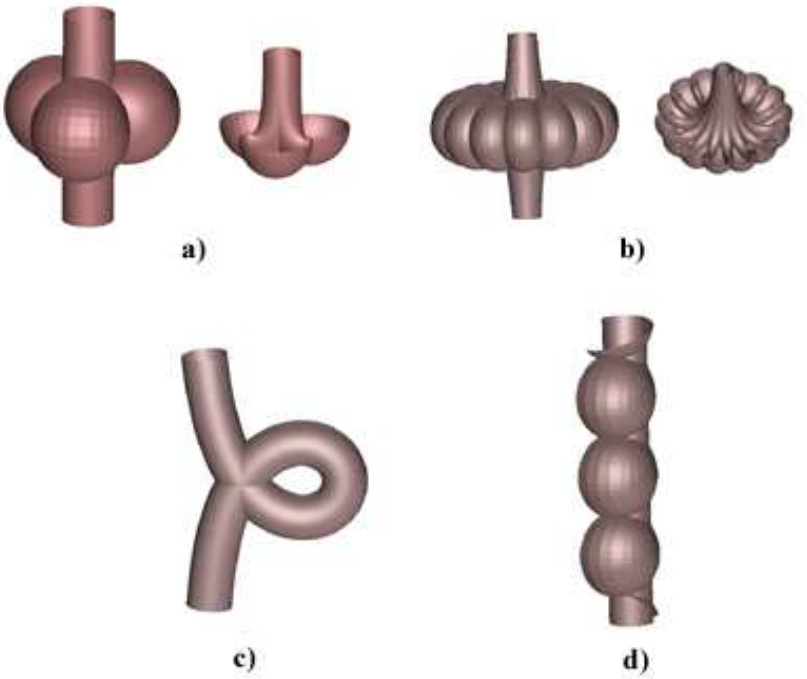


FIGURE 3. Complete Weingarten surfaces $\tilde{X}_{c\gamma}$ which satisfy the relation $-1/2 + H + \gamma K = 0$ and are associated to the cylinder by Ribaucour transformations. (a) 1-periodic cmc surface obtained by considering $\gamma = 0$ and $\sqrt{1-c} = 3/2$; (b) 1-periodic Weingarten surfaces for which $\gamma = 0.2$, $\sqrt{1-c(2\gamma+1)} = 14/13$; (c) tubular surface, $\gamma = -1/2$, $c = -0.1$; (d) cmc surface (not periodic) where $c = 2.8$ and $\gamma = 0$.

from the cylinder by applying the Ribaucour transformation with the appropriate algebraic condition. More precisely:

PROPOSITION 3.3. *Excluding the cylinder, any* $1/2$*-cmc parametrized surface locally associated to the cylinder as in Section 1, is given, up to rigid motions of* $\mathbb{R}^3$*, by*

$$\tilde{X}_c = X + \frac{2}{c(f-g)}(f'X_{u_1} + g'X_{u_2} - gN) \tag{3–5}$$

where N *is the inner unit normal field of the cylinder parametrized by (3–1),* c *is a real constant such that* $c < 0$ *or* $c > 1$*, and the functions* $f(u_1)$ *and* $g(u_2)$ *are given by*

$$f = \varepsilon_1\sqrt{c-1}\sin(\sqrt{c}\,u_1), \qquad g = \varepsilon_2\sqrt{c}\cosh(\sqrt{c-1}\,u_2) \qquad \text{if } c > 1,$$
$$f = \varepsilon_1\sqrt{1-c}\cosh(\sqrt{|c|}\,u_1), \qquad g = \varepsilon_2\sqrt{|c|}\sin(\sqrt{1-c}\,u_2) \qquad \text{if } c < 0.$$

Moreover, one can show that these cmc surfaces correspond to solutions of finite type 1 of the equation $w_{z\bar{z}} + \sinh(2w)/2 = 0$, where z is the complex variable $z = u_1 + iu_2$, in the sense introduced by Pinkall and Sterling in [PS]. Some of the properties of these cmc surfaces are given in the following result.

PROPOSITION 3.4. *Any 1/2-cmc surface $\tilde{X}_c$ given by (3–5) is a complete surface asymptotically close to the cylinder. Moreover,*

(a) *If $c < 0$ and $\sqrt{1-c} = n/m$ is an irreducible rational number, then $\tilde{X}_c$ is an immersion of a cylinder into $\mathbb{R}^3$, with two ends of geometric index m and n isolated points of maximum (respectively minimum) for the Gaussian curvature. The total curvature of $\tilde{X}_c$ is zero, while its total absolute curvature is $8\pi n$. The ends are embedded if and only if $m = 1$. In this case they are cylindrical ends.*

(b) *If $c > 1$ or $c < 0$ and $\sqrt{1-c}$ is not a rational number, then $\tilde{X}_c$ is an immersion of $\mathbb{R}^2$ into $\mathbb{R}^3$, with an infinite number of isolated critical points of its Gaussian curvature.*

The cmc n-bubble surfaces given explicitly by Proposition 3.4(a) were first described by Sievert [S] for $n = 2$ (see also [PS]) and their existence was proved later in [G-B] and [SW]. Moreover, we observe that the cmc surfaces considered in Proposition 3.3, whenever $c < 0$, are called of Enneper type by Wente [W2], since one family of curvature lines is spherical.

Similarly, by using Ribaucour transformations we obtain families of cmc surfaces associated to the Delaunay surfaces. By restricting the range of the parameter c of the Ribaucour transformation, we provide families of complete cmc surfaces which contain an enumerable subset of one periodic surfaces (see [CFT2]).

References

[Bi] L. Bianchi, *Lezioni di geometria differenziale*, Bologna, Zanichelli, 1927.

[BS] F. B. Brito and R. Sa Earp, *On the structure of certain Weingarten surfaces with boundary a circle*, Ann. Fac. Sci. Toulouse 6 (1997), 243–256.

[CHM] M. Callahan, D. Hoffman and W. Meeks III, *Embedded minimal surfaces with an infinite number of ends*, Invent. Math. 96, (1989), 459–505.

[CFT1] A. V. Corro, W. Ferreira and K. Tenenblat, *Minimal surfaces obtained by Ribaucour transformations*, Geometria Dedicata 96 (2003), 117–150.

[CFT2] A. V. Corro, W. Ferreira and K. Tenenblat, *Ribaucour transformations for cmc and linear Weingarten surfaces*, Pac. J. Math. 212 (2003), 265–296.

[Co] C. J. Costa, *Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends*, Bull. Soc. Bras. Mat. 15 (1984), 47–54.

[G-B] K. Große-Brauckmann, *New surfaces of constant mean curvature*, Math. Zeit. 214 (1993), 527–565.

[HM] D. Hoffman and W. Meeks III, *Embedded minimal surfaces of finite topology*, Ann. of Math. 131 (1990), 1–34.

[JM] L. P. Jorge and W. Meeks III, *The topology of complete minimal surfaces of finite total Gaussian curvature* Topology, 2 (1983), 203–221.

[K1] N. Kapouleas, *Complete constant mean curvature surfaces in Euclidean three space*, Ann. of Math. 131 (1990), 239–330.

[K2] N. Kapouleas, *Constant mean curvature surfaces constructed by using Wente tori*, Invent. Math. 119, (1995), 443–518.

[Ka] H. Karcher, *The triply periodic minimal surfaces of A. Schoen and their constant mean curvature companions*, Man. Math. 64 (1989), 291–357.

[L] B. Lawson, *Complete minimal surfaces in S^3*, Ann. Math. 92 (1970), 335–374.

[PS] U. Pinkall and I. Sterling, *On classification of constant mean curvature mean tori*, Ann. Math. 130 (1989), 407–451.

[RS] H. Rosenberg and R. Sa Earp, *The geometry of properly embedded special surfaces in* $\mathbb{R}^3$*; e.g. surfaces satisfying* $aH + bK = 1$*, where* a *and* b *are positive*, Duke Math J. 73 (1994), 291–306.

[S] H. Sievert, *Über die Zentralflächen der Enneperschen Flächen konstanten Krümmungsmasses*, Diss. Tübingen (1886).

[T] K. Tenenblat, Transformations of manifolds and applications to differential equations, Pitman Monographs and Surveys in Pure and Applied Mathematics 93, Addison Wesley Longman, 1998.

[SW] I. Sterling and H. C. Wente, *Existence and classification of cmc multibubbleton of finite and infinite type*, Indiana Univ. Math. J. 42 (1993), 1239–1266.

[W1] H. C. Wente, *Counterexample to a conjecture of H. Hopf*, Pac. J. Math. 121 (1986), 193–243.

[W2] H. C. Wente, *Constant mean curvature immersions of Enneper type*, Memoirs of the AMS 478, 1992.

Departamento de Matemática, Universidade de Brasília, 70910-900, Brasília, DF, Brazil

E-mail address: keti@mat.unb.br

Clay Mathematics Proceedings
Volume 2, 2004

Meromorphic Data for Surfaces of Mean Curvature One in Hyperbolic Space, II

Ricardo Sa Earp and Eric Toubiana

Abstract. We show that given two complex functions E and G there are necessary and sufficient conditions for the existence of a conformal, possibly branched, immersion into hyperbolic space whose euclidean Gauss map is E and hyperbolic Gauss map is G. Furthermore, we use meromorphic data and a representation theorem to give some constructions of complete and conformal mean curvature one immersions into hyperbolic space.

Introduction

In a recent work the authors have given meromorphic data (h, T) for simply connected mean curvature one surfaces in hyperbolic space. This construction led to a representation theorem for any mean curvature one surface in hyperbolic space [**18**]. We were motivated by the following philosophy: Any conformal immersion in the half-space model of hyperbolic space should be expressed in terms of the standard euclidean Gauss E map and the hyperbolic Gauss map G. This is true in general. We will make this more precise in the statement of Theorem 3. Thus, it is possible to compute the coordinate maps of the immersion and its geometric quantities by the means of E and G (see [**18**], [**19**] and the first section of the present paper). Furthermore for mean curvature one immersion any geometric quantity can be written down using solely the euclidean Gauss map E.

Our main goal in this paper is to show that these results are effective to construct examples for which we get control of the branch points. The constructions we made are rather general and one can derive other examples based on them. The first example is a two-parameter family of complete mean curvature one conformal immersions of the twice punctured sphere. One end is regular with finite total curvature and the other end is nonregular and has infinite total curvature. Some of these examples do not admit intrinsic isometry. The second example has two horizontal periodic ends (with the same period) and a third end which is regular with finite total curvature. This surface is invariant under a discrete group of horizontal translations isomorphic to $\mathbb{Z}$. The third example has the same properties except

1991 *Mathematics Subject Classification.* 53A10, 53C42.

Key words and phrases. mean curvature one surfaces, meromorphic data, euclidean Gauss map, hyperbolic Gauss map.

Both authors are supported in part by CNPq, FAPERJ and PRONEX, Brasil.

©2004 Clay Mathematics Institute

that the third end is nonregular and has infinite total curvature. In the fourth example we construct for any polynomial map T a complete and conformal mean curvature one immersion with many horizontal periodic ends. Each immersion is modeled by the sphere punctured at a finite set of points. We note that for any polynomial map T, $(z = \infty)$ is a regular end with finite total curvature. In the last construction we write down explicit examples of complete branched immersions of any compact Riemann surface $\overline{S}$ punctured at a finite set of points.

In this work we focus on the upper half-space model of hyperbolic $\mathbb{H}^3$ space, namely $\mathbb{H}^3 := \{(u, v, w) \in \mathbb{R}^3,\ w > 0\}$. In this model the *euclidean Gauss map* E has a natural meaning. By the way there is a certain affinity between our work and the work of Kenmotsu published in 1979 [**6**] about a Weierstrass representation for constant (nonzero) mean curvature surfaces in euclidean space. As a matter of fact Kenmotsu has considered a simply connected domain U and the following equation on U :

$$E_{z\overline{z}} = 2\frac{\overline{E}E_z E_{\overline{z}}}{1 + E\overline{E}}.$$

Roughly speaking, Kenmotsu said that finding a solution to this equation is tantamount to getting a conformal, possibly branched, immersion with constant mean curvature in euclidean space. But it is difficult to find nontrivial explicit solutions of Kenmotsu's equation. We observe that Ritoré has used the Kenmotsu theorem to construct periodic surfaces with constant mean curvature in $\mathbb{R}^3$, see [**10**]. For some discussions about harmonic maps to the two-sphere, see for example [**5**] and [**2**]. It turns out that we have considered a similar equation for mean curvature one surfaces in hyperbolic space (think $E(z)$ as the euclidean Gauss map and $z \in U$, where U is a simply connected domain):

$$(*) \qquad E_{z\overline{z}} = \frac{\overline{E}E_z E_{\overline{z}}}{1 + E\overline{E}}.$$

We have solved this equation and we have expressed any solution E in terms of meromorphic data (h, T) on U, see Theorem 9. Given any such solution E we have proved that E gives rise to a conformal mean curvature one branched immersion into $\mathbb{H}^3$. The branch points are isolated and are explicitly given by the zeros of $E_{\overline{z}}$, see Theorem 8. Furthermore the height function w, the euclidean Gauss map, the metric and the second fundamental form can be given readily in terms of (h, T). For instance,

$$w = \frac{|h^2 T_z|^2}{|(Th)_z|^2 + |h_z|^2}.$$

Now the horizontal component $u + iv$ is given by one complex integration involving (h, T), that is $\int h^2\, dT$. Therefore, theoretically one can control the branch points in order to construct regular conformal (mean curvature one) immersions. To conclude this introduction we would like to say that we had to look at mean curvature one surfaces in the general context of conformal immersion (with arbitrary mean curvature) into hyperbolic space. In fact, we have derived general formulas for any conformal immersions that are essential for the development of the theory. For instance the following formulas are crucial:

$$u + iv = G - wE, \qquad G_z = wE_z,$$

where $u + iv$ are the horizontal components, w is the height, E is the euclidean Gauss map and G is the hyperbolic Gauss map.

For related results on mean curvature one surfaces in hyperbolic space the reader is referred to Bryant [**1**]; Collin, Hausswirth and Rosenberg [**3**]; Pacard and Pimentel [**8**]; Rosenberg [**11**], Rossman and Sato [**12**], Rossman, Umehara and Yamada [**13**], [**14**], [**15**], [**16**]; Small [**20**]; Umehara and Yamada [**21**], [**22**], [**23**]; see also [**17**].

1. Basic Results for Conformal Immersions into $\mathbb{H}^3$

We consider the half-space model of the hyperbolic 3-space and we denote it by $\mathbb{H}^3$, namely

$$\mathbb{H}^3 = \{(u, v, w) \in \mathbb{R}^3, \quad w > 0\},$$

equipped with the hyperbolic metric

$$\frac{du^2 + dv^2 + dw^2}{w^2}.$$

Throughout this section, $U \subset \mathbb{C}$ will be a domain of the complex plane with coordinate $z = x + iy$, and $X : U \to \mathbb{H}^3$ will be a C^2 conformal immersion of U into $\mathbb{H}^3$.

For any vectors $\vec{u}$ and $\vec{v}$, the notations $\vec{u} \cdot \vec{v}$ and $\langle \vec{u}; \vec{v} \rangle$ stand for the standard euclidean and hyperbolic inner product of $\vec{u}$ and $\vec{v}$.

We set

$$N = \frac{X_x \wedge X_y}{|X_x \wedge X_y|}$$

where $|.|$ stands for the euclidean norm. We call N the *oriented euclidean Gauss map* of X, or more briefly the *euclidean Gauss map* of X, and we denote by N_i, $i = 1, 2, 3$ the coordinate functions of N, that is $N = (N_1, N_2, N_3)$.

DEFINITION 1.1. Let $\Pi : \mathbf{S^2} \to \mathbb{C} \cup \{\infty\}$ be the standard stereographic projection. We set

$$E(z) = (\Pi \circ N)(z) = \frac{N_1 + iN_2}{1 - N_3}(z), \quad \forall z \in U.$$

We also shall call E the oriented euclidean Gauss map of X. Now, let $p = X(z) \in M = X(U)$ be any point on M. Let γ^+ be the half-geodesic issue from p, orthogonal to M and oriented by the normal vector $N(z)$. Let $\omega \in \partial_\infty \mathbb{H}^3 = \mathbb{C} \cup \{\infty\}$ be the asymptotic boundary of γ^+. We then define a map $G : U \to \mathbb{C} \cup \{\infty\}$ setting $G(z) = \omega$. The map G is the well-known *hyperbolic Gauss map* of X (or M), see [**1**].

There is a relationship between the two Gauss maps.

PROPOSITION 1. *Using the notation above we have*

$$E = \frac{G - (u + iv)}{w}.$$

PROOF. Let $z \in U$ be any point. Observe that $G(z) = (u + iv)(z)$ implies $N(z) = (0, 0, -1)$ and, therefore $E(z) = 0 = \big(G - (u + iv)\big)(z)$. Thus, we can assume $G(z) \neq (u + iv)(z)$. We set $P = X(z) = (u, v, w)$, $P_0 = (u, v, 0)$, $G(z) = G_1 + iG_2 = (G_1, G_2, 0)$ and $N = N(z) = (N_1, N_2, N_3)$. Let γ^+ be the half-geodesic issue from P tangent to and oriented by the vector N. Hence, γ^+ is part of a half-circle lying in $\mathbb{H}^3$ with center $C = (c_1, c_2, 0)$. Moreover, γ^+ joins P to G.

By construction, there exists a real number $\lambda \in \mathbb{R}$ such that $C = \lambda P_0 + (1-\lambda)G$. Hence

$$(\top) \qquad u - c_1 = (\lambda - 1)(G_1 - u) \quad \text{and} \quad v - c_2 = (\lambda - 1)(G_2 - v).$$

Furthermore

$$(c_1 - u)^2 + (c_2 - v)^2 + w^2 = (c_1 - G_1)^2 + (c_2 - G_2)^2,$$

since the euclidean distance between C and P equals the euclidean distance between C and G. Using the expressions $(\top)$ above, we infer then

$$\lambda = \frac{(G_1 - u)^2 + (G_2 - v)^2 + w^2}{2\big((G_1 - u)^2 + (G_2 - v)^2\big)},$$

since we are assuming $G \neq u + iv$. Now, observe that:

(i): There exists a *strictly positive* real number $\alpha > 0$ such that $(N_1, N_2, 0) = \alpha \cdot \overrightarrow{P_0 G} = \alpha \cdot (G_1 - u, G_2 - v, 0)$ (because G is the asymptotic boundary of γ^+).

(ii): N and $\overrightarrow{CP}$ are orthogonal vectors.

Thus, we get the system

$$(S) \qquad \begin{cases} N_1(u - c_1) + N_2(v - c_2) + N_3 w = 0, \\ N_1(G_2 - v) - N_2(G_1 - u) = 0. \end{cases}$$

Since $G \neq u + iv$, we can suppose that $G_1 - u \neq 0$, the case $G_2 - v \neq 0$ can be treated similarly. Using $(\top)$ and the expression of λ above, we infer

$$\begin{aligned}
(S) &\Leftrightarrow \begin{cases} N_2 = \dfrac{G_2 - v}{G_1 - u} \cdot N_1 \\ N_1(G_1 - u)(\lambda - 1) + N_2(G_2 - v)(\lambda - 1) + N_3 w = 0 \end{cases} \\
&\Leftrightarrow \begin{cases} N_2 = \dfrac{G_2 - v}{G_1 - u} \cdot N_1 \\ (\lambda - 1)N_1\big((G_1 - u)^2 + (G_2 - v)^2\big) = -w(G_1 - u)N_3 \end{cases} \\
&\Leftrightarrow \begin{cases} N_2 = \dfrac{G_2 - v}{G_1 - u} \cdot N_1 \\ N_1\big(-(G_1 - u)^2 - (G_2 - v)^2 + w^2\big) = -2w(G_1 - u)N_3 \end{cases} \\
&\Leftrightarrow \begin{cases} N_1 = 2\dfrac{w(G_1 - u)}{(G_1 - u)^2 + (G_2 - v)^2 - w^2} \cdot N_3 \\ N_2 = 2\dfrac{w(G_2 - v)}{(G_1 - u)^2 + (G_2 - v)^2 - w^2} \cdot N_3. \end{cases}
\end{aligned}$$

Finally, as $N_1^2 + N_2^2 + N_3^2 = 1$, a simple computation leads to

$$N_3^2 = \left(\frac{\left(\frac{G_1 - u}{w}\right)^2 + \left(\frac{G_2 - v}{w}\right)^2 - 1}{\left(\frac{G_1 - u}{w}\right)^2 + \left(\frac{G_2 - v}{w}\right)^2 + 1} \right)^2.$$

As observed above, we have $(N_1, N_2, 0) = \alpha(G_1 - u, G_2 - v, 0)$ with $\alpha > 0$. Hence,

$$N = \left(\frac{2\left(\frac{G_1-u}{w}\right), 2\left(\frac{G_2-v}{w}\right), \left(\frac{G_1-u}{w}\right)^2 + \left(\frac{G_2-v}{w}\right)^2 - 1}{\left(\frac{G_1-u}{w}\right)^2 + \left(\frac{G_2-v}{w}\right)^2 + 1} \right)$$
$$= \Pi^{-1}\left(\frac{G - (u+iv)}{w}\right).$$

We conclude, therefore, that

$$E = \frac{G - (u+iv)}{w}$$

as desired. □

Using the very definition of E and the fact that X is a conformal immersion we can derive the following.

PROPOSITION 2. *Assume that $E(z) \neq \infty$ for any $z \in U$. We have*

$$G_z = wE_z \quad \text{and} \quad w_{\bar{z}} = \frac{\overline{E}}{1 + E\overline{E}}(G_{\bar{z}} - wE_{\bar{z}}).$$

Furthermore, the (hyperbolic) metric ds^2 induced on U by the immersion X is

$$ds^2 = \frac{|G_{\bar{z}} - wE_{\bar{z}}|^2}{w^2}|dz|^2.$$

We infer the following result.

THEOREM 3. *Let $E : U \to \mathbb{C}$ and $G : U \to \mathbb{C}$ be two C^2 maps. Then, there exists a (possibly branched) C^3 conformal immersion $X : U \to \mathbb{H}^3$ with euclidean (resp. hyperbolic) Gauss map E (resp. G) if and only if E and G satisfy the following conditions:*

$$\frac{G_z}{E_z}(z) \in \mathbb{R}^{*+}, \quad \forall z \in U,$$

$$(\natural) \qquad G_{z\bar{z}}E_z - G_z E_{z\bar{z}} = \frac{\overline{E}E_z}{1 + E\overline{E}}(E_z G_{\bar{z}} - G_z E_{\bar{z}}).$$

Moreover, the induced (pseudo) metric is given by

$$ds^2 = \frac{|E_z G_{\bar{z}} - G_z E_{\bar{z}}|^2}{|G_z|^2}|dz|^2.$$

PROOF. Assume first that E and G are the Gauss maps of a conformal C^2 immersion $X : U \to \mathbb{H}^3$. From Proposition 2 we get $w = \frac{G_z}{E_z}$ and then the first condition is satisfied. Furthermore, substituting w with this expression into the relation $w_{\bar{z}} = \frac{\overline{E}}{1+E\overline{E}}(G_{\bar{z}} - wE_{\bar{z}})$ given in Proposition 2 we easily get relation $(\natural)$.

Conversely, assume that E and G are two C^2 maps $U \to \mathbb{C}$ satisfying the two conditions stated in the premises. We define the C^2 map $X : U \to \mathbb{H}^3$ given by $X = (u, v, w)$ where:

$$w = \frac{G_z}{E_z} \quad \text{and} \quad u + iv = G - wE = \frac{GE_z - EG_z}{E_z}.$$

Since E and G satisfy ($\natural$) we deduce that w satisfies $w_{\bar{z}} = \frac{\overline{E}}{1+E\overline{E}}(G_{\bar{z}} - wE_{\bar{z}})$. We show that X is a conformal map and that E is the oriented euclidean Gauss map of X at regular points. Since $u + iv = G - wE$, Proposition 1 would imply that G is the hyperbolic Gauss map of X at any regular point. We want to show

$$\langle X_x, \Pi^{-1}(E)\rangle = 0, \tag{1–1}$$

$$\langle X_y, \Pi^{-1}(E)\rangle = 0, \tag{1–2}$$

$$\langle X_x, X_x\rangle = \langle X_y, X_y\rangle, \tag{1–3}$$

$$\langle X_x, X_y\rangle = 0. \tag{1–4}$$

We recall that for any C^1 function $f : U \to \mathbb{C}$ we have $f_x = f_z + f_{\bar{z}}$ and $f_y = i(f_z - f_{\bar{z}})$. We are going to use the following relations:

$$(G - wE)_z = G_z - wE_z - w_z E = -\overline{w_{\bar{z}}}E = -\frac{E^2}{1+E\overline{E}}\overline{(G_{\bar{z}} - wE_{\bar{z}})},$$

$$(G - wE)_{\bar{z}} = G_{\bar{z}} - wE_{\bar{z}} - \frac{E\overline{E}}{1+E\overline{E}}(G_{\bar{z}} - wE_{\bar{z}}) = \frac{G_{\bar{z}} - wE_{\bar{z}}}{1+E\overline{E}} = \frac{w_{\bar{z}}}{\overline{E}}.$$

We show (1–1). We have

$$\begin{aligned}
\langle X_x, \Pi^{-1}(E)\rangle &= \left\langle (u_x, v_x, w_x), \left(\frac{2\mathrm{Re}E, 2\mathrm{Im}E, E\overline{E} - 1}{E\overline{E} + 1}\right)\right\rangle \\
&= \mathrm{Re}\left((u_x + iv_x)\cdot\frac{2\overline{E}}{E\overline{E}+1}\right) + w_x\frac{E\overline{E}-1}{E\overline{E}+1} \\
&= \mathrm{Re}\left(((G-wE)_z + (G-wE)_{\bar{z}})\frac{2\overline{E}}{E\overline{E}+1}\right) + (w_z + w_{\bar{z}})\frac{E\overline{E}-1}{E\overline{E}+1} \\
&= \mathrm{Re}\left((-w_z E + \frac{w_{\bar{z}}}{\overline{E}})\frac{2\overline{E}}{E\overline{E}+1}\right) + (w_z + w_{\bar{z}})\frac{E\overline{E}-1}{E\overline{E}+1} \\
&= \mathrm{Re}\frac{2(w_{\bar{z}} - w_z E\overline{E})}{E\overline{E}+1} + (w_z + w_{\bar{z}})\frac{E\overline{E}-1}{E\overline{E}+1} \\
&= 0.
\end{aligned}$$

This shows relation (1–1). The relation (1–2) can be shown in the same way. Now, we have

$$\begin{aligned}
&\langle X_x, X_x\rangle \\
&\quad = \langle (u_x, v_x, w_x), (u_x, v_x, w_x)\rangle \\
&\quad = \mathrm{Re}\,((u+iv)_x(u-iv)_x) + w_x^2 \\
&\quad = ((G-wE)_z + (G-wE)_{\bar z})\,\overline{((G-wE)_z + (G-wE)_{\bar z})} + (w_z + w_{\bar z})^2 \\
&\quad = \frac{(G_{\bar z} - wE_{\bar z}) - E^2\overline{(G_{\bar z} - wE_{\bar z})}}{E\overline{E}+1}\cdot\frac{\overline{(G_{\bar z} - wE_{\bar z}) - E^2\overline{(G_{\bar z} - wE_{\bar z})}}}{E\overline{E}+1} + (w_z + w_{\bar z})^2 \\
&\quad = \frac{|G_{\bar z} - wE_{\bar z}|^2 + (E\overline{E})^2|G_{\bar z} - wE_{\bar z}|^2 - \overline{E}^2(G_{\bar z} - wE_{\bar z})^2 - E^2(\overline{G_{\bar z} - wE_{\bar z}})^2}{(E\overline{E}+1)^2} \\
&\quad + \frac{\overline{E}^2(G_{\bar z} - wE_{\bar z})^2 + E^2(\overline{G_{\bar z} - wE_{\bar z}})^2 + 2E\overline{E}|G_{\bar z} - wE_{\bar z}|^2}{(E\overline{E}+1)^2} \\
&\quad = |G_{\bar z} - wE_{\bar z}|^2.
\end{aligned}$$

In the same way we show that $\langle X_y, X_y\rangle = |G_{\bar z} - wE_{\bar z}|^2$, this implies (1–3). Relation (1–4) can be inferred analogously. At last, as X is a C^2 possibly branched immersion, the induced pseudo-metric is given by

$$ds^2 = \frac{\langle X_x, X_x\rangle}{w^2}|dz|^2 = \frac{|G_{\bar z} - wE_{\bar z}|^2}{w^2}|dz|^2 = \frac{|E_zG_{\bar z} - G_zE_{\bar z}|^2}{|G_z|^2}|dz|^2.$$

This achieves the proof. □

Now, we denote by H (resp. $\widetilde{H}$) the euclidean (resp. hyperbolic) mean curvature of X with respect to N. We denote also by Φ (resp. $\widetilde{\Phi}$) the *Hopf quadratic form* of X in $\mathbb{R}^3$ (resp. $\mathbb{H}^3$). We can show that

PROPOSITION 4. *Assume that $E(z) \neq \infty$ for any $z \in U$. We have*

$$H = \frac{-2E_{\bar z}}{(1+E\overline{E})(G_{\bar z} - wE_{\bar z})}, \tag{1–5}$$

$$2\frac{G_{\bar z}}{1+E\overline{E}} = (1-\widetilde{H})(G_{\bar z} - wE_{\bar z}), \tag{1–6}$$

$$\widetilde{\Phi} = -2E_z\frac{\overline{G_{\bar z} - wE_{\bar z}}}{w(1+E\overline{E})}(dz)^2 = \frac{1}{w}\Phi. \tag{1–7}$$

We easily deduce the following result first shown by R. Bryant, see [**1**].

Corollary 5. *We have:*

(1) Let $p = X(z_0) \in M$ be a point on the surface M such that $E(z_0) \neq \infty$, then

$$p \text{ is an umbilic point of } M \Leftrightarrow G_z(z_0) = 0.$$

Therefore, M is totally umbilic if and only if G is anti-meromorphic.

(2) Let $p = X(z_0) \in M$ be a point on the surface M such that $E(z_0) \neq \infty$, then

$$\widetilde{H}(z_0) = 1 \Leftrightarrow G_{\bar z}(z_0) = 0.$$

Therefore, $\widetilde{H} \equiv 1$ if and only if G is meromorphic.

We deduce that the umbilic points of any non-totally umbilic mean curvature one surface in $\mathbb{H}^3$ are isolated.

PROOF. From Proposition 2 we know that $G_z = wE_z$ on U. Moreover we always have $w > 0$. Therefore we infer that for any $z_0 \in U$ such that $E(z_0) \neq \infty$, we have $G_z(z_0) = 0$ if and only if $E_z(z_0) = 0$, which turns out to occur if and only if $p = X(z_0)$ is an umbilic point. From which we easily deduce the last statement in assertion (1).

In the same way, as $G_{\bar{z}} - wE_{\bar{z}} \neq 0$ on U, see Proposition 2, we deduce from Proposition 4 that G is holomorphic at z_0 if and only if $\widetilde{H}(z_0) = 1$ on U. This achieves the proof of the corollary. □

REMARK 1.1. It is well known that every *positive isometry* (that is isometry preserving the orientation) $J : \mathbb{H}^3 \to \mathbb{H}^3$ extends continuously up to $\partial_\infty \mathbb{H}^3$, the asymptotic boundary of $\mathbb{H}^3$. Moreover, the restriction of J to $\partial_\infty \mathbb{H}^3$, denoted by J_∞ is a Möbius function. Conversely, each Möbius function on $\partial_\infty \mathbb{H}^3$ is the restriction of a unique positive isometry of $\mathbb{H}^3$. Call $\zeta = u + iv$ the coordinate on $\partial_\infty \mathbb{H}^3 = \mathbb{C} \cup \{\infty\}$. Note that every Möbius function f on $\mathbb{C} \cup \{\infty\}$ has one of the two forms

$$f(\zeta) = \lambda e^{i\theta}\zeta + \beta, \qquad \text{or} \qquad f(\zeta) = \alpha + \frac{\lambda e^{i\theta}}{\zeta + \beta}$$

where $\alpha, \beta \in \mathbb{C}$ and $\lambda, \theta \in \mathbb{R}$, $\lambda > 0$.

LEMMA 1. *Let $J : \mathbb{H}^3 \to \mathbb{H}^3$ be a positive isometry and let J_∞ be the restriction of J to $\partial_\infty \mathbb{H}^3$. Let $\widehat{E}$ be the euclidean Gauss map of the immersion $\widehat{X} = J \circ X : U \to \mathbb{H}^3$.*

We have

$$\widehat{E} = e^{i\theta}.E \qquad \text{in the case } J_\infty(\zeta) = \lambda e^{i\theta}\zeta + \beta,$$

$$\widehat{E} = \frac{e^{i\theta}}{G + \beta}\left(G_z \frac{1 + E\overline{E}}{E_z} - E\overline{(G + \beta)}\right) \qquad \text{in the case } J_\infty(\zeta) = \alpha + \frac{\lambda e^{i\theta}}{\zeta + \beta}.$$

PROOF. For any complex number β and any real numbers λ, θ, with $\lambda > 0$, we denote by H_λ the homothety on $\mathbb{H}^3$ with respect to 0 and ratio λ, by T_β the horizontal $(\mathrm{Re}\beta, \mathrm{Im}\beta, 0)$-euclidean translation on $\mathbb{H}^3$, and by R_θ the euclidean rotation on $\mathbb{H}^3$ with respect to the w-axis whose argument is θ. Observe that H_λ, T_β and R_θ are positive isometries of $\mathbb{H}^3$.

Suppose first that $J_\infty(\zeta) = \lambda e^{i\theta}\zeta + \beta$. Therefore, we necessarily have $J = T_\beta \circ H_\lambda \circ R_\theta$. Clearly, T_β and H_λ do not affect the euclidean Gauss map. Moreover the effect of R_θ on the euclidean Gauss map is to rotate it around the w-axis, the argument of this rotation being θ. Thus we have $\widehat{E} = e^{i\theta} \cdot E$.

Now assume that $J_\infty(\zeta) = \alpha + \lambda e^{i\theta}/(\zeta + \beta)$. Let $I : \mathbb{H}^3 \to \mathbb{H}^3$ be the positive isometry of $\mathbb{H}^3$ such that $I_\infty(\zeta) = 1/\zeta$. Then, we have $J = T_\alpha \circ H_\lambda \circ R_\theta \circ I \circ T_\beta$. Note that the euclidean Gauss map, hence E, only changes under the isometries I and R_θ. Set $J \circ X = \widehat{X} = (\widehat{u}, \widehat{v}, \widehat{w})$ and call $\widehat{G}$ the hyperbolic Gauss map of $\widehat{X}$. Thus $\widehat{G} = J_\infty(G)$. Observe also that, necessarily, I is the reflection about the vertical plane $\{v = 0\}$ followed by the reflection about the unit sphere centered at 0, that is

$$I(\zeta, w) = I(u + iv, w) = \frac{(u - iv, w)}{u^2 + v^2 + w^2} = \frac{(\overline{\zeta}, w)}{|\zeta|^2 + w^2}.$$

Hence

$$J(\zeta,w)=(\alpha,0)+\lambda\frac{\big(e^{i\theta}(\overline{\zeta+\beta}),w\big)}{|\zeta+\beta|^2+w^2}\qquad\text{and}\qquad \widehat{G}=\alpha+\frac{\lambda e^{i\theta}}{G+\beta}.$$

Thus

$$\begin{aligned}\widehat{E}&=\frac{1}{\widehat{w}}\big(\widehat{G}-(\widehat{u}+i\widehat{v})\big)\\&=\frac{|\zeta+\beta|^2+w^2}{\lambda w}\cdot\Big((\alpha+\frac{\lambda e^{i\theta}}{G+\beta})-(\alpha+\frac{\lambda e^{i\theta}(\overline{\zeta+\beta})}{|\zeta+\beta|^2+w^2})\Big)\\&=\frac{|\zeta+\beta|^2+w^2}{\lambda w}\cdot\Big(\frac{\lambda e^{i\theta}}{G+\beta}-\frac{\lambda e^{i\theta}(\overline{\zeta+\beta})}{|\zeta+\beta|^2+w^2}\Big)\\&=e^{i\theta}\Big(\frac{|\zeta+\beta|^2+w^2}{w(G+\beta)}-\frac{(\overline{\zeta+\beta})}{w}\Big)\\&=e^{i\theta}\left(\frac{w}{G+\beta}\Big(\frac{|\zeta+\beta|^2}{w^2}+1\Big)-\frac{(\overline{\zeta+\beta})}{w}\right).\end{aligned}$$

But from Proposition 1 we derive

$$\frac{\zeta+\beta}{w}=\frac{G+\beta}{w}-E,$$

thus

$$\begin{aligned}\widehat{E}&=e^{i\theta}\left(\frac{w}{G+\beta}\Big(\Big(\frac{G+\beta}{w}-E\Big)\Big(\frac{(\overline{G+\beta})}{w}-\overline{E}\Big)+1\Big)-\frac{(\overline{G+\beta})}{w}+\overline{E}\right)\\&=e^{i\theta}\Big(\frac{w}{G+\beta}-E\frac{(\overline{G+\beta})}{G+\beta}+w\frac{E\overline{E}}{G+\beta}\Big)\\&=\frac{e^{i\theta}}{G+\beta}\big(w(1+E\overline{E})-E(\overline{G+\beta})\big)\\&=\frac{e^{i\theta}}{G+\beta}\Big(G_z\frac{1+E\overline{E}}{E_z}-E(\overline{G+\beta})\Big)\end{aligned}$$

where we have used the relation $G_z=wE_z$ proved in Proposition 2, this achieves the proof of the lemma. □

2. Representation for Mean Curvature One Surfaces in $\mathbb{H}^3$

Throughout this section we shall use the same notation as before. From now on we assume that $U\subset\mathbb{C}$ is a simply connected domain.

The main equation for the study of conformal mean curvature one immersions in our context is the following.

PROPOSITION 6. *Assume that $\widetilde{H}\equiv 1$ and also that $E(z)\neq\infty$ for any $z\in U$, then*

$$(*)\qquad E_{z\bar{z}}=\frac{\overline{E}}{1+E\overline{E}}E_zE_{\bar{z}}.$$

PROOF. Since X is a conformal mean curvature one immersion, we know from Corollary 5 that G is a holomorphic map. Using the identity $G_{\bar{z}}=0$ in the relation (♮), see Theorem 3, we get

$$G_zE_{z\bar{z}}=G_z\frac{\overline{E}}{1+E\overline{E}}E_zE_{\bar{z}}.$$

Moreover the zeros of G_z are isolated unless G_z is the vanishing map. If $G_z \equiv 0$, the relation $G_z = wE_z$ gives $E_z \equiv 0$. In the other case we can simplify the last equality by G_z. In both cases, E must satisfy the relation $(*)$; this concludes the proof. □

REMARK 2.1. (1) Assume that X is a C^3 conformal immersion. Then, using (1–6), we can show that the mean curvature $\widetilde{H}$ is constant if and only if E and $\widetilde{H}$ satisfy

$$(2+(\widetilde{H}-1)(1+E\overline{E}))(1+E\overline{E})E_{z\bar{z}} - 2(1+(\widetilde{H}-1)(1+E\overline{E}))\overline{E}E_zE_{\bar{z}} = 0.$$

In particular, if $\widetilde{H} \equiv 1$ we get the condition $(*)$.

(2) Let $E : U \to \mathbb{C}$ be a C^2 map and let $\widetilde{H} : U \to \mathbb{R}$ be a C^1 function. It can be shown that there exists a conformal immersion $X : U \to \mathbb{H}^3$ with euclidean Gauss map E and prescribed mean curvature $\widetilde{H}$ if and only if E and $\widetilde{H}$ satisfy a certain relation [**19**, Theorem 1.2]. This is a Kenmotsu type result for prescribed mean curvature immersions in $\mathbb{H}^3$.

(3) Let $E : U \to \mathbb{C}$ be any C^2 map neither holomorphic nor anti-holomorphic. It can be shown that if E satisfies relation $(*)$ the zeros of $E_{\bar{z}}$ are isolated [**18**, Lemma 2–3].

We observe that relation $(*)$ for a conformal immersion is not a necessary condition to have $\widetilde{H} \equiv 1$. Indeed, consider any mean curvature one conformal immersion $X : U \to \mathbb{H}^3$ different from the horospheres. Let $\alpha > 0$ be any positive real number. Consider the vertical translated conformal immersion $Y = X + (0,0,\alpha)$. Then, the (hyperbolic) mean curvature of Y is no longer constant, although Y and X share the same euclidean Gauss map. Nevertheless:

PROPOSITION 7. *Let $X : U \to \mathbb{H}^3$ be a non-totally umbilic C^3 conformal immersion such that $E(z) \neq \infty$ for any $z \in U$ and such that E satisfies $(*)$. Then, one of the three following cases holds:*

(i): *X is a mean curvature one immersion.*

(ii): *There exists a real number $\alpha \in \mathbb{R}^*$ such that $X_1 := X - (0,0,\alpha)$ is a mean curvature one immersion into $\mathbb{H}^3$.*

(iii): *There exists a positive real number $\alpha > 0$ such that $X_2 := -X + (0,0,\alpha)$ is a mean curvature one immersion into $\mathbb{H}^3$.*

PROOF. Differentiating the relation $G_z = wE_z$ with respect to $\bar{z}$ and using Proposition 2 and the fact that E satisfies $(*)$ we get

$$G_{z\bar{z}} = G_{\bar{z}} \frac{\overline{E}}{1+E\overline{E}} E_z.$$

Now, observe that since $\widetilde{H} \not\equiv 1$ we have $G_{\bar{z}} \not\equiv 0$. On any connected component of $U_1 = U - \big(\{z \in U,\ G_{\bar{z}}(z) = 0\} \cup \{z \in U,\ E_{\bar{z}} = 0\}\big)$, we have

$$\frac{G_{z\bar{z}}}{G_{\bar{z}}} = \frac{E_{z\bar{z}}}{E_{\bar{z}}}, \quad \text{so that} \quad \left(\operatorname{Log}\frac{G_{\bar{z}}}{E_{\bar{z}}}\right)_z = 0.$$

Thus, $G_{\bar{z}}/E_{\bar{z}}$ is an anti-holomorphic function on any component of U_1. Furthermore, using the identity $w_{xy} = w_{yx}$ we can show that $G_{\bar{z}}/E_{\bar{z}}$ is a real function. We deduce that $G_{\bar{z}}/E_{\bar{z}}$ is constant on any component of U_1. Therefore, $G_{\bar{z}}/E_{\bar{z}}$ is a

continuous real function on U (with possibly infinite values at some points) which is constant on any connected component of

$$\widetilde{U} = U_1 \cup \mathrm{Int}\big(\{z \in U,\ G_{\bar{z}}(z) = 0\}\big).$$

Since $\{z \in U,\ E_{\bar{z}} = 0\}$ is a discrete set (Remark 2.1), we deduce that $U - \widetilde{U}$ has empty interior. Therefore, the function $G_{\bar{z}}/E_{\bar{z}}$ is constant on U, say, $G_{\bar{z}}/E_{\bar{z}} \equiv \alpha$ for a real number α. Note that $\alpha \neq 0$ because $\widetilde{H} \not\equiv 1$. Moreover, as $ds = \big(|G_{\bar{z}} - wE_{\bar{z}}|/w\big)|dz| = \big(|(w-\alpha)E_{\bar{z}}|/w\big)|dz|$, we infer that $w(z) - \alpha \neq 0$ for any $z \in U$.

Case 1: Suppose that $w - \alpha > 0$ on U. A straightforward calculation shows that $(w-\alpha)E_z$ is a holomorphic map on U. Observe that $X_1 := X + (0, 0, -\alpha)$ is a conformal immersion into $\mathbb{H}^3$ whose hyperbolic Gauss map G_1 satisfies $(G_1)_z = w_1(E_1)_z = (w-\alpha)E_z$ since $E_1 = E$ and $w_1 = w - \alpha$. We deduce that $(G_1)_z$ is a holomorphic map on U. Thus, differentiating with respect to $\bar{z}$ the relation $(G_1)_z = w_1 E_z$ as before, we get

$$0 \equiv (G_1)_{\bar{z}} \cdot \frac{\overline{E}}{1 + E\overline{E}} E_z.$$

Since we assume X not totally umbilic we deduce that G_1 is holomorphic, that is X_1 is a mean curvature one immersion.

Case 2: Suppose that $w - \alpha < 0$ on U. As in the case 1, the function $(\alpha - w)E_z$ is holomorphic on U. Now we set $X_2 := -X + (0, 0, \alpha)$, thus X_2 is a conformal immersion into $\mathbb{H}^3$. We have, as in the case 1, that $(G_2)_z = (\alpha - w)E_z$ is holomorphic on U, where G_2 is the hyperbolic Gauss map of X_2. We conclude as before that X_2 is a mean curvature one immersion into $\mathbb{H}^3$. This concludes the proof. □

Nevertheless we are going to show that any nontrivial solution of $(*)$ gives rise to a mean curvature one conformal immersion into $\mathbb{H}^3$.

THEOREM 8. *Let $U \subset \mathbb{C}$ be a simply connected domain and let $E : U \to \mathbb{C}$ be a nonholomorphic C^2-function satisfying equation $(*)$. We set*

$$U^* = \{z \in U,\ E_{\bar{z}}(z) \neq 0\}$$

($U - U^*$ is discrete; see Remark 2.1). Then, there exists a map $X : U \to \mathbb{H}^3$ such that the restriction of X on U^* is a mean curvature one conformal immersion of U^* into $\mathbb{H}^3$ whose euclidean Gauss map is E. More precisely, we have

$$w(z) = \exp\left(-2\mathrm{Re}\int \frac{\overline{E}E_{\bar{z}}}{1 + E\overline{E}} d\bar{z}\right), \tag{2–1}$$

$$G(z) = \int wE_z dz, \tag{2–2}$$

$$(u + iv)(z) = (G - wE)(z). \tag{2–3}$$

Furthermore, X is uniquely determined up to a positive isometry of $\mathbb{H}^3$. More precisely, if $\widehat{X} : U^* \to \mathbb{H}^3$ is another mean curvature one conformal immersion whose euclidean Gauss map is E, then there exists a positive real number $\lambda > 0$ and a complex number $\alpha \in \mathbb{C}$ such that

$$\widehat{X}(z) = \lambda . X(z) + (\alpha, 0).$$

Moreover, the geometric quantities of X can be expressed in terms of E. Thus, calling ds, $\widetilde{\mathrm{II}}$, $\widetilde{\Phi}$ and $\widetilde{K}$ respectively the metric the second fundamental form, the Hopf quadratic form and the Gauss curvature of X, we have

$$ds = |E_{\bar{z}}| \cdot |dz|, \tag{2–4}$$

$$\widetilde{\mathrm{II}} = 2\mathrm{Re}\Big(\frac{E_z(\overline{E})_z}{1+E\overline{E}}(dz)^2\Big), +|E_{\bar{z}}|^2|dz|^2, \tag{2–5}$$

$$\widetilde{\Phi} = 2\frac{E_z(\overline{E})_z}{1+E\overline{E}}(dz)^2, \tag{2–6}$$

$$\widetilde{K} = \frac{-4}{(1+E\overline{E})^2}\frac{|E_z|^2}{|E_{\bar{z}}|^2}. \tag{2–7}$$

Proof. It is easily seen that the function wE_z is holomorphic, therefore G is holomorphic too. We deduce that E and G satisfy the hypothesis of Theorem 3. Since $G_z/E_z = w$ we deduce that the map $X = (u, v, w) : U \to \mathbb{H}^3$ given by relations (2–1)–(2–3) is a conformal map with euclidean (resp. hyperbolic) Gauss map E (resp. G). Since G is holomorphic the induced metric is given by $ds = |E_{\bar{z}}|.|dz|$ and the restriction of X on U^* is a conformal mean curvature one immersion into $\mathbb{H}^3$. For the proof of the remaining affirmations we refer to [**18**]. □

We infer from Theorem 8 that to produce explicit mean curvature one immersions into $\mathbb{H}^3$ it suffices to produce explicit solutions E of $(*)$. As a matter of fact, the following result shows that the equation $(*)$ is totally solved, more precisely any solution E of $(*)$ can be expressed in terms of meromorphic data.

Theorem 9. *Let $U \subset \mathbb{C}$ be a simply connected domain and let $E : U \to \mathbb{C}$ be a C^2 map neither holomorphic nor anti-holomorphic satisfying equation*

$$(*) \qquad E_{z\bar{z}} = \frac{\overline{E}}{1+E\overline{E}}E_zE_{\bar{z}}.$$

Then there exist meromorphic functions h, T on U such that

$$E = h \cdot \overline{\left(\frac{Th_z + hT_z}{h^2T_z}\right)}\left(T + \overline{\left(\frac{h_z}{Th_z + hT_z}\right)}\right). \tag{2–8}$$

Moreover let $X = (u, v, w) : U \to \mathbb{H}^3$ be the conformal map given by Theorem 8. Then, up to a multiplicative positive constant we have

$$w = \frac{|h^2T_z|^2}{|Th_z + hT_z|^2 + |h_z|^2}, \tag{2–9}$$

$$G_z = h^2T_z. \tag{2–10}$$

Then, up to the same multiplicative positive constant and up to an additive complex constant we have

$$u + iv = G - wE = G - \frac{h^3T_z}{|Th_z + hT_z|^2 + |h_z|^2}\overline{(Th_z + hT_z)}\left(T + \overline{\left(\frac{h_z}{Th_z + hT_z}\right)}\right).$$

Sketch of proof. Assume first that $E_{\bar{z}} \neq 0$ on U. Therefore, we deduce from Theorem 8 that the conformal map $X : U \to \mathbb{H}^3$ given by E is a regular mean curvature one immersion on the whole domain U. Let $Y : U \to \mathbb{R}^3$ be the minimal immersion associated to X, and let $(g, f\,dz)$ be the Weierstrass representation of

Y. We can show that there exists a meromorphic function $\widehat{h}$ defined on the whole U satisfying

$$(\top) \qquad f = -\frac{\widehat{h}g_z\widehat{h}_{zz} - 2\widehat{h}_z^2 g_z - \widehat{h}\widehat{h}_z g_{zz}}{\widehat{h}^2 g_z^2},$$

see [**18**] Theorem 3-9. Set

$$\widehat{E} = \widehat{h}\overline{\left(\frac{T\widehat{h}_z + \widehat{h}T_z}{\widehat{h}^2 T_z}\right)}\left(T + \overline{\left(\frac{\widehat{h}_z}{T\widehat{h}_z + \widehat{h}T_z}\right)}\right).$$

We have:

$$|\widehat{E}_{\bar{z}}| = \frac{|\widehat{h}g_z\widehat{h}_{zz} - 2\widehat{h}_z^2 g_z - \widehat{h}\widehat{h}_z g_{zz}|}{|\widehat{h}^2 g_z^2|}(1+|g|^2) = |f|(1+|g|^2) = |E_{\bar{z}}|$$

since X and Y are isometric immersions. Furthermore it is easily seen that $\widehat{E}$ satisfies $(*)$. Therefore, $\widehat{E}$ is the euclidean Gauss map of a conformal mean curvature one immersion $\widehat{X} : U \to \mathbb{H}^3$. We infer with Theorem 8 that the second fundamental form $\widehat{\mathrm{II}}$ of $\widehat{X}$ is given by

$$\widehat{\mathrm{II}} = \mathrm{Re}\left(\frac{\widehat{h}g_z\widehat{h}_{zz} - 2\widehat{h}_z^2 g_z - \widehat{h}\widehat{h}_z g_{zz}}{\widehat{h}^2 g_z}(dz^2)\right) + ds^2.$$

Since $\widehat{h}$ satisfies $(\top)$ we deduce $\widehat{\mathrm{II}} = \mathrm{Re}\left(-2fg_z(dz)^2\right) + ds^2 = \widetilde{\mathrm{II}}$, where $\widetilde{\mathrm{II}}$ is the second fundamental form of X. Consequently the immersions X and $\widehat{X}$ share the same metric and the same second fundamental form. Therefore there exists a positive isometry $J : \mathbb{H}^3 \to \mathbb{H}^3$ such that $X = J \circ \widehat{X}$. Finally, Lemma 1 and a computation, see Proposition 10, show that there exist meromorphic maps h and T on U such that we have

$$E = h \cdot \overline{\left(\frac{Th_z + hT_z}{h^2 T_z}\right)}\left(T + \overline{\left(\frac{h_z}{Th_z + hT_z}\right)}\right).$$

In the case where the set $\{z, U,\ E_{\bar{z}} \neq 0\}$ is not empty, we refer the reader to [**18**], Theorem 3-12. □

REMARK 2.2. We use the same notation as in Theorem 9. We deduce from Theorem 8 that geometric quantities of the (possibly branched) immersion given by E can be expressed in terms of (h, T). Namely, we have

$$(2\text{–}11) \qquad ds = |E_{\bar{z}}dz| = \frac{|hT_z h_{zz} - 2h_z^2 T_z - hh_z T_{zz}|}{|hT_z|^2}(1+|T|^2)|dz|,$$

$$(2\text{–}12) \qquad \widetilde{\Phi} = 2\frac{hT_z h_{zz} - 2h_z^2 T_z - hh_z T_{zz}}{h^2 T_z}(dz)^2,$$

$$(2\text{–}13) \qquad \widetilde{\mathrm{II}} = \mathrm{Re}(\widetilde{\Phi}) + ds^2,$$

$$(2\text{–}14) \qquad K = -4\frac{|h|^4|T_z|^6}{|hT_z h_{zz} - 2h_z^2 T_z - hh_z T_{zz}|^2(1+|T|^2)^4},$$

and, up to a rigid motion of $\mathbb{R}^3$, we can choose

$$(2\text{–}15) \qquad g = T,$$

$$(2\text{–}16) \qquad f = -\frac{1}{T_z} \cdot \frac{hT_z h_{zz} - 2h_z^2 T_z - hh_z T_{zz}}{h^2 T_z},$$

where (g, fdz) is the Weierstrass representation of the minimal immersion in $\mathbb{R}^3$ associated to X.

PROPOSITION 10. *Let $X : U \to \mathbb{H}^3$ be a conformal mean curvature one immersion and let E (resp. G) be the euclidean (resp. hyperbolic) Gauss map of X. Assume that there exist meromorphic functions h and T on U such that*

$$E = h \cdot \overline{\left(\frac{Th_z + hT_z}{h^2 T_z}\right)} \left(T + \overline{\left(\frac{h_z}{Th_z + hT_z}\right)}\right).$$

Let $J : \mathbb{H}^3 \to \mathbb{H}^3$ be a positive isometry of $\mathbb{H}^3$ and call $\widehat{E}$ the euclidean oriented Gauss map of the immersion $J \circ X$. Denote by J_∞ the restriction of J to $\partial_\infty \mathbb{H}^3$.

Then, we have

$$\widehat{E} = \widehat{h} \cdot \overline{\left(\frac{\widehat{T}\widehat{h}_z + \widehat{h}\widehat{T}_z}{\widehat{h}^2 \widehat{T}_z}\right)} \left(\widehat{T} + \overline{\left(\frac{\widehat{h}_z}{\widehat{T}\widehat{h}_z + \widehat{h}\widehat{T}_z}\right)}\right).$$

where

$$\widehat{h} = h \quad \text{and} \quad \widehat{T} = e^{i\theta} T \quad \text{in the case where} \quad J_\infty(\zeta) = \lambda e^{i\theta} \zeta + \beta,$$

$$\widehat{h} = \frac{hT}{G + \beta} \quad \text{and} \quad \widehat{T} = \frac{e^{i\theta}}{T} \quad \text{in the case where } J_\infty(\zeta) = \alpha + \frac{\lambda e^{i\theta}}{\zeta + \beta}.$$

PROOF. Let $J : \mathbb{H}^3 \to \mathbb{H}^3$ be a positive isometry. Assume first that J_∞ has the form $J_\infty(\zeta) = \lambda e^{i\theta} \zeta + \beta$ where $\lambda > 0$, $\theta \in \mathbb{R}$ and $\beta \in \mathbb{C}$ (see Remark 1.1). Then, using the notation of the proof of Lemma 1, $J = T_\beta \circ H_\lambda \circ R_\theta$, where: T_β is the horizontal $(\mathrm{Re}\beta, \mathrm{Im}\beta, 0)$- euclidean translation; H_λ is the homothety in $\mathbb{H}^3$ with respect to 0 and ratio λ, and R_θ is the euclidean rotation in $\mathbb{H}^3$ with respect to the w-axis whose argument is θ. We get therefore $\widehat{E} = e^{i\theta} \cdot E$, which forces $\widehat{E}$ to have the desired form setting $\widehat{h} = h$ and $\widehat{T} = e^{i\theta} T$.

It remains to consider the case where $J_\infty(\zeta) = \alpha + \lambda e^{i\theta}/(\zeta + \beta)$, where $\alpha, \beta \in \mathbb{C}$ and $\lambda, \theta \in \mathbb{R}$, $\lambda > 0$. In this case it follows from Lemma 1 that

$$\widehat{E} = \frac{e^{i\theta}}{G + \beta}\left(G_z \frac{1 + E\overline{E}}{E_z} - E\overline{(G + \beta)}\right) = \frac{e^{i\theta}}{G + \beta}\left(w(1 + E\overline{E}) - E\overline{(G + \beta)}\right).$$

We infer from relation (2–1) in Theorem 8 that

$$w = \frac{|h^2 T_z|^2}{|Th_z + hT_z|^2 + |h_z|^2} \quad \text{(up to a multiplicative constant)}.$$

Moreover, a computation shows that

$$1 + E\overline{E} = \frac{(1 + |T|^2)}{|hT_z|^2} \cdot \left(|Th_z + hT_z|^2 + |h_z|^2\right).$$

We can therefore infer

$$\begin{aligned}\widehat{E} &= \frac{e^{i\theta}}{G+\beta}\cdot\left(|h|^2(1+|T|^2)-\overline{(G+\beta)}\frac{h}{\overline{h^2T_z}}\left(T\overline{(Th_z+hT_z)}+\overline{h_z}\right)\right)\\ &= \frac{e^{i\theta}hT}{(G+\beta)}\left(\overline{h}\left(\frac{1}{T}+\overline{T}\right)-\overline{\left(\frac{G+\beta}{h^2T_z}\right)}\overline{((Th_z+hT_z)}+\left(\frac{\overline{h_z}}{T}\right)\right)\right)\\ &= \frac{e^{i\theta}hT}{(G+\beta)}\left(\frac{1}{T}\left(\overline{h}-\overline{\left(\frac{h_z(G+\beta)}{h^2T_z}\right)}\right)+\overline{hT}-\overline{\left(\frac{G+\beta}{h^2T_z}\right)}\overline{(Th_z+hT_z)}\right)\end{aligned}$$

and

$$\begin{aligned}\widehat{E} &= \frac{e^{i\theta}hT}{(G+\beta)}\overline{\left(h-\frac{h_z(G+\beta)}{h^2T_z}\right)}\left(\frac{1}{T}+\frac{\overline{hT}-\overline{\left(\frac{G+\beta}{h^2T_z}\right)}\overline{(Th_z+hT_z)}}{\overline{\left(h-\frac{h_z(G+\beta)}{h^2T_z}\right)}}\right)\\ &= \frac{e^{i\theta}hT}{(G+\beta)}\overline{\left(h-\frac{h_z(G+\beta)}{h^2T_z}\right)}\left(\frac{1}{T}+\overline{T}-\overline{\left(\frac{\left(\frac{hT_z(G+\beta)}{h^2T_z}\right)}{\left(h-\frac{h_z(G+\beta)}{h^2T_z}\right)}\right)}\right)\\ &= \frac{e^{i\theta}hT}{(G+\beta)}\overline{\left(h-\frac{h_z(G+\beta)}{h^2T_z}\right)}\left(\frac{1}{T}+\overline{T}-\overline{\left(\frac{hT_z(G+\beta)}{h^3T_z-h_z(G+\beta)}\right)}\right).\end{aligned}$$

Henceforth, it suffices to take $\widehat{h}=\dfrac{hT}{G+\beta}$ and $\widehat{T}=\dfrac{e^{i\theta}}{T}$. This completes the proof. □

Observe that $(*)$ is a *global* equation on any Riemann surface. This leads to the following *global representation.*

THEOREM 11. *Let S be a Riemann surface and let h and T be nonconstant meromorphic functions on S such that $h\neq\frac{1}{\alpha T+\beta}$, for any complex numbers α, β. We set:*

$$w(z)=\frac{|h^2\,dT|^2}{|T\,dh+h\,dT|^2+|\,dh|^2},\qquad E=h\cdot\overline{\left(\frac{T\,dh+h\,dT}{h^2\,dT}\right)}\left(T+\overline{\left(\frac{dh}{T\,dh+h\,dT}\right)}\right)$$

and $S^*=\{z\in S,\ |E_{\overline{z}}|\neq 0,\infty\}$. Assume that

$$\int_\gamma h^2\,dT=0$$

for every closed path $\gamma\subset S$ on which neither h nor T have poles, that is the 1-form $h^2\,dT=h^2T_z\,dz$ has a global primitive G on S. Set:

$$(u+iv)(z):=(G-wE)(z),\quad z\in S^*.$$

Then, the function $X:=(u,\ v,\ w):\ S^*\to\mathbb{H}^3$ defines a mean curvature one conformal immersion whose euclidean Gauss map is E and the hyperbolic Gauss map is G (recall that $S-S^*$ is a discrete set). Furthermore, the geometric quantities of X are given by relations (2–11)–(2–14) *in Remark 2.2.*

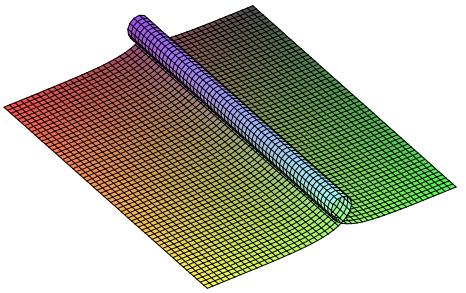


Figure 1

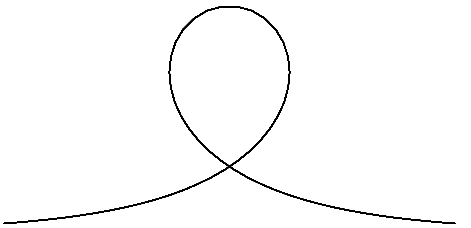


Figure 2

EXAMPLE. Consider $U = \mathbb{C}$ and choose the holomorphic data $h(z) = e^{\gamma z}$ and $T(z) = b + e^z$, where $b, \gamma \in \mathbb{C}$. Thus we have

$$E(z) = e^{\gamma z} e^{-\overline{\gamma z}} \cdot \left(\overline{b\gamma e^{-z} + (1+\gamma)}\right) \cdot \left(b + e^z + \overline{\left(\frac{\gamma}{b\gamma + (1+\gamma)e^z}\right)}\right).$$

Therefore $|E_{\bar{z}}(z)| = |\gamma(1+\gamma)e^{-z}|\left(1 + |b+e^z|^2\right)$. It turns out that (h, T) give rise to a complete mean curvature one immersion $X : \mathbb{C} \to \mathbb{H}^3$ for any complex number b and any $\gamma \in \mathbb{C} - \{-1, 0\}$. These immersions and their geometric features were studied in [**18**]. In particular, when $b = 0$ and $\gamma = -\frac{1}{2}$ the given immersion, say $X_{0,1/2}$, is invariant with respect to a continuous group of horizontal translations, see Figure 1. It is well known that this surface is unique up to an isometry of $\mathbb{H}^3$. Surprisingly the profile curve of this surface was first studied by Poleni in 1729, see [**9**]. Namely, Poleni considered the parametric curve in the (u, w) plane given by

$$c(x) = (u(x), w(x)) = \left(x - 2\tanh x,\ \frac{2}{\cosh x}\right), \quad x \in \mathbb{R},$$

and called it *la courbe des forçats*; see Figure 2.

PROPOSITION 12. *Let* $c(x) = (u(x), w(x)) = \left(x - 2\tanh x,\ 2/\cosh x\right)$, $x \in \mathbb{R}$, *be Poleni's curve. Then the surface in* $\mathbb{H}^3$ *generated by translating* c *along the* v*-axis is a mean curvature one surface. Consequently this surface is the same, up to an isometry of* $\mathbb{H}^3$, *as the surface given by* $X_{0,1/2}$.

PROOF. The surface generated by the curve c can be parametrized as follows, setting $z = x + iy$:

$$X(z) = X(x + iy) = \left(x - 2\tanh x,\ y,\ \frac{2}{\cosh x}\right), \quad z \in \mathbb{C}.$$

It is readily seen that X is a conformal immersion whose oriented euclidean Gauss map is

$$N(z) = \frac{1}{\cosh^2 x}(2\sinh x,\ 0,\ -1 + \sinh^2 x).$$

A computation shows that the euclidean mean curvature with respect to N is given by $H(z) = 1/\cosh x$. Using the relation $\widetilde{H} = wH + N_3$ (which can be established using Proposition 4) we find $\widetilde{H}(z) \equiv 1$ as required. □

REMARK 2.3. When $b = 0$ these immersions are associated to the minimal immersions in $\mathbb{R}^3$ given by the Weierstrass representation $(g, \omega) = (e^z, \gamma(\gamma+1)e^{-z}\,dz)$. We recognize the catenoid-helicoid family. Using the Bryant representation [**1**] and

techniques of Umehara and Yamada [**21**] we have given explicitly in a previous work the mean curvature one conformal immersions associated to the catenoid-helicoid family, see [**17**]. In particular, the surface generated by the Poleni's curve is associated to a catenoid. In the same work, we observed that this surface is the *dual* of the Enneper cousin. We remark also that this surface was considered by Gomes in 1985 ([**4**]) and by Ordóñes in 1995 (see his doctoral thesis [**7**]).

3. Some Constructions of Mean Curvature One Surfaces in $\mathbb{H}^3$

(1) Let $U = \mathbb{C} - \{1\}$. Consider the functions $h(z) = z$ and $T(z) = \alpha e^{z-1}z^{-1} + a$, $z \in U$, where α and a are fixed complex numbers, $\alpha \neq 0$. We have $h^2 T_z = \alpha(z-1)e^{z-1}$, therefore we can choose $G(z) = \alpha(z-2)e^{z-1}$. We observe that the functions w and E given by relations (2–8) and (2–9) are well-defined on U. Therefore, setting $u + iv = G - wE$, we deduce that $X := (u, v, w) : U \to \mathbb{H}^3$ is a well-defined conformal map. Moreover the induced metric, given by (2–11), is

$$ds = \frac{|z^2 e^{-(z-1)}|}{|\alpha(z-1)^2|}\left(1 + \left|\alpha\frac{e^{z-1}}{z} + a\right|^2\right)|dz|.$$

We deduce that X is a conformal and complete mean curvature one immersion with two ends: one end ($z = 1$) is regular (that is G extends meromorphically across the puncture) and has finite total curvature. The other end ($z = \infty$) is not regular and has infinite total curvature. It can be shown that if α and a are chosen so that $\alpha\overline{a} \notin \mathbb{R}$, then X has no symmetry, that is there is no isometry $J : \mathbb{H}^3 \to \mathbb{H}^3$ such that $J(X(U)) = X(U)$. In fact, ds does not admit intrinsic isometry.

(2) Now choose $U = \mathbb{C} - \{z_1, z_2\}$ and $T(z) = az + b$, $h(z) = \dfrac{1}{(z-z_1)(z-z_2)}$, where a, b, z_1 and z_2 are complex numbers, $a \neq 0$ and $z_1 \neq z_2$. Since

$$G_z(z) = (h^2 T_z)(z) = \frac{a}{(z-z_1)^2(z-z_2)^2},$$

we deduce that G is a multivalued map on U. More precisely, G is defined up to an additive complex constant. Observe that E and w are well-defined on U. Since $u + iv = G - wE$, we deduce that $X := (u, v, w)$ is a multivalued conformal map defined up to a horizontal constant vector. That is, the surface $X(U)$ is invariant under one horizontal translation. The induced metric is

$$ds = \frac{2}{|a(z-z_1)(z-z_2)|}(1 + |az+b|^2)|dz|.$$

Therefore, X is a conformal, complete and multivalued mean curvature one immersion into $\mathbb{H}^3$. We observe that X has a regular and well-defined end parametrized by ($z = \infty$). Moreover X has two periodic ends: $z = z_1$ and $z = z_2$. A computation shows that the height function w is given by

$$w(z) = \frac{|a|^2}{|-az^2 - 2bz + az_1z_2 + b(z_1+z_2)|^2 + |2z - (z_1+z_2)|^2}.$$

Thus, each periodic end is asymptotic to a horosphere $\{w \equiv cst\}$.

(3) Now we set $U = \mathbb{C} - \{z_1, z_2\}$ where z_1 and z_2 are the roots of $z^2 - z + 1 = 0$. Choose $T(z) = ze^z$ and $h(z) = -e^{-z}/(z^2 - z + 1)$. The induced metric is

$$ds = \frac{|z|}{|z^2 - z + 1|}\left(\frac{1}{|ze^z|} + |ze^z|\right)|dz|;$$

thus ds is a well-defined complete metric on U. We have

$$G_z = \frac{z+1}{(z^2-z+1)^2}e^{-z};$$

we deduce that G is defined up to a complex constant. That is, G is a multivalued map on U. As before, we deduce that $X := (u, v, w)$ is a multivalued conformal map defined up to a horizontal constant vector. That is, the surface $X(U)$ is invariant under a horizontal translation. More precisely, the ends $(z = z_1)$ and $(z = z_2)$ are periodic ends, that is invariant under a horizontal translation. Note that now $(z = \infty)$ is a well-defined nonregular end. We have

$$w(z) = \frac{1}{|z|^2 + |e^{2z}(z-1)^2|},$$

therefore the ends $(z = z_1)$ and $(z = z_2)$ are asymptotic to a horosphere $\{w \equiv cst\}$.

(4) Let $T : \mathbb{C} \to \mathbb{C}$ be any nonconstant polynomial map. We set

$$h = \frac{-1}{\int zT_z\,dz},$$

thus h is a well-defined meromorphic map on $\mathbb{C}$. We get

$$G_z = \frac{T_z}{(\int zT_z\,dz)^2}.$$

We note that the map G must have an additive period at each zero $z_0 \neq 0$ of the polynomial map $\int zT_z\,dz$.

The metric defined by h and T is

$$ds = \frac{1}{|\int zT_z\,dz|}(1 + |T|^2)|dz|.$$

Let U be the complex plane punctured at the zeros of the polynomial map $\int zT_z\,dz$, $U = \mathbb{C} - \{z, \int zT_z\,dz = 0\}$. Observe that ds is a regular and complete metric on U. Observe also that w and E are well-defined maps on U. In fact we have:

$$w = \frac{|h|^2}{|1 + zTh|^2 + |zh|^2}.$$

Thus the above arguments show that $X = (u, v, w) : U \to \mathbb{H}^3$ is invariant under a finitely generated group of horizontal translations. We deduce therefore that $X = (u, v, w) : U \to \mathbb{H}^3$ is a complete, conformal and multivalued mean curvature one immersion. We note that $(z = \infty)$ is a regular end. For instance, choosing $T(z) = (z-1)^2$ and

$$h(z) = \frac{-3}{z^2(2z-3)},$$

it is easily seen that $z = \frac{3}{2}$ is a periodic end converging to a horizontal horosphere.

Note also that we can do the same construction choosing T to be a rational function on $\mathbb{C}$, getting a possibly branched multivalued immersion.

(5) Now we want to construct (possibly) branched mean curvature one immersions into $\mathbb{H}^3$ with any genus. For that let $\overline{S}$ be any closed Riemann surface. Let $A_1, \ldots, A_n, B$ and G be any meromorphic functions on $\overline{S}$. Let $\alpha_1, \ldots, \alpha_n$ be any real numbers. We set

$$T = A_1^{\alpha_1} \cdots A_n^{\alpha_n} B \quad \text{and} \quad h = \sqrt{\frac{dG}{dT}}.$$

It is easy to check that T and h give rise to well defined functions w and E on $\overline{S}$ with possibly infinite values at some points. The metric given by T and h is

$$ds = |h| \left| d\frac{dh}{h^2\, dT} \right| (1+|T|^2) = \frac{1}{2} \left| \sqrt{\frac{G_z}{T_z}} \right| \left| d\frac{G_{zz}T_z - G_z T_{zz}}{(T_z G_z)^{3/2}} \right| (1+|T|^2).$$

We set $S = \overline{S} - \{ds = \infty\}$. We observe that ds is a well-defined and complete metric on S, possibly branched at a finite set of points. Since G is well-defined on S, we then have a conformal map $X = (u, v, w)$ on S setting $u + iv = G - wE$. That is, $X : S \to \mathbb{H}^3$ is a possibly branched conformal and complete mean curvature one immersion. Note that each end, that is each pole of ds, is regular and has finite total curvature. We observe that each zero and pole of A_j, $j = 1, \ldots, n$, is an end or a branch point of X. This assertion follows from the meromorphic data theorem, see Theorem 9.

References

[1] R. Bryant. *Surfaces of mean curvature one in hyperbolic space*, Astérisque **154–155**, Soc. Math de France, 321–347 (1987).

[2] F.E. Burstall, D. Ferus, F. Pedit and U. Pinkall. *Harmonic tori in symmetric spaces and commuting Hamiltonian systems on loop algebras.* Ann. of Math. (2) **138**:1, 173–212 (1993)

[3] P. Collin, L. Hausswirth and H. Rosenberg. *The geometry of finite topology Bryant surfaces.* Ann. of Math. **153**, 623–659 (2001)

[4] J. M. Gomes. *Sobre hipersuperfícies com curvatura média constante no espaço hiperbólico*, Doctoral thesis, IMPA (1985).

[5] N. J. Hitchin. *Harmonic maps from a 2-torus to the 3-sphere.* J. Differential Geom. **31**:3, 627–710 (1990).

[6] K. Kenmotsu. *Weierstrass formula for surfaces of prescribed mean curvature*, Math. Ann. **245**, 89–99 (1979).

[7] J. Ordóñes. *Superfícies helicoidais com curvatura constante no espaço de formas tridimensionais*, Doctoral Thesis, PUC-Rio (1995).

[8] F. Pacard and A.F. Pimentel. *Attaching handles to Bryant surfaces*, Preprint.

[9] Revue du Palais de la Découverte **45**, p. 106 (1995).

[10] M. Ritoré. *Examples of constant mean curvature surfaces obtained from harmonic maps to the two sphere*, Math. Z. **226**, 127–146 (1997).

[11] H. Rosenberg. *Bryant Surfaces*, Preprint.

[12] W. Rossman and K. Sato. *Constant mean curvature surfaces in hyperbolic 3-space with two ends*, J. Exp. Math. **7**:1, 101–119 (1998).

[13] W. Rossman, M. Umehara and K. Yamada. *Irreducible constant mean curvature 1 surfaces in hyperbolic space with positive genus*, Tôhoku Math. J. **49**, 449–484 (1997).

[14] W. Rossman, M. Umehara and K. Yamada. *A new flux for mean curvature 1 surfaces in hyperbolic 3-space, and applications*, Proc. Amer. Math. Soc. **127**, 2147–2154 (1999).

[15] W. Rossman, M. Umehara and K. Yamada. *Mean curvature 1 surfaces in hyperbolic 3-space with low total curvature I*, Preprint.

[16] W. Rossman, M. Umehara and K. Yamada. *Mean curvature 1 surfaces in hyperbolic 3-space with low total curvature II*, Preprint.

[17] R. Sa Earp and E. Toubiana. *On the geometry of constant mean curvature one surfaces in hyperbolic space*, Illinois J. Math. **45**:2 (2001).

[18] R. Sa Earp and E. Toubiana. *Meromorphic data for mean curvature one surfaces in hyperbolic space*, to appear in Tôhoku Mathematical Journal.

[19] R. Sa Earp and E. Toubiana. *A Weierstrass–Kenmotsu formula for prescribed mean curvature surfaces in hyperbolic space*, Séminaire de Théorie Spectrale et Géométrie de l'Institut Fourier de Grenoble, **19**, 9–23 (2001).

[20] A.J. Small. *Surfaces of constant mean curvature 1 in $\mathbb{H}^3$ and algebraic curves on a quadric*, Proc. AMS **122**:4, 1211–1220 (1994).

[21] M. Umehara and K. Yamada. *Complete surfaces of constant mean curvature-1 in the hyperbolic 3-space*, Annals of Math. **137**, 611–638 (1993).

[22] M. Umehara and K. Yamada. *A parametrization of the Weierstrass formulae and perturbation of some minimal surfaces in R^3 into the hyperbolic 3-space*, J. Reine Angew. Math. **432**, 93–116 (1992).

[23] M. Umehara and K. Yamada. *Surfaces of constant mean curvature c in $H^3(-c^2)$ with prescribed Gauss map*, Math. Ann. **304**, 203–224 (1996).

Pontifí cia Univ. Católica, Rio de Janeiro, Depto de matemática, Rua Marquês de São Vicente 225, 24 453-900 Rio de Janeiro, RJ, Brazil

E-mail address: earp@mat.puc-rio.br

Université Paris VII Département de Mathématiques, 2 Place Jussieu, 75251 Paris Cedex 05, France

E-mail address: toubiana@math.jussieu.fr

Clay Mathematics Proceedings
Volume 2, 2004

Special Lagrangian Submanifolds

Richard Schoen

In this paper we introduce the variational approach to construction of volume minimizing lagrangian submanifolds, and in particular special lagrangian and minimal lagrangian submanifolds. We first describe the geometry of the mean curvature and its connection with the symplectic structure for lagrangian submanifolds. We then formulate the basic existence problems both for compact cycles and submanifolds with boundary. There are two free boundary problems which lead to constructions of minimal and special lagrangian submanifolds. We then describe in some detail the results of [**SW**] concerning the two-dimensional theory. This includes a discussion of the two-dimensional hamiltonian stationary cones and their variational properties. Finally we close with a discussion of the horizontal Plateau problem in the Heisenberg group, and explain why this is the proper setting for the local variational theory.

We thank Weiyang Qiu for taking notes on the lectures and writing a first draft of this paper.

1. Mean Curvature for Lagrangian Submanifolds

Let (M^{2n}, g) be a Riemannian manifold, and let J be a compatible almost complex structure on M —that is, $J : TM \to TM$ is a linear isomorphism such that $J^2 = -\mathrm{Id}$ and $g(J\,\cdot, J\,\cdot) = g(\cdot, \cdot)$. Define a two-form $\omega(X, Y) := g(JX, Y)$. If $d\omega = 0$, then (M, g, J) is called *almost Kähler*. If $\nabla J = 0$ where ∇ denotes the Levi-Civita connection for g, then (M^{2n}, g, J) is called *Kähler* and the two-form ω is then called the Kähler form.

An important example of a Kähler manifold is the flat space $\mathbb{R}^{2n}$ with the standard complex structure J, i.e.,

$$J\Big(\frac{\partial}{\partial x_i}\Big) = \frac{\partial}{\partial y_i}, \qquad J\Big(\frac{\partial}{\partial y_i}\Big) = -\frac{\partial}{\partial x_i},$$

for $i = 1, \ldots, n$. The Kähler form is then the standard symplectic form $\omega = \sum_{i=1}^{n} dx_i \wedge dy_i$.

Definition 1.1. *Let (M^{2n}, g, J) be a Kähler manifold. An n-dimensional submanifold $\Sigma^n \subset M^{2n}$ is said to be lagrangian if $\omega|_\Sigma = 0$ (that is, $\omega(X, Y) = 0$ for any X and Y in $T\Sigma$). This is clearly equivalent to the condition that $J(T_x\Sigma) = (T_x\Sigma)^\perp$ for any $x \in \Sigma$.*

Partially supported by NSF grant DMS-0104163.

©2004 Clay Mathematics Institute

We give a local description of lagrangian submanifolds in $\mathbb{R}^{2n}$. Assume that Σ^n is a graph over some domain Ω in the x-plane, given by

$$y_j = F_j(x), \quad j = 1, \dots n.$$

Then the condition that Σ be lagrangian is equivalent to the condition that the one-form $\sum_{j=1}^n F_j dx_j = 0$ be closed. Locally it is equivalent to the existence of a function $u(x)$ such that $F_j = \partial u/\partial x_j$ for $j = 1, \dots, n$.

The variational problem we will consider is the problem of finding lagrangian submanifolds which minimize volume among "all" lagrangian competitors which agree outside a compact set. Here "all" has different meanings in different situations, depending on the class in which we are minimizing. We will elaborate this point in the later sections of this paper.

We can easily derive the local Euler–Lagrange equation for this variational problem. Assume the submanifold is a graph over a domain Ω in the x-plane. Since it is lagrangian, it may be parametrized as the graph of the gradient of a function $u(x)$. Let Σ_u be the lagrangian graph over Ω defined by

$$y_j = \frac{\partial u}{\partial x_j}, \quad j = 1, \dots, n.$$

Now if Σ_u minimizes volume among lagrangian submanifolds with the same boundary, then $Vol(\Sigma_u) \leq \mathrm{Vol}(\Sigma_v)$ for any function $v(x)$ defined on Ω such that $u = v$ and $\partial u/\partial \nu = \partial v/\partial \nu$ along $\partial\Omega$, where ν is the outer normal vector to $\partial\Omega$. Note that

$$Vol(\Sigma_u) = \int_\Omega \sqrt{\det(g_{ij})}\, dx,$$

where

$$g_{ij} = \delta_{ij} + \sum_{k=1}^n u_{ik} u_{kj}$$

is the induced metric on Σ_u. The lagrangian stationarity of Σ_u then gives

$$\frac{d}{dt} \mathrm{Vol}(\Sigma_{u+t\eta})|_{t=0} = 0,$$

for any $\eta \in C_c^\infty(\Omega)$. A straightforward calculation shows that

$$\sum_{j=1}^n \frac{\partial}{\partial x_j} \left(\triangle_g \left(\frac{\partial u}{\partial x_j} \right) \right) = 0. \tag{1–1}$$

We see that this is a fourth-order elliptic equation of biharmonic type, and the natural Dirichlet boundary condition is to specify u and $\partial u/\partial \nu$ along $\partial\Omega$.

Now we discuss an important geometric interpretation of the Euler–Lagrange equation. Let (M^{2n}, g, J) be a Kähler manifold. Let Σ^n be an oriented lagrangian submanifold. Choose an oriented orthonormal basis $\{e_i\}_{i=1}^n$ on $T_p\Sigma$. Since Σ is Lagrangian, there is an orthogonal basis of (complex valued) $(1,0)$-forms $\omega_1, \dots, \omega_n$ ($\omega_j = \theta_j - \sqrt{-1}\theta_j \circ J$ where $\theta_1, \dots, \theta_n$ is the dual basis to $e_1, \dots, e_n$) such that

$$\omega_j(e_k) = \delta_{jk}.$$

Define an $(n,0)$-form α along Σ by

$$\alpha := \omega_1 \wedge \cdots \wedge \omega_n. \tag{1–2}$$

It is easy to see that $\|\alpha\| = 2^{n/2}$ and in fact $\alpha|_\Sigma$ is the volume form of Σ. A calculation shows that

$$\nabla\alpha = \sqrt{-1}\sigma \cdot \alpha,$$

where σ is a real one-form on Σ. It is not difficult to check that

$$\sigma = H\rfloor\omega,$$

where H is the mean curvature and ω is the Kähler form. From the construction we see that σ is the connection one-form of the restriction of the canonical bundle to Σ with respect to the natural trivialization we have defined. It follows that $d\sigma$ is the restriction of the curvature two-form of the canonical bundle to Σ, that is,

$$d\sigma = \rho|_\Sigma, \tag{1–3}$$

where ρ is the Ricci form on M defined by

$$\rho(X, Y) = \mathrm{Ric}(JX, Y). \tag{1–4}$$

Let Σ be a lagrangian submanifold. We say that Σ is hamiltonian stationary if Σ is stationary under the flow given by all hamiltonian vector fields X with compact support in Σ (A vector field X is called hamiltonian if $X = J\nabla h$ for some smooth function h on M. Clearly the flow of any hamiltonian vector field preserves the Kähler form, and hence the image of a lagrangian submanifold under the flow is still lagrangian). For a hamiltonian stationary submanifold Σ, the first variation formula is

$$0 = \int_\Sigma J\nabla h \cdot H d\mu_\Sigma = \int_\Sigma \langle dh, \sigma\rangle d\mu,$$

for any h with compact support. Therefore we see that

$$\delta\sigma = 0. \tag{1–5}$$

Summarizing these calculations:

PROPOSITION 1.2. *Let $\Sigma^n \subset M^{2n}$ be a hamiltonian stationary lagrangian submanifold. Define a one-form on Σ by $\sigma := H\rfloor\omega$, where H is the mean curvature vector field of Σ and ω is the Kähler form. We then have*

$$d\sigma = \rho|_\Sigma, \quad \delta\sigma = 0, \tag{1–6}$$

where ρ is the Ricci form defined in (1–4).

In the important special case in which M is Kähler–Einstein, we have $\rho = c\omega$ for a constant c. Therefore (1–6) becomes

$$d\sigma = 0, \quad \delta\sigma = 0, \tag{1–7}$$

which are the equations which define σ as a harmonic form on Σ. Since σ is closed, we have $\sigma = d\beta$ locally, or equivalently $H = J\nabla\beta$ where β is called the *lagrangian angle* and is a multi-valued function. Therefore (1–6) becomes the Laplace equation

$$\triangle_\Sigma\beta = 0. \tag{1–8}$$

In the special case in which M is a Calabi–Yau manifold, we have (by the definition of Calabi–Yau manifold) a global parallel $(n, 0)$ form α_0. If Σ is lagrangian, we have $\alpha_0|_\Sigma = e^{-\sqrt{-1}\beta}\alpha$, where α is defined in (1–2). In this case we see that the periods of β are integer multiples of 2π. For Kähler–Einstein manifolds in general, β will have arbitrary real periods. In the flat case $\mathbb{R}^{2n}$, the parallel form α_0 may be chosen to be $dz := dz_1 \wedge \cdots \wedge dz_n$.

Example. If $n = 1$, then any curve $\gamma(s)$ (assume s is an arclength parameter) is lagrangian, and the lagrangian angle β is the angle between the tangent line and positive x-axis. Therefore γ is hamiltonian stationary if and only if $d^2\beta/ds^2 = 0$. Thus the hamiltonian stationary curves in $\mathbb{R}^2$ are lines and circles.

In the Kähler–Einstein case we are particularly interested in the *minimal Lagrangian submanifolds*, which are both minimal and Lagrangian (they form an important special class of hamiltonian stationary submanifolds). If M is a Calabi–Yau manifold, the minimal lagrangian submanifolds are also called *special lagrangian* because they are *calibrated* by the form $Re(e^{-\sqrt{-1}\theta}\alpha_0)$ for some constant θ. The notion of calibration was defined by Harvey and Lawson [**HL**]).

Our main goal is to develop the existence theory for minimal lagrangian and special lagrangian submanifolds. It is of interest to do this for at least three reasons. First, in the SYZ conjecture in mirror symmetry (see the paper of D. Joyce [**J**] in this volume for more details), fibrations of Calabi–Yau manifolds by special lagrangian submanifolds play a key role. Secondly, minimal lagrangian submanifolds are expected to be a very rigid canonical class of representatives for a (hopefully) large part of the homology of a Kähler–Einstein manifold, and as such they should play an important role in understanding the geometry of these manifolds. Thirdly, we hope that similar variational methods can be used to construct other classes of calibrated submanifolds, such as holomorphic submanifolds.

One of the key arguments in the construction of minimal lagrangian submanifolds using variational methods is the following proposition.

PROPOSITION 1.3. *Let M^{2n} be a Kähler–Einstein manifold. If Σ is a* smooth *compact lagrangian submanifold which is stationary for lagrangian variations, then Σ is minimal.*

PROOF. We only give a sketch. Since $H = J\nabla h$ is locally hamiltonian, mean curvature deformation preserves the lagrangian condition but reduces volume unless $H = 0$. This shows that Σ is minimal. □

REMARK 1.4. *In particular we see that if M is Calabi–Yau and Σ is lagrangian stationary, then Σ automatically* minimizes *volume among all cycles in the integral homology class of Σ in $H_n(M;\mathbb{Z})$.*

REMARK 1.5. *Proposition 1.3 suggests that if we minimize volume among all lagrangian competitors, then this minimizer will in fact be minimal. The hard point in carrying out this construction rigorously is that the proposition requires* sufficient smoothness *of the minimizer Σ.*

2. Formulations of the Existence Question

The nature of the existence theory for the Plateau problem depends very much on the class of competitors one considers. For our situation, we consider the following three classes.

1. We can minimize volume among the class of all integral lagrangian currents in a given homology class. We can construct a least volume representative in this class by using the results of Geometric Measure Theory in a straightforward way.
2. If $n = 2$, we have the classical Plateau setting, and we can minimize area among all lagrangian maps (maps u from a smooth surface to M such that $u^*\omega = 0$). Existence in this class can be carried out as well (see [**SW**]).

3. One can minimize mass among the class of integral currents with $\mathbb{Z}_2$ coefficients. This amounts to allowing nonorientable lagrangian competitors. It turns out that the problem here must be formulated very carefully in order to be solvable (see the discussion in [**W**]).

Regardless of the class in which we minimize, the key question is: Are the minimizers sufficiently smooth to show that they are minimal, that is, to carry through the argument of Proposition 1.3?

3. Boundary Conditions

In this section we propose two boundary value problems whose solutions are expected to be minimal lagrangian. These are analogous to the Dirichlet and Neumann conditions for the classical Dirichlet problem for harmonic functions. The major difference is that we need to look for boundary conditions on the competing submanifolds which give rise to Dirichlet or Neumann conditions for the resulting lagrangian angle function β. We describe two such conditions here.

We first mention the classical Plateau boundary condition. Let $\Gamma^{n-1} = \partial\Sigma_0$ for some lagrangian Σ_0. Then we can show that there exists an integral lagrangian current Σ with the least volume among all lagrangian chains with the same boundary Γ. While Σ is hamiltonian stationary, we cannot expect it to be minimal since the mean curvature does not vanish along the boundary. Indeed, there are clearly much more stringent conditions on a boundary Γ that it bound a minimal lagrangian submanifold than that it bound a lagrangian submanifold. In order to get minimal lagrangian submanifolds, we need more flexible boundary conditions; in particular, free boundary conditions.

Dirichlet boundary condition. Assume S is a minimal lagrangian submanifold in a Kähler–Einstein manifold M. For simplicity we assume M is Calabi–Yau with the parallel $(n,0)$ form α, and S is calibrated by $Re(\alpha)$.

Let $[\Sigma_0] \in H_n(M, S; \mathbb{Z})$ be a lagrangian class in the integral relative homology. Assume Σ is a lagrangian chain which minimizes volume among all lagrangian chains in the same relative homology class. We also assume Σ has sufficient regularity (say smooth). The first variation formula gives that

$$\int_\Sigma div_\Sigma(X) d\mu = 0,$$

for any hamiltonian vector field X which is tangent to S along $\partial\Sigma$. Integration by parts gives

$$\int_{\partial\Sigma} \langle X, \nu\rangle ds - \int_\Sigma \langle X, H\rangle d\mu_\Sigma = 0$$

where ν is the outward conormal vector along $\partial\Sigma$. Assuming $X = J\nabla h$, we then have (noticing that $H = J\nabla\beta$)

$$\int_{\partial\Sigma} \langle J\nabla h, \nu\rangle ds - \int_\Sigma \langle \nabla h, \nabla\beta\rangle d\mu_\Sigma = 0.$$

Since $\triangle\beta = 0$ (which follows by using h of compact support on Σ in the above equation), integrating by parts, we see

$$-\int_{\partial\Sigma} J\nu(h) ds - \int_{\partial\Sigma} h\frac{\partial\beta}{\partial\nu} d\mu_\Sigma = 0.$$

Now we must have $\nabla h \perp S$ along S since $J\nabla h$ is tangent to S and S is lagrangian. Thus h is constant along S, say without loss of generality $h = 0$ along S. Then the above equation gives,

$$\int_{\partial\Sigma} J\nu(h)d\mu = 0.$$

This implies that $J\nu$ is tangent to S since the normal derivative of h can be arbitrarily specified, that is, $\nu \perp S$ (since S is lagrangian). Hence we see that $\beta = \pm\pi/2$ along $\partial\Sigma$ since the tangent plane of S (which has $\beta = 0$) along $\partial\Sigma$ has an $n-1$ dimensional subspace in common with that of Σ, and ν is orthogonal to S. This implies that $H = J\nabla\beta$ is tangent to S, hence is an admissible variation. Using the mean curvature as a variational vector field, we see from the first variation formula that Σ is in fact minimal. Therefore we have shown:

PROPOSITION 3.1. *Assume S is a minimal lagrangian submanifold in a Kähler–Einstein manifold M. Let Σ be a* smooth *lagrangian submanifold which satisfies:*

1. $\partial\Sigma \subset S$;
2. Σ *is a critical point of the volume functional in the class of all lagrangian submanifolds in the relative homology class* $[\Sigma] \in H_n(M, S; Z)$.

It then follows that Σ is minimal.

Neumann boundary condition. In the previous section we considered the free boundary problem with constraint manifold being a special lagrangian submanifold. In this section we consider the case where the constraint submanifold S is a complex hypersurface (real codimension 2). As in the previous section we have the first variation formula for the free boundary condition

$$\int_{\partial\Sigma} J\nu(h)ds - \int_{\partial\Sigma} h\frac{\partial\beta}{\partial\nu}d\mu_\Sigma = 0.$$

First note that we may choose h arbitrary on $\partial\Sigma$ and extend it so that ∇h is tangent to $\partial\Sigma$. We then get $J\nu(h) = 0$ on $\partial\Sigma$ because $J\nu \perp \Sigma$. We therefore conclude that

$$\frac{\partial\beta}{\partial\nu} = 0.$$

This is the Neumann type boundary condition for β.

Secondly, we may take h arbitrary on S and extend it so that ∇h is tangent to S. Then we get $\int_{\partial\Sigma} J\nu(h) = 0$ for any h on S. Thus, we see $J\nu \perp S$, and since S is complex, we see that

$$\nu \perp S.$$

Therefore the mean curvature vector gives rises to an admissible deformation, and we have the following proposition.

PROPOSITION 3.2. *Assume S is a complex hypersurface in a Kähler–Einstein manifold M^{2n}. Let Σ be a* smooth *lagrangian submanifold which satisfies:*

1. $\partial\Sigma \subset S$;
2. Σ *is a critical point of the volume functional in the class of all lagrangian submanifolds in the relative homology class* $[\Sigma] \in H_n(M, S; Z)$.

It then follows that Σ is minimal.

REMARK 3.3. *In both the Dirichlet and the Neumann cases, if M is Calabi–Yau, then we can show that the critical lagrangian submanifold is calibrated and hence minimizes volume in its relative homology class. In the Dirichlet case, Σ*

is calibrated by $Im(\alpha)$ since S is calibrated by $Re\alpha$. In the Neumann case, Σ is calibrated by $Re(\alpha)$ since S is a complex hypersurface.

Finally we point out that all of the existence arguments apply in the context of integral currents or in the context of maps (when $n = 2$). J. Wolfson [**W**] discusses the solvability issue in the class of $\mathbb{Z}_2$ currents in his lectures in this volume.

4. The Two-Dimensional Homology Problem

We now describe the existence theory in the special case of surfaces in a four-dimensional manifold. Let (M^4, g, J) be a compact almost Kähler manifold. The first question we must ask is: Given an integral homology class $\sigma \in H^2(M, \mathbb{Z})$, does it admit a lagrangian representative? The following proposition answers this question completely.

PROPOSITION 4.1. *Let (M^4, g, J) be a compact almost Kähler manifold. Let $\sigma \in H^2(M, \mathbb{Z})$ be an integral homology class. There exists a piecewise smooth lagrangian representative Σ in σ if and only if $\int_\sigma \omega = 0$, where ω is the Kähler form.*

PROOF. If σ admits a lagrangian representative, then obviously $\int_\sigma \omega = 0$. To prove the converse, assume $\int_\sigma \omega = 0$.

Step 1. It is well known that a curve Γ bounds a piecewise smooth lagrangian disk in R^4 if and only if $\int_\Gamma \eta = 0$ where $\eta = \sum_{j=1}^2 x_j dy_j$ is a primitive of the standard Kähler form ω. (Gromov [**G**] and Allcock [**A**] in fact gave a quantitative version where they proved that the lagrangian disk Σ may be chosen to satisfy the area bound $Area(\Sigma) \leq cLength(\Gamma)^2$.)

Step 2. We choose a smooth, embedded representative Σ_0 of σ. Let $N_r(\Sigma_0)$ be a small neighborhood of Σ_0. Since $\int_{\Sigma_0} \omega = 0$, it follows that $\omega = d\eta_0$ on Σ_0. Since Σ_0 is a retraction of $N_r(\Sigma_0)$, by the homotopy formula for differential forms there exists a one-form η on $N_r(\Sigma_0)$ such that $\omega = d\eta$.

Step 3. Choose a fine triangulation of Σ_0, and let e be an edge of the one-skeleton. We can assume $\int_e \eta$ small by making the triangulation sufficiently fine. We may then perturb each edge e, fixing the vertices to get a new edge $\tilde{e}$ with $\int_{\tilde{e}} \eta = 0$. (For example, we can perturb e to $\tilde{e}$ so that $e - \tilde{e} = \partial S$ where S is symplectic with correctly chosen $\int_S \omega$.)

Step 4. By Step 1, we may fill in each perturbed triangle with a lagrangian disk to get a nearby Σ in σ which is piecewise smooth and lagrangian. □

Once we have a lagrangian competitor it is possible to find one of least area (see [**SW**]).

PROPOSITION 4.2. *Let (M^4, g, J) be a compact almost Kähler manifold. Let $\sigma \in H^2(M, \mathbb{Z})$ be an integral homology class such that $\int_\sigma \omega = 0$. There exists a least area lagrangian representative Σ in σ which is the image of a map from a smooth closed surface.*

Now the key question is : Is this least area lagrangian Σ a (branched) smooth immersion?

The answer in general is no for the following reason. Assume M is Kähler, and recall that if Σ is a smooth Lagrangian immersion, then the one-form on Σ defined

by $\sigma_H := H\rfloor\omega$(where H is the mean curvature) satisfies

$$d\sigma_H = \rho,$$

where ρ is the Ricci form defined by $\rho(X,Y) = \mathrm{Ric}(JX,Y)$. Therefore Stokes' theorem(Σ has no boundary) implies

$$c_1(\Sigma) = \frac{1}{2\pi}\int_\Sigma \rho = 0,$$

where c_1 is the first Chern class of M. Thus we see a clear obstruction to the existence of a smooth lagrangian representative of a homology class; namely, if $c_1(\sigma) \neq 0$, then there is no *smooth* lagrangian representative of σ. Therefore, in general any least area lagrangian representative must have singularities worse than branch points (it can be seen that branch points cannot account for this discrepancy (see [**SW**])).

Hamiltonian stationary cones. To study singularity issues, it is natural to look for cones which are hamiltonian stationary. Let γ be a curve in $S^3 \subset \mathbb{R}^4$, and let $C(\gamma)$ denote the cone over γ in $\mathbb{R}^4$. We want to find all cones $C(\gamma)$ which are hamiltonian stationary.

Let the curve in S^3 be parametrized as $\gamma(\theta)$. For the cone $C(\gamma)$, we see that the tangent space at a point $\gamma(\theta)$ is spanned by the vector $\gamma(\theta)$ and $\gamma'(\theta)$, that is, $T_{\gamma(\theta)}C(\gamma) = \mathrm{span}(\gamma',\gamma)$. Therefore $C(\gamma)$ is Lagrangian if and only if $\gamma' \perp J\gamma$, or equivalently, γ is a legendrian curve in S^3 with the standard contact structure, where the horizontal space at a point $P \in S^3$ is given by $H_P := (JP)^\perp$. So in this terminology, $C(\gamma)$ is Lagrangian if and only if $\gamma'(\theta) \in H_{\gamma(\theta)}$.

Now let $\pi : S^3 \to S^2$ be the Hopf map defined by:

$$(z_1,z_2) \in S^3 \mapsto \frac{z_1}{z_2} \in S^2 = \mathbb{C}\cup\{\infty\}.$$

Note that the preimage of any point in S^2 is a great circle in S^3. More precisely

$$\pi^{-1}(z) = \{e^{\sqrt{-1}\theta}(z_1,z_2) : (z_1,z_2) \in \pi^{-1}(z)\}.$$

So clearly the horizontal space at $P \in S^3$ is just the orthogonal complement of the tangent space of the Hopf fiber of $\pi(P)$:

$$H_p = (T_P(\pi^{-1}(\pi(P))))^\perp.$$

We can show that $C(\gamma)$ is hamiltonian stationary if and only if $\pi(\gamma)$ is a circle in S^2, which is equivalent to the condition that γ is a critical point for the length functional among legendrian curves. A little work shows that in order that γ be a closed horizontal geodesic for the horizontal length functional, $\pi(\gamma)$ must be a circle on S^2 satisfying a certain rationality condition, and vice versa. This condition is the condition that the circle $\pi(\gamma)$ divide S^2 into regions whose areas are rationally related.

A direct computation gives an explicit description of the hamiltonian stationary cones in R^4.

THEOREM 4.3. *Assume $\gamma(\theta)$ is a closed curve in S^3 such that the cone $C(\gamma)$ over γ is hamiltonian stationary. Then, after a unitary transformation, we obtain*

$$\gamma(\theta) = \frac{1}{\sqrt{p+q}}(\sqrt{q}e^{\sqrt{-1}p\theta}, \sqrt{-1}\sqrt{p}e^{-\sqrt{-1}q\theta}),$$

for $0 \leq \theta \leq 2\pi$, where p and q are positive integers such that $(p,q) = 1$ (we will refer to such a cone as being of (p,q) type). Moreover any cone of (p,q) type is hamiltonian stationary.

REMARK 4.4. *For any cone of (p,q) type, the Lagrangian angle β is given by*

$$\beta = (p-q)\theta.$$

Thus the local "Maslov index" of $C(\gamma)$ is $p-q$.

All the (p,q) cones are hamiltonian stationary, but it is not clear whether they are area minimizing among lagrangian competitors. The following proposition says that most of them are not area minimizing.

PROPOSITION 4.5. *If $|p-q| > 1$, then there is a hamiltonian function $h \in C_c^\infty(\mathbb{R}^4 - \{0\})$ such that*

$$Area(F_t(\Sigma)) < Area(\Sigma) - ct^2$$

for small $t > 0$ and a constant $c > 0$. Here Σ is the cone of (p,q) type and F_t is the one-parameter group of diffeomorphisms generated by the hamiltonian vector field $J\nabla h$. In particular the cones are not minimizing if $|p-q| > 1$.

The idea of the proof (given in [**SW**]) is to separate variables: let $h(r,\theta) = f(r)\sin(m\theta)$ and study the second variation of area (which is associated to a fourth order differential operator).

The same second variation analysis implies the following.

COROLLARY 4.6. *If $|p-q| = 1$, then the cone $C(\gamma)$ of (p,q) type is stable for any hamiltonian deformation with compact support in $\mathbb{R}^4 - \{0\}$.*

Mean curvature deformation (which would generally reduce area) does not work for the cones $C(\gamma)$. The reason is that the hamiltonian β for the mean curvature vector is multi-valued near the origin, and hence cannot be cut off to make it an allowable hamiltonian variation. If we do mean curvature deformation on a subregion of the cone away from the origin, the area is reduced, but the deformed boundary curves are no longer zero area curves, and hence cannot be filled with lagrangian disks.

5. Main Regularity Theorem

We now state the main regularity result of [**SW**].

THEOREM 5.1. *Let (M^4, g, J) be an almost Kähler manifold. Let $\sigma \in H_2(M,\mathbb{Z})$ be an integral homology class which admits a lagrangian representative. Assume $\Sigma \in \sigma$ (given as the image under a weakly conformal map u of a possibly disconnected surface) has least area among all lagrangian competitors (in an appropriate class of maps) representing σ. Then u is a global Lipschitz map and is a smooth immersion away from a finite set of points $\{P_1, \ldots, P_k, Q_1, \ldots, Q_l\}$, where*

1. *P_j are branch points;*
2. *Q_j are nonflat singular points with tangent cone of nonzero Maslov index $\mathrm{Ind}(Q_j) \in \mathbb{Z}$. In fact $\mathrm{Ind}(Q_j) = 1$ or -1. Moreover*

$$\sum_{j=1}^{l} \mathrm{Ind}(Q_j) = c_1(\sigma), \tag{5–1}$$

where c_1 is the first Chern class of M.

PROOF. We refer the reader to [**SW**] for the detailed proof. Here we only prove the index count (5–1) in the Kähler case by a different argument than that given in [**SW**]. On $\Sigma \setminus \{Q_1, \ldots, Q_l\}$, we have $d\sigma_H = \rho$. The standard formula for the first Chern class gives

$$c_1(\sigma) = \frac{1}{2\pi} \int_\Sigma \rho.$$

Since ρ is an ambient smooth form, we see that if ε is small, we have

$$\frac{1}{2\pi} \int_\Sigma \rho \approx \frac{1}{2\pi} \int_{\Sigma \setminus (B_\varepsilon(Q_1) \cup \cdots \cup B_\varepsilon(Q_2))} \rho = -\frac{1}{2\pi} \sum_{j=1}^{l} \int_{\partial B_\varepsilon(Q_j)} \sigma_H.$$

For ε small, we can prove that

$$\partial B_\varepsilon(Q_j) \approx \partial(C_j \cap B_\varepsilon(Q_j)),$$

where C_j is the tangent cone at Q_j. Moreover,

$$\int_{\partial(C_j \cap B_\varepsilon(Q_j))} \sigma_H = -2\pi \operatorname{Ind}(Q_j).$$

Therefore as ε goes to 0, we see that

$$\sum_{j=1}^{l} \operatorname{Ind}(Q_j) = c_1(\sigma). \qquad \square$$

COROLLARY 5.2. *At least one of the (p, q) type cones for $|p - q| = 1$ in $\mathbb{R}^4$ is minimizing among all Lagrangian competitors.*

PROOF. We need only construct a lagrangian homology class σ in some Kähler manifold such that the first Chern class acting on σ is nonzero. Then we see that any area minimizer in this lagrangian homology class must have a nonflat singularity. Hence the tangent cone at such a singular point is minimizing. $\square$

REMARK 5.3. *This is a very indirect way of proving one of the cones is minimizing. The question is still open as to which of the cones with $|p - q| = 1$ type is minimizing.*

We have seen that the first Chern class is an obstruction to the existence of a smooth Lagrangian representative of a given homology class. In the Kähler–Einstein case, this obstruction disappears since the first Chern class is a multiple of the Kähler class which vanishes on any Lagrangian homology class; that is,

$$c_1(\sigma) = c \int_\sigma \omega = 0.$$

Therefore the above index count gives that

$$\sum_{j=1}^{l} \operatorname{Ind}(Q_j) = 0.$$

In particular, there are an even number of nonflat singularities, and the number of singularities of Maslov index $+1$ is the same as the number of those of Maslov index -1.

REMARK 5.4. *There does exist an example of a lagrangian T^2 in a $K3$ surface which is hamiltonian stationary but not minimal (see* [**SW**]*).*

Therefore, it will be necessary to have a global argument to show that no such nonflat singularities exist for *minimizers* (not merely hamiltonian stationary). We believe that there will be a proof in this direction.

We should mention that similar arguments also apply to the boundary value problems described previously.

It is shown in [**SW**] that the $|p-q|=1$ cones do not minimize area if we allow nonorientable surfaces. J. Wolfson's lecture [**W**] discusses this in more detail.

6. More on Variational Theory

In this section we discuss the monotonicity formula which is a key to the regularity theorem. Since regularity results are essentially local, we will assume our ambient manifold is $\mathbb{R}^{2n}$.

It turns out to be important to derive the monotonicity formula in the contact space $\mathbb{R}^{2n+1}$. We consider the *horizontal Plateau problem in the contact space* $(\mathbb{R}^{2n+1},\theta)$, where $\mathbb{R}^{2n+1}$ has coordinates (x,y,φ) and $\theta = d\varphi - 2^{-1}(\sum_{i=1}^{n} x_i dy_i - y_i dx_i)$ is the standard contact one-form. The horizontal space H_P at a point P is the kernel of θ in $T_P\mathbb{R}^{2n+1}$.

Consider the projection $\pi : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n}$ along the φ direction, $\pi(x,y,\varphi) := (x,y)$. Clearly for any $P \in \mathbb{R}^{2n+1}$, $d\pi_P$ restricted to the horizontal space H_P is a linear isomorphism, and we define the metric dx^2+dy^2 on H_P so that it becomes an isometry (note that it is a nondegenerate metric on H_P although it is degenerate on the whole tangent space).

We call a k-dimensional submanifold Σ^k in $\mathbb{R}^{2n+1}$ horizontal if $\theta|_\Sigma = 0$, or equivalently $T_P\Sigma \subset H_P$ for any $P \in \Sigma$. If $k=n$, we call Σ *legendrian.*

There is a natural relationship between legendrian submanifolds in $\mathbb{R}^{2n+1}$ and lagrangian submanifolds in $\mathbb{R}^{2n}$. Let $\pi : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n}$ be the projection along φ direction defined above. If $\Sigma^k \subset \mathbb{R}^{2n+1}$ is horizontal, then we can show that $\Sigma_1 := \pi(\Sigma)$ is isotropic in $\mathbb{R}^{2n}$ (that is, $\omega|_{\Sigma_1} = 0$). In particular, if $\Sigma^n \subset \mathbb{R}^{2n+1}$ is legendrian , then $\Sigma_1 := \pi(\Sigma)$ is lagrangian in $\mathbb{R}^{2n}$.

To state the converse, we need the notion of exactness, which plays an important role in the theory. An isotropic submanifold $\Sigma_1 \subset \mathbb{R}^{2n}$ is *exact* if the one-form $\eta := 2^{-1}\big(\sum_{j=1}^{n} x_j dy_j - y_j dx_j\big)$ is exact on Σ_1. If Σ_1 is exact, then there is a function φ on Σ_1 such that $\eta = d\varphi$. Using this φ we can lift Σ_1 to a horizontal submanifold in $\mathbb{R}^{2n+1}$.

For $k \le n$, one can study the k-dimensional horizontal Plateau problem; to seek a least volume representative among all k-dimensional horizontal submanifolds with a given boundary.

Example: $k=1$. Given two points $P_i = (x^{(i)}, y^{(i)}, \varphi^{(i)})$, $i=1,2$ in $\mathbb{R}^{2n+1}$, we want to find the horizontal curve of least length connecting them. Assume γ is a horizontal curve connecting P_0 and P_1. Then

$$\varphi_1 - \varphi_0 = \int_{\pi(\gamma)} \eta = \int_{\pi(\gamma)+l} \eta,$$

where l is the straight line in $\mathbb{R}^{2n}$ connecting $\pi(P_1)$ and $\pi(P_0)$ with the correct orientation. Since $\int_{\pi(\gamma)+l}\eta$ is the symplectic area enclosed by $\pi(\gamma)$ and l, we see that the problem of finding the horizontal curve γ of least length becomes the problem of finding the curve $\pi(\gamma)$ from $\pi(P_1)$ to $\pi(P_0)$ of least length among all curves such that $\pi(\gamma)+l$ has the specific enclosed area $\varphi_1-\varphi_0$. Straightforward

arguments show that such a curve in $\mathbb{R}^{2n}$ is a circular arc in the complex line containing l.

DEFINITION 6.1. *A diffeomorphism $F : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n+1}$ is called* contact *if $F^*\theta = f(x, y, \varphi)\theta$ for some smooth function $f(x, y, \varphi)$ in $\mathbb{R}^{2n+1}$.*

Clearly contact diffeomorphisms preserve the horizontal condition, that is, the image of any horizontal submanifold under any contact diffeomorphism remains horizontal.

The key advantage to working in $\mathbb{R}^{2n+1}$ is that the class of contact transformations is larger than the class of symplectic transformations. For example, a dilation is not a symplectic transformation in $\mathbb{R}^{2n}$, but the dilation defined as

$$(x, y, \varphi) \mapsto (\lambda x, \lambda y, \lambda^2 \varphi)$$

is a contact transformation in $\mathbb{R}^{2n+1}$.

Similar to the hamiltonian vector fields in $\mathbb{R}^{2n}$, we can define contact vector fields for any ambient function h on $\mathbb{R}^{2n+1}$; e.g., the function φ generates the dilations. Using these contact vector fields, we can derive the key monotonicity formula for contact stationary surfaces in $\mathbb{R}^{2n+1}$. The case $k = 2$ and $n = 2$ is done by Schoen and Wolfson [**SW**]. The case $k = 2$ and $n \geq 3$ is done using a similar approach by Weiyang Qiu [**Q**].

References

[A] Allcock, D., *An isoperimetric inequality for the Heisenberg groups*, Geometric and Functional Analisys **8** (1998), 219–233.

[G] Gromov, M., *Carnot–Carathéodory spaces seen from within*, Progress in Mathematics **144** (1996), 79–323.

[HL] Harvey, R., Lawson, H. B., *Calibrated geometries*, Acta Math. **148** (1982), 47–157.

[J] Joyce, D., *Special lagrangian fibrations and the SYZ conjecture*, pp. 667–695 in this volume.

[Q] Qiu, W. Y., *Interior regularity of the solution to the isotropically constrained Plateau problem*, preprint.

[SW] Schoen, R., and Wolfson, J., *Minimizing area among Lagrangian surfaces: the mapping problem*, J. Diff. Geom. **58** (2001), 1–86.

[W] Wolfson, J., *Variational problems in Lagrangian geometry: $\mathbb{Z}_2$ currents*, pp. 697–704 in this volume.

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, STANFORD, CA 94305, UNITED STATES

E-mail address: schoen@math.stanford.edu

Clay Mathematics Proceedings
Volume 2, 2004

Lectures on Special Lagrangian Geometry

Dominic Joyce

1. Introduction

Calabi–Yau m-folds (M, J, ω, Ω) are compact complex manifolds (M, J) of complex dimension m, equipped with a Ricci-flat Kähler metric g with Kähler form ω, and a holomorphic $(m, 0)$-form Ω of constant length $|\Omega|^2 = 2^m$. Using Algebraic Geometry and Yau's solution of the Calabi Conjecture, one can construct them in huge numbers. String Theorists (a species of theoretical physicist) are very interested in Calabi–Yau 3-folds, and have made some extraordinary conjectures about them, in the subject known as *Mirror Symmetry*.

Special Lagrangian submanifolds, or *SL m-folds*, are a distinguished class of real m-dimensional minimal submanifolds that may be defined in $\mathbb{C}^m$, or in Calabi–Yau m-folds, or more generally in *almost Calabi–Yau m-folds*. They are calibrated with respect to the m-form $\mathrm{Re}\,\Omega$. They are fairly rigid and well-behaved, so that compact SL m-folds N occur in smooth moduli spaces of dimension $b^1(N)$, for instance. They are important in String Theory, and are expected to play a rôle in the eventual explanation of Mirror Symmetry.

This article is intended as an introduction to special Lagrangian geometry, and a survey of the author's research on the singularities of SL m-folds, of directions in which the subject might develop in the next few years, and of possible applications of it to Mirror Symmetry and the SYZ Conjecture.

Sections 2 and 3 discuss general properties of special Lagrangian submanifolds of $\mathbb{C}^m$, and ways to construct examples. Then Section 4 defines Calabi–Yau and almost Calabi–Yau manifolds, and their special Lagrangian submanifolds. Section 5 discusses the deformation and obstruction theory of *compact* SL m-folds, and properties of their moduli spaces.

In Section 6 we describe a theory of *isolated conical singularities* in compact SL m-folds. Finally, Section 7 briefly introduces String Theory, Mirror Symmetry and the *SYZ Conjecture*, a conjectural explanation of Mirror Symmetry of Calabi–Yau 3-folds, and discusses mathematical progress towards clarifying and proving the conjecture.

Acknowledgments. This article is based on lecture courses given to the Summer School on Symplectic Geometry in Nordfjordeid, Norway, in June 2001, and to the Clay Institute's Summer School on the Global Theory of Minimal Surfaces, MSRI,

©2004 Clay Mathematics Institute

California, in July 2001. I would like to thank the organizers of these for inviting me to speak.

Many people have helped me develop my ideas on special Lagrangian geometry; amongst them I would particularly like to thank Robert Bryant, Mark Gross, Mark Haskins, Nigel Hitchin, Ian McIntosh, Richard Thomas, and Karen Uhlenbeck.

2. Special Lagrangian Submanifolds in $\mathbb{C}^m$

We begin by defining *calibrations* and *calibrated submanifolds*, following Harvey and Lawson [**11**].

DEFINITION 2.1. Let (M, g) be a Riemannian manifold. An *oriented tangent k-plane* V on M is a vector subspace V of some tangent space T_xM to M with $\dim V = k$, equipped with an orientation. If V is an oriented tangent k-plane on M, then $g|_V$ is a Euclidean metric on V, so combining $g|_V$ with the orientation on V gives a natural *volume form* vol_V on V, which is a k-form on V.

Now let φ be a closed k-form on M. We say that φ is a *calibration* on M if for every oriented k-plane V on M we have $\varphi|_V \leqslant \mathrm{vol}_V$. Here $\varphi|_V = \alpha \cdot \mathrm{vol}_V$ for some $\alpha \in \mathbb{R}$, and $\varphi|_V \leqslant \mathrm{vol}_V$ if $\alpha \leqslant 1$. Let N be an oriented submanifold of M with dimension k. Then each tangent space T_xN for $x \in N$ is an oriented tangent k-plane. We say that N is a *calibrated submanifold* if $\varphi|_{T_xN} = \mathrm{vol}_{T_xN}$ for all $x \in N$.

It is easy to show that calibrated submanifolds are automatically *minimal submanifolds* [**11**, Th. II.4.2]. Here is the definition of special Lagrangian submanifolds in $\mathbb{C}^m$, taken from [**11**, §III].

DEFINITION 2.2. Let $\mathbb{C}^m \cong \mathbb{R}^{2m}$ have complex coordinates $(z_1, \ldots, z_m)$ and complex structure I, and define a metric g', Kähler form ω' and complex volume form Ω' on $\mathbb{C}^m$ by

$$g' = |\mathrm{d}z_1|^2 + \cdots + |\mathrm{d}z_m|^2, \quad \omega' = \frac{i}{2}(\mathrm{d}z_1 \wedge \mathrm{d}\bar{z}_1 + \cdots + \mathrm{d}z_m \wedge \mathrm{d}\bar{z}_m), \quad \text{and} \quad \Omega' = \mathrm{d}z_1 \wedge \cdots \wedge \mathrm{d}z_m. \tag{2–1}$$

Then $\mathrm{Re}\,\Omega'$ and $\mathrm{Im}\,\Omega'$ are real m-forms on $\mathbb{C}^m$. Let L be an oriented real submanifold of $\mathbb{C}^m$ of real dimension m. We call L a *special Lagrangian submanifold* in $\mathbb{C}^m$, or *SL m-fold* for short, if L is calibrated with respect to $\mathrm{Re}\,\Omega'$, in the sense of Definition 2.1.

In fact there is a more general definition involving a *phase* $\mathrm{e}^{i\theta}$: if $\theta \in [0, 2\pi)$, we say that L is *special Lagrangian with phase* $\mathrm{e}^{i\theta}$ if it is calibrated with respect to $\cos\theta\ \mathrm{Re}\,\Omega' + \sin\theta\ \mathrm{Im}\,\Omega'$. But we will not use this.

We shall identify the family $\mathcal{F}$ of tangent m-planes in $\mathbb{C}^m$ calibrated with respect to $\mathrm{Re}\,\Omega'$. The subgroup of $\mathrm{GL}(2m, \mathbb{R})$ preserving g', ω' and Ω' is the Lie group $\mathrm{SU}(m)$ of complex unitary matrices with determinant 1. Define a real vector subspace U in $\mathbb{C}^m$ to be

$$U = \bigl\{(x_1, \ldots, x_m) : x_j \in \mathbb{R}\bigr\} \subset \mathbb{C}^m, \tag{2–2}$$

and let U have the usual orientation. Then U is calibrated with respect to $\mathrm{Re}\,\Omega'$.

Furthermore, any oriented real vector subspace V in $\mathbb{C}^m$ calibrated with respect to $\mathrm{Re}\,\Omega'$ is of the form $V = \gamma \cdot U$ for some $\gamma \in \mathrm{SU}(m)$. Therefore $\mathrm{SU}(m)$ acts transitively on $\mathcal{F}$. The stabilizer subgroup of U in $\mathrm{SU}(m)$ is the subset of matrices

in $\mathrm{SU}(m)$ with real entries, which is $\mathrm{SO}(m)$. Thus $\mathcal{F} \cong \mathrm{SU}(m)/\mathrm{SO}(m)$, and we prove:

PROPOSITION 2.3. *The family $\mathcal{F}$ of oriented real m-dimensional vector subspaces V in $\mathbb{C}^m$ with $\mathrm{Re}\,\Omega'|_V = \mathrm{vol}_V$ is isomorphic to $\mathrm{SU}(m)/\mathrm{SO}(m)$, and has dimension $\frac{1}{2}(m^2+m-2)$.*

The dimension follows because

$$\dim \mathrm{SU}(m) = m^2 - 1 \quad \text{and} \quad \dim \mathrm{SO}(m) = \tfrac{1}{2}m(m-1).$$

It is easy to see that $\omega'|_U = \mathrm{Im}\,\Omega'|_U = 0$. As $\mathrm{SU}(m)$ preserves ω' and $\mathrm{Im}\,\Omega'$ and acts transitively on $\mathcal{F}$, it follows that $\omega'|_V = \mathrm{Im}\,\Omega'|_V = 0$ for any $V \in \mathcal{F}$. Conversely, if V is a real m-dimensional vector subspace of $\mathbb{C}^m$ and $\omega'|_V = \mathrm{Im}\,\Omega'|_V = 0$, then V lies in $\mathcal{F}$, with some orientation. This implies an alternative characterization of special Lagrangian submanifolds, [**11**, Cor. III.1.11]:

PROPOSITION 2.4. *Let L be a real m-dimensional submanifold of $\mathbb{C}^m$. Then L admits an orientation making it into a special Lagrangian submanifold of $\mathbb{C}^m$ if and only if $\omega'|_L \equiv 0$ and $\mathrm{Im}\,\Omega'|_L \equiv 0$.*

Note that an m-dimensional submanifold L in $\mathbb{C}^m$ is called *Lagrangian* if $\omega'|_L$ vanishes identically. (This is a term from symplectic geometry, and ω' is a symplectic structure.) Thus special Lagrangian submanifolds are Lagrangian submanifolds satisfying the extra condition that $\mathrm{Im}\,\Omega'|_L \equiv 0$, which is how they get their name.

2.1. Special Lagrangian 2-folds in $\mathbb{C}^2$ and the quaternions. The smallest interesting dimension, $m = 2$, is a special case. Let $\mathbb{C}^2$ have complex coordinates (z_1, z_2), complex structure I, and metric g', Kähler form ω' and holomorphic 2-form Ω' defined in (2–1). Define real coordinates (x_0, x_1, x_2, x_3) on $\mathbb{C}^2 \cong \mathbb{R}^4$ by $z_0 = x_0 + ix_1$, $z_1 = x_2 + ix_3$. Then

$$g' = \mathrm{d}x_0^2 + \cdots + \mathrm{d}x_3^2, \qquad \omega' = \mathrm{d}x_0 \wedge \mathrm{d}x_1 + \mathrm{d}x_2 \wedge \mathrm{d}x_3,$$
$$\mathrm{Re}\,\Omega' = \mathrm{d}x_0 \wedge \mathrm{d}x_2 - \mathrm{d}x_1 \wedge \mathrm{d}x_3 \quad \text{and} \quad \mathrm{Im}\,\Omega' = \mathrm{d}x_0 \wedge \mathrm{d}x_3 + \mathrm{d}x_1 \wedge \mathrm{d}x_2.$$

Now define a *different* set of complex coordinates (w_1, w_2) on $\mathbb{C}^2 = \mathbb{R}^4$ by $w_1 = x_0 + ix_2$ and $w_2 = x_1 - ix_3$. Then $\omega' - i\,\mathrm{Im}\,\Omega' = \mathrm{d}w_1 \wedge \mathrm{d}w_2$.

But by Proposition 2.4, a real 2-submanifold $L \subset \mathbb{R}^4$ is special Lagrangian if and only if $\omega'|_L \equiv \mathrm{Im}\,\Omega'|_L \equiv 0$. Thus, L is special Lagrangian if and only if $(\mathrm{d}w_1 \wedge \mathrm{d}w_2)|_L \equiv 0$. But this holds if and only if L is a *holomorphic curve* with respect to the complex coordinates (w_1, w_2).

Here is another way to say this. There are *two different* complex structures I and J involved in this problem, associated to the two different complex coordinate systems (z_1, z_2) and (w_1, w_2) on $\mathbb{R}^4$. In the coordinates $(x_0, \ldots, x_3)$, I and J are given by

$$I\Big(\frac{\partial}{\partial x_0}\Big) = \frac{\partial}{\partial x_1}, \quad I\Big(\frac{\partial}{\partial x_1}\Big) = -\frac{\partial}{\partial x_0}, \quad I\Big(\frac{\partial}{\partial x_2}\Big) = \frac{\partial}{\partial x_3}, \quad I\Big(\frac{\partial}{\partial x_3}\Big) = -\frac{\partial}{\partial x_2},$$
$$J\Big(\frac{\partial}{\partial x_0}\Big) = \frac{\partial}{\partial x_2}, \quad J\Big(\frac{\partial}{\partial x_1}\Big) = -\frac{\partial}{\partial x_3}, \quad J\Big(\frac{\partial}{\partial x_2}\Big) = -\frac{\partial}{\partial x_0}, \quad J\Big(\frac{\partial}{\partial x_3}\Big) = \frac{\partial}{\partial x_1}.$$

The usual complex structure on $\mathbb{C}^2$ is I, but a 2-fold L in $\mathbb{C}^2$ is special Lagrangian if and only if it is holomorphic with respect to the alternative complex structure J. This means that special Lagrangian 2-folds are already very well understood, so we generally focus our attention on dimensions $m \geqslant 3$.

We can express all this in terms of the *quaternions* $\mathbb{H}$. The complex structures I, J anticommute, so that $IJ = -JI$, and $K = IJ$ is also a complex structure on $\mathbb{R}^4$, and $\langle 1, I, J, K \rangle$ is an algebra of automorphisms of $\mathbb{R}^4$ isomorphic to $\mathbb{H}$.

2.2. Special Lagrangian submanifolds in $\mathbb{C}^m$ as graphs. In symplectic geometry, there is a well-known way of manufacturing *Lagrangian* submanifolds of $\mathbb{R}^{2m} \cong \mathbb{C}^m$, which works as follows. Let $f : \mathbb{R}^m \to \mathbb{R}$ be a smooth function, and define

$$\Gamma_f = \left\{ \left(x_1 + i \frac{\partial f}{\partial x_1}(x_1, \ldots, x_m), \ldots, x_m + i \frac{\partial f}{\partial x_m}(x_1, \ldots, x_m) \right) : x_1, \ldots, x_m \in \mathbb{R} \right\}.$$

Then Γ_f is a smooth real m-dimensional submanifold of $\mathbb{C}^m$, with $\omega'|_{\Gamma_f} \equiv 0$. Identifying $\mathbb{C}^m \cong \mathbb{R}^{2m} \cong \mathbb{R}^m \times (\mathbb{R}^m)^*$, we may regard Γ_f as the graph of the 1-form $\mathrm{d}f$ on $\mathbb{R}^m$, so that Γ_f is the *graph of a closed 1-form*. Locally, but not globally, every Lagrangian submanifold arises from this construction.

Now by Proposition 2.4, a special Lagrangian m-fold in $\mathbb{C}^m$ is a Lagrangian m-fold L satisfying the additional condition that $\operatorname{Im} \Omega'|_L \equiv 0$. We shall find the condition for Γ_f to be a special Lagrangian m-fold. Define the *Hessian* $\operatorname{Hess} f$ of f to be the $m \times m$ matrix

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1}^{m}$$

of real functions on $\mathbb{R}^m$. Then it is easy to show that $\operatorname{Im} \Omega'|_{\Gamma_f} \equiv 0$ if and only if

$$\operatorname{Im} \det{}_{\mathbb{C}} \big(I + i \operatorname{Hess} f \big) \equiv 0 \quad \text{on } \mathbb{C}^m. \tag{2–3}$$

This is a *nonlinear second-order elliptic partial differential equation* upon the function $f : \mathbb{R}^m \to \mathbb{R}$.

2.3. Local discussion of special Lagrangian deformations. Suppose L_0 is a special Lagrangian submanifold in $\mathbb{C}^m$ (or, more generally, in some (almost) Calabi–Yau m-fold). What can we say about the family of *special Lagrangian deformations* of L_0, that is, the set of special Lagrangian m-folds L that are "close to L_0" in a suitable sense? Essentially, deformation theory is one way of thinking about the question "how many special Lagrangian submanifolds are there in $\mathbb{C}^m$"?

Locally (that is, in small enough open sets), every special Lagrangian m-fold looks quite like $\mathbb{R}^m$ in $\mathbb{C}^m$. Therefore deformations of special Lagrangian m-folds should look like special Lagrangian deformations of $\mathbb{R}^m$ in $\mathbb{C}^m$. So, we would like to know what special Lagrangian m-folds L in $\mathbb{C}^m$ close to $\mathbb{R}^m$ look like.

Now $\mathbb{R}^m$ is the graph Γ_f associated to the function $f \equiv 0$. Thus, a graph Γ_f will be close to $\mathbb{R}^m$ if the function f and its derivatives are small. But then $\operatorname{Hess} f$ is small, so we can approximate equation (2–3) by its *linearization*. For

$$\operatorname{Im} \det{}_{\mathbb{C}} \big(I + i \operatorname{Hess} f \big) = \operatorname{Tr} \operatorname{Hess} f + \text{higher order terms}.$$

Thus, when the second derivatives of f are small, equation (2–3) reduces approximately to $\operatorname{Tr} \operatorname{Hess} f \equiv 0$. But

$$Tr \operatorname{Hess} f = \frac{\partial^2 f}{(\partial x_1)^2} + \cdots + \frac{\partial^2 f}{(\partial x_m)^2} = -\Delta f,$$

where Δ is the *Laplacian* on $\mathbb{R}^m$.

Hence, the small special Lagrangian deformations of $\mathbb{R}^m$ in $\mathbb{C}^m$ are approximately parametrized by small *harmonic functions* on $\mathbb{R}^m$. Actually, because adding a constant to f has no effect on Γ_f, this parametrization is degenerate. We can get

round this by parametrizing instead by $\mathrm{d}f$, which is a closed and coclosed 1-form. This justifies the following:

PRINCIPLE. *Small special Lagrangian deformations of a special Lagrangian m-fold L are approximately parametrized by closed and coclosed 1-forms α on L.*

This is the idea behind McLean's Theorem, Theorem 5.1 below.

We have seen using (2–3) that the deformation problem for special Lagrangian m-folds can be written as an *elliptic equation*. In particular, there are the same number of equations as functions, so the problem is neither overdetermined nor underdetermined. Therefore we do not expect special Lagrangian m-folds to be very few and very rigid (as would be the case if (2–3) were overdetermined), nor to be very abundant and very flabby (as would be the case if (2–3) were underdetermined).

If we think about Proposition 2.3 for a while, this may seem surprising. For the set $\mathcal{F}$ of special Lagrangian m-planes in $\mathbb{C}^m$ has dimension $\frac{1}{2}(m^2+m-2)$, but the set of all real m-planes in $\mathbb{C}^m$ has dimension m^2. So the special Lagrangian m-planes have codimension $\frac{1}{2}(m^2-m+2)$ in the set of all m-planes.

This means that the condition for a real m-submanifold L in $\mathbb{C}^m$ to be special Lagrangian is $\frac{1}{2}(m^2-m+2)$ real equations on each tangent space of L. However, the freedom to vary L is the sections of its normal bundle in $\mathbb{C}^m$, which is m real functions. When $m \geqslant 3$, there are more equations than functions, so we would expect the deformation problem to be *overdetermined*.

The explanation is that because ω' is a *closed* 2-form, submanifolds L with $\omega'|_L \equiv 0$ are much more abundant than would otherwise be the case. So the closure of ω' is a kind of integrability condition necessary for the existence of many special Lagrangian submanifolds, just as the integrability of an almost complex structure is a necessary condition for the existence of many complex submanifolds of dimension greater than 1 in a complex manifold.

3. Constructions of SL m-folds in $\mathbb{C}^m$

We now describe five methods of constructing special Lagrangian m-folds in $\mathbb{C}^m$, drawn from papers by the author [**15, 16, 18, 19, 20, 21, 22, 23**], Bryant [**1**], Castro and Urbano [**2**], Goldstein [**4, 5**], Harvey [**10**, p. 139–143], Harvey and Lawson [**11**, §III], Haskins [**12**], Lawlor [**32**], Ma and Ma [**33**], McIntosh [**35**] and Sharipov [**40**]. These yield many examples of singular SL m-folds, and so hopefully will help in understanding general singularities of SL m-folds in Calabi–Yau m-folds.

3.1. SL m-folds with large symmetry groups. Here is a method used in [**18**] (and also by Harvey and Lawson [**11**, §III.3], Haskins [**12**] and Goldstein [**4, 5**]) to construct examples of SL m-folds in $\mathbb{C}^m$. The group $\mathrm{SU}(m) \ltimes \mathbb{C}^m$ acts on $\mathbb{C}^m$ preserving all the structure g', ω', Ω', so that it takes SL m-folds to SL m-folds in $\mathbb{C}^m$. Let G be a Lie subgroup of $\mathrm{SU}(m) \ltimes \mathbb{C}^m$ with Lie algebra $\mathfrak{g}$, and N a connected G-invariant SL m-fold in $\mathbb{C}^m$.

Since G preserves the symplectic form ω' on $\mathbb{C}^m$, one can show that it has a *moment map* $\mu : \mathbb{C}^m \to \mathfrak{g}^*$. As N is Lagrangian, one can show that μ is constant on N, that is, $\mu \equiv c$ on N for some $c \in Z(\mathfrak{g}^*)$, the *center* of $\mathfrak{g}^*$.

If the orbits of G in N are of codimension 1 (that is, dimension $m-1$), then N is a 1-parameter family of G-orbits $\mathcal{O}_t$ for $t \in \mathbb{R}$. After reparametrizing the variable t, it can be shown that the special Lagrangian condition is equivalent to an ODE in t upon the orbits $\mathcal{O}_t$.

Thus, we can construct examples of cohomogeneity one SL m-folds in $\mathbb{C}^m$ by solving an ODE in the family of $(m-1)$-dimensional G-orbits $\mathcal{O}$ in $\mathbb{C}^m$ with $\mu|_{\mathcal{O}} \equiv c$, for fixed $c \in Z(\mathfrak{g}^*)$. This ODE usually turns out to be *integrable*.

Now suppose N is a *special Lagrangian cone* in $\mathbb{C}^m$, invariant under a subgroup $G \subset \mathrm{SU}(m)$ which has orbits of dimension $m-2$ in N. In effect the symmetry group of N is $G \times \mathbb{R}_+$, where $\mathbb{R}_+$ acts by *dilations*, as N is a cone. Thus, in this situation too the symmetry group of N acts with cohomogeneity one, and we again expect the problem to reduce to an ODE.

One can show that $N \cap \mathcal{S}^{2m-1}$ is a 1-parameter family of G-orbits $\mathcal{O}_t$ in $\mathcal{S}^{2m-1} \cap \mu^{-1}(0)$ satisfying an ODE. By solving this ODE we construct SL cones in $\mathbb{C}^m$. When $G = \mathrm{U}(1)^{m-2}$, the ODE has many *periodic solutions* which give large families of distinct SL cones on T^{m-1}. In particular, we can find many examples of SL T^2-cones in $\mathbb{C}^3$.

3.2. Evolution equations for SL m-folds. The following method was used in [**15**] and [**16**] to construct many examples of SL m-folds in $\mathbb{C}^m$. A related but less general method was used by Lawlor [**32**], and completed by Harvey [**10**, p. 139–143].

Let P be a real analytic $(m-1)$-dimensional manifold, and χ a nonvanishing real analytic section of $\Lambda^{m-1}TP$. Let $\{\phi_t : t \in \mathbb{R}\}$ be a 1-parameter family of real analytic maps $\phi_t : P \to \mathbb{C}^m$. Consider the ODE

$$\left(\frac{\mathrm{d}\phi_t}{\mathrm{d}t}\right)^b = (\phi_t)_*(\chi)^{a_1 \ldots a_{m-1}} (\mathrm{Re}\,\Omega')_{a_1 \ldots a_{m-1} a_m} g'^{a_m b}, \tag{3–1}$$

using the index notation for (real) tensors on $\mathbb{C}^m$, where g'^{ab} is the inverse of the Euclidean metric g'_{ab} on $\mathbb{C}^m$.

It is shown in [**15**, §3] that if the ϕ_t satisfy (3–1) and $\phi_0^*(\omega') \equiv 0$, then $\phi_t^*(\omega') \equiv 0$ for all t, and

$$N = \bigl\{\phi_t(p) : p \in P,\, t \in \mathbb{R}\bigr\}$$

is an SL m-fold in $\mathbb{C}^m$ wherever it is nonsingular. We think of (3–1) as an *evolution equation*, and N as the result of evolving a 1-parameter family of $(m-1)$-submanifolds $\phi_t(P)$ in $\mathbb{C}^m$.

Here is one way to understand this result. Suppose we are given $\phi_t : P \to \mathbb{C}^m$ for some t, and we want to find an SL m-fold N in $\mathbb{C}^m$ containing the $(m-1)$-submanifold $\phi_t(P)$. As N is Lagrangian, a necessary condition for this is that $\omega'|_{\phi_t(P)} \equiv 0$, and hence $\phi_t^*(\omega') \equiv 0$ on P.

The effect of equation (3–1) is to flow $\phi_t(P)$ in the direction in which $\mathrm{Re}\,\Omega'$ is "largest". The result is that $\mathrm{Re}\,\Omega'$ is "maximized" on N, given the initial conditions. But $\mathrm{Re}\,\Omega'$ is maximal on N exactly when N is calibrated with respect to $\mathrm{Re}\,\Omega'$, that is, when N is special Lagrangian. The same technique also works for other calibrations, such as the associative and coassociative calibrations on $\mathbb{R}^7$, and the Cayley calibration on $\mathbb{R}^8$.

Now (3–1) evolves amongst the infinite-dimensional family of real analytic maps $\phi : P \to \mathbb{C}^m$ with $\phi^*(\omega') \equiv 0$, so it is an *infinite-dimensional* problem, and thus difficult to solve explicitly. However, there are *finite-dimensional* families $\mathcal{C}$ of maps $\phi : P \to \mathbb{C}^m$ such that evolution stays in $\mathcal{C}$. This gives a *finite-dimensional* ODE, which can hopefully be solved fairly explicitly. For example, if we take G to be a Lie subgroup of $\mathrm{SU}(m) \ltimes \mathbb{C}^m$, P to be an $(m-1)$-dimensional homogeneous space G/H, and $\phi : P \to \mathbb{C}^m$ to be G-equivariant, we recover the construction of Section 3.1.

But there are also other possibilities for $\mathcal{C}$ which do not involve a symmetry assumption. Suppose P is a submanifold of $\mathbb{R}^n$, and χ the restriction to P of a linear or affine map $\mathbb{R}^n \to \Lambda^{m-1}\mathbb{R}^n$. (This is a strong condition on P and χ.) Then we can take $\mathcal{C}$ to be the set of restrictions to P of linear or affine maps $\mathbb{R}^n \to \mathbb{C}^m$.

For instance, set $m = n$ and let P be a quadric in $\mathbb{R}^m$. Then one can construct SL m-folds in $\mathbb{C}^m$ with few symmetries by evolving quadrics in Lagrangian planes $\mathbb{R}^m$ in $\mathbb{C}^m$. When P is a quadric cone in $\mathbb{R}^m$ this gives many SL cones on products of spheres $\mathcal{S}^a \times \mathcal{S}^b \times \mathcal{S}^1$.

3.3. Ruled special Lagrangian 3-folds. A 3-submanifold N in $\mathbb{C}^3$ is called *ruled* if it is fibered by a 2-dimensional family $\mathcal{F}$ of real lines in $\mathbb{C}^3$. A *cone* N_0 in $\mathbb{C}^3$ is called *two-sided* if $N_0 = -N_0$. Two-sided cones are automatically ruled. If N is a ruled 3-fold in $\mathbb{C}^3$, we define the *asymptotic cone* N_0 of N to be the two-sided cone fibered by the lines passing through 0 and parallel to those in $\mathcal{F}$.

Ruled SL 3-folds are studied in [**19**], and also by Harvey and Lawson [**11**, §III.3.C, §III.4.B] and Bryant [**1**, §3]. Each (oriented) real line in $\mathbb{C}^3$ is determined by its *direction* in $\mathcal{S}^5$ together with an orthogonal *translation* from the origin. Thus a ruled 3-fold N is determined by a 2-dimensional family of directions and translations.

The condition for N to be special Lagrangian turns out [**19**, §5] to reduce to two equations, the first involving only the direction components, and the second *linear* in the translation components. Hence, if a ruled 3-fold N in $\mathbb{C}^3$ is special Lagrangian, then so is its asymptotic cone N_0. Conversely, the ruled SL 3-folds N asymptotic to a given two-sided SL cone N_0 come from solutions of a linear equation, and so form a *vector space*.

Let N_0 be a two-sided SL cone, and let $\Sigma = N_0 \cap \mathcal{S}^5$. Then Σ is a *Riemann surface*. Holomorphic vector fields on Σ give solutions to the linear equation (though not all solutions) [**19**, §6], and so yield new ruled SL 3-folds. In particular, each SL T^2-cone gives a 2-dimensional family of ruled SL 3-folds, which are generically diffeomorphic to $T^2 \times \mathbb{R}$ as immersed 3-submanifolds.

3.4. Integrable systems. Let N_0 be a special Lagrangian cone in $\mathbb{C}^3$, and set $\Sigma = N_0 \cap \mathcal{S}^5$. As N_0 is calibrated, it is minimal in $\mathbb{C}^3$, and so Σ is minimal in $\mathcal{S}^5$. That is, Σ is a *minimal Legendrian surface* in $\mathcal{S}^5$. Let $\pi : \mathcal{S}^5 \to \mathbb{CP}^2$ be the Hopf projection. One can also show that $\pi(\Sigma)$ is a *minimal Lagrangian surface* in $\mathbb{CP}^2$.

Regard Σ as a *Riemann surface*. Then the inclusions $\iota : \Sigma \to \mathcal{S}^5$ and $\pi \circ \iota : \Sigma \to \mathbb{CP}^2$ are *conformal harmonic maps*. Now harmonic maps from Riemann surfaces into $\mathcal{S}^n$ and $\mathbb{CP}^m$ are an *integrable system*. There is a complicated theory for classifying them in terms of algebro-geometric "spectral data", and finding "explicit" solutions. In principle, this gives all harmonic maps from T^2 into $\mathcal{S}^n$ and $\mathbb{CP}^m$. So, the field of integrable systems offers the hope of a *classification* of all SL T^2-cones in $\mathbb{C}^3$.

For a good general introduction to this field, see Fordy and Wood [**3**]. Sharipov [**40**] and Ma and Ma [**33**] apply this integrable systems machinery to describe minimal Legendrian tori in $\mathcal{S}^5$, and minimal Lagrangian tori in $\mathbb{CP}^2$, respectively, giving explicit formulae in terms of Prym theta functions. McIntosh [**35**] provides a more recent, readable, and complete discussion of special Lagrangian cones in $\mathbb{C}^3$ from the integrable systems perspective.

The families of SL T^2-cones constructed by U(1)-invariance in Section 3.1, and by evolving quadrics in Section 3.2, turn out to come from a more general, very explicit, "integrable systems" family of conformal harmonic maps $\mathbb{R}^2 \to \mathcal{S}^5$ with Legendrian image, involving two commuting, integrable ODEs, described in [**20**]. So, we can fit some of our examples into the integrable systems framework.

However, we know a good number of other constructions of SL m-folds in $\mathbb{C}^m$ which have the classic hallmarks of integrable systems — elliptic functions, commuting ODEs, and so on — but which are not yet understood from the point of view of integrable systems.

3.5. Analysis and U(1)-invariant SL 3-folds in $\mathbb{C}^3$. Next we summarize the author's three papers [**21, 22, 23**], which study SL 3-folds N in $\mathbb{C}^3$ invariant under the U(1)-action

$$\mathrm{e}^{i\theta} : (z_1, z_2, z_3) \mapsto (\mathrm{e}^{i\theta} z_1, \mathrm{e}^{-i\theta} z_2, z_3) \quad \text{for } \mathrm{e}^{i\theta} \in \mathrm{U}(1). \tag{3–2}$$

These three papers are surveyed in [**24**]. Locally we can write N in the form

$$\begin{aligned} N = \big\{ (z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 z_2 = v(x,y) + iy, \quad z_3 = x + iu(x,y), \\ |z_1|^2 - |z_2|^2 = 2a, \quad (x,y) \in S \big\}, \end{aligned} \tag{3–3}$$

where S is a domain in $\mathbb{R}^2$, $a \in \mathbb{R}$ and $u, v : S \to \mathbb{R}$ are continuous.

Here we may take $|z_1|^2 - |z_2|^2 = 2a$ to be one of the equations defining N as $|z_1|^2 - |z_2|^2$ is the *moment map* of the U(1)-action (3–2), and so $|z_1|^2 - |z_2|^2$ is constant on any U(1)-invariant Lagrangian 3-fold in $\mathbb{C}^3$. Effectively (3–3) just means that we are choosing $x = \mathrm{Re}(z_3)$ and $y = \mathrm{Im}(z_1 z_2)$ as local coordinates on the 2-manifold $N/\mathrm{U}(1)$. Then we find [**21**, Prop. 4.1]:

PROPOSITION 3.1. *Let S, a, u, v and N be as above. Then*

(a) *If $a = 0$, then N is a (possibly singular) SL 3-fold in $\mathbb{C}^3$ if u, v are differentiable and satisfy*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial v}{\partial x} = -2\big(v^2 + y^2\big)^{1/2} \frac{\partial u}{\partial y}, \tag{3–4}$$

except at points $(x,0)$ in S with $v(x,0) = 0$, where u, v need not be differentiable. The singular points of N are those of the form $(0,0,z_3)$, where $z_3 = x + iu(x,0)$ for $(x,0) \in S$ with $v(x,0) = 0$.

(b) *If $a \neq 0$, then N is a nonsingular SL 3-fold in $\mathbb{C}^3$ if and only if u, v are differentiable in S and satisfy*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial v}{\partial x} = -2\big(v^2 + y^2 + a^2\big)^{1/2} \frac{\partial u}{\partial y}. \tag{3–5}$$

Now (3–4) and (3–5) are *nonlinear Cauchy–Riemann equations*. Thus, we may treat $u + iv$ as a holomorphic function of $x + iy$. Many of the results in [**21, 22, 23**] are analogues of well-known results in elementary complex analysis.

In [**21**, Prop. 7.1] we show that solutions $u, v \in C^1(S)$ of (3–5) come from a potential $f \in C^2(S)$ satisfying a second-order quasilinear elliptic equation.

PROPOSITION 3.2. *Let S be a domain in $\mathbb{R}^2$ and $u, v \in C^1(S)$ satisfy* (3–5) *for $a \neq 0$. Then there exists $f \in C^2(S)$ with $\frac{\partial f}{\partial y} = u$, $\frac{\partial f}{\partial x} = v$ and*

$$P(f) = \left(\left(\frac{\partial f}{\partial x} \right)^2 + y^2 + a^2 \right)^{-1/2} \frac{\partial^2 f}{\partial x^2} + 2 \frac{\partial^2 f}{\partial y^2} = 0. \tag{3–6}$$

This f is unique up to addition of a constant, $f \mapsto f + c$. Conversely, all solutions of (3–6) *yield solutions of* (3–5).

In the following result, a condensation of [**21**, Th. 7.6] and [**22**, Th.s 9.20 & 9.21], we prove existence and uniqueness for the *Dirichlet problem* for (3–6).

THEOREM 3.3. *Suppose S is a strictly convex domain in $\mathbb{R}^2$ invariant under $(x, y) \mapsto (x, -y)$, and $\alpha \in (0, 1)$. Let $a \in \mathbb{R}$ and $\phi \in C^{3,\alpha}(\partial S)$. Then if $a \neq 0$ there exists a unique solution f of* (3–6) *in $C^{3,\alpha}(S)$ with $f|_{\partial S} = \phi$. If $a = 0$ there exists a unique $f \in C^1(S)$ with $f|_{\partial S} = \phi$, which is twice weakly differentiable and satisfies* (3–6) *with weak derivatives. Furthermore, the map $C^{3,\alpha}(\partial S) \times \mathbb{R} \to C^1(S)$ taking $(\phi, a) \mapsto f$ is continuous.*

Here a domain S in $\mathbb{R}^2$ is *strictly convex* if it is convex and the curvature of ∂S is nonzero at each point. Also domains are by definition compact, with smooth boundary, and $C^{3,\alpha}(\partial S)$ and $C^{3,\alpha}(S)$ are *Hölder spaces* of functions on ∂S and S. For more details see [**21, 22**].

Combining Propositions 3.1 and 3.2 and Theorem 3.3 gives existence and uniqueness for a large class of U(1)-invariant SL 3-folds in $\mathbb{C}^3$, with boundary conditions, and including *singular* SL 3-folds. It is interesting that this existence and uniqueness is *entirely unaffected* by singularities appearing in S°.

Here are some other areas covered in [**21, 22, 23**]. Examples of solutions u, v of (3–4) and (3–5) are given in [**21**, §5]. In [**22**] we give more precise statements on the regularity of singular solutions of (3–4) and (3–6). In [**21**, §6] and [**23**, §7] we consider the zeroes of $(u_1, v_1) - (u_2, v_2)$, where (u_j, v_j) are (possibly singular) solutions of (3–4) and (3–5).

We show that if $(u_1, v_1) \not\equiv (u_2, v_2)$, then the zeroes of $(u_1, v_1) - (u_2, v_2)$ in S° are *isolated*, with a positive integer *multiplicity*, and that the zeroes of $(u_1, v_1) - (u_2, v_2)$ in S° can be counted with multiplicity in terms of boundary data on ∂S. In particular, under some boundary conditions we can show $(u_1, v_1) - (u_2, v_2)$ has no zeroes in S°, so that the corresponding SL 3-folds do not intersect. This will be important in constructing U(1)-invariant SL fibrations in Section 7.5.

In [**23**, §9–§10] we study singularities of solutions u, v of (3–4). We show that either $u(x, -y) \equiv u(x, y)$ and $v(x, -y) \equiv -v(x, y)$, so that u, v are singular all along the x-axis, or else the singular points of u, v in S° are all *isolated*, with a positive integer *multiplicity*, and one of two *types*. We also show that singularities exist with every multiplicity and type, and multiplicity n singularities occur in codimension n in the family of all U(1)-invariant SL 3-folds.

3.6. Examples of singular special Lagrangian 3-folds in $\mathbb{C}^3$. We shall now describe four families of SL 3-folds in $\mathbb{C}^3$, as examples of the material of Sections 3.1–3.4. They have been chosen to illustrate different kinds of singular behavior of SL 3-folds, and also to show how nonsingular SL 3-folds can converge to a singular SL 3-fold, to serve as a preparation for our discussion of singularities of SL m-folds in Section 6.

Our first example derives from Harvey and Lawson [**11**, §III.3.A], and is discussed in detail in [**17**, §3].

EXAMPLE 3.4. Define a subset L_0 in $\mathbb{C}^3$ by

$$L_0 = \bigl\{(re^{i\theta_1}, re^{i\theta_2}, re^{i\theta_3}) : r \geqslant 0, \quad \theta_1, \theta_2, \theta_3 \in \mathbb{R}, \quad \theta_1 + \theta_2 + \theta_3 = 0\bigr\}.$$

Then L_0 is a *special Lagrangian cone* on T^2. An alternative definition is

$$L_0 = \bigl\{(z_1,z_2,z_3) \in \mathbb{C}^3 : |z_1| = |z_2| = |z_3|,\ \mathrm{Im}(z_1z_2z_3) = 0,\ \mathrm{Re}(z_1z_2z_3) \geqslant 0\bigr\}.$$

Let $t > 0$, write $\mathcal{S}^1 = \bigl\{\mathrm{e}^{i\theta} : \theta \in \mathbb{R}\bigr\}$, and define a map $\phi_t : \mathcal{S}^1 \times \mathbb{C} \to \mathbb{C}^3$ by

$$\phi_t : (e^{i\theta}, z) \mapsto \bigl((|z|^2 + t^2)^{1/2}\mathrm{e}^{i\theta}, z, e^{-i\theta}\bar{z}\bigr).$$

Then ϕ_t is an *embedding*. Define $L_t = \mathrm{Image}\,\phi_t$. Then L_t is a nonsingular special Lagrangian 3-fold in $\mathbb{C}^3$ diffeomorphic to $\mathcal{S}^1 \times \mathbb{R}^2$. An equivalent definition is

$$L_t = \bigl\{(z_1,z_2,z_3) \in \mathbb{C}^3 : |z_1|^2 - t^2 = |z_2|^2 = |z_3|^2,\ \mathrm{Im}(z_1z_2z_3) = 0,\quad \mathrm{Re}(z_1z_2z_3) \geqslant 0\bigr\}.$$

As $t \to 0_+$, the nonsingular SL 3-fold L_t converges to the singular SL cone L_0. Note that L_t is *asymptotic* to L_0 at infinity, and that $L_t = t\,L_1$ for $t > 0$, so that the L_t for $t > 0$ are all homothetic to each other. Also, each L_t for $t \geqslant 0$ is invariant under the T^2 subgroup of SU(3) acting by

$$(z_1,z_2,z_3) \mapsto (\mathrm{e}^{i\theta_1}z_1, \mathrm{e}^{i\theta_2}z_2, \mathrm{e}^{i\theta_3}z_3) \text{ for } \theta_1,\theta_2,\theta_3 \in \mathbb{R} \text{ with } \theta_1+\theta_2+\theta_3 = 0,$$

and so fits into the framework of Section 3.1. By [**21**, Th. 5.1] the L_a may also be written in the form (3–3) for continuous $u, v : \mathbb{R}^2 \to \mathbb{R}$, as in Section 3.5.

Our second example is adapted from Harvey and Lawson [**11**, §III.3.B].

Example 3.5. For each $t > 0$, define

$$L_t = \bigl\{(\mathrm{e}^{i\theta}x_1, \mathrm{e}^{i\theta}x_2, \mathrm{e}^{i\theta}x_3) : x_j \in \mathbb{R},\ \theta \in (0,\pi/3),\ x_1^2+x_2^2+x_3^2 = t^2(\sin 3\theta)^{-2/3}\bigr\}.$$

Then L_t is a nonsingular embedded SL 3-fold in $\mathbb{C}^3$ diffeomorphic to $\mathcal{S}^2 \times \mathbb{R}$. As $t \to 0_+$ it converges to the singular union L_0 of the two SL 3-planes

$$\Pi_1 = \bigl\{(x_1,x_2,x_3) : x_j \in \mathbb{R}\bigr\} \quad \text{and} \quad \Pi_2 = \bigl\{(\mathrm{e}^{i\pi/3}x_1, \mathrm{e}^{i\pi/3}x_2, \mathrm{e}^{i\pi/3}x_3) : x_j \in \mathbb{R}\bigr\},$$

which intersect at 0. Note that L_t is invariant under the action of the Lie subgroup SO(3) of SU(3), acting on $\mathbb{C}^3$ in the obvious way, so again this comes from the method of Section 3.1. Also L_t is asymptotic to L_0 at infinity.

Our third example is taken from [**18**, Ex. 9.4 & Ex. 9.5].

Example 3.6. Let a_1, a_2 be positive, coprime integers, and set $a_3 = -a_1 - a_2$. Let $c \in \mathbb{R}$, and define

$$L_c^{a_1,a_2} = \bigl\{(\mathrm{e}^{ia_1\theta}x_1, \mathrm{e}^{ia_2\theta}x_2, i\mathrm{e}^{ia_3\theta}x_3) : \theta \in \mathbb{R},\ x_j \in \mathbb{R},\ a_1x_1^2 + a_2x_2^2 + a_3x_3^2 = c\bigr\}.$$

Then $L_c^{a_1,a_2}$ is an SL 3-fold, which comes from the "evolving quadrics" construction of Section 3.2. It is also symmetric under the U(1)-action

$$(z_1,z_2,z_3) \mapsto (\mathrm{e}^{ia_1\theta}z_1, \mathrm{e}^{ia_2\theta}z_2, i\mathrm{e}^{ia_3\theta}z_3) \quad \text{for } \theta \in \mathbb{R},$$

but this is not a necessary feature of the construction; these are just the easiest examples to write down.

When $c = 0$ and a_3 is odd, $L_0^{a_1,a_2}$ is an embedded special Lagrangian cone on T^2, with one singular point at 0. When $c = 0$ and a_3 is even, $L_0^{a_1,a_2}$ is two opposite embedded SL T^2-cones with one singular point at 0.

When $c > 0$ and a_3 is odd, $L_c^{a_1,a_2}$ is an embedded 3-fold diffeomorphic to a nontrivial real line bundle over the Klein bottle. When $c > 0$ and a_3 is even, $L_c^{a_1,a_2}$ is an embedded 3-fold diffeomorphic to $T^2 \times \mathbb{R}$. In both cases, $L_c^{a_1,a_2}$ is a *ruled*

SL 3-fold, as in Section 3.3, since it is fibered by hyperboloids of one sheet in $\mathbb{R}^3$, which are ruled in two different ways.

When $c<0$ and a_3 is odd, $L_c^{a_1,a_2}$ is an immersed copy of $\mathcal{S}^1\times\mathbb{R}^2$. When $c<0$ and a_3 is even, $L_c^{a_1,a_2}$ is two immersed copies of $\mathcal{S}^1\times\mathbb{R}^2$.

All the singular SL 3-folds we have seen so far have been *cones* in $\mathbb{C}^3$. Our final example, taken from [**16**], has more complicated singularities which are not cones. They are difficult to describe in a simple way, so we will not say much about them. For more details, see [**16**].

EXAMPLE 3.7. In [**16**, §5] the author constructed a family of maps $\Phi:\mathbb{R}^3\to\mathbb{C}^3$ with special Lagrangian image $N=\operatorname{Image}\Phi$. It is shown in [**16**, §6] that generic Φ in this family are immersions, so that N is nonsingular as an immersed SL 3-fold, but in codimension 1 in the family they develop isolated singularities.

Here is a rough description of these singularities, taken from [**16**, §6]. Taking the singular point to be at $\Phi(0,0,0)=0$, one can write Φ as

$$\begin{aligned}\Phi(x,y,t)=&\bigl(x+\tfrac{1}{4}g'(\mathbf{u},\mathbf{v})t^2\bigr)\,\mathbf{u}+\bigl(y^2-\tfrac{1}{4}|\mathbf{u}|^2t^2\bigr)\,\mathbf{v}\\&+2yt\,\mathbf{u}\times\mathbf{v}+O\bigl(x^2+|xy|+|xt|+|y|^3+|t|^3\bigr),\end{aligned}\tag{3–7}$$

where $\mathbf{u},\mathbf{v}$ are linearly independent vectors in $\mathbb{C}^3$ with $\omega'(\mathbf{u},\mathbf{v})=0$, and $\times:\mathbb{C}^3\times\mathbb{C}^3\to\mathbb{C}^3$ is defined by

$$(r_1,r_2,r_3)\times(s_1,s_2,s_3)=\tfrac{1}{2}(\bar r_2\bar s_3-\bar r_3\bar s_2,\bar r_3\bar s_1-\bar r_1\bar s_3,\bar r_1\bar s_2-\bar r_2\bar s_1).$$

The next few terms in the expansion (3–7) can also be given very explicitly, but we will not write them down as they are rather complex, and involve further choices of vectors $\mathbf{w},\mathbf{x},\ldots$.

What is going on here is that the lowest order terms in Φ are a *double cover* of the special Lagrangian plane $\langle\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}\rangle_{\mathbb{R}}$ in $\mathbb{C}^3$, *branched* along the real line $\langle\mathbf{u}\rangle_{\mathbb{R}}$. The branching occurs when $y=t=0$. Higher order terms deviate from the 3-plane $\langle\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}\rangle_{\mathbb{R}}$, and make the singularity isolated.

4. Almost Calabi–Yau Geometry

Calabi–Yau m-folds (M,J,ω,Ω) are compact complex m-folds (M,J) equipped with a Ricci-flat Kähler metric g with Kähler form ω, and a holomorphic $(m,0)$-form Ω of constant length $|\Omega|^2=2^m$. Then $\operatorname{Re}\Omega$ is a *calibration* on (M,g), and the corresponding calibrated submanifolds are called *special Lagrangian m-folds*. They are a natural generalization of the idea of special Lagrangian submanifolds in $\mathbb{C}^m$.

However, we will actually define and study special Lagrangian submanifolds in the much larger class of *almost Calabi–Yau manifolds* (M,J,ω,Ω), in which g is not required to be Ricci-flat, and Ω not required to have constant length. Apart from greater generality, this has the advantage that by restricting to a *generic* almost Calabi–Yau manifold one can (the author believes) much simplify the singular behavior of the special Lagrangian submanifolds within it.

The idea of extending special Lagrangian geometry to almost Calabi–Yau manifolds appears in the work of Goldstein [**4**, §3.1], Bryant [**1**, §1], who uses the term "special Kähler" instead of "almost Calabi–Yau", and the author [**25**].

4.1. Calabi–Yau and almost Calabi–Yau manifolds. Here is our definition of Calabi–Yau and almost Calabi–Yau manifolds.

DEFINITION 4.1. Let $m \geqslant 2$. An *almost Calabi–Yau m-fold*, or *ACY m-fold* for short, is a quadruple (M, J, ω, Ω) such that (M, J) is a compact m-dimensional complex manifold, ω is the Kähler form of a Kähler metric g on M, and Ω is a nonvanishing holomorphic $(m, 0)$-form on M.

We call (M, J, ω, Ω) a *Calabi–Yau m-fold*, or *CY m-fold* for short, if in addition ω and Ω satisfy

$$\omega^m/m! = (-1)^{m(m-1)/2}(i/2)^m \Omega \wedge \bar{\Omega}. \tag{4–1}$$

Furthermore, g is Ricci-flat and its holonomy group is a subgroup of $\mathrm{SU}(m)$.

This is not the usual definition of a Calabi–Yau manifold, but is essentially equivalent to it. Using Yau's proof of the Calabi Conjecture [**42**], [**14**, §5] one can prove:

THEOREM 4.2. *Let (M, J) be a compact complex manifold with trivial canonical bundle K_M, admitting Kähler metrics. Then in each Kähler class on M there is a unique Ricci-flat Kähler metric g, with Kähler form ω. Given such g and ω, there exists a holomorphic section Ω of K_M, unique up to change of phase $\Omega \mapsto \mathrm{e}^{i\theta}\Omega$, such that (M, J, ω, Ω) is a Calabi–Yau manifold.*

Thus, to find examples of Calabi–Yau manifolds all one needs is complex manifolds (M, J) satisfying certain essentially topological conditions. Using algebraic geometry one can construct very large numbers of such complex manifolds, particularly in complex dimension 3, and thus Calabi–Yau manifolds are very abundant. For a review of such constructions, and other general properties of Calabi–Yau manifolds, see [**14**, §6].

4.2. SL m-folds in almost Calabi–Yau m-folds. Next, we define *special Lagrangian m-folds* in almost Calabi–Yau m-folds.

DEFINITION 4.3. Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold, and N a real m-dimensional submanifold of M. We call N a *special Lagrangian submanifold*, or *SL m-fold* for short, if $\omega|_N \equiv \mathrm{Im}\,\Omega|_N \equiv 0$. It easily follows that $\mathrm{Re}\,\Omega|_N$ is a nonvanishing m-form on N. Thus N is orientable, with a unique orientation in which $\mathrm{Re}\,\Omega|_N$ is positive.

Let (M, J, ω, Ω) be a Calabi–Yau m-fold, with metric g. Then equation (4–1) ensures that for each $x \in M$ there exists an isomorphism $T_xM \cong \mathbb{C}^m$ that identifies g_x, ω_x and Ω_x with the flat versions g', ω', Ω' on $\mathbb{C}^m$ in (2–1). From Proposition 2.4 we then deduce:

PROPOSITION 4.4. *Let (M, J, ω, Ω) be a Calabi–Yau m-fold, with metric g, and N a real m-submanifold of M. Then N is special Lagrangian, with the natural orientation, if and only if it is calibrated with respect to* $\mathrm{Re}\,\Omega$.

Thus, in the Calabi–Yau case Definition 4.3 is equivalent to the conventional definition of special Lagrangian m-folds in Calabi–Yau m-folds, which is that they should be calibrated with respect to $\mathrm{Re}\,\Omega$, as in Definition 2.2. In the almost Calabi–Yau case, we can still interpret SL m-folds as calibrated submanifolds, but with respect to a conformally rescaled metric $\tilde{g}$. We explain how in the next proposition, which is easily proved using Proposition 2.4.

PROPOSITION 4.5. *Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold with metric g, define $f : M \to (0, \infty)$ by $f^{2m}\omega^m/m! = (-1)^{m(m-1)/2}(i/2)^m \Omega \wedge \bar{\Omega}$, and let $\tilde{g}$ be the conformally equivalent metric $f^2 g$ on M. Then $\operatorname{Re}\Omega$ is a calibration on the Riemannian manifold $(M, \tilde{g})$.*

A real m-submanifold N in M is special Lagrangian in (M, J, ω, Ω) if and only if it admits an orientation for which it is calibrated with respect to $\operatorname{Re}\Omega$ in $(M, \tilde{g})$. In particular, special Lagrangian m-folds in M are minimal in $(M, \tilde{g})$.

Thus, we can give an equivalent definition of SL m-folds in terms of calibrated geometry. Nonetheless, in the author's view the definition of SL m-folds in terms of the vanishing of closed forms is more fundamental than the definition in terms of calibrated geometry, at least in the almost Calabi–Yau case, and so should be taken as the primary definition.

One important reason for considering SL m-folds in almost Calabi–Yau rather than Calabi–Yau m-folds is that they have much stronger *genericness properties*. There are many situations in geometry in which one uses a genericity assumption to control singular behavior.

For instance, pseudo-holomorphic curves in an arbitrary almost complex manifold may have bad singularities, but the possible singularities in a generic almost complex manifold are much simpler. In the same way, it is reasonable to hope that in a *generic* Calabi–Yau m-fold, compact SL m-folds may have better singular behavior than in an arbitrary Calabi–Yau m-fold.

But because Calabi–Yau manifolds come in only finite-dimensional families, choosing a generic Calabi–Yau structure is a fairly weak assumption, and probably will not help very much. However, almost Calabi–Yau manifolds come in *infinite-dimensional* families, so choosing a generic almost Calabi–Yau structure is a much more powerful thing to do, and will probably simplify the singular behavior of compact SL m-folds considerably. We will return to this idea in Section 6.

5. Compact SL m-folds in ACY m-folds

In this section we shall discuss *compact* special Lagrangian submanifolds in almost Calabi–Yau manifolds. Here are three important questions which motivate work in this area.

1. Let N be a compact special Lagrangian m-fold in a fixed almost Calabi–Yau m-fold (M, J, ω, Ω). Let $\mathcal{M}_N$ be the moduli space of *special Lagrangian deformations* of N, that is, the connected component of the set of special Lagrangian m-folds containing N. What can we say about $\mathcal{M}_N$? For instance, is it a smooth manifold, and of what dimension?

2. Let $\{(M, J_t, \omega_t, \Omega_t) : t \in (-\epsilon, \epsilon)\}$ be a smooth 1-parameter family of almost Calabi–Yau m-folds. Suppose N_0 is an SL m-fold in $(M, J_0, \omega_0, \Omega_0)$. Under what conditions can we extend N_0 to a smooth family of special Lagrangian m-folds N_t in $(M, J_t, \omega_t, \Omega_t)$ for $t \in (-\epsilon, \epsilon)$?

3. In general the moduli space $\mathcal{M}_N$ in Question 1 will be noncompact. Can we enlarge $\mathcal{M}_N$ to a compact space $\overline{\mathcal{M}}_N$ by adding a "boundary" consisting of *singular* special Lagrangian m-folds? If so, what is the nature of the singularities that develop?

Briefly, these questions concern the *deformations* of special Lagrangian m-folds, *obstructions* to their existence, and their *singularities* respectively. The local answers to Questions 1 and 2 are well understood, and we shall discuss them in this section. Question 3 is the subject of Sections 6–7.

5.1. Deformations of compact special Lagrangian m-folds. The deformation theory of compact SL m-folds N was studied by McLean [**36**], who proved the following result in the Calabi–Yau case. Because McLean's proof only relies on the fact that $\omega|_N \equiv \mathrm{Im}\,\Omega|_N \equiv 0$, it also applies equally well to SL m-folds in almost Calabi–Yau m-folds.

THEOREM 5.1. *Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold, and N a compact special Lagrangian m-fold in M. Then the moduli space $\mathcal{M}_N$ of special Lagrangian deformations of N is a smooth manifold of dimension $b^1(N)$, the first Betti number of N.*

SKETCH OF PROOF. Suppose for simplicity that N is an embedded submanifold. There is a natural orthogonal decomposition $TM|_N = TN \oplus \nu$, where $\nu \to N$ is the *normal bundle* of N in M. As N is Lagrangian, the complex structure $J : TM \to TM$ gives an isomorphism $J : \nu \to TN$. But the metric g gives an isomorphism $TN \cong T^*N$. Composing these two gives an isomorphism $\nu \cong T^*N$.

Let T be a small *tubular neighborhood* of N in M. Then we can identify T with a neighborhood of the zero section in ν. Using the isomorphism $\nu \cong T^*N$, we have an identification between T and a neighborhood of the zero section in T^*N. This can be chosen to identify the Kähler form ω on T with the natural symplectic structure on T^*N. Let $\pi : T \to N$ be the obvious projection.

Under this identification, submanifolds N' in $T \subset M$ which are C^1 close to N are identified with the graphs of small smooth sections α of T^*N. That is, submanifolds N' of M close to N are identified with 1-*forms* α on N. We need to know: which 1-forms α are identified with *special Lagrangian* submanifolds N'?

Well, N' is special Lagrangian if $\omega|_{N'} \equiv \mathrm{Im}\,\Omega|_{N'} \equiv 0$. Now $\pi|_{N'} : N' \to N$ is a diffeomorphism, so we can push $\omega|_{N'}$ and $\mathrm{Im}\,\Omega|_{N'}$ down to N, and regard them as functions of α. Calculation shows that

$$\pi_*\bigl(\omega|_{N'}\bigr) = \mathrm{d}\alpha \quad \text{and} \quad \pi_*\bigl(\mathrm{Im}\,\Omega|_{N'}\bigr) = F(\alpha, \nabla\alpha),$$

where F is a nonlinear function of its arguments. Thus, the moduli space $\mathcal{M}_N$ is locally isomorphic to the set of small 1-forms α on N such that $\mathrm{d}\alpha \equiv 0$ and $F(\alpha, \nabla\alpha) \equiv 0$.

Now it turns out that F satisfies $F(\alpha, \nabla\alpha) \approx \mathrm{d}(*\alpha)$ when α is small. Therefore $\mathcal{M}_N$ is locally approximately isomorphic to the vector space of 1-forms α with $\mathrm{d}\alpha = \mathrm{d}(*\alpha) = 0$. But by Hodge theory, this is isomorphic to the de Rham cohomology group $H^1(N, \mathbb{R})$, and is a manifold with dimension $b^1(N)$.

To carry out this last step rigorously requires some technical machinery: one must work with certain *Banach spaces* of sections of T^*N, $\Lambda^2 T^*N$ and $\Lambda^m T^*N$, use *elliptic regularity results* to prove that the map $\alpha \mapsto \bigl(\mathrm{d}\alpha, F(\alpha, \nabla\alpha)\bigr)$ has *closed image* in these Banach spaces, and then use the *Implicit Function Theorem for Banach spaces* to show that the kernel of the map is what we expect. □

5.2. Obstructions to the existence of compact SL m-folds. Next we address Question 2 above. First, observe that if (M, J, ω, Ω) is an almost Calabi–Yau m-fold and N a compact SL m-fold in M, then $\omega|_N \equiv \mathrm{Im}\,\Omega|_N \equiv 0$, and thus

$[\omega|_N]$ and $[\operatorname{Im}\Omega|_N]$ are zero in $H^2(N,\mathbb{R})$ and $H^m(N,\mathbb{R})$. But $[\omega|_N]$ and $[\operatorname{Im}\Omega|_N]$ are unchanged under continuous variations of N in M. So we deduce:

LEMMA 5.2. *Let (M,J,ω,Ω) be an almost Calabi–Yau m-fold, and N a compact real m-submanifold in M. Then a necessary condition for N to be isotopic to a special Lagrangian submanifold N' in M is that $[\omega|_N]=0$ in $H^2(N,\mathbb{R})$ and $[\operatorname{Im}\Omega|_N]=0$ in $H^m(N,\mathbb{R})$.*

This gives a simple, necessary topological condition for an isotopy class of m-submanifolds in an almost Calabi–Yau m-fold to contain an SL m-fold. Our next result, following from Marshall [**34**, Th. 3.2.9], shows that locally, this is also a *sufficient* condition for an SL m-fold to persist under deformations of the almost Calabi–Yau structure.

THEOREM 5.3. *Let $\bigl\{(M,J_t,\omega_t,\Omega_t):t\in(-\epsilon,\epsilon)\bigr\}$ be a smooth family of almost Calabi–Yau m-folds. Let N_0 be a compact SL m-fold in $(M,J_0,\omega_0,\Omega_0)$, and suppose that $[\omega_t|_{N_0}]=0$ in $H^2(N_0,\mathbb{R})$ and $[\operatorname{Im}\Omega_t|_{N_0}]=0$ in $H^m(N_0,\mathbb{R})$ for all $t\in(-\epsilon,\epsilon)$. Then N_0 extends to a smooth family $\bigl\{N_t:t\in(-\delta,\delta)\bigr\}$, where $0<\delta\leqslant\epsilon$ and N_t is a compact SL m-fold in $(M,J_t,\omega_t,\Omega_t)$.*

This is proved using similar techniques to Theorem 5.1, though McLean did not prove it. Note that the condition $[\operatorname{Im}\Omega_t|_{N_0}]=0$ for all t can be satisfied by choosing the phases of the Ω_t appropriately, and if the image of $H_2(N,\mathbb{Z})$ in $H_2(M,\mathbb{R})$ is zero, then the condition $[\omega|_N]=0$ holds automatically.

Thus, the obstructions $[\omega_t|_{N_0}]=[\operatorname{Im}\Omega_t|_{N_0}]=0$ in Theorem 5.3 are actually fairly mild restrictions, and special Lagrangian m-folds should be thought of as pretty stable under small deformations of the almost Calabi–Yau structure.

REMARK. The deformation and obstruction theory of compact special Lagrangian m-folds are *extremely well-behaved* compared to many other moduli space problems in differential geometry. In other geometric problems (such as the deformations of complex structures on a complex manifold, or pseudo-holomorphic curves in an almost complex manifold, or instantons on a Riemannian 4-manifold, and so on), the deformation theory often has the following general structure.

There are vector bundles E,F over a compact manifold M, and an elliptic operator $P:C^\infty(E)\to C^\infty(F)$, usually first-order. The kernel $\operatorname{Ker}P$ is the set of *infinitesimal deformations*, and the cokernel $\operatorname{Coker}P$ the set of *obstructions*. The actual moduli space $\mathcal{M}$ is locally the zeros of a nonlinear map $\Psi:\operatorname{Ker}P\to\operatorname{Coker}P$.

In a *generic* case, $\operatorname{Coker}P=0$, and then the moduli space $\mathcal{M}$ is locally isomorphic to $\operatorname{Ker}P$, and so is locally a manifold with dimension $\operatorname{ind}(P)$. However, in nongeneric situations $\operatorname{Coker}P$ may be nonzero, and then the moduli space $\mathcal{M}$ may be nonsingular, or have an unexpected dimension.

However, SL m-folds do not follow this pattern. Instead, the obstructions are *topologically determined*, and the moduli space is *always* smooth, with dimension given by a topological formula. This should be regarded as a minor mathematical miracle.

5.3. Natural coordinates on the moduli space $\mathcal{M}_N$. Let N be a compact SL m-fold in an almost Calabi–Yau m-fold (M,J,ω,Ω). Theorem 5.1 shows that the moduli space $\mathcal{M}_N$ has dimension $b^1(N)$. By Poincaré duality $b^1(N)=b^{m-1}(N)$. Thus $\mathcal{M}_N$ has the same dimension as the de Rham cohomology groups $H^1(M,\mathbb{R})$ and $H^{m-1}(M,\mathbb{R})$.

We shall construct natural local diffeomorphisms Φ from $\mathcal{M}_N$ to $H^1(N,\mathbb{R})$, and Ψ from $\mathcal{M}_N$ to $H^{m-1}(N,\mathbb{R})$. These induce two natural *affine structures* on $\mathcal{M}_N$, and can be thought of as two *natural coordinate systems* on $\mathcal{M}_N$. The material of this section can be found in Hitchin [**13**, §4].

Here is how to define Φ and Ψ. Let U be a connected and simply-connected open neighborhood of N in $\mathcal{M}_N$. We will construct smooth maps $\Phi : U \to H^1(N,\mathbb{R})$ and $\Psi : U \to H^{m-1}(N,\mathbb{R})$ with $\Phi(N) = \Psi(N) = 0$, which are local diffeomorphisms.

Let $N' \in U$. Then as U is connected, there exists a smooth path $\gamma : [0,1] \to U$ with $\gamma(0) = N$ and $\gamma(1) = N'$, and as U is simply-connected, γ is unique up to isotopy. Now γ parametrizes a family of submanifolds of M diffeomorphic to N, which we can lift to a smooth map $\Gamma : N \times [0,1] \to M$ with $\Gamma(N \times \{t\}) = \gamma(t)$.

Consider the 2-form $\Gamma^*(\omega)$ on $N \times [0,1]$. As each fiber $\gamma(t)$ is Lagrangian, we have $\Gamma^*(\omega)|_{N\times\{t\}} \equiv 0$ for each $t \in [0,1]$. Therefore we may write $\Gamma^*(\omega) = \alpha_t \wedge \mathrm{d}t$, where α_t is a closed 1-form on N for $t \in [0,1]$. Define

$$\Phi(N') = \textstyle\left[\int_0^1 \alpha_t \,\mathrm{d}t\right] \in H^1(N,\mathbb{R}).$$

That is, we integrate the 1-forms α_t with respect to t to get a closed 1-form $\int_0^1 \alpha_t \,\mathrm{d}t$, and then take its cohomology class.

Similarly, write $\Gamma^*(\operatorname{Im}\Omega) = \beta_t \wedge \mathrm{d}t$, where β_t is a closed $(m-1)$-form on N for $t \in [0,1]$, and define $\Psi(N') = \left[\int_0^1 \beta_t \,\mathrm{d}t\right] \in H^{m-1}(N,\mathbb{R})$. Then Φ and Ψ are independent of choices made in the construction (exercise). We need to restrict to a simply-connected subset U of $\mathcal{M}_N$ so that γ is unique up to isotopy. Alternatively, one can define Φ and Ψ on the universal cover $\widetilde{\mathcal{M}}_N$ of $\mathcal{M}_N$.

6. Singularities of Special Lagrangian m-folds

Now we move on to Question 3 of Section 5, and discuss the *singularities* of special Lagrangian m-folds. We can divide it into two sub-questions:

3(a) What kinds of singularities are possible in singular special Lagrangian m-folds, and what do they look like?

3(b) How can singular SL m-folds arise as limits of nonsingular SL m-folds, and what does the limiting behavior look like near the singularities?

These questions are addressed in the author's series of papers [**26, 27, 28, 29, 30**] for *isolated conical singularities* of SL m-folds, that is, singularities locally modelled on an SL cone C in $\mathbb{C}^m$ with an isolated singularity at 0. We now explain the principal results. Readers of the series are advised to begin with the final paper [**30**], which surveys the others.

6.1. Special Lagrangian cones.

We define *SL cones*, and some notation.

Definition 6.1. A (singular) SL m-fold C in $\mathbb{C}^m$ is called a *cone* if $C = tC$ for all $t > 0$, where $tC = \{t\mathbf{x} : \mathbf{x} \in C\}$. Let C be a closed SL cone in $\mathbb{C}^m$ with an isolated singularity at 0. Then $\Sigma = C \cap \mathcal{S}^{2m-1}$ is a compact, nonsingular $(m-1)$-submanifold of $\mathcal{S}^{2m-1}$, not necessarily connected. Let g_Σ be the restriction of g' to Σ, where g' is as in (2–1).

Set $C' = C \setminus \{0\}$. Define $\iota : \Sigma \times (0,\infty) \to \mathbb{C}^m$ by $\iota(\sigma, r) = r\sigma$. Then ι has image C'. By an abuse of notation, *identify* C' with $\Sigma \times (0,\infty)$ using ι. The *cone metric* on $C' \cong \Sigma \times (0,\infty)$ is $g' = \iota^*(g') = \mathrm{d}r^2 + r^2 g_\Sigma$. For $\alpha \in \mathbb{R}$, we say that a function $u : C' \to \mathbb{R}$ is *homogeneous of order* α if $u \circ t \equiv t^\alpha u$ for all $t > 0$. Equivalently, u is homogeneous of order α if $u(\sigma, r) \equiv r^\alpha v(\sigma)$ for some function $v : \Sigma \to \mathbb{R}$.

In [**26**, Lem. 2.3] we study *homogeneous harmonic functions* on C'.

LEMMA 6.2. *In the situation of Definition 6.1, let $u(\sigma,r)\equiv r^\alpha v(\sigma)$ be a homogeneous function of order α on $C'=\Sigma\times(0,\infty)$, for $v\in C^2(\Sigma)$. Then*

$$\Delta u(\sigma,r)=r^{\alpha-2}\big(\Delta_\Sigma v-\alpha(\alpha+m-2)v\big),$$

where Δ, Δ_Σ are the Laplacians on (C',g') and (Σ,g_Σ). Hence, u is harmonic on C' if and only if v is an eigenfunction of Δ_Σ with eigenvalue $\alpha(\alpha+m-2)$.

Following [**26**, Def. 2.5], we define:

DEFINITION 6.3. In the situation of Definition 6.1, suppose $m>2$ and define

$$\mathcal{D}_\Sigma=\big\{\alpha\in\mathbb{R}:\alpha(\alpha+m-2)\text{ is an eigenvalue of }\Delta_\Sigma\big\}. \tag{6–1}$$

Then $\mathcal{D}_\Sigma$ is a countable, discrete subset of $\mathbb{R}$. By Lemma 6.2, an equivalent definition is that $\mathcal{D}_\Sigma$ is the set of $\alpha\in\mathbb{R}$ for which there exists a nonzero homogeneous harmonic function u of order α on C'.

Define $m_\Sigma:\mathcal{D}_\Sigma\to\mathbb{N}$ by taking $m_\Sigma(\alpha)$ to be the multiplicity of the eigenvalue $\alpha(\alpha+m-2)$ of Δ_Σ, or equivalently the dimension of the vector space of homogeneous harmonic functions u of order α on C'. Define $N_\Sigma:\mathbb{R}\to\mathbb{Z}$ by

$$N_\Sigma(\delta)=-\sum_{\alpha\in\mathcal{D}_\Sigma\cap(\delta,0)}m_\Sigma(\alpha)\ \rangle\text{if }\delta<0,\text{ and }\rangle N_\Sigma(\delta)=\sum_{\alpha\in\mathcal{D}_\Sigma\cap[0,\delta]}m_\Sigma(\alpha)\ \rangle\text{if }\delta\geqslant 0. \tag{6–2}$$

Then N_Σ is monotone increasing and upper semicontinuous, and is discontinuous exactly on $\mathcal{D}_\Sigma$, increasing by $m_\Sigma(\alpha)$ at each $\alpha\in\mathcal{D}_\Sigma$. As the eigenvalues of Δ_Σ are nonnegative, we see that $\mathcal{D}_\Sigma\cap(2-m,0)=\varnothing$ and $N_\Sigma\equiv 0$ on $(2-m,0)$.

We define the *stability index* of C, and *stable* cones [**27**, Def. 3.6].

DEFINITION 6.4. Let C be an SL cone in $\mathbb{C}^m$ for $m>2$ with an isolated singularity at 0, let G be the Lie subgroup of $\mathrm{SU}(m)$ preserving C, and use the notation of Definitions 6.1 and 6.3. Then [**27**, eq. (8)] shows that

$$m_\Sigma(0)=b^0(\Sigma),\quad m_\Sigma(1)\geqslant 2m\quad\text{and}\quad m_\Sigma(2)\geqslant m^2-1-\dim G. \tag{6–3}$$

Define the *stability index* $\text{s-ind}(C)$ to be

$$\text{s-ind}(C)=N_\Sigma(2)-b^0(\Sigma)-m^2-2m+1+\dim G. \tag{6–4}$$

Then $\text{s-ind}(C)\geqslant 0$ by (6–3), as $N_\Sigma(2)\geqslant m_\Sigma(0)+m_\Sigma(1)+m_\Sigma(2)$ by (6–2). We call C *stable* if $\text{s-ind}(C)=0$.

Here is the point of these definitions. By the Principle in Section 2.3, homogeneous harmonic functions v of order α on C' correspond to infinitesimal deformations $\mathrm{d}v$ of C' as an SL m-fold in $\mathbb{C}^m$, which grow like $O(r^{\alpha-1})$. Hence, $N_\Sigma(\lambda)$ is effectively the dimension of a space of *infinitesimal deformations* of C' as an SL m-fold, which grow like $O(r^{\alpha-1})$ for $\alpha\in[0,\lambda]$, when $\lambda\geqslant 0$.

For $\lambda=2$ this space of harmonic functions, or infinitesimal deformations of C', contains some from obvious geometrical sources: locally constant functions on C', and infinitesimal deformations of C' from translations in $\mathbb{C}^m$, or $\mathfrak{su}(m)$ rotations. We get $\text{s-ind}(C)$ by subtracting these obvious geometrical deformations from $N_\Sigma(2)$. Hence, $\text{s-ind}(C)$ is the dimension of a space of *excess infinitesimal deformations* of C' as an SL m-fold, with growth between $O(r^{-1})$ and $O(r)$, which do not arise from infinitesimal automorphisms of $\mathbb{C}^m$.

We shall see in Section 6.2 that s-ind(C) is the dimension of an *obstruction space* to deforming an SL m-fold X with a conical singularity with cone C, and that if C is *stable*, the deformation theory of X simplifies.

6.2. Special Lagrangian m-folds with conical singularities. Now we can define *conical singularities* of SL m-folds, following [**26**, Def. 3.6].

DEFINITION 6.5. Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold for $m > 2$. Suppose X is a compact singular SL m-fold in M with singularities at distinct points $x_1, \ldots, x_n \in X$, and no other singularities.

Fix isomorphisms $\upsilon_i : \mathbb{C}^m \to T_{x_i}M$ for $i = 1, \ldots, n$ such that $\upsilon_i^*(\omega) = \omega'$ and $\upsilon_i^*(\Omega) = a_i\Omega'$, where ω', Ω' are as in (2–1) and $a_1, \ldots, a_n > 0$. Let $C_1, \ldots, C_n$ be SL cones in $\mathbb{C}^m$ with isolated singularities at 0. For $i = 1, \ldots, n$ let $\Sigma_i = C_i \cap \mathcal{S}^{2m-1}$, and let $\mu_i \in (2, 3)$ with

$$(2, \mu_i] \cap \mathcal{D}_{\Sigma_i} = \varnothing, \quad \text{where } \mathcal{D}_{\Sigma_i} \text{ is defined in (6–1).} \tag{6–5}$$

Then we say that X has a *conical singularity* or *conical singular point* at x_i, with *rate* μ_i and *cone* C_i for $i = 1, \ldots, n$, if the following holds.

By Darboux's Theorem there exist embeddings $\Upsilon_i : B_R \to M$ for $i = 1, \ldots, n$ satisfying $\Upsilon_i(0) = x_i$, $\mathrm{d}\Upsilon_i|_0 = \upsilon_i$ and $\Upsilon_i^*(\omega) = \omega'$, where B_R is the open ball of radius R about 0 in $\mathbb{C}^m$ for some small $R > 0$. Define $\iota_i : \Sigma_i \times (0, R) \to B_R$ by $\iota_i(\sigma, r) = r\sigma$ for $i = 1, \ldots, n$.

Define $X' = X \setminus \{x_1, \ldots, x_n\}$. Then there should exist a compact subset $K \subset X'$ such that $X' \setminus K$ is a union of open sets $S_1, \ldots, S_n$ with $S_i \subset \Upsilon_i(B_R)$, whose closures $\bar{S}_1, \ldots, \bar{S}_n$ are disjoint in X. For $i = 1, \ldots, n$ and some $R' \in (0, R]$ there should exist a smooth $\phi_i : \Sigma_i \times (0, R') \to B_R$ such that $\Upsilon_i \circ \phi_i : \Sigma_i \times (0, R') \to M$ is a diffeomorphism $\Sigma_i \times (0, R') \to S_i$, and

$$\bigl|\nabla^k(\phi_i - \iota_i)\bigr| = O(r^{\mu_i - 1 - k}) \quad \text{as } r \to 0 \text{ for } k = 0, 1. \tag{6–6}$$

Here ∇ is the Levi-Civita connection of the cone metric $\iota_i^*(g')$ on $\Sigma_i \times (0, R')$, $|\,.\,|$ is computed using $\iota_i^*(g')$. If the cones $C_1, \ldots, C_n$ are *stable* in the sense of Definition 6.4, then we say that X has *stable conical singularities*.

We show in [**26**, Th.s 4.4 & 5.5] that if (6–6) holds for $k = 0, 1$ and some μ_i satisfying (6–5), then we can choose a natural ϕ_i for which (6–6) holds for *all* $k \geqslant 0$, and for *all* rates μ_i satisfying (6–5). Thus the number of derivatives required in (6–6) and the choice of μ_i both make little difference. We choose $k = 0, 1$ in (6–6), and some μ_i in (6–5), to make the definition as weak as possible.

Suppose we did not require (6–5), and that $\alpha \in (2, \mu_i) \cap \mathcal{D}_{\Sigma_i}$. Then there exists a homogeneous harmonic function v on C_i' of order α. By the Principle in Section 2.3, dv yields an infinitesimal deformation of C_i' as an SL m-fold in $\mathbb{C}^m$, growing like $O(r^{\alpha-1})$. Locally this gives a way to deform X into an SL m-fold which would *not* satisfy (6–6), as $\alpha < \mu_i$. Effectively, v acts as an *obstruction* to deforming X through SL m-folds satisfying Definition 6.5. So the point of (6–5) is to reduce to a minimum the obstructions to existence and deformation of SL m-folds with isolated conical singularities.

In [**27**] we study the *deformation theory* of compact SL m-folds with conical singularities, generalizing Theorem 5.1 in the nonsingular case. Following [**27**, Def. 5.4], we define the space $\mathcal{M}_X$ of compact SL m-folds $\hat{X}$ in M with conical singularities deforming a fixed SL m-fold X with conical singularities.

DEFINITION 6.6. Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold and X a compact SL m-fold in M with conical singularities at $x_1, \ldots, x_n$ with identifications $\upsilon_i : \mathbb{C}^m \to T_{x_i}M$ and cones $C_1, \ldots, C_n$. Define the *moduli space* $\mathcal{M}_X$ *of deformations of* X to be the set of $\hat{X}$ such that

(i) $\hat{X}$ is a compact SL m-fold in M with conical singularities at $\hat{x}_1, \ldots, \hat{x}_n$ with cones $C_1, \ldots, C_n$, for some $\hat{x}_i$ and identifications $\hat{\upsilon}_i : \mathbb{C}^m \to T_{\hat{x}_i}M$.

(ii) There exists a homeomorphism $\hat{\imath} : X \to \hat{X}$ with $\hat{\imath}(x_i) = \hat{x}_i$ for $i = 1, \ldots, n$ such that $\hat{\imath}|_{X'} : X' \to \hat{X}'$ is a diffeomorphism and $\hat{\imath}$ and $\imath$ are isotopic as continuous maps $X \to M$, where $\imath : X \to M$ is the inclusion.

In [**27**, Def. 5.6] we define a *topology* on $\mathcal{M}_X$, and explain why it is the natural choice. We will not repeat the complicated definition here. In [**27**, Th. 6.10] we describe $\mathcal{M}_X$ near X, in terms of a smooth map Φ between the *infinitesimal deformation space* $\mathcal{I}_{X'}$ and the *obstruction space* $\mathcal{O}_{X'}$.

THEOREM 6.7. *Suppose (M, J, ω, Ω) is an almost Calabi–Yau m-fold and X a compact SL m-fold in M with conical singularities at $x_1, \ldots, x_n$ and cones $C_1, \ldots, C_n$. Let $\mathcal{M}_X$ be the moduli space of deformations of X as an SL m-fold with conical singularities in M, as in Definition 6.6. Set $X' = X \setminus \{x_1, \ldots, x_n\}$.*

Then there exist natural finite-dimensional vector spaces $\mathcal{I}_{X'}$, $\mathcal{O}_{X'}$ such that $\mathcal{I}_{X'}$ is the image of $H^1_{\rm cs}(X', \mathbb{R})$ in $H^1(X', \mathbb{R})$ and $\dim \mathcal{O}_{X'} = \sum_{i=1}^n$ s-ind(C_i), where s-ind(C_i) is the stability index of Definition 6.4. There exists an open neighbourhood U of 0 in $\mathcal{I}_{X'}$, a smooth map $\Phi : U \to \mathcal{O}_{X'}$ with $\Phi(0) = 0$, and a map $\Xi : \{u \in U : \Phi(u) = 0\} \to \mathcal{M}_X$ with $\Xi(0) = X$ which is a homeomorphism with an open neighbourhood of X in $\mathcal{M}_X$.

If the C_i are *stable*, then $\mathcal{O}_{X'} = \{0\}$ and we deduce [**27**, Cor. 6.11]:

COROLLARY 6.8. *Suppose (M, J, ω, Ω) is an almost Calabi–Yau m-fold and X a compact SL m-fold in M with stable conical singularities, and let $\mathcal{M}_X$ and $\mathcal{I}_{X'}$ be as in Theorem 6.7. Then $\mathcal{M}_X$ is a smooth manifold of dimension* $\dim \mathcal{I}_{X'}$.

This has clear similarities with Theorem 5.1. Here is another simple condition for $\mathcal{M}_X$ to be a manifold near X, [**27**, Def. 6.12].

DEFINITION 6.9. Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold and X a compact SL m-fold in M with conical singularities, and let $\mathcal{I}_{X'}, \mathcal{O}_{X'}, U$ and Φ be as in Theorem 6.7. We call X *transverse* if the linear map $\mathrm{d}\Phi|_0 : \mathcal{I}_{X'} \to \mathcal{O}_{X'}$ is surjective.

If X is transverse, then $\{u \in U : \Phi(u) = 0\}$ is a manifold near 0, so Theorem 6.7 yields [**27**, Cor. 6.13]:

COROLLARY 6.10. *Suppose (M, J, ω, Ω) is an almost Calabi–Yau m-fold and X a transverse compact SL m-fold in M with conical singularities, and let $\mathcal{M}_X, \mathcal{I}_{X'}$ and $\mathcal{O}_{X'}$ be as in Theorem 6.7. Then $\mathcal{M}_X$ is, near X, a smooth manifold of dimension* $\dim \mathcal{I}_{X'} - \dim \mathcal{O}_{X'}$.

We would like to conclude that by choosing a sufficiently generic perturbation ω^s we can make $\mathcal{M}^s_X$ smooth everywhere. This is the idea of the following conjecture, [**27**, Conj. 9.5]:

CONJECTURE 6.11. *Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold, X a compact SL m-fold in M with conical singularities, and $\mathcal{I}_{X'}, \mathcal{O}_{X'}$ be as in Theorem*

6.7. Then for a second category subset of Kähler forms $\hat{\omega}$ in the Kähler class of ω, the moduli space $\hat{\mathcal{M}}_X$ of compact SL m-folds $\hat{X}$ with conical singularities in $(M,J,\hat{\omega},\Omega)$ isotopic to X consists of transverse $\hat{X}$, and so is a smooth manifold of dimension $\dim \mathcal{I}_{X'} - \dim \mathcal{O}_{X'}$.

For a partial proof of this, see [**27**, Th.s 9.1 & 9.3]. Basically, we can prove the conjecture for $\hat{\omega}$ close to ω and $\hat{X} \in \hat{\mathcal{M}}_X$ close to X, or more generally, close to a fixed compact subset of the moduli space $\mathcal{M}_X$ in (M,J,ω,Ω).

6.3. Asymptotically conical SL m-folds. The local models for how to *desingularize* compact SL m-folds with isolated conical singularities are *Asymptotically Conical* SL m-folds L in $\mathbb{C}^m$, so we discuss these briefly. Here is the definition, [**26**, Def. 7.1].

Definition 6.12. Let C be a closed SL cone in $\mathbb{C}^m$ with isolated singularity at 0 for $m>2$, and let $\Sigma = C \cap \mathcal{S}^{2m-1}$, so that Σ is a compact, nonsingular $(m-1)$-manifold, not necessarily connected. Let g_Σ be the metric on Σ induced by the metric g' on $\mathbb{C}^m$ in (2–1), and r the radius function on $\mathbb{C}^m$. Define $\iota : \Sigma \times (0,\infty) \rightarrow \mathbb{C}^m$ by $\iota(\sigma,r) = r\sigma$. Then the image of ι is $C \setminus \{0\}$, and $\iota^*(g') = r^2 g_\Sigma + \mathrm{d}r^2$ is the cone metric on $C \setminus \{0\}$.

Let L be a closed, nonsingular SL m-fold in $\mathbb{C}^m$. We call L *Asymptotically Conical (AC)* with *rate* $\lambda < 2$ and *cone* C if there exists a compact subset $K \subset L$ and a diffeomorphism $\varphi : \Sigma \times (T,\infty) \rightarrow L \setminus K$ for some $T>0$, such that

$$\bigl|\nabla^k(\varphi - \iota)\bigr| = O(r^{\lambda-1-k}) \quad \text{as } r \rightarrow \infty \text{ for } k = 0, 1.$$

Here $\nabla, |\,.\,|$ are computed using the cone metric $\iota^*(g')$.

The deformation theory of Asymptotically Conical SL m-folds in $\mathbb{C}^m$ has been studied independently by Pacini [**37**] and Marshall [**34**]. Pacini's results are earlier, but Marshall's are more complete.

Definition 6.13. Suppose L is an Asymptotically Conical SL m-fold in $\mathbb{C}^m$ with cone C and rate $\lambda < 2$, as in Definition 6.12. Define the *moduli space* $\mathcal{M}_L^\lambda$ *of deformations of* L *with rate* λ to be the set of AC SL m-folds $\hat{L}$ in $\mathbb{C}^m$ with cone C and rate λ, such that $\hat{L}$ is diffeomorphic to L and isotopic to L as an Asymptotically Conical submanifold of $\mathbb{C}^m$. One can define a natural *topology* on $\mathcal{M}_L^\lambda$.

The following result can be deduced from Marshall [**34**, Th. 6.2.15] and [**34**, Table 5.1]. (See also Pacini [**37**, Th. 2 & Th. 3].)

Theorem 6.14. *Let L be an Asymptotically Conical SL m-fold in $\mathbb{C}^m$ with cone C and rate $\lambda < 2$, and let $\mathcal{M}_L^\lambda$ be as in Definition 6.13. Set $\Sigma = C \cap \mathcal{S}^{2m-1}$, and let $\mathcal{D}_\Sigma, N_\Sigma$ be as in Section 6.1 and $b^k(L), b^k_{\mathrm{cs}}(L)$ be the Betti numbers in ordinary and compactly-supported de Rham cohomology $H^k(L,\mathbb{R})$, $H^k_{\mathrm{cs}}(L,\mathbb{R})$. Then*

(a) *If $\lambda \in (0,2) \setminus \mathcal{D}_\Sigma$, then $\mathcal{M}_L^\lambda$ is a manifold with*

$$\dim \mathcal{M}_L^\lambda = b^1(L) - b^0(L) + N_\Sigma(\lambda). \tag{6–7}$$

Note that if $0 < \lambda < \min\bigl(\mathcal{D}_\Sigma \cap (0,\infty)\bigr)$, then $N_\Sigma(\lambda) = b^0(\Sigma)$.

(b) *If $\lambda \in (2-m,0)$, then $\mathcal{M}_L^\lambda$ is a manifold of dimension $b^1_{\mathrm{cs}}(L) = b^{m-1}(L)$.*

This is the analogue of Theorems 5.1 and 6.7 for AC SL m-folds. If $\lambda \in (2-m,2) \setminus \mathcal{D}_\Sigma$, the deformation theory for L with rate λ is *unobstructed* and $\mathcal{M}_L^\lambda$ is a *smooth manifold* with a given dimension.

6.4. Desingularizing singular SL m-folds. Suppose (M, J, ω, Ω) is an almost Calabi–Yau m-fold, and X a compact SL m-fold in M with conical singularities at $x_1, \ldots, x_n$ and cones $C_1, \ldots, C_n$. In [**28, 29**] we study *desingularizations* of X, realizing X as a limit of a family of compact, nonsingular SL m-folds $\tilde{N}^t$ in M for small $t > 0$.

Here is the basic method. Let $L_1, \ldots, L_n$ be *Asymptotically Conical* SL m-folds in $\mathbb{C}^m$, as in Section 6.3, with L_i asymptotic to the cone C_i at infinity. We shrink L_i by a small factor $t > 0$, and glue tL_i into X at x_i for $i = 1, \ldots, n$ to get a 1-parameter family of compact, nonsingular *Lagrangian* m-folds N^t in (M, ω) for small $t > 0$.

Then we show using analysis that when t is sufficiently small we can deform N^t to a compact, nonsingular *special* Lagrangian m-fold $\tilde{N}^t$ via a small Hamiltonian deformation. This $\tilde{N}^t$ depends smoothly on t, and as $t \to 0$ it converges to the singular SL m-fold X, in the sense of currents.

Our simplest desingularization result is [**28**, Th. 6.13].

THEOREM 6.15. *Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold and X a compact SL m-fold in M with conical singularities at $x_1, \ldots, x_n$ and cones $C_1, \ldots, C_n$. Let $L_1, \ldots, L_n$ be Asymptotically Conical SL m-folds in $\mathbb{C}^m$ with cones $C_1, \ldots, C_n$ and rates $\lambda_1, \ldots, \lambda_n$. Suppose $\lambda_i < 0$ for $i = 1, \ldots, n$, and $X' = X \setminus \{x_1, \ldots, x_n\}$ is connected.*

Then there exists $\epsilon > 0$ and a smooth family $\{\tilde{N}^t : t \in (0, \epsilon]\}$ of compact, nonsingular SL m-folds in (M, J, ω, Ω), such that $\tilde{N}^t$ is constructed by gluing tL_i into X at x_i for $i = 1, \ldots, n$. In the sense of currents, $\tilde{N}^t \to X$ as $t \to 0$.

The theorem contains two *simplifying assumptions*: that $\lambda_i < 0$ for all i, and that X' is connected. These avoid two kinds of *obstructions* to desingularizing X using the L_i. For the first, the L_i have *cohomological invariants* $Y(L_i)$ in $H^1(\Sigma_i, \mathbb{R})$ derived from the relative cohomology class of ω'. If $\lambda_i < 0$, then $Y(L_i) = 0$. But if $\lambda_i \geqslant 0$ and $Y(L_i) \neq 0$, then there are obstructions to the existence of N^t as a *Lagrangian* m-fold. That is, we can only define N^t if the $Y(L_i)$ satisfy an equation.

For the second, if X' is not connected, then there is an analytic obstruction to deforming N^t to $\tilde{N}^t$, because the Laplacian Δ on functions on N^t has *small eigenvalues*. Again, the L_i have cohomological invariants $Z(L_i)$ in $H^{m-1}(\Sigma_i, \mathbb{R})$ derived from the relative cohomology class of $\operatorname{Im}\Omega'$, and we can only deform N^t to $\tilde{N}^t$ if the $Z(L_i)$ satisfy an equation.

In the obstructed cases we prove generalizations of Theorem 6.15 showing that SL desingularizations $\tilde{N}^t$ exist when $Y(L_i), Z(L_i)$ satisfy equations, and also generalize the results to *families* of almost Calabi–Yau m-folds. As the details are complicated we will not give them, but we refer the reader to [**28, 29**] and [**30**, §7].

6.5. The index of singularities of SL m-folds. We now consider the *boundary* $\partial\mathcal{M}_N$ of a moduli space $\mathcal{M}_N$ of SL m-folds.

DEFINITION 6.16. Let (M, J, ω, Ω) be an almost Calabi–Yau m-fold, N a compact, nonsingular SL m-fold in M, and $\mathcal{M}_N$ the moduli space of deformations of N in M. Then $\mathcal{M}_N$ is a smooth manifold of dimension $b^1(N)$, in general noncompact. We can construct a natural *compactification* $\overline{\mathcal{M}}_N$ as follows.

Regard $\mathcal{M}_N$ as a moduli space of special Lagrangian *integral currents* in the sense of Geometric Measure Theory, as discussed in [**26**, §6]. Let $\overline{\mathcal{M}}_N$ be the closure of $\mathcal{M}_N$ in the space of integral currents. As elements of $\mathcal{M}_N$ have uniformly

bounded volume, $\overline{\mathcal{M}}_N$ is *compact*. Define the *boundary* $\partial\mathcal{M}_N$ to be $\overline{\mathcal{M}}_N \setminus \mathcal{M}_N$. Then elements of $\partial\mathcal{M}_N$ are *singular SL integral currents*.

In good cases, say if (M, J, ω, Ω) is suitably generic, it seems reasonable that $\partial\mathcal{M}_N$ should be divided into a number of *strata*, each of which is a moduli space of singular SL m-folds with singularities of a particular type, and is itself a manifold with singularities. In particular, some or all of these strata could be moduli spaces $\mathcal{M}_X$ of SL m-folds with isolated conical singularities, as in Section 6.2.

Let $\mathcal{M}_N$ be a moduli space of compact, nonsingular SL m-folds N in (M, J, ω, Ω), and $\mathcal{M}_X$ a moduli space of singular SL m-folds in $\partial\mathcal{M}_N$ with singularities of a particular type, and $X \in \mathcal{M}_X$. Following [**30**, §8.3], we (loosely) define the *index* of the singularities of X to be $\mathrm{ind}(X) = \dim \mathcal{M}_N - \dim \mathcal{M}_X$, provided $\mathcal{M}_X$ is smooth near X. Note that $\mathrm{ind}(X)$ depends on N as well as X.

In [**30**, Th. 8.10] we use the results of [**27, 28, 29**] to compute $\mathrm{ind}(X)$ when X is *transverse* with conical singularities, in the sense of Definition 6.9. Here is a simplified version of the result, where we assume that $H^1_{\mathrm{cs}}(L_i, \mathbb{R}) \to H^1(L_i, \mathbb{R})$ is surjective to avoid a complicated correction term to $\mathrm{ind}(X)$ related to the obstructions to defining N^t as a Lagrangian m-fold.

THEOREM 6.17. *Let X be a compact, transverse SL m-fold in (M, J, ω, Ω) with conical singularities at $x_1, \ldots, x_n$ and cones $C_1, \ldots, C_n$. Let $L_1, \ldots, L_n$ be AC SL m-folds in $\mathbb{C}^m$ with cones $C_1, \ldots, C_n$, such that the natural projection $H^1_{\mathrm{cs}}(L_i, \mathbb{R}) \to H^1(L_i, \mathbb{R})$ is surjective. Construct desingularizations N of X by gluing AC SL m-folds $L_1, \ldots, L_n$ in at $x_1, \ldots, x_n$, as in Section 6.4. Then*

$$\mathrm{ind}(X) = 1 - b^0(X') + \sum_{i=1}^{n} b^1_{\mathrm{cs}}(L_i) + \sum_{i=1}^{n} s\text{-}ind(C_i). \tag{6–8}$$

If the cones C_i are not *rigid*, for instance if $C_i \setminus \{0\}$ is not connected, then (6–8) should be corrected, as in [**30**, §8.3]. If Conjecture 6.11 is true, then for a *generic* Kähler form ω, *all* compact SL m-folds X with conical singularities are transverse, and so Theorem 6.17 and [**30**, Th. 8.10] allow us to calculate $\mathrm{ind}(X)$.

Now singularities with *small index* are the most commonly occurring, and so arguably the most interesting kinds of singularity. Also, as $\mathrm{ind}(X) \leqslant \dim \mathcal{M}_N$, for various problems it will only be necessary to know about singularities with index up to a certain value.

For example, in [**17**] the author proposed to define an invariant of almost Calabi–Yau 3-folds by counting special Lagrangian homology 3-spheres (which occur in 0-dimensional moduli spaces) in a given homology class, with a certain topological weight. This invariant will only be interesting if it is essentially conserved under deformations of the underlying almost Calabi–Yau 3-fold. During such a deformation, nonsingular SL 3-folds can develop singularities and disappear, or new ones appear, which might change the invariant.

To prove the invariant is conserved, we need to show that it is unchanged along generic 1-parameter families of almost Calabi–Yau 3-folds. The only kinds of singularities of SL homology 3-spheres that arise in such families will have index 1. Thus, to resolve the conjectures in [**17**], we only have to know about index 1 singularities of SL 3-folds in almost Calabi–Yau 3-folds.

Another problem in which the index of singularities will be important is the *SYZ Conjecture*, to be discussed in Section 7. This has to do with dual 3-dimensional

families $\mathcal{F}, \hat{\mathcal{F}}$ of SL 3-tori in (almost) Calabi–Yau 3-folds $M, \hat{M}$. If $M, \hat{M}$ are generic, then the only kinds of singularities that can occur at the boundaries of $\mathcal{F}, \hat{\mathcal{F}}$ are of index 1, 2 or 3. So, to study the SYZ Conjecture in the generic case, we only have to know about singularities of SL 3-folds with index 1, 2 and 3.

7. The SYZ Conjecture and SL Fibrations

Mirror Symmetry is a mysterious relationship between pairs of Calabi–Yau 3-folds $M, \hat{M}$, arising from a branch of physics known as *String Theory*, and leading to some very strange and exciting conjectures about Calabi–Yau 3-folds, many of which have been proved in special cases.

The *SYZ Conjecture* is an attempt to explain Mirror Symmetry in terms of dual "fibrations" $f : M \to B$ and $\hat{f} : \hat{M} \to B$ of $M, \hat{M}$ by special Lagrangian 3-folds, including singular fibers. We give brief introductions to String Theory, Mirror Symmetry, and the SYZ Conjecture, and then a short survey of the state of mathematical research into the SYZ Conjecture, biased in favor of the author's own interests.

7.1. String theory and Mirror Symmetry. String Theory is a branch of high-energy theoretical physics in which particles are modeled not as points but as 1-dimensional objects — "strings" — propagating in some background space-time S. String theorists aim to construct a *quantum theory* of the string's motion. The process of quantization is extremely complicated, and fraught with mathematical difficulties that are as yet still poorly understood.

The most popular version of String Theory requires the universe to be 10-dimensional for this quantization process to work. Therefore, String Theorists suppose that the space we live in looks locally like $S = \mathbb{R}^4 \times M$, where $\mathbb{R}^4$ is Minkowski space, and M is a compact Riemannian 6-manifold with radius of order 10^{-33}cm, the Planck length. Since the Planck length is so small, space then appears to macroscopic observers to be 4-dimensional.

Because of supersymmetry, M has to be a *Calabi–Yau 3-fold*. Therefore String Theorists are very interested in Calabi–Yau 3-folds. They believe that each Calabi–Yau 3-fold M has a quantization, which is a *Super Conformal Field Theory* (SCFT), a complicated mathematical object. Invariants of M such as the Dolbeault groups $H^{p,q}(M)$ and the number of holomorphic curves in M translate to properties of the SCFT.

However, two entirely different Calabi–Yau 3-folds M and $\hat{M}$ may have the *same* SCFT. In this case, there are powerful relationships between the invariants of M and of $\hat{M}$ that translate to properties of the SCFT. This is the idea behind *Mirror Symmetry* of Calabi–Yau 3-folds.

It turns out that there is a very simple automorphism of the structure of an SCFT — changing the sign of a U(1)-action — which does *not* correspond to a classical automorphism of Calabi–Yau 3-folds. We say that M and $\hat{M}$ are *mirror* Calabi–Yau 3-folds if their SCFTs are related by this automorphism. Then one can argue using String Theory that

$$H^{1,1}(M) \cong H^{2,1}(\hat{M}) \quad \text{and} \quad H^{2,1}(M) \cong H^{1,1}(\hat{M}).$$

Effectively, the mirror transform exchanges even- and odd-dimensional cohomology. This is a very surprising result!

More involved String Theory arguments show that, in effect, the Mirror Transform exchanges things related to the complex structure of M with things related to the symplectic structure of $\hat{M}$, and vice versa. Also, a generating function for the number of holomorphic rational curves in M is exchanged with a simple invariant to do with variation of complex structure on $\hat{M}$, and so on.

Because the quantization process is poorly understood and not at all rigorous — it involves non-convergent path-integrals over horrible infinite-dimensional spaces — String Theory generates only conjectures about Mirror Symmetry, not proofs. However, many of these conjectures have been verified in particular cases.

7.2. Mathematical interpretations of Mirror Symmetry. In the beginning (the 1980s), Mirror Symmetry seemed mathematically completely mysterious. But there are now two complementary conjectural theories, due to Kontsevich and Strominger–Yau–Zaslow, which explain Mirror Symmetry in a fairly mathematical way. Probably both are true, at some level.

The first proposal was due to Kontsevich [**31**] in 1994. This says that for mirror Calabi–Yau 3-folds M and $\hat{M}$, the derived category $D^b(M)$ of coherent sheaves on M is equivalent to the derived category $D^b(\mathrm{Fuk}(\hat{M}))$ of the Fukaya category of $\hat{M}$, and vice versa. Basically, $D^b(M)$ has to do with M as a complex manifold, and $D^b(\mathrm{Fuk}(\hat{M}))$ with $\hat{M}$ as a symplectic manifold, and its Lagrangian submanifolds. We shall not discuss this here.

The second proposal, due to Strominger, Yau and Zaslow [**41**] in 1996, is known as the *SYZ Conjecture*. Here is an attempt to state it.

The SYZ Conjecture. *Suppose M and $\hat{M}$ are mirror Calabi–Yau 3-folds. Then (under some additional conditions) there should exist a compact topological 3-manifold B and surjective, continuous maps $f : M \to B$ and $\hat{f} : \hat{M} \to B$, such that*

(i) *There exists a dense open set $B_0 \subset B$, such that for each $b \in B_0$, the fibers $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are nonsingular special Lagrangian 3-tori T^3 in M and $\hat{M}$. Furthermore, $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are in some sense dual to one another.*

(ii) *For each $b \in \Delta = B \setminus B_0$, the fibers $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are expected to be singular special Lagrangian 3-folds in M and $\hat{M}$.*

We call f and $\hat{f}$ *special Lagrangian fibrations*, and the set of singular fibers Δ is called the *discriminant*. In part (i), the nonsingular fibers of f and $\hat{f}$ are supposed to be *dual tori*. What does this mean?

On the topological level, we can define duality between two tori $T, \hat{T}$ to be a choice of isomorphism $H^1(T,\mathbb{Z}) \cong H_1(\hat{T},\mathbb{Z})$. We can also define duality between tori equipped with flat Riemannian metrics. Write $T = V/\Lambda$, where V is a Euclidean vector space and Λ a *lattice* in V. Then the dual torus $\hat{T}$ is defined to be V^*/Λ^*, where V^* is the dual vector space and Λ^* the dual lattice. However, there is no notion of duality between non-flat metrics on dual tori.

Strominger, Yau and Zaslow argue only that their conjecture holds when $M, \hat{M}$ are close to the "large complex structure limit". In this case, the diameters of the fibers $f^{-1}(b), \hat{f}^{-1}(b)$ are expected to be small compared to the diameter of the base space B, and away from singularities of $f, \hat{f}$, the metrics on the nonsingular fibers are expected to be approximately flat.

So, part (i) of the SYZ Conjecture says that for $b \in B \setminus B_0$, $f^{-1}(b)$ is approximately a flat Riemannian 3-torus, and $\hat{f}^{-1}(b)$ is approximately the dual flat Riemannian torus. Really, the SYZ Conjecture makes most sense as a statement about the limiting behavior of *families* of mirror Calabi–Yau 3-folds M_t, $\hat{M}_t$ which approach the "large complex structure limit" as $t \to 0$.

7.3. The symplectic topological approach to SYZ. The most successful approach to the SYZ Conjecture so far could be described as *symplectic topological*. In this approach, we mostly forget about complex structures, and treat $M, \hat{M}$ just as *symplectic manifolds*. We mostly forget about the 'special' condition, and treat $f, \hat{f}$ just as *Lagrangian fibrations*. We also impose the condition that B is a *smooth* 3-manifold and $f : M \to B$ and $\hat{f} : \hat{M} \to B$ are *smooth maps*. (It is not clear that $f, \hat{f}$ can in fact be smooth at every point, though.)

Under these simplifying assumptions, Gross [**6, 7, 8, 9**], Ruan [**38, 39**], and others have built up a beautiful, detailed picture of how dual SYZ fibrations work at the global topological level, in particular for examples such as the quintic and its mirror, and for Calabi–Yau 3-folds constructed as hypersurfaces in toric 4-folds, using combinatorial data.

7.4. Local geometric approach and SL singularities. There is also another approach to the SYZ Conjecture, begun by the author in [**23, 25**], and making use of the ideas and philosophy set out in Section 6. We could describe it as a *local geometric* approach.

In it we try to take the special Lagrangian condition seriously from the outset, and our focus is on the local behavior of special Lagrangian submanifolds, and especially their singularities, rather than on global topological questions. Also, we are interested in what fibrations of *generic* (almost) Calabi–Yau 3-folds might look like.

One of the first-fruits of this approach has been the understanding that for *generic* (almost) Calabi–Yau 3-folds M, special Lagrangian fibrations $f : M \to B$ will not be smooth maps, but only piecewise smooth. Furthermore, their behavior at the singular set is rather different to the smooth Lagrangian fibrations discussed in Section 7.3.

For smooth special Lagrangian fibrations $f : M \to B$, the discriminant Δ is of codimension 2 in B, and the typical singular fiber is singular along an $\mathcal{S}^1$. But in a generic special Lagrangian fibration $f : M \to B$ the discriminant Δ is of codimension 1 in B, and the typical singular fiber is singular at finitely many points.

One can also show that if $M, \hat{M}$ are a mirror pair of generic (almost) Calabi–Yau 3-folds and $f : M \to B$ and $\hat{f} : \hat{M} \to B$ are dual special Lagrangian fibrations, then in general the discriminants Δ of f and $\hat{\Delta}$ of $\hat{f}$ cannot coincide in B, because they have different topological properties in the neighborhood of a certain kind of codimension 3 singular fiber.

This contradicts part (ii) of the SYZ Conjecture, as we have stated it in Section 7.2. In the author's view, these calculations support the idea that the SYZ Conjecture in its present form should be viewed primarily as a limiting statement, about what happens at the "large complex structure limit", rather than as simply being about pairs of Calabi–Yau 3-folds. A similar conclusion is reached by Mark Gross in [**9**, §5].

7.5. U(1)-invariant SL fibrations in $\mathbb{C}^3$. We finish by describing work of the author in [**23**, §8] and [**25**], which aims to describe what the singularities of SL fibrations of *generic* (almost) Calabi–Yau 3-folds look like, providing they exist.

This proceeds by first studying SL fibrations of subsets of $\mathbb{C}^3$ invariant under the U(1)-action (3–2), using the ideas of Section 3.5. For a brief survey of the main results, see [**24**]. Then we argue that the kinds of singularities we see in codimension 1 and 2 in generic U(1)-invariant SL fibrations in $\mathbb{C}^3$, also occur in codimension 1 and 2 in SL fibrations of generic (almost) Calabi–Yau 3-folds.

Following [**23**, Def. 8.1], we use the results of Section 3.5 to construct a family of SL 3-folds $N_{\boldsymbol{\alpha}}$ in $\mathbb{C}^3$, depending on boundary data $\Phi(\boldsymbol{\alpha})$.

DEFINITION 7.1. Let S be a strictly convex domain in $\mathbb{R}^2$ invariant under $(x,y)\mapsto(x,-y)$, let U be an open set in $\mathbb{R}^3$, and $\alpha\in(0,1)$. Suppose

$$\Phi : U \to C^{3,\alpha}(\partial S)$$

is a continuous map such that if $(a,b,c)\neq(a,b',c')$ in U, then $\Phi(a,b,c)-\Phi(a,b',c')$ has exactly one local maximum and one local minimum in ∂S.

For $\boldsymbol{\alpha}=(a,b,c)\in U$, let $f_{\boldsymbol{\alpha}}\in C^{3,\alpha}(S)$ or $C^1(S)$ be the unique (weak) solution of (3–6) with $f_{\boldsymbol{\alpha}}|_{\partial S}=\Phi(\boldsymbol{\alpha})$, which exists by Theorem 3.3. Define

$$u_{\boldsymbol{\alpha}}=\frac{\partial f_{\boldsymbol{\alpha}}}{\partial y}\quad\text{and}\quad v_{\boldsymbol{\alpha}}=\frac{\partial f_{\boldsymbol{\alpha}}}{\partial x}.$$

Then $(u_{\boldsymbol{\alpha}},v_{\boldsymbol{\alpha}})$ is a solution of (3–5) in $C^{2,\alpha}(S)$ if $a\neq 0$, and a weak solution of (3–4) in $C^0(S)$ if $a=0$. Also $u_{\boldsymbol{\alpha}},v_{\boldsymbol{\alpha}}$ depend continuously on $\boldsymbol{\alpha}\in U$ in $C^0(S)$, by Theorem 3.3.

For each $\boldsymbol{\alpha}=(a,b,c)$ in U, define $N_{\boldsymbol{\alpha}}$ in $\mathbb{C}^3$ by

$$\begin{aligned} N_{\boldsymbol{\alpha}}=\big\{(z_1,z_2,z_3)\in\mathbb{C}^3 : {}& z_1z_2=v_{\boldsymbol{\alpha}}(x,y)+iy,\quad z_3=x+iu_{\boldsymbol{\alpha}}(x,y),\\ & |z_1|^2-|z_2|^2=2a,\quad (x,y)\in S^\circ\big\}. \end{aligned}\tag{7–1}$$

Then $N_{\boldsymbol{\alpha}}$ is a noncompact SL 3-fold without boundary in $\mathbb{C}^3$, which is nonsingular if $a\neq 0$, by Proposition 3.1.

In [**23**, Th. 8.2] we show that the $N_{\boldsymbol{\alpha}}$ are the fibers of an *SL fibration*.

THEOREM 7.2. *In the situation of Definition 7.1, if $\boldsymbol{\alpha}\neq\boldsymbol{\alpha}'$ in U, then*

$$N_{\boldsymbol{\alpha}}\cap N_{\boldsymbol{\alpha}'}=\varnothing.$$

There exists an open set $V\subset\mathbb{C}^3$ and a continuous, surjective map $F:V\to U$ such that $F^{-1}(\boldsymbol{\alpha})=N_{\boldsymbol{\alpha}}$ for all $\boldsymbol{\alpha}\in U$. Thus, F is a special Lagrangian fibration of $V\subset\mathbb{C}^3$, which may include singular fibers.

It is easy to produce families Φ satisfying Definition 7.1. For example [**23**, Ex. 8.3], given any $\phi\in C^{3,\alpha}(\partial S)$ we may define $U=\mathbb{R}^3$ and $\Phi:\mathbb{R}^3\to C^{3,\alpha}(\partial S)$ by $\Phi(a,b,c)=\phi+bx+cy$. So this construction produces very large families of U(1)-invariant SL fibrations, including singular fibers, which can have any multiplicity and type.

Here is a simple, explicit example. Define $F : \mathbb{C}^3 \to \mathbb{R} \times \mathbb{C}$ by

$$F(z_1, z_2, z_3) = (a, b), \quad \text{where} \quad 2a = |z_1|^2 - |z_2|^2$$

$$\text{and} \quad b = \begin{cases} z_3, & a = z_1 = z_2 = 0, \\ z_3 + \bar{z}_1 \bar{z}_2 / |z_1|, & a \geqslant 0,\ z_1 \neq 0, \\ z_3 + \bar{z}_1 \bar{z}_2 / |z_2|, & a < 0. \end{cases} \tag{7–2}$$

This is a piecewise-smooth SL fibration of $\mathbb{C}^3$. It is not smooth on $|z_1| = |z_2|$.

The fibers $F^{-1}(a, b)$ are T^2-cones singular at $(0, 0, b)$ when $a = 0$, and nonsingular $\mathcal{S}^1 \times \mathbb{R}^2$ when $a \neq 0$. They are isomorphic to the SL 3-folds of Example 3.4 under transformations of $\mathbb{C}^3$, but they are assembled to make a fibration in a novel way.

As a goes from positive to negative the fibers undergo a surgery, a Dehn twist on $\mathcal{S}^1$. The reason why the fibration is only piecewise-smooth, rather than smooth, is really this topological transition, rather than the singularities themselves. The fibration is not differentiable at every point of a singular fiber, rather than just at singular points, and this is because we are jumping from one moduli space of SL 3-folds to another at the singular fibers.

I claim that F is a local model for codimension one singularities of SL fibrations of generic almost Calabi–Yau 3-folds. The reason for this is that these T^2-cone singularities are *stable*, as in Definition 6.4, so SL 3-folds X with these singularities form *smooth moduli spaces* $\mathcal{M}_X$ by Corollary 6.8.

The singularities are automatically *transverse*, as in Definition 6.9, so we can apply [**30**, Th. 8.10] to compute the *index* $\mathrm{ind}(X)$ of the singularities, as in Section 6.5. This is done in detail in [**30**, §10]. If the topology of X is suitably chosen, then $\mathrm{ind}(X) = 1$, so $\mathcal{M}_X$ has codimension one in $\mathcal{M}_N$. The singular behavior is stable under small exact perturbations of the underlying almost Calabi–Yau structure.

I also have a U(1)-invariant model for codimension two singularities, described in [**25**], in which two of the codimension one T^2-cones come together and cancel out. I conjecture that it too is a typical codimension two singular behavior in SL fibrations of generic almost Calabi–Yau 3-folds. I do not expect codimension three singularities in generic SL fibrations to be locally U(1)-invariant, and so this approach will not help.

References

[1] R. L. Bryant, *Second order families of special Lagrangian 3-folds*, math.DG/0007128, 2000.

[2] I. Castro and F. Urbano, *New examples of minimal Lagrangian tori in the complex projective plane*, Manuscripta math. 85 (1994), 265–281.

[3] A. P. Fordy and J. C. Wood, editors, *Harmonic Maps and Integrable Systems*, Aspects of Math. E23, Vieweg, Wiesbaden, 1994.

[4] E. Goldstein, *Calibrated fibrations*, Communications in Analysis and Geometry 10 (2002), 127–150. math.DG/9911093.

[5] E. Goldstein, *Minimal Lagrangian tori in Kähler–Einstein manifolds*, math.DG/0007135, 2000.

[6] M. Gross, *Special Lagrangian fibrations I: Topology*. In M.-H. Saito, Y. Shimizu, and K. Ueno, editors, *Integrable Systems and Algebraic Geometry*, pages 156–193, World Scientific, Singapore, 1998. alg-geom/9710006.

[7] M. Gross, *Special Lagrangian fibrations II: Geometry*. In *Differential Geometry inspired by String Theory*, Surveys in Differential Geometry 5, pages 341–403, International Press, Boston, MA, 1999. math.AG/9809072.

[8] M. Gross, *Topological mirror symmetry*, Invent. math. 144 (2001), 75–137. math.AG/9909015.

[9] M. Gross, *Examples of special Lagrangian fibrations.* In K. Fukaya, Y.-G. Oh, K. Ono and G. Tian, editors, *Symplectic geometry and mirror symmetry (Seoul, 2000)*, pages 81–109, World Scientific, Singapore, 2001. math.AG/0012002.
[10] R. Harvey, *Spinors and calibrations*, Academic Press, San Diego, 1990.
[11] R. Harvey and H. B. Lawson, *Calibrated geometries*, Acta Mathematica 148 (1982), 47–157.
[12] M. Haskins, *Special Lagrangian Cones*, math.DG/0005164, 2000.
[13] N. J. Hitchin, *The moduli space of Special Lagrangian submanifolds*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. 25 (1997), 503–515. dg-ga/9711002.
[14] D. D. Joyce, *Compact Manifolds with Special Holonomy*, OUP, Oxford, 2000.
[15] D. D. Joyce, *Constructing special Lagrangian m-folds in $\mathbb{C}^m$ by evolving quadrics*, Math. Ann. 320 (2001), 757–797. math.DG/0008155.
[16] D. D. Joyce, *Evolution equations for special Lagrangian 3-folds in $\mathbb{C}^3$*, Ann. Global Anal. Geom. 20 (2001), 345–403. math.DG/0010036.
[17] D. D. Joyce, *On counting special Lagrangian homology 3-spheres*, pages 125–151 in *Topology and Geometry: Commemorating SISTAG*, editors A.J. Berrick, M.C. Leung and X.W. Xu, Contemporary Mathematics 314, A.M.S., Providence, RI, 2002. hep-th/9907013.
[18] D. D. Joyce, *Special Lagrangian m-folds in $\mathbb{C}^m$ with symmetries*, Duke Math. J. 115 (2002), 1–51. math.DG/0008021.
[19] D. D. Joyce, *Ruled special Lagrangian 3-folds in $\mathbb{C}^3$*, Proc. L.M.S. 85 (2002), 233–256. math.DG/0012060.
[20] D. D. Joyce, *Special Lagrangian 3-folds and integrable systems*, math.DG/0101249, 2001. To appear in volume 1 of the Proceedings of the Mathematical Society of Japan's 9th International Research Institute on *Integrable Systems in Differential Geometry*, Tokyo, 2000.
[21] D. D. Joyce, *U(1)-invariant special Lagrangian 3-folds. I. Nonsingular solutions*, math.DG/0111324, 2001. To appear in Advances in Mathematics.
[22] D. D. Joyce, *U(1)-invariant special Lagrangian 3-folds. II. Existence of singular solutions*, math.DG/0111326, 2001.
[23] D. D. Joyce, *U(1)-invariant special Lagrangian 3-folds. III. Properties of singular solutions*, math.DG/0204343, 2002.
[24] D. D. Joyce, *U(1)-invariant special Lagrangian 3-folds in $\mathbb{C}^3$ and special Lagrangian fibrations*, Turkish Math. J. 27 (2003), 99–114. math.DG/0206016.
[25] D. D. Joyce, *Singularities of special Lagrangian fibrations and the SYZ Conjecture*, math.DG/0011179, 2000. To appear in Communication in Analysis and Geometry.
[26] D. D. Joyce, *Special Lagrangian submanifolds with conical singularities. I. Regularity*, math.DG/021124, 2002.
[27] D. D. Joyce, *Special Lagrangian submanifolds with conical singularities. II. Moduli spaces*, math.DG/0211295, 2002.
[28] D. D. Joyce, *Special Lagrangian submanifolds with isolated conical singularities. III. Desingularization, the unobstructed case*, math.DG/0302355, 2003.
[29] D. D. Joyce, *Special Lagrangian submanifolds with isolated conical singularities. IV. Desingularization, obstructions and families*, math.DG/0302356, 2003.
[30] D. D. Joyce, *Special Lagrangian submanifolds with isolated conical singularities. V. Survey and applications*, math.DG/0303272, 2003. To appear in the Journal of Differential Geometry.
[31] M. Kontsevich, *Homological Algebra of Mirror Symmetry.* In *Proceedings of the International Congress of Mathematicians (Zürich, 1994)*, pages 120–139, Birkhäuser, Basel, 1994. alg-geom/9411018.
[32] G. Lawlor, *The angle criterion*, Invent. math. 95 (1989), 437–446.
[33] Hui Ma and Yujie Ma, *Totally real minimal tori in $\mathbb{CP}^2$*, math.DG/0106141, 2001.
[34] S.P. Marshall, *Deformations of special Lagrangian submanifolds*, DPhil thesis, University of Oxford, 2002.
[35] I. McIntosh, *Special Lagrangian cones in $\mathbb{C}^3$ and primitive harmonic maps*, J. London Math. Soc. 67 (2003), 769–789. math.DG/0201157.
[36] R. C. McLean, *Deformations of calibrated submanifolds*, Communications in Analysis and Geometry 6 (1998), 705–747.
[37] T. Pacini, *Flows and Deformations of Lagrangian Submanifolds in Kaehler–Einstein Geometry*, PhD thesis, University of Pisa, 2002.
[38] W.-D. Ruan, *Lagrangian tori fibration of toric Calabi–Yau manifold I*, math.DG/9904012, 1999.

[39] W.-D. Ruan, *Lagrangian torus fibration and mirror symmetry of Calabi–Yau hypersurface in toric variety*, math.DG/0007028, 2000.
[40] R. A. Sharipov, *Minimal tori in the five-dimensional sphere in* $\mathbb{C}^3$, Theoretical and Mathematical Physics 87 (1991), 363–369. Revised version: math.DG/0204253.
[41] A. Strominger, S.-T. Yau, and E. Zaslow, *Mirror symmetry is T-duality*, Nuclear Physics B479 (1996), 243–259. hep-th/9606040.
[42] S.-T. Yau. *On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equations. I*, Comm. pure appl. math. 31 (1978), 339–411.

Lincoln College, Oxford, OX1 3DR, Great Britain
E-mail address: dominic.joyce@lincoln.ox.ac.uk

Clay Mathematics Proceedings
Volume 2, 2004

Variational Problems in Lagrangian Geometry: $\mathbb{Z}_2$-Currents

Jon Wolfson

Introduction

In this article we discuss aspects of the existence theory for extrema of the lagrangian variational problem introduced in [S-W] and described in the article of R. Schoen in this volume [Sc]. We will discuss questions about the regularity of the extrema only as they relate to the existence question.

Recall the set-up. We let (X, ω) be a compact symplectic manifold of dimension $2n$ equipped with a compatible metric g. In particular we could take (X, ω, g) to be Kähler. An n-dimensional submanifold Σ of X is called *lagrangian* if $\omega_{|\Sigma} = 0$. An n-dimensional cycle is lagrangian if each n-simplex is lagrangian. We say a homology class $\alpha \in H_n(X, \mathbb{Z})$ is lagrangian if it can be represented by a lagrangian cycle. We consider the variational problem of finding extrema of volume among the lagrangian cycles (lagrangian currents, lagrangian maps, etc.) representing the lagrangian homology class α. It is also possible to formulate other variational problems, such as boundary value problems, a homotopy problem, etc., but for simplicity we will emphasis the homology problem. To be rigorous it is necessary to specify precisely the class of lagrangians among which extrema are sought.

Mapping problem. When the domain manifold is 2-dimensional, because the energy of a conformal map equals the area of its image, it is possible to formulate a mapping problem. Let Σ be a Riemann surface and consider the maps $f : \Sigma \to X$ in $W^{1,2}(\Sigma, X)$. These are maps such that f and Df are in $L^2(\Sigma, \mathbb{R}^N)$ for some isometric embedding of X into $\mathbb{R}^N$. Note that $f^*\omega$ is an L^1-valued 2-form on Σ. We say that a map $f \in W^{1,2}(\Sigma, X)$ is *weakly lagrangian* if $f^*\omega = 0$ almost everywhere. An important ingredient in formulating the lagrangian mapping problem is the following compactness result [S-W]: The set of weakly lagrangian maps in $W^{1,2}(\Sigma, X)$ satisfying a uniform energy bound is closed in the weak topology. As a consequence a minimizing sequence of lagrangian maps in $W^{1,2}(\Sigma, X)$ has a weakly lagrangian limit in $W^{1,2}(\Sigma, X)$. If the maps in the sequence represent a homology class then, as in the classical unconstrained case, the limit map may not represent the same class. However there are well-known techniques available to understand and handle this phenomenon. The main results for this problem were obtained in [S-W] and are surveyed in [Sc].

The author was partially supported by NSF grant DMS-0304587.

©2004 Clay Mathematics Institute

Integral currents. In general it is necessary to use currents in the formulation of our variational problem. The class of currents that are the geometric measure theory generalization of oriented submanifolds are the integral currents. There are two natural topologies for integral currents, the weak topology and the flat norm topology. A sequence $\{T_i\}$ of n-currents converges to the n-current T in the weak topology if for every C^∞ n-form ϕ on X with compact support

$$\lim_{i\to\infty} T_i(\phi) = T(\phi).$$

The sequence converges to T in the flat norm topology if there is a sequence of $n+1$- integral currents P_i and a sequence of n-integral currents Q_i with

$$\lim_{i\to\infty} (|P_i| + |Q_i|) = 0,$$

satisfying:

$$T_i - T = \partial P_i + Q_i,$$

where $|P|$ denotes the mass (or volume) of the integral current P. Two currents are close in the flat norm topology if they are the boundary (up to a small n-volume current) of a current with small $(n+1)$-volume. We say an integral current T is lagrangian if for every C^∞ $(n-2)$-form ϕ on X with compact support,

$$T(\omega \wedge \phi) = 0.$$

It follows easily that if a sequence of lagrangian integral currents converges weakly to an integral current, the limit current is lagrangian. From the basic theory of currents (see [Si]) this implies the same result using the flat norm topology. Applying these compactness results to a minimizing sequence of lagrangian integral currents (representing a homology class) we can find a minimizer that is also a lagrangian integral current and represents the homology class. Unfortunately, in the unconstrained case, there are serious difficulties in the regularity theory of minimizers when the codimension of the minimizer is greater than 1. These problems persist in the constrained (lagrangian) problem. Accordingly we seek a larger class of currents.

$\mathbb{Z}_2$-currents. The $\mathbb{Z}_2$-currents are the geometric measure theory generalization of the unoriented (in particular, nonorientable) submanifolds. Since there are more comparisons available than for integral currents, the regularity theory for minimizers is better than that for integral currents [F]. Because the currents are unoriented, the weak topology is not available so we must use the flat norm topology. We say a $\mathbb{Z}_2$-current is lagrangian if its approximate tangent planes are lagrangian planes almost everywhere. To use the lagrangian $\mathbb{Z}_2$-currents in our variational problem we seek to establish a sequential compactness result in the flat norm topology. This problem is the subject of the next section.

1. Lagrangian $\mathbb{Z}_2$-Currents

In $\mathbb{R}^4$ set $\lambda = \frac{1}{2}\sum_{j=1}^2 (x_i dy_i - y_i dx_i)$. Then $d\lambda = \sum_{j=1}^2 dx_i \wedge dy_i = \omega$ is the standard symplectic form. If γ is a closed curve in $\mathbb{R}^4$ the quantity $\int_\gamma \lambda$ is called the *period* of the curve. A well-known fact in symplectic geometry is that a closed curve γ in $\mathbb{R}^4$ spans a lagrangian disc if and only if its period vanishes. The following isoperimetric inequality due to Allcock [A] and Gromov [G] is a quantitative version of this result.

THEOREM 1.1. *Let γ be a closed curve in $\mathbb{R}^4$ satisfying $\int_\gamma \lambda = 0$. Then there is a lagrangian disc D spanning γ with,*

$$|D| \leq c|\gamma|^2,$$

where c is a universal constant.

Qiu [Q] proved the following nonorientable isoperimetric inequality using ideas based on Allcock's proof.

THEOREM 1.2. *Let γ be any closed curve in $\mathbb{R}^4$. Then there is a lagrangian Möbius band M spanning γ with,*

$$|M| \leq c|\gamma|^2,$$

where c is a universal constant.

This isoperimetric inequality has the following interesting application. Recall ([Sc], [S-W]) that the 2-dimensional lagrangian cones in $\mathbb{R}^4 \simeq \mathbb{C}^2$ that are stationary under hamiltonian variations are parameterized by a pair of relatively prime integers $p, q \geq 1$. Explicitly the (p,q)-cone in $\mathbb{C}^2$ with coordinates (z_1, z_2) is:

$$C_{p,q} = \frac{1}{\sqrt{p+q}} \left(r\sqrt{q}e^{i\sqrt{\frac{p}{q}}s}, ir\sqrt{p}e^{-i\sqrt{\frac{q}{p}}s} \right), \tag{1.1}$$

where $0 \leq s \leq 2\pi\sqrt{pq}$ and $r \geq 0$. The difference $p-q$ is the Maslov index of the cone. If $|p-q| > 1$, the cone is unstable for compactly supported hamiltonian variations and hence these cones are not tangent cones on lagrangian minimizers. However, for at least one pair (p,q) with $|p-q| = 1$, the cone is a minimizer among oriented lagrangian comparisons and occurs as a tangent cone on a minimizer. (This will be discussed in more detail below.) Using the nonorientable isoperimetric inequality the following result is obtained in [S-W]:

THEOREM 1.3. *For any (p,q) with $(p,q) \neq (1,1)$, the cone $C_{p,q}$ does not minimize area among nonorientable lagrangian comparison surfaces.*

OUTLINE OF THE PROOF. Deform the annular region of the cone from $r = 1$ to $r = \rho$ by the mean curvature H. Since $H \approx 1/r$ this decreases area by $c\int\int_1^\rho H^2 r dr ds = c' \ln \rho$. The mean curvature is a symplectic deformation so the deformed region remains lagrangian however, because H is not hamiltonian, the curves $r = 1$ to $r = \rho$ are deformed to curves that no longer have zero period. Thus the deformed region cannot be joined to the undeformed regions of the cone by annular lagrangian strips. However they can be joined by nonorientable lagrangians. Using the isoperimetric inequality such nonorientable lagrangians can be constructed having bounded area (independent of ρ). Therefore, for ρ sufficiently large, the new lagrangian surface has smaller area. □

Suppose that the ambient manifold X is Kähler–Einstein (i.e., $\mathrm{Ric} = \mathrm{R}\omega$). On a smoothly immersed lagrangian the mean curvature vector H is an infinitesimal symplectic motion and thus an admissible variation in the lagrangian variational problem. Using this observation and a first variation argument (see [Sc]) it follows that a *regular* lagrangian minimizer is a classical minimal submanifold ($H = 0$). In the mapping problem the regularity theorem for minimizers ([S-W]) shows that a minimizer is an immersion except at isolated points. These points are either branch

points or singularities with tangent cones $C_{p,q}$ (1.1). The existence of nonplanar tangent cones that are minimizing (among *oriented* lagrangian comparisons) implies that a minimizer for the mapping problem may have a singular point with such a tangent cone. For such a minimizer we are then unable to use H as a variation and therefore we cannot conclude that the minimizer is a classical minimal surface. This same problem occurs in the variational problem using 2-dimensional lagrangian integral currents. However, Theorem 1.3 implies that there are no nonplanar tangent cones that are minimizing (among *unoriented* lagrangian comparisons). Using this observation and the regularity theory developed for the mapping problem [S-W], we conclude that a 2-dimensional lagrangian $\mathbb{Z}_2$ minimizer is regular (has no singularities) and therefore H is an admissible variation. What is lacking is a result establishing the existence of a lagrangian $\mathbb{Z}_2$-current minimizer. Unfortunately, the following result of Qiu [Q], also an application of the nonorientable isoperimetric inequality, shows that the existence theory for lagrangian $\mathbb{Z}_2$-current minimizers is problematic.

THEOREM 1.4. *The set of lagrangian $\mathbb{Z}_2$-currents in $\mathbb{R}^4$ is dense in the flat norm topology in the set of all $\mathbb{Z}_2$-currents.*

PROOF. Throughout the proof c will denote universal constants. Fleming [Fl] shows that the polyhedral chains are dense, in the flat norm topology, in the $\mathbb{Z}_2$-currents. Therefore, it suffices to prove the theorem for a planar unit square P in $\mathbb{R}^4$. Divide P into N^2 subsquares P_j, $j = 1, \dots, N^2$, so that each P_j has sides of length $1/N$. By Theorem 1.2 for each P_j there is a lagrangian Möbius band M_j with $\partial M_j = \partial P_j$ and

$$|M_j| \leq c|\partial P_j|^2 = c\Big(\frac{4}{N}\Big)^2.$$

Set $M = \sum_j M_j$. Since interior boundaries cancel we have $\partial P = \partial M$. Note that,

$$|M| \leq \sum_j |M_j| \leq N^2 c\Big(\frac{4}{N}\Big)^2 = 16c.$$

Since $\partial(M_j - P_j) = 0$ by the classical isoperimetric inequality there is a three-dimensional $\mathbb{Z}_2$-current T_j with $\partial T_j = M_j - P_j$ and

$$|T_j| \leq c|M_j - P_j|^{3/2}.$$

But $|M_j - P_j| \leq |M_j| + |P_j| \leq c/N^2$, so

$$|T_j| \leq \frac{c}{N^3}.$$

Set $T = \sum_j T_j$. Then $\partial T = M - P$ and

$$|T| \leq N^2 \frac{c}{N^3} = \frac{c}{N}.$$

It follows that as $N \to \infty$, M approximates P in the flat norm topology. □

Notice that the proof implies that a bounded domain in $\mathbb{C} \subset \mathbb{R}^4$ can be approximated by lagrangians satisfying a uniform area bound. Another curious consequence of Qiu's theorem is that in a symplectic 4-manifold X every class in $H_2(X, \mathbb{Z}_2)$ is lagrangian, that is, can be represented by a lagrangian cycle. By comparison a class $\alpha \in H_2(X, \mathbb{Z})$ is lagrangian if and only if $\int_\alpha \omega = 0$.

For our purposes the main consequence of Qiu's theorem is that the limit of a sequence of lagrangian $\mathbb{Z}_2$-currents with uniformly bounded volumes (masses) may

not be lagrangian. It remains possible that a *minimizing* sequence of lagrangian $\mathbb{Z}_2$-currents is always lagrangian. However Qiu's result suggests otherwise. Note that the construction, in the proof of the theorem, introduces curves γ on the lagrangians for which $\int_\gamma \lambda \neq 0$, where $\lambda = \frac{1}{2}\sum_{j=1}^{2}(x_i dy_i - y_i dx_i)$ is the Liouville form. An immersed lagrangian $\Sigma \subset \mathbb{R}^{2n}$ is called *exact* if $\int_\tau \lambda = 0$, for every closed curve τ on Σ. That is, the approximating lagrangians constructed in the proof of Theorem 1.4 are not exact. Thus it is possible that, restricting to exact lagrangians, sequential compactness holds.

Recall that an equivalent formulation of exactness, one more convenient for our purposes, is given as follows: Consider $\mathbb{R}^{2n+1} = \{(x, y, \varphi)\}$. Denote by π the projection $\mathbb{R}^{2n+1} \to \mathbb{R}^{2n}$ $(x, y, \varphi) \mapsto (x, y)$. On $\mathbb{R}^{2n+1}$ define the *contact* 1-form $\eta = d\varphi - \frac{1}{2}\sum_{j=1}^{n}(x_i dy_i - y_i dx_i)$. The pair $(\mathbb{R}^{2n+1}, \eta)$ is called a *contact manifold.* The hyperplane distribution defined by $\eta = 0$ is called the *contact distribution.* An n-dimensional submanifold S is called *legendrian* if $\eta_{|S} = 0$ or equivalently if it is everywhere tangent to the contact distribution. An n-dimensional $\mathbb{Z}_2$-current is called *legendrian* if its approximate tangent planes lie in the contact distribution almost everywhere. It is easy to see that if S is legendrian, then $\pi(S) \subset \mathbb{R}^{2n}$ be an exact lagrangian. Conversely, a smoothly immersed exact lagrangian $\Sigma \subset \mathbb{R}^{2n}$ has a legendrian lift.

For a point $p \in \mathbb{R}^{2n+1}$, denote the contact $2n$-plane at p by H_p. Then $d\pi_p : H_p \to \mathbb{R}^{2n}$ is an isomorphism. Define a metric g on the contact distribution by requiring that $d\pi_p$ be an isometry for every $p \in \mathbb{R}^{2n+1}$. The metric g is a Riemannian metric on the contact distribution but defines a degenerate (Carnot) metric on $\mathbb{R}^{2n+1}$. The importance of the legendrians for our purposes is due to the following theorem.

THEOREM 1.5. *Let $\{\Sigma_i\}$ be a sequence of legendrian $\mathbb{Z}_2$-currents in $\mathbb{R}^{2n+1}$ that satisfy $|\Sigma_i| + |\partial\Sigma_i| < C$ (using the metric g). There is a subsequence that converges in the flat norm topology to a legendrian $\mathbb{Z}_2$-current.*

PROOF. Introduce a Riemannian metric, depending on a parameter, on $\mathbb{R}^{2n+1}$ as follows. Let V be the Reeb vector field associated to η, that is, the vector field satisfying $i_V\, d\eta = 0$ and $\eta(V) = 1$. Define a metric g_1 on $\mathbb{R}^{2n+1}$ so that $d\pi$ restricted to the contact subspace $H \subset T\mathbb{R}^{2n+1}$ is an isometry and so that V is orthogonal to P and of unit length. Thus, $g_1 = g + \eta^2$. Introduce the family of metrics $g_\varepsilon = g + \varepsilon^{-2}\eta^2$.

Consider the sequence of legendrian $\mathbb{Z}_2$-currents Σ_i. Choose a sequence $\varepsilon_j \downarrow 0$. For each j, using the metric g_{ε_j}, choose a subsequence $\{\Sigma_{i_j}\}$ that converges in the flat norm topology to a $\mathbb{Z}_2$-current. Let Σ_0 denote the limit of a diagonal subsequence of $\{\Sigma_{i_j}\}$. It follows that, for any ε_j, $\mathrm{Vol}_{\varepsilon_j}(\Sigma_0)$ is bounded. If Σ_0 is not legendrian, there is a set $E \subset \Sigma_0$ of positive measure whose tangent planes are not subspaces of the contact planes. This implies that $\mathrm{Vol}_{\varepsilon_j}(E) \to \infty$ as $\varepsilon_j \to 0$, a contradiction. □

REMARK 1. The theorem shows that the set of legendrian $\mathbb{Z}_2$-currents has a suitable compactness property. It is remarkable that to formulate the monotonicity formula of [S-W] for lagrangian minimizers in $\mathbb{R}^{2n}$ it is necessary to introduce the legendrian lifts to the contact manifold $\mathbb{R}^{2n+1}$. There is no obvious reason for this coincidence.

REMARK 2. The theorem is also true for sequences of legendrian integral currents. However in the case of legendrian integral currents, convergence to a legendrian current can be easily proved in the weak topology. Convergence in the flat norm topology then follows.

An apparent conclusion of Theorem 1.5 is that minimizers among legendrian $\mathbb{Z}_2$-currents can be found. However, Theorem 1.5 is a local result and it is not clear how to generalize the notion of legendrian to manifolds. Indeed if ω is not an exact form, in general, periods cannot be defined. Of course, ω is locally exact and locally, by the Darboux theorem, every symplectic manifold looks like $\mathbb{R}^{2n}$. One could try to exploit this to define a class of "locally exact" lagrangians. However, fixing a covering of a symplectic manifold X by Darboux neighborhoods, it is not difficult to construct a sequence of lagrangians that are exact in every neighborhood of the cover but that converge in the flat norm topology to a lagrangian lying in a Darboux ball that is not exact. What is needed is a global version of exactness, that is, a global version of legendrian.

2. Global Exactness

Let (X, ω) be a compact symplectic manifold of dimension $2n$. Suppose that the symplectic form ω is integral, that is, the homology class $[\omega]$ is an integral class. Then there is an hermitian line bundle L over X with unitary connection η such that the curvature of η is ω [K]. Let L also denote the total space of the associated S^1 principal bundle and let $\pi : L \to X$ denote the bundle projection. On L the one-form η is globally well-defined and satisfies $d\eta = \pi^*\omega$. Thus η is a contact one-form. The horizontal distribution of the connection η is the contact distribution. Define a metric g on the horizontal distribution by requiring that $d\pi_p : H_p \to T_p(X)$ be an isometry for each $p \in L$, where H_p is the horizontal $2n$-plane at p. Then g is a degenerate (Carnot) metric on the contact manifold L. We say that an n-dimensional smooth submanifold S is *legendrian* if $\eta_{|S} = 0$ or, equivalently, if $T_p(S) \subset H_p$ for each $p \in S$. We say that a $\mathbb{Z}_2$-current is *legendrian* if its approximate tangent planes lie in the horizontal distribution almost everywhere.

Using the same argument as in the proof of Theorem 1.5 it follows that a sequence of legendrian $\mathbb{Z}_2$-currents that satisfy a uniform mass bound for the metric g (as in Theorem 1.5) has a subsequence that converges in the flat norm topology to a legendrian $\mathbb{Z}_2$-current. Thus

THEOREM 2.1. *Let $\mathcal{L}_\alpha$ be the set of legendrian $\mathbb{Z}_2$-cycles in L that represent the homology class $\alpha \in H_n(L, \mathbb{Z}_2)$. A sequence of currents in $\mathcal{L}_\alpha$ that minimizes volume has a subsequence that converges in the flat norm topology to a current in $\mathcal{L}_\alpha$.*

Theorem 2.1 shows that there is a global existence theory for legendrian $\mathbb{Z}_2$ minimizers. This existence theory can be applied to a class a in $H_n(X, \mathbb{Z}_2)$ that has a lift to a legendrian class α in $H_n(L, \mathbb{Z}_2)$. The push forward by π of the legendrian minimizer representing α is a lagrangian $\mathbb{Z}_2$-cycle that minimizes volume among $\mathbb{Z}_2$-cycles that represent a and that have legendrian lifts. We will call the lagrangian $\mathbb{Z}_2$ minimizer an "exact lagrangian" $\mathbb{Z}_2$ minimizer or, by abuse of nomenclature, a legendrian $\mathbb{Z}_2$ minimizer.

We remark that the minimizer of Theorem 2.1 is contact stationary in the sense of [S-W]. That is, the minimizer is stationary with respect to the contact

transformations of L. These are the diffeomorphisms of L that preserve the contact distribution. In the case that the currents are 2-dimensional, the contact stationary currents satisfy a monotonicity inequality [S-W]. Using this it can be shown that the minimizers satisfy a regularity result similar to that satisfied by the minimizers of the mapping problem. In particular, the minimizers are (nonorientable) immersed surfaces except at isolated points which are singularities with $(p, p+1)$ tangent cone. The regularity question for the 2-dimensional legendrian minimizers is thus reduced to the question of whether or not the 2-dimensional $(p, p+1)$ cones are minimizing among $\mathbb{Z}_2$ legendrian comparisons. (We saw above in Theorem 1.3 that they are not minimizing among $\mathbb{Z}_2$ lagrangian comparisons. However the comparisons used in the proof of Theorem 1.3 have nontrivial periods and are therefore not legendrian comparisons.)

In [S-W] it is shown that at least one of the $(p, p+1)$ cones is minimizing among orientable lagrangian comparisons. Why such a cone exists and why it is plausible to believe that such minimizing cones do not exist among nonorientable legendrian comparisons can be seen as a consequence of the following topological reasoning: Let $\ell : \Sigma \to N$ be a lagrangian immersion where Σ is a surface (orientable or not) and N is a symplectic 4-manifold. Then there is a splitting $\ell^* TN \simeq T\Sigma \oplus T^*\Sigma$ of $\ell^* TN$ into a pair of lagrangian subbundles over Σ. But then,

$$\ell^* TN \simeq T\Sigma \oplus T^*\Sigma \simeq T\Sigma \otimes \mathbb{C}, \tag{2.1}$$

where the isomorphisms are of symplectic (or equivalently almost complex) vector bundles over Σ. It follows that $c_1(\ell^* TN) \in H^2(\Sigma, \mathbb{Z})$ is an element of order 2 and so vanishes if Σ is orientable. Suppose $\alpha \in H_2(N, \mathbb{Z})$ is a lagrangian homology class and $c_1(N)(\alpha) \neq 0$. Then the minimizer among orientable lagrangians representing α cannot be a (branched) immersion and hence at least one cone singularity must occur on the minimizer. Thus at least one of the $(p, p+1)$ cones must be a minimizer among oriented lagrangian comparisons. On the other hand if Σ is nonorientable then $c_1(\ell^* TN)$ need not vanish. In fact since it is an element of order 2 we can replace $c_1(N)$ with its mod 2 reduction, $w_2(N)$, the second Stiefel–Whitney class. Using (2.1) and the Whitney sum formula we have:

$$\begin{aligned} \ell^* w_2(N) &= w_2(T\Sigma \oplus T^*\Sigma) \\ &= w_2(T\Sigma) + w_1(T\Sigma) \cdot w_1(T^*\Sigma) + w_2(T^*\Sigma) \\ &= w_1^2(\Sigma). \end{aligned}$$

Now suppose $\alpha \in H_2(N, \mathbb{Z}_2)$ is a legendrian homology class and $w_2(N)(\alpha) \neq 0$. If the minimizer among $\mathbb{Z}_2$ legendrians representing α is immersed, the Stiefel–Whitney number w_1^2 of the minimizer must be nonzero. What was an obstruction to regularity in the orientable case becomes a restriction on topology in the nonorientable case. Philosophically, singularities in the orientable case are replaced with Möbius bands in the nonorientable case. The problem is to show that this philosophy is correct. If this can be done, the introduction of legendrian $\mathbb{Z}_2$-currents will be justified.

References

[A] Allcock, D., An isoperimetric inequality for the Heisenberg groups, GAFA **8** (1998), 219–233.

[F] Federer, H., The singular sets of area minimizing rectifiable currents with codimension one and of area minimizing flat chains modulo two with arbitrary codimension, Bull. AMS **76** (1970), 767–771.

[Fl] Fleming, W., Flat chains over a finite coefficient group, Trans. AMS **121** (1966), 160–186.

[G] Gromov, M., Carnot–Carathéodory spaces seen from within *Sub-Riemannian Geometry*, 79–323, Progress in Math. 144, Birkhäuser, Basel, 1996.

[K] Kostant, B., Quantization and unitary representations, Lecture Notes in Mathematics 170, Springer, New York.

[Q] Qiu, W., Non-orientable lagrangian surfaces with controlled area, Math. Research Letters **8** (2001).

[Sc] Schoen, R., Special lagrangian submanifolds, pp. 655–666 in this volume.

[S-W] Schoen, R., and Wolfson, J., Minimizing area among lagrangian surfaces: The mapping problem, J. Diff. Geom. **58** (2001), 1–86.

[Si] Simon, Leon, Lectures on Geometric Measure Theory, Proc. of the Centre for Math. Analysis 3, Australian National University, 1983.

Department of Mathematics, Michigan State University, East Lansing, MI 48824
E-mail address: wolfson@math.msu.edu

Clay Mathematics Proceedings
Volume 2, 2004

Minimal Surfaces and the Topology of Three-Manifolds

Joel Hass

Abstract. This article, based on three lectures given at MSRI in July 2001, discusses old and new results relating minimal surface theory and the topology of 3-manifolds. The goal is to give a picture of how minimal surfaces have become an important tool in the study of 3-manifolds.

1. Introduction

In this paper we discuss some old and new results relating minimal surface theory and the topology of 3-manifolds. We have selected only a few of the many wonderful results in this area, but hopefully enough to give the flavor of its ideas. For the sake of exposition we keep the discussion somewhat informal, and don't try to give completely general statements. Detailed proofs and general statements can be found in the cited references. The article is based on three lectures given in July, 2001 at the Clay Mathematics Institute 2001 Summer School on the Global Theory of Minimal Surfaces.

Throughout this paper, F will refer to a compact surface and M to a closed 3-dimensional Riemannian manifold, one that is compact and has no boundary. All maps and spaces will be assumed C^∞ unless otherwise noted.

2. Existence

Establishing the existence of least area surfaces was an important problem in the 19th and 20th centuries. The modern formulation is usually ascribed to the Belgian physicist Plateau [**55**], who studied soap films experimentally and instigated the exploration of their mathematical properties. The Plateau problem had various formulations, depending on how close to the soap bubble model one wanted to stay. A common version was stated as follows:

Plateau's Problem: Let Γ be a rectifiable, simple closed curve in $\mathbb{R}^3$. Show that Γ bounds a smoothly immersed spanning disk of least area.

1991 *Mathematics Subject Classification.* Primary 57N10, 53A10; Secondary 68Q25.

Key words and phrases. Minimal surface, 3-manifold, complexity of algorithms, normal surface, area.

Partially supported by NSF grant DMS-0072348.

©2004 Clay Mathematics Institute

To formulate this more precisely, define $\mathcal{F}$ to be the family of piecewise smooth maps $f : D^2 \to R^3$ such that $f|_{\partial D^2}$ is a homeomorphism from ∂D^2 to Γ. Let $\mathcal{I} = inf\{Area(f) : f \in \mathcal{F}\}$. The problem is to find a map in $\mathcal{F}$ with area $\mathcal{I}$.

This statement allows for solutions that are immersed disks, possibly self-intersecting, and thus is not a faithful model of the behavior of soap films in space. However this formulation is well suited to the study of 3-manifolds, as well as to solution by analytic methods.

There were many partial results, and a complete solution in $\mathbb{R}^n$ was obtained by Douglas and Radó around 1930. See [**58**] for a historical summary. The two techniques are rather different. Radó developed a method based on minimizing energy rather than area. The energy of a map $f(u_1, u_2)$ from the disk to $\mathbb{R}^k$ is given by the Dirichlet integral,

$$\int\int_D \sum_{k=1}^{n} \Big(\frac{\partial f_k}{\partial u_1}\Big)^2 + \Big(\frac{\partial f_k}{\partial u_2}\Big)^2 \, du_1 \, du_2 \ .$$

Minimizing with respect to fixed boundary values parameterizing Γ leads to a harmonic map. Minimizing energy among all boundary values monotonically parameterizing Γ gives a least area map. The resulting map is conformal, except possibly at a finite number of singularities, called *branch points*. See [**7**], [**47**] for a more extensive discussion. Douglas took a different approach, and minimized what is now called the Douglas Integral [**9**]. This proved harder to extend to general metrics and topologies than the Dirichlet integral, though Douglas achieved some success in this direction.

Morrey extended the existence results for minimal disks to null-homotopic curves in Riemannian manifolds that are homogeneously regular [**46**].

THEOREM 2.1 (Douglas, Rado, Morrey). *Let Γ be a rectifiable, null-homotopic, simple curve in a compact Riemannian 3-manifold. Then Γ bounds a disk of least area.*

The assumption that the curve is embedded can be removed [**23**].

There is an alternative approach to existence results based on Geometric Measure Theory, completed for surfaces in 3-dimensional manifolds by Federer and Fleming [**12**], [**11**]. With further work of Hardt and Simon [**20**] on boundary regularity, this gave a solution of a version of Plateau's problem more closely modeling a physical soap film. We will concentrate on the previous formulation, which is well suited to investigation of problems in low-dimensional topology.

We fix some terminology. An immersion $f : F \to M$ of a compact surface to a Riemannian manifold is *minimal* if it has zero mean curvature. More generally, we consider minimal surfaces that fail to be immersions at isolated branch points. A map of a surface is *least area* if it has no larger area than any homotopic map. A map of a surface is *area minimizing* if it is homologically minimizing, so that the area of its image is no larger than that of any surface in the same homology class. If the surface is noncompact, we can use the same definitions, but two competing surfaces are required to agree off a compact subset.

3. Branch Points

Minimal surfaces in $\mathbb{R}^n$ are immersed except at isolated singular points, where their structure can be understood through the properties of harmonic functions,

which form the coordinate functions. These surfaces, which include the solutions to the Plateau problem in $\mathbb{R}^3$ produced by Douglas and Radó, are "conformal branched immersions".

For least area surfaces in $\mathbb{R}^3$ it was unknown whether singularities actually do occur. Osserman showed that branch points do not exist in the interior of least area surfaces. He established that least area surfaces have immersed images, using an argument similar to one used by Dehn in his study of generic maps of surfaces into 3-manifolds [**48**], [**8**]. Osserman's work was supplemented by further contributions of Gulliver and Alt, who established also that no "false" branch points can occur in a least area surface [**15**], [**3**]. These are singularities of a map due to a bad parameterization rather then a singularity in the image, These methods apply equally well in a general Riemannian 3-manifold.

THEOREM 3.1. *A least area surface in a Riemannian 3-manifold is immersed in its interior.*

IDEA OF PROOF. Suppose a least area surface has an interior branch point. The local self-intersections of the surface near a branched point can be analyzed, and reveal that there is a curve of double points leading into the branched point. Cutting and pasting along that double curve near the branch point leads to a homotopic surface of the same genus and area. The resulting surface has a "fold" along the double curve, a curve of points where it fails to be smoothly immersed. But minimal surfaces have only isolated singularities, and so there is a smaller surface in the homotopy class, contradicting the least area hypothesis. □

The possible existence of branch points at the boundary of a least area surface remains open, though important special cases have been shown to be free of such boundary singularities. The structure of a boundary branch point is described in [**17**]. In [**16**], Gulliver constructs a minimal surface with a boundary branch point which cannot be eliminated using the methods that were used for interior branch points. It is known that boundary branch points do not exist in many settings, such as when a curve lies on the boundary of its convex hull, or is analytic [**17**]. To the extent that this question remains open, the version of the Plateau problem stated above is not completely settled.

4. Embeddedness

Osserman asked the following question: If Γ is a simple closed curve on the 2-sphere and D is a least area disk spanning Γ, is D embedded? This stimulated a series of important results. Special cases followed from work of Radó [**58**], Gulliver–Spruck [**18**], Tomi–Tromba [**69**] and Almgren–Simon [**2**]. A general positive solution was obtained by Meeks and Yau, who proved the following very general result [**41**].

THEOREM 4.1 (Meeks–Yau Dehn's Lemma). *If D is a least area disk whose boundary is a simple closed curve on the boundary of a Riemannian 3-manifold with mean convex boundary, then D is embedded.*

Mean convex is a generalization of convex that allows for manifolds whose boundaries have positive (or nonnegative) mean curvature. Mean convex is a sufficient condition to allow construction of a minimizing sequence whose elements don't bump against the boundary, a key step in establishing existence of area minimizers.

The Meeks–Yau Dehn's Lemma had profound implications in 3-manifold topology. It is not obvious that ∂D bounds any embedded disk, least area or not. That it does is the content of the somewhat notorious Dehn's Lemma, first stated by Dehn in 1920 [**8**]. Dehn's argument contained a gap, and the result was finally proved in 1957 by Papakyriakopoulos [**49**]. The existence of a least area embedded disk leads to a stronger, equivariant version of Dehn's Lemma. In this setting we have a finite group G acting on a 3-manifold M. By averaging a metric, we can assume that the group acts by isometries. A curve or surface in M is *equivariant* if its translates under the group are either disjoint from it or coincide with it.

THEOREM 4.2 (Meeks–Yau Equivariant Dehn's Lemma). *Suppose M is a Riemannian 3-manifold with mean convex boundary and G is a finite group acting as isometries on M. Suppose also that Γ is a null-homotopic curve on ∂M that is equivariant. If D is a least area disk with boundary Γ, then D is embedded and equivariant.*

As an application, we show how to classify $\mathbb{Z}_3$ actions on the solid torus

$$S^1 \times D^2 = \{(\theta_1, \theta_2, r) : 0 \leq \theta_1 < 2\pi,\ 0 \leq \theta_2 < 2\pi,\ 0 \leq r \leq 1\}.$$

An action is *standard* if it is conjugate to an action where a generator τ of $\mathbb{Z}_3$ rotates one or both circles of the torus. That is $\tau(\theta_1, \theta_2, r) = (\omega\theta_1, \mu\theta_2, r)$ where ω, μ are cube roots of unity.

THEOREM 4.3. *Suppose $\mathbb{Z}_3$ acts on $S^1 \times D^2$. Then the action is standard.*

SKETCH OF PROOF. Take any metric on $S^1 \times D^2$. Average it over the group action to get a new metric on which the group acts as isometries. Let β be a geodesic on the boundary torus that is shortest among all nontrivial curves on the torus that bound disks in the solid torus. The three translates β, $\tau(\beta)$, $\tau^2(\beta)$ are embedded and either disjoint or equal, as otherwise one could find a homotopic curve shorter than β. Take a least area disk D with boundary β. By the Meeks–Yau Equivariant Dehn's Lemma, the three translates β, $\tau(D)$, $\tau^2(D)$ are embedded and either disjoint or equal. Together they split the solid torus into either one or three balls. If three, the three balls are rotated one to another by τ. If one, the action of the group takes this ball to itself. Actions of $\mathbb{Z}_3$ on the 3-ball were shown to be standard in the proof of the Smith Conjecture, by a combination of results including another application of the Meeks–Yau Dehn's Lemma [**45**]. It follows in each case that the action on the solid torus is standard. □

Much more general groups are shown to act standardly on many 3-manifolds by similar techniques [**39**].

The basic idea in proving the embeddedness results such as Theorem 4.1 is to find "parallel" subsurfaces of an immersed disk to exchange, by cutting and pasting along suitable double curves. The study of such double curves is a notoriously difficult area in 3-manifold topology. If one is fortunate enough to have a "product region" as in Figure 1, then one can interchange the two pieces, smooth off, and reduce area without changing topology or homotopy class. In Figure 1 the product region is a ball, homeomorphic to a disk $\times$ I. It is not clear that one can always find such regions in an immersed surface with excess intersections, and in fact one sometimes cannot. The difficulty arises from immersions with intersecting double curves, which lead to triple points where three sheets of the surface cross. These

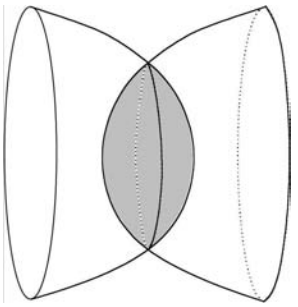


FIGURE 1. A product region cannot occur in a least area immersion.

triple points are difficult to analyze. A covering space technique introduced by Papakyriakopoulos allows us to avoid the problem. A sequence of covering spaces of neighborhoods of the immersed disk is constructed that unwrap the immersed disk until one of two things happens. Either a cover is constructed in which a lift to that cover does indeed contain a product piece, or a smaller area disk is constructed in some cover by taking a component of the boundary of a slight thickening of the immersed disk. In either case, a smaller disk can be constructed in a cover, and projected down to give a smaller disk in the original space. By constructing a smaller area surface with the same boundary, one arrives at a contradiction if the original disk was not embedded.

5. Closed Surfaces

Closed minimal surfaces, compact and without boundary are not present in $\mathbb{R}^3$, but they do play a key role in the study of 3-manifolds. Since they don't have boundary, we need to impose some topological condition to stop a minimizing surface from collapsing to a point. We first look at minimal spheres, which are treated somewhat differently from higher genus surfaces. We consider a closed Riemannian 3-manifold that has nonzero second homotopy group $\pi_2(M)$. An existence theorem in this context was established by Sacks and Uhlenbeck. Let $\mathcal{F}$ be the family of smooth maps $f : S^2 \to M$ such that f is not homotopic to a point, or equivalently, f represents a nontrivial element of $\pi_2(M)$. (We are suppressing mention of a base point for simplicity.) Set $\mathcal{I} = \inf\{\text{Area}(f) : f \in \mathcal{F}\}$.

Sacks and Uhlenbeck showed that while a particular element of the second homotopy group may or may not contain a least area representative, there exists an element of smallest area among *all* nontrivial 2-spheres.

THEOREM 5.1 (Sacks–Uhlenbeck). *If $\pi_2(M)$ is nontrivial, there is map in $\mathcal{F}$ having area $\mathcal{I}$.*

The question of whether *every* representative of $\pi_2(M)$ has a minimal representative, asked by Eells and Lemaire [**10**, pp. 417], remains open. This is equivalent to asking whether there is a harmonic representative in every homotopy class, since the notion of harmonic map and minimal surface coincide for a 2-sphere domain.

OPEN PROBLEM 1. *Let M be a closed Riemannian manifold and $f : S^2 \to M$ a map that is homotopically nontrivial. Does f have a minimal representative?*

Probably the answer is no. One approach to constructing a Riemannian 3-manifold with a homotopy class containing no minimal representative is to take a 3-manifold and remove two small balls far away from each other. Then glue in a nontrivial 3-manifold, such as a punctured 3-torus, to replace each of the removed balls. Consider a minimal 2-sphere representing the homotopy class given by the sum in $\pi_2(M)$ of the boundaries of the two removed balls. It seems unlikely that a minimal 2-sphere could span the gap between the balls if the metric is arranged so that they are sufficiently far away from each other. This can be made rigorous for manifolds with convex boundary. For example, construct a Riemannian manifold by taking a unit ball in $\mathbb{R}^3$, removing radius ϵ balls around $(-1/2, 0, 0)$ and $(1/2, 0, 0)$, with ϵ small, and changing the metric near these balls to make a convex boundary. Then no minimal 2-sphere can enclose both the balls. A catenoid of appropriate size would be tangent to such a 2-sphere while not crossing it, contradicting the maximum principle for minimal surfaces [**56**]. In a closed manifold, an example of a homotopy class of 2-spheres with no minimal representative has not yet been found.

There is an important topological result about embedding spheres in 3-manifolds, the Sphere Theorem, also due to Papakyriakopoulos [**49**]. It states that if $\pi_2(M)$ is nontrivial there is an embedded, homotopically nontrivial 2-sphere in M, and its proof is similar in flavor to that of Dehn's Lemma, though more involved. Meeks and Yau extended the Sacks–Uhlenbeck Theorem, and proved least area and equivariant versions of the Sphere Theorem [**41**].

THEOREM 5.2 (Meeks–Yau Sphere Theorem). *If M is a Riemannian manifold which is closed or compact with mean convex boundary, and if $\pi_2(M)$ is nontrivial, then there is a generating set of $\pi_2(M)$, as a $\pi_1(M)$-module, consisting of minimal 2-spheres. Each 2-sphere is least area in its homotopy class, is either embedded or double covers an embedded RP^2, and any pair of 2-spheres is disjoint.*

Given a finite group acting as isometries on a 3-manifold, the Meeks–Yau sphere theorem has an equivariant version, similar to the equivariant Dehn's Lemma. This led to further new results on classifying group actions on 3-manifolds, including groups acting on $\mathbb{R}^3$ [**40**].

Another important result that had implications to 3-manifold topology was stated by W. Meeks, L. Simon and S.-T. Yau [**43**]. Its proof is based more on Geometric Measure Theory than the other results we've discussed.

THEOREM 5.3. *Let F be an embedded surface in a Riemannian 3-manifold M. Then after a series of compressions, isotopies, and collapsings of the boundaries of I-bundles to their cores, F can be realized as a union (possibly empty) of disjoint embedded minimal surfaces.*

The phenomenon of collapsing I-bundles can be seen, one dimension down, in the Möbius band, where a curve isotopic to the boundary can be homotoped to double cover a central core curve. With an appropriate Riemannian metric, the shortest curve can be forced to double cover the core. In three dimensions this can occur even when the surface and the 3-manifold are both orientable.

A 3-manifold is called *irreducible* if any 2-sphere is the boundary of a 3-ball, so that M contains no nontrivial embedded 2-spheres. Cutting along such nontrivial 2-spheres is an important first step in the (not yet completed) classification of

3-manifolds. Theorem 5.3 has the following important consequence, obtained by minimizing among all spheres not bounding balls [**43**].

THEOREM 5.4 (Meeks–Simon–Yau). *Let M be an irreducible orientable 3-manifold M and $\tilde{M}$ a covering space of M. Then $\tilde{M}$ is irreducible.*

We now turn to higher genus surfaces. A common approach in trying to understand 3-manifolds is to cut them open along surfaces into simpler building blocks, and to understand the ways that these pieces are recombined to form the original manifold. For the process to be useful, the cutting surfaces should reflect the global nature of the 3-manifold, as otherwise the building blocks can be more complicated than the original manifold. Experience shows that it is counterproductive to cut open along a surface with lots of knotted tubes, or with complicated self-intersections. Two classes of surfaces are commonly used to cut up 3-manifolds. *Incompressible surfaces* are embedded and contain no trivial tubes or handles, while *Heegaard surfaces* cut the 3-manifold into two handlebodies. Even with these surfaces, careful choices must be made for the cutting procedure to be useful. See the article of Rubinstein in this volume for a discussion of Heegaard minimal surfaces [**60**]. We look here at incompressible (also called π_1-*injective*) surfaces, defined to be surfaces (other than 2-spheres) whose fundamental groups inject into the fundamental group of M. The basic existence theorem for this class of surfaces was established by Schoen–Yau [**63**].

THEOREM 5.5 (Schoen–Yau). *If a closed Riemannian manifold M contains an incompressible surface of genus $g \geq 1$, then a smallest such surface exists and is immersed.*

The proof of the Schoen–Yau theorem actually shows that a least area surface exists in the homotopy class of each incompressible surface, if $\pi_2(M) = 1$. An embedding theorem for maps of incompressible surfaces was proven by Freedman–Hass–Scott [**13**].

THEOREM 5.6. *The least area surface homotopic to an embedded, incompressible, orientable surface in a closed, orientable, irreducible Riemannian manifold is either embedded or double covers an embedded one-sided surface.*

More generally, a collection of such least area surfaces is shown to minimize its self-intersections [**13**]. These intersection properties were not previously known even for shortest geodesics on a surface with a general metric. Applications to 3-manifolds of this result include a theorem of Peter Scott about the topological rigidity of Seifert Fiber spaces [**65**]. This class of 3-manifolds includes manifolds that contain immersed, but not embedded, incompressible tori. Waldhausen had proved in [**71**] that a 3-manifold M containing an incompressible embedded surface is *topologically rigid*, that is, any irreducible 3-manifold with isomorphic fundamental group is homeomorphic to M. Waldhausen's methods failed for Seifert Fiber spaces that contain immersed, but not embedded, incompressible tori, but Scott was able to extend the result to these manifolds through the use of least area immersed tori.

Another important theorem of Waldhausen states that if an irreducible manifold contains an embedded, incompressible surface, the universal cover of the manifold is homeomorphic to $\mathbb{R}^3$. This was extended by Hass–Rubinstein–Scott using

least area techniques, removing the embeddedness condition from Waldhausen's assumptions [**28**].

THEOREM 5.7. *The universal cover of an irreducible manifold containing an immersed, incompressible surface is homeomorphic to* $\mathbb{R}^3$.

6. Some Open Problems on Minimal Surfaces in 3-Manifolds

We discuss here a small number of open problems relating minimal surface theory and 3-dimensional topology. More extensive lists of problems can be found in [**38**] and [**73**].

The famous Poincaré Conjecture asserts that a closed, simply connected 3-manifold is homeomorphic to the 3-sphere. This is equivalent (by removing a ball) to the claim that a compact 3-manifold that is simply connected and has 2-sphere boundary is homeomorphic to a 3-ball. A minimal surface approach to the Poincaré conjecture was suggested in the 1980s by M. Freedman and S.-T. Yau. We will sketch this plan, which has not been completed to date.

Suppose E is a *fake 3-ball*, a simply connected 3-manifold with boundary a 2-sphere that is not homeomorphic to a ball. Take a Riemannian metric on E in which the boundary is convex and in which there are only finitely many disjoint embedded minimal 2-spheres. Such a metric exists by a theorem of White [**72**]. The boundary 2-sphere cannot be isotoped to a point in E, so it is isotopic to a stable minimal 2-sphere by Theorem 5.3. Take a stable minimal 2-sphere not bounding a ball that does not contain any other such 2-sphere inside it, and let N be the 3-manifold bounded by this 2-sphere. Then N is also a fake ball with smooth, mean convex boundary. Now consider a curve γ on ∂N. The two disks on ∂N bounded by γ are each stable minimal disks. We expect to be able to find an unstable, immersed minimal disk in the interior of N that lies on a family of disks connecting these two stable disks.

OPEN PROBLEM 2. *Among all such curves* γ, *is the smallest area unstable minimal disk embedded?*

This looks intriguingly similar to the Meeks–Yau Dehn's Lemma, which states that a least area minimal disk exists and is embedded for any such γ.

If the answer is yes, then N is divided into two pieces, with common boundary the embedded minimal disk. Each of these regions is simply connected, but at least one is not a ball, since N is not a ball and the union of two balls along a disk is also a ball. Let P be a component that is a fake ball. The boundary of P is mean convex, so we can apply the Meeks–Simon–Yau Theorem to find a stable minimal 2-sphere inside P, contradicting our hypothesis that ∂N was innermost among such 2-spheres. The Poincaré conjecture would follow.

We now turn to a collection of questions concerning one-sided surfaces. A surface immersed in a 3-manifold is *one-sided* if its normal bundle is nontrivial. Otherwise it is *two-sided*. A map of a surface $f : F \to M$ is π_1*-injective* if it induces an injective homomorphism from $\pi_1(F)$ to $\pi_1(M)$.

CONJECTURE 3. *Let* $f : F \to M$ *be a one-sided, least area,* π_1*-injective map of a closed surface* F *to an irreducible Riemannian 3-manifold* M. *If* f *is homotopic to an embedding,* f *is embedded.*

This is known for two-sided surfaces [**13**] (up to the phenomenon of double covering a one-sided embedding) and for both one-sided and two-sided geodesics on surfaces [**27**]. It follows from Theorem 5.3 that there is a one-sided surface homotopic to f that is embedded or a 2-1 cover of an embedding and that minimizes area among all such surfaces. This surface minimizes area in the isotopy class of f (allowing maps that factor through double coverings to be included in this class). However it is not known whether this surface minimizes area in its homotopy class. There exists a least area map in the homotopy class of f by [**63**], and conceivably in some manifolds there is a least area map whose image crosses itself that is smaller than any embedding.

Closely related is the following:

CONJECTURE 4. *Let $f : F \to M$ be a one-sided map from a closed surface F to a Riemannian manifold M that induces an isomorphism of fundamental groups. If f has smallest area among such maps, then f is embedded.*

Here we assume that f induces an isomorphism of fundamental groups, rather than just a π_1-injective homomorphism. However we do not assume that f is homotopic to an embedding, nor that M is irreducible. Existence of a least area map is known as is the existence of a least area embedding among all embeddings [**63**], [**43**]. Conjecture 4 has some interesting implications, including the topological Conjecture 5. Conjecture 4 implies that if a 3-manifold has fundamental group $\mathbb{Z}_2$, there is a least area embedded RP^2 inside it. If the manifold is also irreducible, the boundary of a neighborhood of this RP^2 is a 2-sphere that bounds a 3-ball, necessarily on the other side. The union of a neighborhood of the embedded RP^2 and the 3-ball is homeomorphic to RP^3, which implies a solution to the following open problem in 3-manifold theory.

CONJECTURE 5 (No fake irreducible RP^3's). *If M is an irreducible 3-manifold homotopy equivalent to RP^3, then M is homeomorphic to RP^3.*

Conjecture 5 follows from Conjecture 4, since a minimal RP^2 would be embedded, and irreducibility would then imply that M is homeomorphic to RP^3.

A related topological conjecture is:

CONJECTURE 6. *Every homotopy 3-sphere admits a free involution.*

This is certainly true for the 3-sphere, and thus follows from the Poincaré conjecture. Together with Conjecture 5, Conjecture 6 would imply the Poincaré Conjecture in dimension 3.

Closely related is the following conjecture.

CONJECTURE 7. *Let M be irreducible and homotopy equivalent to RP^3, with the boundary of M consisting of two copies of RP^2. Let A be a least area annulus whose boundary curves are generators of the fundamental group of each of the two boundary RP^2's. Then A is embedded.*

The above manifold gives an h-cobordism of RP^2 to itself. A corollary of Conjecture 7 is the following topological conjecture.

CONJECTURE 8. *There are no fake h-cobordisms of RP^2. Equivalently, there is no fake irreducible $RP^2 \times I$.*

The following case of the above type of question is perhaps the simplest, but still seems quite hard.

CONJECTURE 9 (Meeks). *Let M be a Riemannian manifold diffeomorphic to $S^1 \times D^2$, the solid torus, with mean convex boundary. Let F be a least area Möbius band bounding an embedded $(2,1)$ curve on the boundary torus. Then F is embedded.*

Even the special cases where M is a submanifold of $\mathbb{R}^3$, where M has the product metric, or where M is the standard solid torus of revolution are unknown. Some special cases that can be solved are when the boundary curve is invariant under a circle action or when the total curvature of the boundary curve is at most 6π. Two other special cases of Conjecture 4 are the following.

CONJECTURE 10. *Let $f : RP^2 \to RP^3$ be a least area π_1-injective map, where RP^3 is given an arbitrary Riemannian metric. Then f is an embedding.*

This conjecture has a generalization to more complicated types of surfaces. A generalized projective plane is a disk except for a curve of singularities along its boundary where locally p-surfaces meet, see [**5**]. It is not known in what generality least area singular surfaces of this type exist.

OPEN PROBLEM 11. *Let S be a least area, π_1-injective generalized projective plane in a homotopy lens space. Then S is embedded.*

Also of interest are non-π_1-injective incompressible surfaces. These embedded surfaces contain immersed curves that bound disks with embedded interiors in their complement, but no embedded curves of this type. They are always one-sided. Examples occur in some lens spaces and many other 3-manifolds, and are of interest in the study of 3-manifolds containing no incompressible surfaces.

OPEN PROBLEM 12. *Let F be a non-π_1-injective incompressible surface in a Riemannian 3-manifold M. Is a least area surface homotopic to F embedded?*

7. Minimal Foliations

Given a foliation of a 3-manifold M, under certain conditions it is possible to find a Riemannian metric on M in which every leaf is a minimal surface. This leads to interesting results in both 3-manifolds and minimal surface theory [**22**].

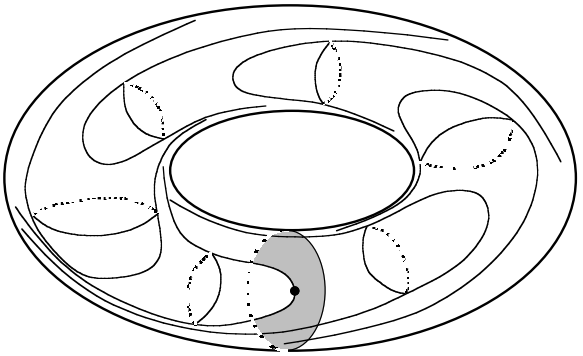


FIGURE 2. The existence of both minimal leaves and a minimal disk violates the maximum principle.

One obstruction to the existence of such a metric is the existence of Reeb components. Suppose that each Reeb leaf is a minimal surface in some metric,

including the boundary torus. Consider a least area compressing meridian disk for the solid torus. Such a disk can be obtained by solving the Plateau problem for its boundary, since the boundary of the solid torus is mean convex. The disk must be tangent to one of the Reeb leaves at some point, while lying locally to one side of this leaf. This violates the maximum principle for minimal surfaces [**56**] [**38**].

The absence of Reeb components is almost sufficient to allow the construction of a metric in which leaves are minimal. A closed curve in M that is never tangent to a leaf of F is called a *closed transversal.* The necessary condition was given by Sullivan [**67**]: every compact leaf of the foliation intersects some closed transversal. Such a foliation is called *taut.* See also Harvey and Lawson's work on calibrations [**21**].

THEOREM 7.1. *Let M be a closed 3-manifold and let F be a codimension-one foliation of M. Then the following conditions are equivalent:*

(1) *Every compact leaf of F intersects a closed transversal.*
(2) *M admits a metric in which each leaf of F is a minimal surface.*
(3) *M admits a metric in which each leaf of F is area minimizing in its homology class.*

If the leaves are noncompact, area minimizing in a homology class means that each compact subsurface is minimizing in its homology class (rel boundary).

Sometimes such a taut metric is known, for example in the case of a product metric on a $\{\text{surface}\} \times S^1$. But often it is mysterious. An intriguing case is the hyperbolic one. It is known that there are many hyperbolic manifolds that fiber over S^1 with fiber a surface. Such a 3-manifold has a taut foliation consisting of all the fibers, and therefore has a metric in which each leaf is minimal.

OPEN PROBLEM 13. *Is there a 3-manifold that fibers over the circle in which each fiber is a minimal surface in the hyperbolic metric?*

Thurston observed that the fibers of some hyperbolic bundles over the circle cannot all be minimal surfaces. Namely, if there is a very short geodesic in the manifold, then a leaf going into the Margulis tube around this geodesic could not be minimizing in its homology class. A calculation shows that a smaller area homologous surface could be formed by cutting across a torus around the short geodesic, which looks something like a horotorus. However existence of a minimal foliation would force minimizing leaves into the Margulis tube.

An equivalent question asks whether there can be a family of stable minimal surfaces in a hyperbolic 3-manifold. If so, the family forms an analytic set [**4**], that closes up to create a bundle of minimal surfaces in some covering space.

8. The Spherical Space-Form Problem

The classification program for 3-manifolds starts by cutting a 3-manifold open along 2-spheres. The resulting prime pieces are uniquely determined [**36**] [**44**]. Understanding the pieces with finite fundamental group has two aspects:

(1) What are the simply connected closed 3-manifolds?
(2) What are the manifolds which arise as their quotients?

The former problem is the Poincaré Conjecture and the second is the Spherical Space-Form Problem. The Spherical Space-Form Problem itself has two parts:

(1) What are the groups that can act on S^3 (or on a homotopy three sphere)?

(2) Is the action of a group on S^3 standard, i.e., conjugate to a fixed point free action of the rotation group $SO(4)$ on the round 3-sphere?

There are many restrictions known on the groups that can act freely on a homotopy 3-sphere, but a classification is still not achieved [**35**]. We will consider here the aspect of the problem that asks whether a subgroup of the rotation group $SO(4)$ that does act freely on S^3 always acts standardly. It suffices to consider cyclic subgroups to deal with most group actions, so we will restrict to this case. Livesay proved that $\mathbb{Z}_2$ acts standardly on S^3, and Rubinstein extended this to $\mathbb{Z}_{2^k}$ and $\mathbb{Z}_{3^k 2^l}$ for all $k, l \in \mathbb{N}$. An action of $\mathbb{Z}_5$, however, is still not known to be standard.

The problem admits an attack by minimal surface theory, as proposed by Pitts and Rubinstein. The Pitts–Rubinstein program involves finding an unknotted curve γ that is taken to itself by each element of the group. If such an invariant curve exists, then a small neighborhood of the curve is an invariant unknotted solid torus, since the action is smooth. Since γ is unknotted, the complement of this solid torus is also a solid torus. Group actions on a solid torus are understood, via the Meeks–Yau Dehn's Lemma and the proof of the Smith Conjecture [**40**] [**45**], as we saw in Theorem 4.3. This leads to a proof that the action on the 3-sphere is standard.

Instead of searching for an invariant curve, we can look for an invariant solid torus, or for an invariant unknotted torus. Pitts and Rubinstein pointed out that if there is no invariant, unknotted torus, then one expects to find a sequence of unknotted minimal tori in the 3-sphere with unbounded index. Although this argument is not worked out completely, the idea is roughly as follows: By averaging a metric we can assume that the group acts as isometries. The space of unknotted tori $\mathcal{T}$ has known homotopy type, from the work of Hatcher on the Smale Conjecture [**30**]. If there is no fixed point of the group G acting on $\mathcal{T}$, then $\mathcal{T}/G$ has cohomology in infinitely many dimensions, and this in turn implies that a Morse function on $\mathcal{T}$ has critical points of unbounded index. The area function is defined on $\mathcal{T}/G$, and accordingly is expected to have critical points of unbounded index. More explicit arguments for why these higher index surfaces should exist can be made, though a rigorous proof is not yet available. If it were, the standardness of many group actions on the 3-sphere would be implied by a positive answer to the following question.

Open Problem 14. *Given a metric on the 3-sphere, is there an upper bound on the index of the unknotted, embedded, minimal tori in the 3-sphere?* [1]

9. Normal Surfaces

In the piecewise linear (PL) context, where one has a triangulated 3-manifold, an attempt to push the surface around until it becomes as simple as possible gives rise to what is called a *normal surface*. Normal surfaces are discrete analogs of minimal surfaces.

A triangulated 3-manifold is a decomposition of a 3-manifold into a union of tetrahedra, that intersect one another along lower-dimensional simplices. We do not

[1] Since this paper was written there has been progress on Open Problem 14. Colding and Hingston [**6**] have shown how to construct a metric with embedded minimal 2-tori of arbitrarily high Morse index on any 3-dimensional manifold. Hass, Norbury and Rubinstein [**25**] construct a smooth Riemannian metric on any 3-manifold with the property that there are genus zero embedded minimal surfaces of arbitrarily high Morse index.

restrict to a combinatorial triangulation, so it's not forbidden for two tetrahedra to intersect along several faces or edges.

Definitions: *Normal triangles* are disks in a 3-simplex that meet three edges and three faces of the 3-simplex, and *normal quadrilaterals* are disks in a 3-simplex that meet four edges and four faces of the 3-simplex. An *elementary disk* is a normal triangle or quadrilateral. A *normal surface* in a triangulated 3-manifold is an embedded surface in M that intersects each 3-simplex in a disjoint union of elementary disks.

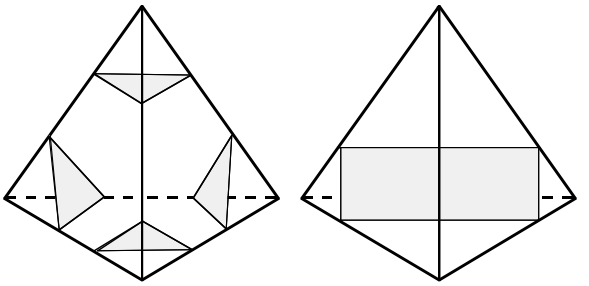


FIGURE 3. Pieces of a normal surface are triangles and quadrilaterals.

All four types of normal triangle can coexist disjointly in a 3-simplex. However as soon as one type of normal quadrilateral is around, the other two types of normal quadrilateral cannot be present, or else an intersection is forced.

A *compression* of a surface is an operation that squeezes away a nontrivial handle. Specifically, suppose that there is a disk whose interior misses the surface and whose boundary is a homotopically nontrivial curve on the surface. A thickening of this disk has boundary an annulus on the surface and two disks in the complement of the surface. A compression removes the annulus from the surface and replaces it with the two disks, resulting in a surface whose genus has dropped.

Any embedded surface can be compressed and isotoped to a union (possibly empty) of normal surfaces. To see this we introduce a measure of how complex a surface is relative to a given triangulation. The *weight* $w(F)$ of a surface F is the number of times it intersects the 1-skeleton of M. Weight gives an analog of area in the PL context. It gives the area in the limiting case when all area measure is concentrated near the 1-skeleton. Starting with an arbitrary surface, Kneser showed that one can carry out a series of compressions and weight decreasing isotopies until the surface becomes normal or empty.

THEOREM 9.1. *Let F be an embedded surface in M. Then after a series of compressions, isotopies and removal of trivial 2-spheres, F becomes isotopic to a union (possibly empty) of disjoint normal surfaces.*

This result, implicit in the work of Kneser [**36**] and Haken [**19**], is the PL analog of Meeks–Simon–Yau's Theorem 5.3, except that in the Riemannian setting the metric may force the minimizer to be a double cover, while in the PL setting this type of collapse is unnecessary. Theorem 9.1 is proved by simplifying the intersections of a surface and the 1-skeleton, and compressing inside tetrahedra, until either the surface becomes normal or nothing is left [**24**].

Normal surfaces play a key role in the construction of algorithms to answer topological questions such as whether a knot is trivial and whether a 3-manifold is reducible. They are discrete objects described by a finite amount of data, the number of each type of triangle and quadrilateral that occurs in each tetrahedron. This data can be manipulated algebraically with the techniques of integer linear programming. In the 1990s this method was extended to solve an important problem, when Hyam Rubinstein discovered an algorithm to determine if a 3-manifold is homeomorphic to the 3-sphere [**59**]. A proof that this algorithm does indeed recognize the 3-sphere, based on thin position methods, was given by Abigail Thompson [**68**].

The idea behind the 3-sphere recognition algorithm is an insight from minimal surface theory. Suppose that M is a closed 3-manifold and we wish to determine whether it is a 3-sphere. We can assume that the homology of M is trivial, so that M is a homology sphere, as homology is straightforward to compute effectively. We then proceed as follows: Take a generic metric on M and pick a maximal collection of stable minimal 2-spheres in this metric. M is a 3-sphere if and only if each complementary piece is a ball, possibly with some further balls removed. If a complementary piece is a ball, then we can sweep it out by 2-spheres and find an unstable embedded 2-sphere, by a result of Simon and Smith, which in turn is based on earlier work of Pitts, Almgren and Simon [**50**] [**2**] [**66**]. See also Jost [**34**]. If the complementary piece is not homeomorphic to a ball, then it does not contain an unstable minimal 2-sphere. If it did we could push this unstable sphere to either side to get a slightly smaller 2-sphere. This smaller 2-sphere could then be deformed to a union of stable minimal 2-spheres or points by the Meeks–Simon–Yau Theorem. Any stable minimal 2-sphere lies on the boundary since our original collection was maximal. Thus this 2-sphere can be pushed out to the boundary on either side, and that implies the complementary piece is homeomorphic to a punctured 3-ball. Thus the problem of recognizing the 3-sphere could be solved if we knew all about the location of stable and unstable minimal 2-spheres in M.

In the Riemannian setting this is no easier, and maybe harder, than the original question. But in the PL setting, Rubinstein realized that stable and unstable minimal 2-spheres had PL analogs that could be found by a systematic procedure, leading to an algorithm. The analog of stable minimal surfaces were the normal surfaces. The analog of unstable minimal surfaces were modifications of normal surfaces introduced by Rubinstein, called *almost normal* surfaces. These look just like normal surfaces except in a single tetrahedron, where they either have an octagonal piece or a pair of normal disks connected by a tube. See Figure 4.

Schoen's curvature bound theorem states that a stable minimal surface in a Riemannian 3-manifold has principal curvatures that are bounded by a constant depending only on the geometry of the manifold [**62**]. It follows that an index-one minimal surface can have large principal curvature near one, but not two, points. Two disjoint neighborhoods, each containing a point of high principal curvature in a minimal surface, are each unstable, and lead to two independent regions of instability, implying that the index is at least 2. The octagon or tube correspond in the PL context to one of these regions of instability.

The above discussion attempts to give the intuition behind almost normal surfaces. Once the analogies and definitions are established, one can work purely in the PL context and show that

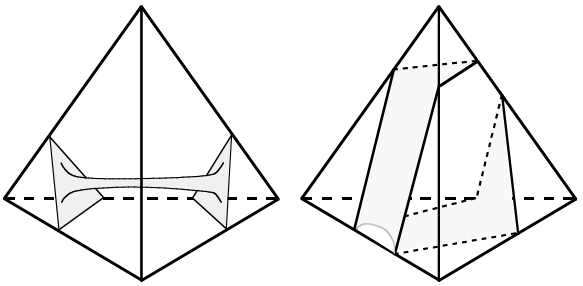


FIGURE 4. Pieces of an almost normal surface.

(1) There is a maximal collection of disjoint normal 2-spheres in a 3-manifold.
(2) The complementary regions of this collection are punctured balls if and only if they contain an almost normal 2-sphere.
(3) The search for normal and almost normal spheres admits an algorithmic solution following the ideas of Haken [**19**].

M is homeomorphic to a 3-sphere if and only if each complementary region is a punctured ball, so the above combine to give an algorithm to recognize the 3-sphere. A closely related question is still open:

OPEN PROBLEM 15. *Give an algorithm to determine if a 3-manifold is simply connected.*

10. Area and Computational Complexity

Recent work of Agol, Hass and Thurston [**1**] has shown an intriguing relation between the area of surfaces and the computational complexity of algorithms. The main result of [**1**] studies the complexity of algorithms to calculate the minimal genus of a spanning surface for a curve in a 3-manifold. The manifolds and curves are PL, allowing finite combinatorial descriptions. The 3-manifold is described as a collection of tetrahedra, with faces identified, and the curve as a collection of edges. Formally, the problem is described as a question with a yes or no answer.

Problem: 3-MANIFOLD KNOT GENUS
INSTANCE: A triangulated 3-dimensional manifold M, a knot K in the 1-skeleton of M,and a natural number g.
QUESTION: Does the knot K have genus $g(K) \leq g$?

In [**1**] it is shown that 3-MANIFOLD KNOT GENUS is NP-complete, and thus in the same complexity class as a large collection of extensively studied problems in other fields [**14**].

Area calculation problems seem at first to be less likely to fall into such classes, since area takes on real values that vary continuously as a surface is deformed. However we can recast the problem of determining minimal area into a form where its computational complexity can be measured. As with genus, we take a curve in a triangulated 3-manifold, but rather than ask whether it bounds a surface of genus bounded by k, we ask whether it bounds a surface of area less than k, where k is an integer. To describe a Riemannian metric on a 3-manifold with a finite amount of data requires limiting to a special class of metrics. We will use piecewise flat metrics, constructed from collections of tetrahedra and triangular prisms with flat

metrics whose faces are identified by isometries. The curvature of such metrics is concentrated along their edges and vertices. A particular manifold in this class is described by a triangulation with a length assigned to each edge. The edge lengths of each tetrahedron and prism are rational numbers, and agree for identified edges. We do not require that the total angle around an edge is 2π, nor do we make any metric conditions at a vertex. Such piecewise flat metrics are described by a finite set of data, and can continuously approximate an arbitrary Riemannian metric. A curve is given as a collection of edges in the 1-skeleton of M. We refer to the following yes-no question as MINIMAL SPANNING AREA.

Problem: MINIMAL SPANNING AREA
INSTANCE: A 3-dimensional metrized PL manifold M, a 1-dimensional curve K in the 1-skeleton of M, and an integer k.
QUESTION: Does K bound a surface of area A satisfying $A \leq k$?

In analyzing its complexity, the size of an instance of this problem is taken to be the number of tetrahedra in the triangulation plus the log of the sum of the edge lengths plus $\log k$. We show that this problem is as hard as any NP problem.

THEOREM 10.1. *MINIMAL SPANNING AREA is NP-hard.*

SKETCH OF PROOF. We reduce in polynomial time an arbitrary instance of a known NP-complete problem to an instance of MINIMAL SPANNING AREA. We thus show that MINIMAL SPANNING AREA is at least as hard, up to polynomial time reduction, as this problem, which is called ONE-IN-THREE SAT.

Problem: ONE-IN-THREE SAT
INSTANCE: A set U of variables and a collection C of clauses over U such that each clause $c \in C$ has $|c| = 3$.
QUESTION: Is there a truth assignment for U such that each clause in C has exactly one true literal?

Schaefer [**61**] established that ONE-IN-THREE SAT is NP-complete.

A general instance will have n variables and m clauses. As an example, consider the expression

$$(x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3)$$

with three variables and two clauses. A truth assignment setting x_2 to true and x_1, x_3 to false sets exactly one literal in each of the two clauses to be true, so ONE-IN-THREE SAT is satisfied.

We first set up a 2-dimensional version of MINIMAL SPANNING AREA. Given a boolean expression as above we construct a certain metrized 2-complex, a curve K in this 2-complex, and an integer k. The expression admits a satisfying assignment if and only if K bounds a surface of area less than k. For the expression above this metrized complex is shown in Figure 5.

The surface branches off a punctured 2-sphere that has one boundary component called K, one additional boundary component on the left for each variable x_i, and one boundary component on the right for each clause. Each branching circle of the surface on the left corresponds to a variable x_i or $\bar{x}_i$, and a surface spanning K will go over one or the other of the two entering handles with odd algebraic degree. Each branching circle of the surface on the right corresponds to one of the clauses of the expression, and an odd number of the three entering handles is covered with odd degree by a spanning surface for K.

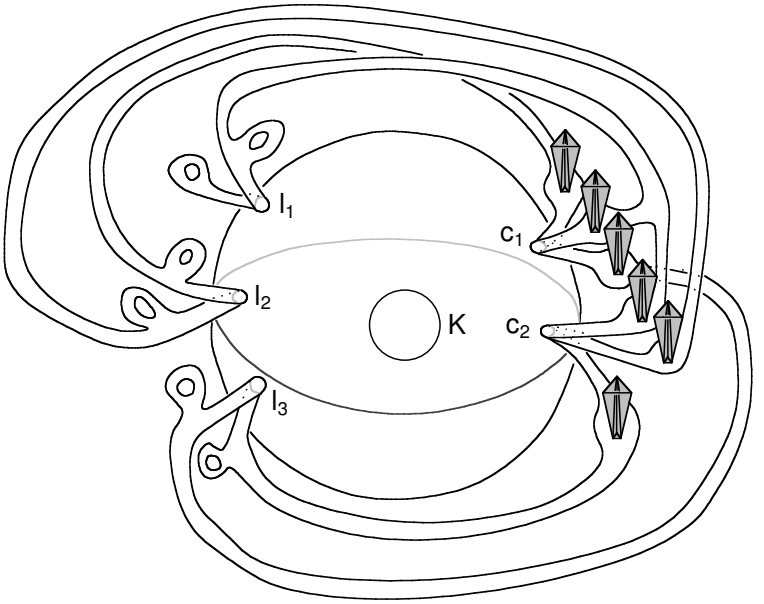


FIGURE 5. A branched surface with boundary curve K corresponding to the boolean expression $(x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3)$.

The metric on the branched surface is chosen so that each triangle in the triangulation is flat, and triangles that are not shaded have total area less than $1/2$. The protruding shaded prisms each have area close to 1, and total area close to the number of clauses m.

A subsurface of the complex that is a spanning surface for K meets each boundary component of the punctured 2-sphere an odd number of times. It thus goes over at least one shaded prism disk of the three associated to each clause curve c_i, and has area at least m, where m is the number of clauses (two in our example). Any spanning surface meeting more than one shaded prism disk associated to a single clause curve c_i must have area greater than $m + 1$. Furthermore, a satisfying assignment for ONE-IN-THREE SAT leads to an embedded spanning surface with the satisfying values of the variables selecting the branches of the surface. It therefore picks out a spanning surface having area less than $m + 1$. So such a surface exists if and only if the expression has a satisfying assignment. Thus an instance of ONE-IN-THREE SAT can be reduced to an instance of MINIMAL SPANNING AREA in polynomial time, and MINIMAL SPANNING AREA is NP-hard.

There is one final point to address. We have worked with a branched surface rather than a 3-manifold. To pass to a 3-manifold we thicken each triangle in the branched surface to a triangular prism with the product metric. In this metric projection is area nonincreasing and a least area surface spanning K must lie on the branched surface. Passage to a closed manifold is obtained by a doubling construction. See [**1**] for details. □

References

[1] I. Agol, J. Hass and W. P. Thurston, The computational complexity of Knot Genus and Spanning Area, preprint, (2001).

[2] F. J. Almgren Jr. and L. Simon, Existence of embedded solutions of Plateau's Problem, Annali Scuola Normale Superiore Pisa, Serie IV, Vol. VI (1979), 447–495.

[3] H. W. Alt, Verzweigungspunkte von H-Flachen. I. (German) Math. Z. 127 (1972), 333–362.

[4] R. Bohme and F. Tomi, Zur Struktur der Lösungsmenge des Plateauproblems (German) Math. Z. 133 (1973), 1–29.

[5] F. Bonahon and J. P. Otal, Scindements de Heegaard des espaces lenticulaires, Ann. Sci. Ecole Norm. Sup. (4) 16:3 (1983), 451–466 (1984).

[6] T. H. Colding and N. Hingston, Metrics without Morse index bounds. (English. English summary) Duke Math. J. 119 , no. 2, (2003) 345–365.

[7] R. Courant, Dirichlet's principle, conformal mapping, and minimal surfaces. With an appendix by M. Schiffer. Reprint of the 1950 original. Springer, New York, 1977.

[8] M. Dehn, Uber die Topologie der dreidimensionalen Raumes Math. Ann. 69 (1920) 137–168.

[9] J. Douglas, Solution of the problem of Plateau, Trans. AMS 33 (1931), 263–321.

[10] J. Eells and L. Lemaire, Another report on harmonic maps. Bull. London Math. Soc. 20:5 (1988), 385–524.

[11] W. H. Fleming, On the oriented Plateau problem. Rend. Circ. Mat. Palermo (2) 11 1962 69–90.

[12] H. Federer, and W. H. Fleming. Normal and integral currents. Ann. of Math. (2) 72 1960 458–520.

[13] M. Freedman, J. Hass and P. Scott, Least area incompressible surfaces in 3-manifolds, Invent Math. 7 (1983), 609–642.

[14] M. Garey and D. S. Johnson, Computers and intractability. A guide to the theory of NP-completeness, W. H. Freeman and Co., San Francisco, 1979.

[15] R. Gulliver, Regularity of minimizing surfaces of prescribed mean curvature, Ann. of Math. (2) 97 (1973), 275–305.

[16] R. Gulliver, A minimal surface with an atypical boundary branch point. Differential geometry, 211–228, Surveys Pure Appl. Math., 52, Longman Sci. Tech., Harlow, 1991.

[17] R. Gulliver and F. Leslie, On boundary branch points of minimizing surfaces, Arch. Rat. Mech. Anal. 52 (1973), 20–25.

[18] R. Gulliver and J. Spruck, On embedded minimal surfaces, Ann. of Math. (2) 103:2 (1976), 331–347.

[19] W. Haken, Theorie der Normalflächen: Ein Isotopiekriterium für den Kreisknoten, Acta Math., 105 (1961) 245–375.

[20] R. Hardt and L. Simon, Boundary regularity and embedded solutions for the oriented Plateau problem. Ann. of Math. (2) 110 (1979).

[21] R. Harvey and H. B. Lawson, Calibrated geometries, Acta Math. 148 (1982), 47–157.

[22] J. Hass, Minimal surfaces in foliated manifolds, Comment. Math. Helv. 61:1 (1986), 1–32.

[23] J. Hass, Singular curves and the Plateau problem, International J. of Math. 2, (1991) 1–16.

[24] J. Hass, Algorithms for recognizing knots and 3-manifolds, Chaos, Solitons and Fractals, 9 (1998) 569–581.

[25] J. Hass, P. Norbury and J. H. Rubinstein, Minimal spheres of arbitrarily high Morse index. Comm. Anal. Geom. 11 , no. 3, (2003) 425–439.

[26] J. Hass, J. C. Lagarias and N. Pippenger, The computational complexity of Knot and Link problems, Journal of the ACM, 46 (1999) 185–211.

[27] J. Hass and J. H. Rubinstein, One-sided closed geodesics on surfaces, Mich. Math. J. 33 (1986) 155–168.

[28] J. Hass, J. H. Rubinstein and P. Scott, Compactifying coverings of 3-manifolds, J. Differential Geometry 30, (1989) 817–832.

[29] J. Hass and P. Scott, The existence of least area surfaces in 3-manifolds, Trans. Amer. Math. Soc. 310 (1988), 87–114.

[30] A. Hatcher, A proof of a Smale conjecture, $\mathrm{Diff}(S^3) \simeq \mathrm{O}(4)$. Ann. of Math. (2) 117:3 (1983), 553–607.

[31] J. Hempel, 3-manifolds, Annals of Math. Studies, 86, Princeton University Press 1976.

[32] W. Jaco, Lectures on the topology of 3-manifolds, Conf. Board Math. Sci. Publications, 43 Amer. Math. Soc. Providence, RI, 1977.
[33] W. Jaco and J. H. Rubinstein, PL Equivariant Surgery and Invariant Decompositions of 3-Manifolds, Advances in Math., 73 (1989) 149–191.
[34] J. Jost, Embedded minimal surfaces in manifolds diffeomorphic to the three-dimensional ball or sphere. J. Differential Geom. 30:2 (1989), 555–577.
[35] R. Kirby, Problems in low-dimensional topology, Geometric topology (Athens, GA, 1993), AMS/IP Stud. Adv. Math., 2.1, Amer. Math. Soc., Providence, RI, 1997.
[36] H. Kneser, Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten, Jahresber. Math. Verein. 28 (1929), 248–260.
[37] H. B. Lawson, Lectures on minimal submanifolds, 9 Publish or Perish, Berkeley 1980.
[38] W. Meeks, Lectures on Plateau's problem, IMPA, Rio de Janeiro, 1978.
[39] W. Meeks and G. P. Scott. Finite group actions on 3-manifolds. Invent. Math. 86:2 (1986), 287–346.
[40] W. Meeks and S.-T. Yau, Group actions on $\mathbb{R}^3$. The Smith conjecture (New York, 1979), 167–179,
[41] W. Meeks and S.-T. Yau, Topology of three-dimensional manifolds and the embedding problems in minimal surface theory. Ann. of Math. (2) 112:3 (1980), 441–484.
[42] W. Meeks and S.-T. Yau, The classical Plateau problem and the topology of three-dimensional manifolds, The embedding of the solution given by Douglas–Morrey and an analytic proof of Dehn's lemma, Topology 21:4 (1982), 409–442.
[43] W. Meeks, L. Simon and S.-T. Yau, Embedded minimal surfaces, exotic spheres, and manifolds with positive Ricci curvature. Ann. of Math. (2) 116:3 (1982), 621–659.
[44] J. Milnor, Morse theory, Annals of Math. Studies, 51 Princeton University Press, 1963.
[45] Morgan, John W. The Smith conjecture (New York, 1979), 3–6, Pure Appl. Math., 112, Academic Press, Orlando, FL, 1984.
[46] C. B. Morrey, The problem of Plateau on a Riemannian Manifold, Annals of Math. 49 (1948) 807–851.
[47] R. Osserman, A survey of minimal surfaces. Second edition, Dover Publications, Inc., New York, 1986.
[48] R. Osserman, A proof of the regularity everywhere of the classical solution to Plateau's problem, Annals of Math. 91 (1970), 550–569.
[49] C. D. Papakyriakopoulos, Dehn's lemma and the asphericity of knots, Annals of Math. 66 (1957), 1–26.
[50] J. Pitts, Existence and regularity of minimal surfaces on Riemannian manifolds, Lecture Notes, Princeton University Press, Princeton, 1981.
[51] J. Pitts and J. H. Rubinstein, Existence of minimal surfaces of bounded topological type in 3-manifolds, Proc. Centre Math. Applic., Australian National University 10, (1985), 163–176.
[52] J. Pitts and J. H. Rubinstein, Applications of minimax to minimal surfaces and the topology of 3-manifolds, Proc. Centre Math. Applic., Australian National University 12, (1987), 137–170.
[53] J. Pitts and J. H. Rubinstein, The topology of minimal surfaces in Seifert fibered spaces, Mich. Math. Journal 42 (1995), 525–535.
[54] J. Pitts and J. H. Rubinstein, Equivariant minimax and minimal surfaces in geometric 3-manifolds, Bulletin Amer. Math. Soc. 19 (1988), 303–309.
[55] J. Plateau, Statique experimentale et theorique des liquides soumis auz seules forces molecularies, Gathier-Villars, Paris, 1873.
[56] M. H. Protter and H. F. Weinberger, Maximum principles in differential equations. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1967.
[57] T. Radó, The problem of least area and the problem of Plateau, Math. Z. 32 (1930), 763–796.
[58] T. Radó, On the Problem of Plateau. Chelsea Publishing Co., New York, N. Y., 1951.
[59] J. Rubinstein, An algorithm to recognize the 3-sphere, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zurich, 1994), Birkhauser, Basel, 601–611.
[60] J. Rubinstein, Minimal surfaces in geometric 3-manifolds, pp. 723–746 in this volume.
[61] T. J. Schaefer, The complexity of satisfiability problems, Proc 10th Ann. ACM. Symp. on Theory of Computing, ACM, NY, (1978) 216–226.

[62] R. Schoen, Curvature estimates for stable minimal surfaces in three-dimensional manifolds, Seminar on minimal submanifolds, Ann of Math. Studies 103, Princeton University Press 1983. 111–126.
[63] R. Schoen and S.-T. Yau, Existence of incompressible surfaces and the topology of 3-dimensional manifolds with non-negative scalar curvature, Annals of Math. 119 (1979), 127–142.
[64] P. Scott, The geometries of 3-manifolds, Bull. London Math. Soc. 15 (1983), 401–487.
[65] P. Scott, There are no fake Seifert fibre spaces with infinite π_1, Ann. of Math. (2) 117:1 (1983), 35–70.
[66] F. Smith, On the existence of embedded minimal 2-spheres in the 3-sphere, endowed with an arbitrary Riemannian metric, PhD thesis, supervisor L. Simon, University of Melbourne, 1982.
[67] D. Sullivan, A homological characterization of foliations consisting of minimal surfaces, Comment. Math. Helv. 54:2 (1979), 218–223.
[68] A. Thompson, Thin position and the recognition problem for S^3, Math. Res. Lett. 1 (1994), 613–630.
[69] F. Tomi and A. Tromba, Extreme curves bound embedded minimal surfaces of the type of the disc. Math. Z. 158:2 (1978), 137–145.
[70] W. Thurston, Three-dimensional geometry and topology, Vol. 1. Edited by Silvio Levy. Princeton Mathematical Series, 35. Princeton University Press, Princeton, NJ, 1997.
[71] F. Waldhausen, On irreducible 3-manifolds which are sufficiently large, Ann. of Math. (2) 87 (1968), 56–88.
[72] B. White, The space of minimal submanifolds for varying Riemannian metrics, Indiana Math. Journal 40 (1991), 161–200.
[73] S. T. Yau, Open problems in geometry, Proc. Symp. Pure Appl. Math., 54, I Amer. Math. Soc. Providence, RI, 1993.

Department of Mathematics, University of California, Davis, California 95616
E-mail address: hass@math.ucdavis.edu

Clay Mathematics Proceedings
Volume 2, 2004

Minimal Surfaces in Geometric 3-Manifolds

J. Hyam Rubinstein

Abstract. In these notes, we study the existence and topology of closed minimal surfaces in 3-manifolds with geometric structures. In some cases, it is convenient to consider wider classes of metrics, as similar results hold for such classes. Also we briefly diverge to consider embedded minimal 3-manifolds in 4-manifolds with positive Ricci curvature, extending an argument of Lawson to this case.

In the topology of 3-manifolds, surfaces of two types play special roles, namely Heegaard splittings and incompressible surfaces. We will see that such surfaces also occur naturally as minimal surfaces, and often are essentially the only types of embedded minimal surfaces which can occur. In the most important class of 3-manifolds with complete hyperbolic metrics of finite volume, we will show that the existence of many different types of minimal surfaces is closely related to the thick/thin decomposition of Margulis. Dehn surgery along knots and links in the 3-sphere generically gives such hyperbolic 3-manifolds and when the surgery coefficients are large enough, we find that the resulting Margulis tubes about short closed geodesics give a type of barrier when forming minimal surfaces.

Applications include new lower bounds for the genus of irreducible and strongly irreducible Heegaard splittings and of incompressible surfaces in terms of the injectivity radius of the thick part of the manifold. Also we construct "knotted" minimal surfaces inside handlebodies with minimal boundary in hyperbolic 3-manifolds, answering a question of S.-T. Yau. Another question of Yau is partially answered — we give infinitely many distinct closed minimal immersions in any hyperbolic 3-manifold with a complete metric of finite volume. Previous applications include describing the boundary slopes of properly immersed incompressible surfaces in complete finite volume hyperbolic 3-manifolds with cusps and giving the overall structure of Heegaard splittings of bounded genus in sequences of 3-manifolds obtained by all Dehn surgeries on a fixed link.

Finally we discuss briefly ideas of B. Andrews, C. Epstein, K. Uhlenbeck, W. Thurston on minimal surfaces with small principal curvature. This includes some joint work with I. Aitchison and S. Matsumoto on a new method to explicitly construct such surfaces in some knot and link complements.

1991 *Mathematics Subject Classification.* Primary 57M25, 57N10.

Key words and phrases. Minimal surfaces, geometric 3-manifolds, incompressible surfaces, Heegaard splittings, Dehn surgery.

©2004 Clay Mathematics Institute

1. Introduction

A good general reference on minimal surfaces is the recent set of notes [**8**]. General existence results for embedded and immersed closed surfaces in Riemannian 3-manifolds are given in [**49, 31, 19, 36, 37**]. We quickly summarize this material as it is crucial in all subsequent discussions. We also quickly review some basic topics from the topology of 3-manifolds — see [**20, 21, 50**].

For simplicity, all 3-manifolds M will be orientable and either compact or the interior of a compact manifold. We will work in the smooth category. All 3-manifolds will be *irreducible* unless stated otherwise, i.e., every embedded 2-sphere bounds a 3-ball. (By results of Kneser and Milnor, compact 3-manifolds can be decomposed essentially uniquely into connected sums of irreducible 3-manifolds.)

DEFINITION 1.1. A mapping ϕ of a closed surface S different from the 2-sphere or projective plane into M is called incompressible or π_1-injective, if the induced map $\pi_1(\phi) : \pi_1(S) \to \pi_1(M)$ is one-to-one.

A basic result of [**49**] and [**11**] (see also [**31, 19**]) is the following:

THEOREM 1.2. *Given an incompressible surface $\phi : S \to M$ in a Riemannian 3-manifold, there is a least area map ρ in the homotopy class of ϕ which is a minimal immersion. Moreover if ϕ is an embedding, so is ρ. In fact, one only needs the map to be injective on simple loops, i.e., that there be no simple closed essential curves in the kernel of the induced map of ϕ on π_1.*

The opposite type of surface in a 3-manifold to an incompressible one is a 'completely compressible' one. The most important type of the latter is a Heegaard splitting. We consider the cases of closed, compact and the interior of compact manifolds.

DEFINITION 1.3. If M is closed, a Heegaard splitting is a decomposition of M into the union of two handlebodies H_+, H_- glued along their common boundary $S = \partial H_+ = \partial H_-$. S is called a Heegaard surface for M.

If M is compact with nonempty boundary or is the interior of a compact manifold with boundary, then a Heegaard splitting is similarly a representation of M as a union of compression bodies. Compression bodies are obtained by attaching one-handles to a product $L \times [0,1]$ or $L \times (0,1]$ respectively, along the surface $L \times \{1\}$. The common boundary is again called a Heegaard surface for M.

A closely related concept is of a one-sided Heegaard splitting [**41**].

DEFINITION 1.4. A one-sided Heegaard splitting for a closed M is a decomposition into the union of a handlebody and a twisted I bundle over a nonorientable surface S', along the common boundary surface. A similar notion occurs also for compact manifolds and interiors of compact manifolds, using compression bodies instead of handlebodies. The surface S' is called a one-sided Heegaard surface.

One-sided splittings occur naturally by taking a geometrically incompressible nonorientable surface S' representing a nonzero element of $H_2(M, \mathbf{Z}_2)$ in an irreducible 3-manifold with no embedded orientable incompressible surfaces. A surface S' is geometrically incompressible, if there is no embedded disk D with $D \cap S' = \partial D$ is a non-contractible loop on S' [**41**].

We need to consider some key properties and results about Heegaard decompositions.

DEFINITION 1.5. A meridian disk for a handlebody H is a properly embedded disk with ∂D a non-contractible loop in ∂H. A splitting (S, H_+, H_-) for M is called irreducible, if there is no pair of meridian disks D_+, D_- for H_+, H_- which meet transversely in a single point. If every meridian disk for H_+ meets every disk for H_-, the splitting is called strongly irreducible. One often calls a reducible splitting one where there is a pair of meridian disks D_+, D_- with $\partial D_+ = \partial D_-$.

It is easy to show that a strongly irreducible splitting is either irreducible, or else the manifold is S^3 and the splitting is the standard genus one decomposition.

Trivial handles can be added to Heegaard splittings by taking an embedded arc ℓ in one handlebody, say H_+, which is boundary parallel (that is, there is an embedded disk D with $\ell \subset \partial D$ and $\partial D \setminus \ell$ included in the Heegaard surface S) and adding a small regular neighborhood of ℓ to H_+. This increases the genus of the handlebody by 1. On the other hand, it is easy to see the result of removing the open neighborhood of ℓ from H_- also gives a handlebody with one extra handle, since D gives a meridian disk for this handle. This process is called stabilization. We can also say that irreducible splittings are not stabilized, since the condition of having a pair of disks D_+, D_- meeting at one point is precisely the same as having added a neighborhood of ℓ, which can be viewed as a neighborhood of say D_-.

Certainly minimal genus Heegaard splittings of 3-manifolds are always irreducible, since a stabilized splitting can be reduced in genus by throwing away the trivial handle.

Haken proved that if M is a nontrivial connected sum, then any Heegaard splitting of M can also be decomposed into splittings of the factors. So it suffices to restrict attention to irreducible M. The concept and existence of strongly irreducible splittings were introduced by Casson and Gordon [**6**] in the following important result.

THEOREM 1.6. *If M is closed and irreducible, then any irreducible Heegaard splitting is strongly irreducible or else M has an embedded incompressible surface.*

DEFINITION 1.7. A closed irreducible 3-manifold is called Haken if it contains an embedded incompressible surface. Otherwise it is called non-Haken.

The minimax technique for finding (unstable) minimal surfaces in 3-manifolds was introduced in [**35, 47, 36, 37**]. The key situation is when we are dealing with strongly irreducible Heegaard splittings and M has a bumpy metric, in the sense of White [**53**].

THEOREM 1.8. [**37**] *Suppose that M is closed and has a bumpy Riemannian metric and S is a strongly irreducible Heegaard splitting. Then either S is isotopic to a minimal surface of index 1, or S is isotopic to the boundary of a regular neighborhood $N(S')$ of a one-sided splitting surface S', with a single handle added by taking the boundary of a regular neighborhood of an arc ℓ which is an I fiber in the I bundle structure on $N(S')$. In the latter case, S' is geometrically incompressible and is isotopic to a stable minimal surface in M. We can view S as collapsing onto S' with multiplicity 2.*

SKETCH OF PROOF. By an extension of the methods of [**35, 47**], one can take a limit of a sequence of surfaces S_n, all in the isotopy class of the Heegaard surface S, approaching the minimax bound in area and show that a smooth minimal surface

is the result. Moreover this surface can be formed topologically by handle pinching and collapsing sheets onto multiply covered components.

A handle pinch (or disk surgery) is obtained by taking an embedded disk D with $D \cap S = \partial D$ a non-contractible loop on S. Let $N(D)$ be a small product neighborhood of D. The annulus $A = N(D) \cap S$ is then replaced by the two parallel disks $\partial N(D) \setminus int(A)$. Handle pinching is associated with the curvature on the sequence of surfaces blowing up at a point or a finite number of points.

Now the assumption that M is strongly irreducible means that only pinching occurs *on one side of the Heegaard surface* S, or else a single handle pinch takes place, followed by collapsing onto a nonorientable surface S' with multiplicity 2. In the latter case, it is easy to check that S' is geometrically incompressible and is a one-sided Heegaard splitting as claimed. The fact that S' can be isotoped to a stable minimal surface follows by [**31**].

In the former case, we need the additional information that the limit surface has components which have total index 1. A bumpy metric is required for this step. (Otherwise we can only show that the minimax limit surface has nullity. A bumpy metric implies that minimal surfaces have no Jacobi fields, so the index must be nonzero). Now it can be observed that the components of the limit surface all have multiplicity 1 (possibly throwing away some nesting spheres) and all bound disjoint balls or handlebodies which are core for one of the original handlebodies, say H_+ for S. By this we mean that the complement of the interiors of these handlebodies in H_+ is a compression body.

Moreover, by Casson–Gordon [**6**], if all these small open handlebodies are removed from M, we obtain a new manifold M' with incompressible minimal boundary of index 1. The unstable component can be pushed into the interior of M' and shrunk to least area by Meeks–Simon–Yau [**31**]. Define M'' by removing the open collar between the stable surface just constructed and the boundary. Next, apply a new minimax procedure in M'' using the compression body decomposition defined by S. (Note that we do not quite have a compression body decomposition defined by S in M'', since the unstable component might be a sphere so a ball is removed from the compression body. This does not affect the argument).

We can iterate this procedure — since the manifold is strictly decreasing in size at each step, we obtain a nested sequence of compact sets which has a limit intersection. The boundary of this limit compact set is a smooth minimal surface and the limiting process contradicts the assumption that the metric is bumpy [**53**]. We conclude that after a finite number of steps the procedure terminates, i.e., no handles are pinched except possibly to split off some spheres, and a minimal surface isotopic to S has been obtained. □

COROLLARY 1.9. *If M has an arbitrary Riemannian metric, the conclusion of the previous theorem holds, except that the minimal surface has index at most 1, as it may have nonzero nullity.*

PROOF. It suffices to show that if a sequence ρ_n of bumpy metrics is chosen converging to the original metric ρ, the resulting sequence of minimal surfaces (in the case of no collapsing to a nonorientable surface with multiplicity 2) converges smoothly to a minimal surface in the metric ρ. We can apply the technique of [**53**] since in either case, there are a priori area bounds on the sequence of minimal surfaces by the fact that the minimax areas for the metrics ρ_n converge to the minimax area for ρ. Since the minimal surfaces cannot converge to a multiply

covered component (by strong irreducibility, such pinching is ruled out, except in the special case of nonorientable surfaces covered twice), the smooth convergence follows by [**53**].

Note that the limiting minimal surface cannot have index more than 1, but can also have nullity. □

To complete this section, we discuss the notion of telescoping of an irreducible but not strongly irreducible splitting S as in [**46**] (see also [**42**]). The idea is that given such a splitting, there is a collection of disjoint incompressible surfaces $S_1, \ldots, S_k$ so that if M is split open along these surfaces, then there are strongly irreducible compression body decompositions $S_{k+1}, \ldots, S_m$ for the components. The original splitting surface S is then formed by tubing together all these surfaces in a standard way.

Now we can apply the previous theorem to this situation in a natural way. In particular, we can first isotope all the incompressible surfaces $S_1, \ldots, S_k$ to be least area [**31**] and either to be disjoint or to coincide. Moreover, then the compression body decompositions $S_{k+1}, \ldots, S_m$ can be isotoped to be either minimal surfaces of index 1, disjoint from the incompressible minimal surfaces (which act as barriers by the maximum principle), or collapse onto geometrically incompressible nonorientable surfaces with multiplicity 2, with a single handle being pinched out in the process. These latter surfaces can also be isotoped to be minimal in the complement of the incompressible minimal surfaces.

EXAMPLE 1. An interesting example to illustrate these ideas is the Heegaard splittings of a hyperbolic surface bundle M over a circle. By [**50**] a pseudo-Anosov gluing ϕ of a surface S of genus at least 2 to itself can be used as the monodromy for such a bundle. Let S be isotoped to be least area by [**31**] and let d be the length of a shortest non-contractible closed curve isotopic onto S. Next, take a cyclic cover $\tilde{M}$ of M, with monodromy ϕ^n for some large n. We can choose n so that the distance D around the bundle between the two sides of S is as big as we like. Now assume that $\tilde{M}$ has a strongly irreducible splitting $\tilde{S}$ of genus g. By Corollary 1.9, $\tilde{S}$ is isotopic to a minimal surface or collapses onto a minimal one-sided surface with multiplicity 2. Now using the coarea formula plus slicing, we see that the area of $\tilde{S}$ or the minimal one-sided surface can be made arbitrarily large. For if a slice by an S-fiber contains a non-contractible curve, its length is at least d. So if we have a long interval of such fibers, the resulting area is large. On the other hand, if a long interval of slices has only contractible curves and no handles, we see that this part of $\tilde{S}$ lifts to the universal covering of $\tilde{M}$ and by monotonicity, has large area. Finally, if there are many handles in such a long interval of slices, the genus of $\tilde{S}$ is large. So we conclude by Gauss–Bonnet in all cases that the genus of any strongly irreducible splitting of $\tilde{M}$ must be arbitrarily large.

Notice we have actually proved more by telescoping, namely that any irreducible splitting must telescope into incompressible surfaces and strongly irreducible splittings, where any component of genus which is not large must be a copy of S. So we conclude (by an easy argument) that the only irreducible splitting of $\tilde{M}$ which does not have large genus must be the standard splitting, i.e., two parallel copies of S joined by two tubes running around the bundle in the two product regions $S \times I$, so that the tubes have cores isotopic to $x \times I$ and $y \times I$, for two points x, y on S.

In the third section on hyperbolic 3-manifolds, we will extend these results to the case of complete hyperbolic metrics of finite volume on manifolds with cusps. In particular we will show that strongly irreducible compression body decompositions can still be isotoped to index 1 closed minimal surfaces or collapse to nonorientable surfaces, i.e., do not move out into the cusps to become non-compact. We will also extend the above example to manifolds with infinite first homology, so that a shortest loop of infinite order in homology is long relative to the injectivity radius of the manifold.

2. Three-Manifolds with Nonnegative Curvature and Seifert Fibered Spaces

We discuss first the geometries $\mathbf{S^3}$, $\mathbf{R^3}$, and $\mathbf{S^2} \times \mathbf{R}$, which have sectional curvatures ≥ 0. So it is convenient to extend the discussion to the more general classes of metrics where the Ricci curvature or scalar curvature is nonnegative.

Two fundamental results about the first case were given in [**24**] and [**9**].

THEOREM 2.1. [**24**] *Suppose that M is a closed orientable irreducible 3-manifold with a Riemannian metric of positive Ricci curvature and let S be a closed orientable minimal surface. If M is split open along S into two components P, Q, then the induced maps $\pi_1(S) \to \pi_1(P)$ and $\pi_1(S) \to \pi_1(Q)$ are onto. Consequently P, Q are both handlebodies.*

THEOREM 2.2. [**9**] *Suppose that M is a closed orientable irreducible 3-manifold with a Riemannian metric of positive Ricci curvature and let S be an embedded closed minimal surface. Then there is an area bound for S in terms of the genus and a lower bound for Ricci curvature. As a consequence, the minimal surfaces of bounded genus form a compact set.*

We now discuss a variation on the proof of [**24**] which gives a little more information. In particular, this raises interesting issues about minimal 3-manifolds in 4-manifolds of positive Ricci curvature.

Lawson uses a second variation argument of Frankel in the following way. Cut M open along S into two domains X, Y (by Mayer's Theorem we know that M has finite fundamental group so any orientable surface must separate M). Then the universal coverings $\tilde{X}, \tilde{Y}$ of X, Y are shown to have connected boundaries, i.e., the induced maps $\pi_1(S) \to \pi_1(X)$ and $\pi_1(S) \to \pi_1(Y)$ are shown to be onto. For if there were two separate components S_1, S_2 of say $\partial\tilde{X}$, then one can connect them by a shortest geodesic arc ℓ which is orthogonal to both surfaces. But Frankel's formula says that such an arc must be unstable, i.e., have second variation negative in some direction. Hence this gives a contradiction.

Lawson then goes on to use Dehn's lemma and the Loop Theorem of Papakyriakopoulos to show that X, Y are both handlebodies and so S is a Heegaard surface for M. By taking double covers, it follows then that any nonorientable minimal surface defines a one-sided Heegaard splitting.

We sketch why the conclusion of Lawson can be obtained without using the 3-manifold techniques, but rather by showing directly that X, Y have one-dimensional spines, i.e., retract onto graphs. To do this, we show that there is a height function (Morse function) on each of X, Y so that S is a level surface and there are no local minima or index 1 critical points. So critical points are of index 2 or 3 and the dual

handle decomposition has only 0- and 1-handles, so gives a handlebody structure on X, Y.

Consider the distance function d of points in X to $S = \partial X$. The Frankel-Lawson argument shows that this function has no local minima on the focal set of S. So we use the standard argument [**32**] of approximating d by a Morse function f. It is easy to check that this can be done so that f has no critical points of index 0 or 1, since d has no local minima and an index 1 critical point would correspond to a local minimum on the focal set. So this completes the sketch.

Note however this method works in all dimensions. In dimensions higher than 4, the conclusion about fundamental groups enables one to eliminate 1-handles by surgery theory. However this cannot be done smoothly in dimension 4 and only topologically in the presence of suitable fundamental group to apply Freedman's techniques. So we have proved:

THEOREM 2.3. *Suppose that W is a compact orientable 4-manifold with a metric of positive Ricci curvature and M is a closed orientable embedded minimal 3-manifold. Then the complementary domains X, Y for M in W both have 2-dimensional spines, or dual handle presentations with only 2-,3- and 4-handles attached to a thickened copy of M.*

Remarks

1. Notice we have shown that the focal set of an orientable minimal surface in an orientable 3-manifold with positive Ricci curvature retracts onto a graph. Similarly in the corresponding case for minimal 3-manifolds in 4-manifolds, the focal set retracts onto a 2-complex. It would be interesting to find other properties of such focal sets.

2. A well-known source of possible examples is given by the Mazur construction [**27**]. Start with a 0-handle (4-ball) and attach first n 1-handles and then n 2-handles which cancel algebraically but not geometrically. Let W^4 be the resulting contractible 4-manifold, its boundary $\partial W = M^3$ is a homology 3-sphere. Then the 5-manifold $W \times I$ has the same collection of handles which now cancel geometrically, giving the 5-ball. One concludes that the boundary $S^4 = 2W^4$. Note that the spines of the two identical complementary domains of M^3 in S^4 are 2-complexes. In Mazur's original construction, these spines are dunce hats, obtained by gluing together the three sides of a triangle with orientations a, a, a^{-1}. Notice that there is a simple foliation of B^5 by copies of W^4; the natural question is can one get an unstable minimal embedding of $W^4 \subset B^5$ so that W^4 meets S^4 orthogonally in a minimal copy of M^3.

3. The examples of White [**52**] show that finding embedded minimal hypersurfaces, with control on the topology, in dimension greater than 3, is very difficult. We note an interesting example in [**26**].

EXAMPLE 2. Lawson gives an action of $SO(3)$ on S^4 with regular orbits which have stabilizers $\mathbf{Z}_2 \times \mathbf{Z}_2$, so are copies of the quaternionic space $M^3 = S^3/\Gamma$, where Γ is the group of eight units in the quaternions with integer entries. There are two exceptional orbits with $SO(2)$ as stabilizer which are minimally embedded copies of RP^2. So we get the focal sets of the orbit of maximum volume, which is a minimal embedding of M^3, are both copies of the projective plane. The two complementary regions of this minimal hypersurface are disk bundles over the projective plane.

We next discuss the case of nonnegative scalar curvature. The fundamental result was proved in [**49**].

THEOREM 2.4. *Suppose M is a closed Riemannian 3-manifold with a metric of nonnegative scalar curvature. If M has an immersed π_1-injective orientable surface S, then S must be a torus and M has a flat metric. The only stable orientable minimal surfaces immersed in M are spheres and tori.*

Using this, it is fairly easy to get a good topological picture of embedded minimal surfaces [**43, 28**].

THEOREM 2.5. *Suppose M is a closed Riemannian 3-manifold with a metric of nonnegative scalar curvature. If M is not flat and S is an embedded orientable minimal surface, there is a collection of embedded 2-spheres in the complement of S so that S together with some of these spheres bounds a punctured handlebody on each side (a punctured Heegaard splitting). The spheres split M into pieces, chosen so that none are punctured 3-balls. In the non-flat case, nonorientable minimal surfaces together with spheres give punctured one-sided Heegaard splittings. In the flat case, either an orientable minimal surface gives a Heegaard splitting or it is a flat torus. Similarly a nonorientable minimal surface gives a one-sided Heegaard splitting or a flat Klein bottle.*

Remarks

1. In the non-flat case, one can use Dehn's lemma and the Loop Theorem to show that an orientable minimal surface S must compress on both sides to spheres, since otherwise there would be disjoint stable minimal surfaces which are not spheres. For nonorientable minimal surfaces the argument is similar.

2. For the flat case, the same argument shows that either a surface can be completely compressed (defining a Heegaard splitting which may be one-sided) or there is a disjoint flat torus. (There are no minimal spheres in flat manifolds, by Gauss–Bonnet.) In the latter case, by translation and the maximum principle, one sees the original minimal surface is a flat torus or Klein bottle. This result was first proved by Meeks [**29**], using a second variation argument in the style of Lawson. He showed that for an orientable closed minimal surface in an orientable closed flat 3-manifold, either the fundamental group maps onto the complementary regions or the surface is a flat torus. In this case, second variation of a shortest geodesic arc may be zero or positive. So he used parallel translation of a piece of the minimal surface along the geodesic to give a contradiction, by the maximum principle.

To complete this section, we discuss minimal surfaces in Seifert fibered spaces. Note that there are eight geometries of Thurston (see [**45**] for an excellent discussion of these metrics). They are $\mathbf{S^3}, \mathbf{S^2} \times \mathbf{R}, \mathbf{R^3}, \widetilde{\mathbf{PSL}(\mathbf{2}, \mathbf{R})}, \mathbf{H^2} \times \mathbf{R}, \mathbf{Nil}, \mathbf{Solv}, \mathbf{H^3}$. Closed manifolds modeled on these geometries have Seifert fibered space structures, except for the most interesting hyperbolic case and **Solv**. By passing to a twofold cover, we will always assume for simplicity that a closed orientable Seifert fibered space has an S^1-action, where there are a finite number of orbits with finite nontrivial stabilizers (called *exceptional* orbits) and all other orbits are called regular. To understand the topological possibilities for embedded minimal surfaces, it turns out to be only necessary to assume that the metric is chosen so that the S^1-action is by isometries.

The first key result was proved in [**16**]:

THEOREM 2.6. *Suppose that an orientable Seifert fibered space has a metric which is invariant under the S^1-action. Then any orientable incompressible surface is either horizontal or vertical, i.e., is either transverse everywhere to the S^1-orbits or is a union of regular orbits. Vertical surfaces are tori or Klein bottles.*

The general case for minimal surfaces which are not necessarily incompressible was given in [**38**]:

THEOREM 2.7. *If M is a closed orientable Seifert fibered space with an S^1-invariant metric and S is a closed embedded minimal surface, there are two compression bodies bounded by S and a collection of disjoint vertical embedded tori, if S is orientable. If S is nonorientable, the boundary of a small regular neighborhood of S together with some vertical embedded tori bound one compression body. The other components of the complement of these tori are sub-Seifert fibered spaces of M.*

Remark: The situation is rather similar to telescoping, described in the previous section.

Note that many interesting geometric examples of closed embedded minimal surfaces in spherical spaces form (with constant sectional curvature $+1$) or in flat manifolds or Seifert fibered spaces can be found in [**25**], [**23**], [**29, 30, 17, 38, 39**], for example.

We include here a brief account of minimal surfaces in **Solv** geometry, in response to a question of P. Shalen. In [**45**], p. 470, this geometry is described as a split extension of $\mathbf{R}$ acting on $\mathbf{R^2}$. There is a horizontal foliation by planes, which can easily be seen to be minimal. In fact, with coordinates (x, y, z) where the planes are $z = c$, multiplication is given by

$$(x, y, z)(x', y', z') = (x + e^{-z}x', y + e^{z}y', z + z')$$

and the metric is

$$ds^2 = e^{2z}dx^2 + e^{-2z}dy^2 + dz^2.$$

Now it is easy to see that the horizontal planes are all minimal, by showing their mean curvatures are zero. We can use the isometry ϕ fixing the origin taking $(x, y, z) \to (y, x, -z)$. This maps the normal $(0, 0, 1)$ to the plane P given by $z = 0$ to the opposite normal $(0, 0, -1)$. Clearly the principal directions on P at the origin must be invariant or interchanged by ϕ. In the latter case, since the signs of the principal curvatures are reversed, we see that their sum must be zero and the plane has zero mean curvature at the origin. In the former case, both principal curvatures would be zero and the plane would be flat. This is actually not the case, but we don't need this fact.

Since **Solv** acts isometrically on itself by left multiplication, we conclude that P and all other planes $z = c$ are minimal in the **Solv** metric. Now given any embedded closed orientable minimal surface S which is not a torus in a closed orientable 3-manifold M with the **Solv** metric, it follows by the maximum principle that any tangencies with the induced foliation of M by minimal tori must be saddles. Now it is easy to use standard 3-manifold techniques to deduce the following theorem.

THEOREM 2.8. *Let M be closed orientable with the* **Solv** *geometry. If S is a closed orientable embedded minimal surface in M, then either S is part of the foliation of M by minimal tori, or S is a Heegaard surface for M.*

3. Three-Manifolds with Complete Hyperbolic Metrics of Finite Volume

By the results of Thurston [**50**], we know that the most common type of geometric structure is a hyperbolic metric. Moreover Thurston's hyperbolization Theorem shows that all orientable closed irreducible atoroidal Haken 3-manifolds have hyperbolic metrics. Also for compact, orientable, irreducible, atoroidal, non-Seifert fibered manifolds with incompressible tori boundary components, there is a complete hyperbolic metric of finite volume on the interior of the manifold. (Here atoroidal means that any incompressible torus is boundary parallel.) By Mostow rigidity, this hyperbolic metric is unique. Furthermore, by recent results [**13, 14**] any irreducible manifold homotopy equivalent to such a hyperbolic manifold is homeomorphic to it. So the hyperbolic metric is a topological invariant of the manifold in this strong sense.

Many interesting invariants of the hyperbolic structure have been extensively studied, such as volume, Chern–Simons invariants, geodesic length spectrum etc. However, minimal surfaces may give a lot of new information about the hyperbolic geometry of 3-manifolds. In particular, area, second fundamental form, and principal curvatures are natural invariants, although difficult to calculate explicitly. The aim in this section is to suggest some preliminary ideas in this direction, coming from elementary bounds on minimal surfaces — for example, using monotonicity and curvature. We refer to [**1, 51, 54, 48**] for useful background information.

A first key observation is that for a closed orientable minimal surface immersed or embedded in a hyperbolic 3-manifold, by the Gauss equation, the induced metric has Gaussian curvature $\kappa \leq -1$. So by Gauss–Bonnet, there is an area bound in terms of the genus g, namely $A \leq 4\pi(g-1)$. In particular, there are no minimal 2-spheres, projective planes, Klein bottles or tori. This type of estimate works also in 3-manifolds with all sectional curvatures strictly negative, which occur when doing Gromov–Thurston Dehn filling [**15, 5, 33, 18**].

A consequence is compactness of the space of minimal surfaces of bounded genus in closed hyperbolic or negatively curved 3-manifolds [**1, 54**]. So any infinite sequence of such surfaces has a convergent subsequence. At a finite number of points, the curvature can blow up and handles can pinch; away from these points the convergence is smooth. Such pinching only happens when the limit surface occurs with multiplicity. We note that compactness also works on non-compact hyperbolic 3-manifolds with complete metrics of finite volume, so long as the minimal surfaces can be bounded away from the cusps.

Note that a cusp is defined by taking a region $R = \{(x, y, z) : z \geq z_0 > 0\}$ in the upper half space model of $\mathbf{H^3}$ and dividing out by a parabolic group $\Gamma = \mathbf{Z} \times \mathbf{Z}$ of isometries of the form $(x, y, z) \to (x + c, y + c', z)$ generating a lattice in the horizontal planes (horospheres). Such horospheres project to horotori in M. Now if an embedded minimal surface S of genus at most g meets the horotorus coming from $z = z_0$ in parallel essential curves $C_1, C_2, \ldots$ of length at most L, then it is easy to bound the "depth", that is, the maximum value of z for the horotori intersected by S. For we can find a foliation of R by minimal strips bounded by lines in the homotopy class of the lifts of $C_1, C_2, \ldots$ and use the maximum principle, since the distance between lifts of $C_1, C_2, \ldots$ is bounded, depending only on the homotopy class of the curves and not on the surface S, since S is embedded.

Therefore to complete the discussion, we must first show that a sequence of embedded minimal surfaces of bounded genus cannot meet the horotorus $z = z_0$ in longer and longer essential curves and must then eliminate inessential curves. The first fact is easy and is proved even for immersed surfaces in [**18**]. The coarea formula and the area bound above give the result. It is also straightforward to eliminate long tubes, i.e., regions between horotori $z_1 \leq z \leq z_0$ where $z_0 - z_1$ is large and the surface S meets every horotoral level in homotopically trivial curves. For such a tube lifts to the universal cover and the monotonicity formula gives that the area of the tube grows exponentially with the size of $z_0 - z_1$. (A similar argument was used in [**33**]). By our genus bound, there is a bounded number of critical points of intersection of the surface and the horotori (there are no local maxima relative to the z coordinate by the maximum principle). So all cases are now done.

Notice that Freedman twisting of immersed essential surfaces about a cusp [**12**] indicate that this result cannot be extended to immersed minimal surfaces. One can construct such surfaces which are separable — that is, lift to embeddings in finite sheeted covers, and so can be isotoped to be closed and minimal. However it is easy to see that the surfaces will penetrate deeper and deeper into the cusp and in the limit will give a non-compact complete minimal surface [**7**]. We will construct some examples in the next section.

We summarize:

THEOREM 3.1. *Let M^3 be a complete hyperbolic 3-manifold of finite volume. Then the collection of closed embedded minimal surfaces of bounded genus forms a compact space.*

We note that the minimax construction for strongly irreducible Heegaard splittings, extends to the case of compression body decompositions of complete hyperbolic 3-manifolds of finite volume. By essentially the same argument as above, given a critical sequence of sweepouts with minimax surfaces converging in the varifold metric to a smooth minimal surface S (possibly nonorientable with multiplicity 2), we can show that S is closed. For the regularity argument in [**36**], [**47**], is local. So we need only show that this critical sequence only penetrates a bounded distance into any cusps. But the argument follows the same lines as above to conclude this.

The minimax construction plus telescoping now enables us to conclude finiteness of the collection of irreducible and strongly irreducible Heegaard splittings of bounded genus of either closed or complete finite volume hyperbolic 3-manifolds. (see [**22**] for a different solution of this problem — the Waldhausen conjecture). For given any infinite sequence of such splittings, they can be isotoped to be closed minimal surfaces — possibly nonorientable surfaces with multiplicity 2 — (the pieces of the splitting in the case of telescoping) and so converge. Now in the case of pinching occurring, the limit surface must have multiplicity. So either the surfaces converge to some nonorientable components with multiplicity 2 or to a surface with multiplicity 1 with no handles pinching, using strong irreducibility (of the pieces in case of telescoping).

THEOREM 3.2. [**40**] *Let M be a complete finite volume hyperbolic 3-manifold. Then there are finitely many irreducible Heegaard splittings of bounded genus up to isotopy.*

Our next idea is to relate the Heegaard genus of a complete finite volume hyperbolic 3-manifold M to its geometry. The Heegaard genus is the smallest genus of all Heegaard splittings of M. In particular, the genus should reflect the smallest cross-sectional area under a sweepout.

DEFINITION 3.3. A sweepout of a 3-manifold consists of a pseudo-isotopy S_t, where $0 \leq t \leq 1$ and S_0, S_1 are graphs, S_t is embedded for $0 < t < 1$. (For many purposes it is useful to allow handles of S_t to pinch and regrow during the sweepout at finitely many values of t; see [**42**]).

Let us first consider the case of a closed hyperbolic 3-manifold M. Denote the injectivity radius (half the length of the shortest closed geodesic) by ρ. We give a simple lower bound for the Heegaard genus of M. By monotonicity, the area of a minimal surface S passing through the center of a ball of radius ρ is at least $2\pi(\cosh\rho - 1)$, which is the area of a totally geodesic disk of this radius. On the other hand, we know by Gauss–Bonnet as at the beginning of this section, that the area A of S satisfies $A \leq 4\pi(g-1)$. Putting these together gives

$$g \geq \frac{\cosh\rho + 1}{2}.$$

Now this estimate can be improved by eliminating any very short closed geodesics from consideration. So we investigate the thick/thin decomposition of M [**50**]. Let $C_1, C_2, \ldots, C_k$ be a collection of disjoint short closed geodesic simple loops in M. Take tubular neighborhoods of these loops and expand out at the same rate until the boundaries bump into each other. Let $N_1, N_2, \ldots, N_k$ denote these maximal embedded neighborhoods. Keep expanding and flatten out into a 2-complex X. So the complement $M \setminus (N_1 \cup N_2 \cup \cdots \cup N_k)$ is a neighborhood of X which we denote by $N(X)$. The solid tori N_i are foliated by tori with inward pointing mean curvature vectors, by taking tori at constant distance from C_i. Let ρ' denote the injectivity radius of loops based at $N(X)$, i.e., we take half the length of the shortest geodesic loop based at some point of $N(X)$ as ρ'. Now the same estimate as before applies, namely $g \geq \frac{1}{2}(\cosh\rho' + 1)$. We need only observe that any minimal surface S must meet X. For otherwise S would lie in one of the open solid tori complementary domains of $N(X)$. But these solid tori are foliated by tori with inward pointing mean curvature vectors, so by the maximum principle, it follows that there are no closed minimal surfaces in such regions.

Note the same argument works well for non-compact complete finite volume hyperbolic 3-manifolds with cusps. For we can remove any short closed geodesics and expand out both the cusps and neighborhoods of the short geodesics simultaneously to again find a neighborhood $N(X)$ of a 2-complex X and compute ρ'. In both cases we will refer to ρ' as the injectivity radius of the thick part of M.

THEOREM 3.4. *Let M be a closed or complete finite volume hyperbolic 3-manifold. Let ρ' be the injectivity radius of the thick part of M. Then the minimal Heegaard genus g of M satisfies*

$$g \geq \frac{\cosh\rho' + 1}{2}.$$

Remarks

1. Note that if the smallest genus Heegaard splitting is irreducible but not strongly irreducible, i.e., telescoping occurs, this estimate can be improved. The

same is true if the minimal surface is obtained by collapsing onto a nonorientable surface with multiplicity 2.

2. The same estimate also works for bounding the smallest genus of embedded or immersed closed π_1-injective surfaces in M. For by [**49**], such surfaces can be homotoped to least area immersions. It is easy to show as above that such least area maps remain away from the cusps and so give closed minimal immersions. So the monotonicity bounds apply.

3. An interesting question about Heegaard genus g is whether the rank r of $\pi_1(M)$, i.e., the smallest number of generators, can be arbitrarily smaller than g, i.e., whether $g - r$ can be arbitrarily large. (It is obvious that $g \geq r$.) However it is hard to find a good upper bound for r.

EXAMPLE 3. We can now extend Example 1. Let M be a closed hyperbolic 3-manifold with an element α of infinite order in $H_1(M, \mathbf{Z})$ or M complete finite volume hyperbolic with cusps and $H_1(M, \mathbf{Z})$ has an element α of infinite order which is non-cuspidal. Assume that the length of a shortest loop in the homology class of α is L and ρ' is the injectivity radius of the thick part of M. Then the smallest genus g' of a strongly irreducible Heegaard splitting for M satisfies $4\pi(g' - 1) \geq 2L\rho' - \epsilon$ and so $g' \geq L\rho'/(2\pi)$.

The reason is that we can map M to a circle S^1 by the argument of Stallings, sending the class α to a generator of the homology of S^1. Now perturb the minimal surface S corresponding to a strongly irreducible splitting (or the corresponding nonorientable surface with multiplicity 2) into general position relative to this map. If the level sets on S are always essential, the estimate follows by the coarea formula. If there are only inessential curves at some levels, this cannot occur except for a very short interval, using monotonicity (see previously — we add in a small correction term ϵ to take account of this case.) So this gives the result.

We conclude that if L is large, then any strongly irreducible splitting must have large genus. By taking cyclic covers, lifting α to a finite multiple, we can make L as large as we like.

Our next observation is that some information can be obtained on the Heegaard genus of finite sheeted coverings of M with a complete hyperbolic metric of finite volume. It is well-known that $\pi_1(M)$ is residually finite. So any nonzero element β of $\pi_1(M)$ can be "unwound" so that some nontrivial multiple lifts to a finite sheeted covering. In this way, we can choose a sequence of finite sheeted coverings to ensure that the injectivity radius (of the whole manifold or of the thick part) increases as much as desired. In this way, by Theorem 3.4, it follows that the Heegaard genus of this covering space also increases to any desired lower bound.

On the other hand, suppose that S is a strongly irreducible splitting of M. Let $D_1, D_2, \ldots, D_k$ and $D_1', D_2', \ldots, D_s'$ be two complete systems of meridian disks for the compression bodies on either side of S, i.e., if the compression bodies are cut open along these disks, the result is products of closed surfaces and intervals or 3-balls. Now if the total intersection number of these disks is N and we take a d-fold covering space, then we get kd and sd disks respectively which meet in dN points. Now if $d^2ks > dN$, i.e., $d > N/(ks)$, then clearly some of the lifted disks in one compression body do not meet some of the disks on the other side, i.e., the lifted splitting is no longer strongly irreducible. Hence either it is irreducible and telescopes or has some trivial handles and destabilizes. In either case we can find

new minimal surfaces in the lifted manifold, by using Corollary 1.9. (Collapsing onto a nonorientable minimal surface with multiplicity 2 is included as a possibility. We will see however that there can be a problem with telescoping — see the next paragraph.) Projecting back to M, this leads to a sequence of immersed minimal surfaces which are not coverings of each other.

To complete this discussion, we need to consider the case where no new strongly irreducible Heegaard splittings come from this covering construction, because telescoping may not produce any interesting surfaces. In this case, some finite sheeted covering of M is a surface bundle $\tilde{M}$ over a circle. For then, telescoping only gives two copies of the fibering surface and no strongly irreducible splitting surfaces of the pieces. Moreover any finite sheeted cover of $\tilde{M}$ is again a bundle so we may be stuck with no further new minimal surfaces. It is easy to see this is the only case where telescoping fails to come up with new strongly irreducible compression body decompositions between incompressible surfaces.

Now an easy way of proceeding is to pass to a finite sheeted covering $\bar{M}$ of $\tilde{M}$ where the monodromy map is trivial on $H_1(F, \mathbf{Z}_2)$, where F is the fiber of the bundle structure. Since the monodromy for $\tilde{M}$ induces a finite order permutation on this homology, by taking a suitable cyclic cover, we can find $\bar{M}$. Now the rank of $H_1(\bar{M}, \mathbf{Z}_2)$ is at least the genus of F. So we get a collection of incompressible orientable or nonorientable embedded surfaces in $\bar{M}$, depending on whether the dual homology classes come from classes in $H_1(\bar{M}, \mathbf{Z})$ or not. These can all be isotoped to least area maps by [**49**]. To get more, we can pass to a higher covering, for instance by taking the kernel of the map from $\pi_1(F) \to H_1(F, \mathbf{Z}_2)$ and forming the induced covering space, since the monodromy of the bundle $\bar{M}$ preserves this kernel.

If we find new nonorientable surfaces in the higher covering space, these cannot cover previously found nonorientable surfaces since all the ones in $\bar{M}$ lift to orientable surfaces in the new covering space. If incompressible orientable surfaces are found which are not fibers of bundles structures, then telescoping gives new minimal surfaces as previously. Finally, if new fiberings are found, again these cannot cover previously found fiberings and we have proved:

THEOREM 3.5. *Let M be a 3-manifold with a complete hyperbolic metric of finite volume. Then M admits infinitely many distinct minimal immersions of closed surfaces.*

Remark: In [**55**], the question is asked whether all closed Riemannian 3-manifolds admit infinitely many embedded closed minimal surfaces. It would be good to at least answer this question for the case of hyperbolic 3-manifolds.

To complete this section, we mention two recent applications of minimal surfaces to questions about hyperbolic 3-manifolds. In [**18**], a question of P. Shalen about boundary slopes of immersed proper π_1-injective and π_1-∂-injective surfaces in 3-manifolds is answered. In particular it is shown that for a complete hyperbolic 3-manifold with finite volume and one cusp, there is a uniform bound on the number of boundary slopes of such surfaces of genus at most g, which is quadratic in g. The main idea is to use area bounds, estimating the amount of area inside a maximal cusp as a function of the length of the boundary slope. (The boundary slope of a proper surface is the homology class of the boundary curves on the boundary torus — we are assuming here that the surface has a single boundary slope, even

though it is only immersed. This is natural when considering closed immersed incompressible surfaces arising from Dehn filling on knots.) A similar result is also obtained for hyperbolic 3-manifolds with several cusps.

In [**33**], the irreducible Heegaard splittings of the collection of Dehn fillings on a complete hyperbolic 3-manifold M with finite volume and several cusps is studied. (Recall that Dehn filling means, remove an open horotoroidal neighborhood of each cusp and glue in a solid torus.) In [**50**], it is shown that except for a finite number of exceptions for each cusp, almost all Dehn fillings give a closed hyperbolic 3-manifold N. The main result of [**33**] is that except for finitely many exceptions at each cusp, there are finitely many surfaces embedded in M so that any Heegaard splitting of genus at most g of N is isotopic to one of these surfaces. Moreover each of these surfaces is either a Heegaard splitting of M or there is a unique curve on the surface isotopic into one of the cusps. Again the key idea is to use area bounds — Dehn filling has the property that if the meridian curve for the glued-in solid torus is very long on a maximal horotorus, the area of the meridian disk becomes very large. So any irreducible Heegaard surface cannot cross the solid torus in a meridian disk without having large genus. (This works well in case of telescoping as well as strong irreducibility.) We conclude any splitting surface of bounded genus must lie outside the solid torus, i.e., is in M as claimed.

This result shows that small genus Heegaard splittings of nearly all the hyperbolic 3-manifolds obtained by Dehn filling on several cusps of M come from the same surfaces in M. Thurston and Jorgensen [**50**] have shown that complete hyperbolic 3-manifolds of finite volume form a well-ordered set using volume as the order. Limit points correspond precisely to the collection of all Dehn fillings on a cusped manifold, which converge (in the Gromov–Hausdorff metric) to the cusped manifold.

4. Surfaces with Small Principal Curvatures and the Topology of Minimal Surfaces in Hyperbolic 3-Manifolds

In this section, we first look at immersed closed surfaces with small principal curvature in complete hyperbolic 3-manifolds of finite volume. In [**51**] and [**10**], some properties of such surfaces were discussed. Here small principal curvature means either ≤ 1 or < 1 depending on circumstances. A key difference between the two cases occurs for 3-manifolds with cusps. It is not difficult to show that an immersed closed surface S with principal curvatures < 1 cannot have any essential loops homotopic into the cusps. (These are often called accidental parabolics.) However we will give many examples of such surfaces with principal curvatures ≤ 1 with accidental parabolics. Note that Freedman twisting [**12**] is associated with accidental parabolics.

The main result about surfaces of small principal curvature is that they are π_1-injective. Such surfaces are of great interest in 3-manifold topology — a fundamental problem — the Waldhausen–Thurston conjecture states that every aspherical 3-manifold has such surfaces. In fact, one approach to Thurston's geometrization conjecture is to try to show the existence of π_1-injective surfaces which are separable, i.e., lift to embeddings in finite sheeted covering spaces (see for example [**7**]). It is not hard to prove that if a π_1-injective surface immersed in a hyperbolic 3-manifold with cusps has no accidental parabolics, then the surface remains π_1-injective under most Dehn fillings (see [**4**] for a proof of Thurston). So this gives

a good way of constructing such examples. Moreover, Thurston has conjectured that all complete finite volume hyperbolic 3-manifolds have immersed surfaces with principal curvature < 1. Note also that under sufficiently large hyperbolic Dehn filling, a surface with principal curvature bounded below 1, will have its metric perturbed only a small amount, so the surface will remain in this class.

THEOREM 4.1. [**51, 10**] Suppose that M is a complete hyperbolic 3-manifold with finite volume. Let S be an immersed closed surface with all principal curvatures at most 1. Then S is π_1-injective and lifts to an embedding in the covering space M_S corresponding to the subgroup $\pi_1(S)$ of $\pi_1(M)$. The Gauss or normal map gives a foliation of M_S by copies of S at constant distance along normal geodesics. Finally, if S has accidental parabolics, S has points with principal curvature exactly 1.

PROOF. There are several ways of proving this; we use the foliation of the universal cover by horospheres, since this will be useful for the new method of constructing examples in [**3**]. Assume first that the principal curvatures of S are < 1. Choose a lift $\tilde{S}$ of S to the universal covering space $\mathbf{H^3}$ and a point x on the sphere at infinity, which is not in the limit set of $\tilde{S}$. To find such a point is easy — just choose a horosphere touching $\tilde{S}$ at some point. Then it is easy to see that the center of this horosphere is such a point x. Now the foliation of $\mathbf{H^3}$ by horospheres centered at x will have tangencies on S which can be minima but never maxima nor saddles, since the principal curvatures of S are less than those of the horospheres. We conclude that $\tilde{S}$ must be a plane, with only one local minimum relative to this foliation. Moreover it also follows that $\tilde{S}$ is embedded. For the intersection curves with the horospheres are always embedded loops and there are no self intersections.

If we allow principal curvatures up to 1, accidental parabolics can arise. Note that if there is an essential loop C on S homotopic into a cusp, and if C is shrunk to a geodesic loop on S, then this loop lifts to a line in a horosphere in $\mathbf{H^3}$. Let $\mathbf{Z}$ denote the subgroup of $\pi_1(S) \subset \pi_1(M)$ generated by C. Let M_Z denote the corresponding covering of M and let S_Z be the covering of S. Then S_Z is an annulus with center C and M_Z is the quotient of $\mathbf{H^3}$ by the action of $\mathbf{Z}$. We can foliate M_Z by the horoannuli with center at the fixed point of the action of $\mathbf{Z}$ on the sphere at infinity. It is easy to see that there must be a tangency of S_Z and one of these horoannuli along the geodesic representative of C lifted to S_Z.

If there are no accidental parabolics and S is not a horotorus, there must be points where at least one of the principal curvatures is < 1. So we can choose a horosphere touching $\tilde{S}$ at such a point x and use the horospheres with center at x as the foliation. Although now the foliation can touch $\tilde{S}$ in regions such as disks or graphs, one can still show there is a single such (simply connected) region which corresponds to a minimum and no other critical points. Consequently $\tilde{S}$ is an embedded plane. As in the previous paragraph, if S has accidental parabolics, it follows that $\tilde{S}$ and one of the horospheres touch along a line with ends at the center of the horosphere. At any rate, the same conclusion works that $\tilde{S}$ is an embedded plane.

Finally the observation about the foliation of M_S follows easily — there are clearly no focal conjugate points, by looking at Jacobi fields along a geodesic ray. If two normal geodesics of the same length end at a common point y, then by taking metric spheres centered at y we can touch $\tilde{S}$ at some point in a saddle or local maximum, contradicting the principal curvature ≤ 1 assumption. □

Ben Andrews has described a beautiful heat flow method for deforming any surface with all principal curvatures < 1 to a canonical minimal surface with the same property. In fact, the minimal surface has the amazing property that it has the smallest maximum principal curvature in its homotopy class. So the flow proceeds by "flattening" the surfaces as well as decreasing their area.

Note it is easy to see that any minimal surface S with principal curvatures ≤ 1 is unique in its homotopy class. For lifting to the covering space M_S, the foliation by surfaces at a constant distance from a lift of S all have inward pointing mean curvature. So any other minimal surface S' homotopic to S would lift to M_S and be touched by one of these surfaces, contradicting the maximum principle.

To complete this discussion, we sketch the main idea of [**3**] (see [**2**] for background). Thurston [**50**] described a beautiful triangulation of the figure-8 knot space, i.e., the result M of removing this knot from S^3. He took two regular ideal tetrahedra, with all dihedral angles $\frac{\pi}{3}$ between their faces and glued together the faces to form M. In particular, this exhibits the complete finite volume hyperbolic metric on M in a simple way.

Now Haken's theory of normal surfaces can be viewed as a form of discrete minimal surface theory [**42**]. A normal surface consists of elementary disks properly embedded in each tetrahedron. There are two types of disks, triangles cutting off vertices and quadrilaterals separating opposite pairs of edges in a tetrahedron. Our idea is to find an analogue of small principal curvature for such immersed normal surfaces. We follow the outline in Theorem 4.1 by using foliations by horospheres to bound the bending of immersed normal surfaces.

THEOREM 4.2. [**3**] *Let S be a closed surface which is tiled by triangles and quadrilaterals satisfying the following conditions:*

- *Every vertex of the tiling is of degree 6.*
- *No two quadrilaterals share a common edge.*
- *Every vertex has an even number of quadrilaterals.*

Then after possibly passing to a small finite sheeted covering of S we get a surface $\tilde{S}$ which immerses as a normal surface in the figure-8 knot space M and is π_1-injective. Moreover this surface can be deformed to a smooth surface with principal curvatures ≤ 1 and has accidental parabolics if and only if there is an essential annulus of triangles in the tiling. If there are no accidental parabolics, the surface can be deformed to have principal curvatures < 1.

SKETCH OF PROOF. The main ideas of the proof are as follows. Firstly one can locally immerse such a tiling into M. However to match up around essential loops, the surface may not glue up as one may have different normal arc types at a given face of the triangulation around such a loop. However this is corrected by passing to a covering space.

Next, there is a convenient way to place all the elementary disks, i.e., triangles and quadrilaterals. All the vertices are placed at the midpoints of the edges of the triangulation, i.e., the points of symmetry (all edges of the triangulation are infinitely long geodesics). The edges of the elementary disks are geodesic arcs in the faces between the vertices.

Finally, by studying a neighborhood of an elementary disk (i.e., all elementary disks with a vertex or edge in common with a given disk), it can be shown that given the three conditions of the theorem, that such a neighborhood can be placed

"outside" of a horosphere. For example, given a triangle, its three vertices lie on a unique horosphere. All the other vertices of such a neighborhood then lie outside this horosphere. For a quadrilateral, one takes any three vertices and checks that the fourth lies outside automatically as do the other vertices of the neighborhood. In this way, we can see that a foliation by horospheres can only touch the surface at local minima. These local minima must include elementary triangular disks. So the conclusions of Theorem 4.1 apply.

Note that the eight triangles at the corners of the two tetrahedra form a horotorus in M. So any annulus of triangles in the tiling gives an accidental parabolic. It is easy to check that the absence of such an annulus means that there are no accidental parabolics. Finally we can round off the edges and vertices of the surface to form a smooth surface which still has principal curvatures ≤ 1. More work is required in the case of no parabolics to deform the surface to have all principal curvatures < 1; the main idea is to slightly bend regions of triangles to lie on surfaces of constant principal curvature less than 1. □

REMARKS. The same method works with other link complements with regular cell decompositions, such as the Whitehead link and Borromean rings, which are divided into regular ideal octahedra with all dihedral angles $\frac{\pi}{2}$. In theory the method should work for general ideal cell decompositions and triangulations, but it is much more difficult to characterize the tilings of surfaces and to locate the vertices so conveniently. Possibly the class of simple 2-bridge knot and link complements could be analyzed.

Our final topic is the topology of embedded minimal surfaces in hyperbolic 3-manifolds. In particular, we consider two related questions. By Corollary 1.9, we know that any strongly irreducible Heegaard splitting can either be isotoped to be minimal, or else collapses onto a nonorientable minimal surface with multiplicity 2 and the resulting surface defines a one-sided Heegaard splitting. So the first question, asked by S.-T. Yau, is whether there can be a disjoint minimal surface inside a handlebody bounded by a minimal surface. The second question is can such a surface be knotted, i.e., can we find embedded closed orientable minimal surfaces which are neither incompressible nor bound handlebodies.

We sketch a construction giving both types of examples. The key idea is to form a barrier by using a big tube about a short closed geodesic, i.e., the thin part of the hyperbolic manifold.

THEOREM 4.3. [**44**] *Hyperbolic 3-manifolds which are complete of finite volume or closed, can be constructed with arbitrarily many disjoint closed embedded minimal surfaces, which are neither incompressible nor bound handlebodies nor compression bodies (in the case of cusps). Moreover such surfaces can be found inside handlebodies bounded by minimal surfaces.*

SKETCH OF PROOF. The basic idea is to choose two simple closed curves C, C' on a strongly irreducible Heegaard surface S so that C, C' are "disk busting" in the handlebodies, H_+, H_- respectively, bounded by S. By this we mean that C meets every meridian disk for H_+ and similarly for C' and H_-. Then push C into X and C' into Y. We also require that $M \setminus (C \bigcup C')$ be hyperbolic, that is, irreducible and atoroidal.

Next do large longitudinal Dehn surgery along both C and C'. It is easy to see that the handlebodies H_+, H_- remain handlebodies, only the gluing between their

boundaries along S is changed. Moreover one can obtain that the new Heegaard splitting is still strongly irreducible if desired, following an argument of Casson-Gordon [**34**]. This is useful if we want to find stable and unstable minimal surfaces which are parallel. (Note longitudinal surgery means that the new meridian disk for the glued-in solid torus meets a longitude for the original solid torus once.)

Now we will form a barrier surface using S and the big tubes about C, C'. The intuitive idea is that in the limit of Dehn surgeries, C, C' become cusps and so S is now an incompressible surface. In this case we can find a stable minimal surface isotopic to S. Note that the property that C, C' are disk busting comes in here, since we want no compressing disks for S remaining when these curves are removed.

Remove smaller open tubes about the core circles C, C', of say half the radius of the big tubes, to form a new manifold $\tilde{M}$ with two boundary tori T, T'. We can find disjoint compact properly embedded incompressible and boundary incompressible surfaces $\bar{S}, \bar{S}'$ with two boundary essential boundary curves on T, T' respectively. These are formed by pushing an annulus on S out into the tubes about C, C' respectively. (The disk busting assumption is needed here.)

Now as in [**18**], we can taper off the metric to make the two boundary tori totally geodesic in $\tilde{M}$. We can then isotope $\bar{S}, \bar{S}'$ to be stable minimal surfaces which have geodesic boundary curves and meet $\partial\tilde{M}$ orthogonally. Now a simple area estimate shows that inside a collar of the boundary, the surfaces $\bar{S}, \bar{S}'$ meet each horotorus in two curves which are close to being geodesics. So we can choose least area annuli A, A' between two such curves for $\bar{S}, \bar{S}'$ respectively so that if we cut off $\bar{S}, \bar{S}'$ by removing two annuli and replacing them by A, A', then new closed surfaces S_1, S_2 isotopic to S are formed. These are piecewise least area and form barriers. To justify this, we must check that the angles between the annuli and the surfaces are less than π. This can be done by knowing that the horotori are met in curves close to geodesics. In particular, at this stage explicit estimates can be obtained on the size of Dehn surgery required to perform the construction, depending only on the genus of S. So we can isotope S to a stable minimal surface in the region between these surfaces.

Now this argument can be applied to find new types of minimal surfaces, as well as many parallel surfaces inside handlebodies. Choose any surface U inside one of the handlebodies, say H_+, to bound a smaller handlebody X, so that there is a compression body Y between U and S. Now perform exactly the same procedure, using X, Y in place of H_+, H_-. We see that after longitudinal Dehn surgery of curves pushed off V, that S still defines a strongly irreducible compression body splitting for $M \setminus \operatorname{int} X$. So U can be isotoped to be a stable minimal surface inside the handlebody bounded by the minimal surface isotopic to S.

Finally, we can choose a separating compressible surface V which does not bound a handlebody on either side. Again loops C, C' can be selected to be disk busting for any compressing disks on each side of V. The same construction will give a new 3-manifold for which V still does not bound handlebodies on either side (longitudinal Dehn surgery does not affect the homeomorphism type of the two sides of V, only their gluing along V). Here V can be isotoped to a stable minimal surface by the previous barrier argument.

It is not difficult now to find many disjoint embedded minimal surfaces of different types, using this procedure a number of times. □

Acknowledgments

I thank I. Agol, I. Aitchison, J. Hass, J. Schultens and S.-T. Yau for helpful conversations. J. Pitts is a joint author for much of the background material for these notes.

References

[1] M. Anderson, Curvature estimates for minimal surfaces in 3-manifolds, Ann. Scient. École Norm. Sup. 18 (1985), 89–105.
[2] I. Aitchison, S. Matsumoto and J.H. Rubinstein, Surfaces in the figure-8 knot: complement, Journal of Knot theory and its ramifications, 7 (1998), 1005–1025.
[3] I. Aitchison, S. Matsumoto and J.H. Rubinstein, Immersed incompressible surfaces of small curvature in the figure-8 knot complement, in preparation.
[4] I. Aitchison and J.H. Rubinstein, Incompressible surfaces and the topology of 3-manifolds, J. Aust. Math Soc. Series A, 55 (1993), 1–22.
[5] S. Bleiler and C. Hodgson, Spherical space forms and Dehn filling, Topology 35 (1996), 809–833.
[6] A. Casson and C. Gordon, Reducing Heegaard splittings of 3-manifolds, Topology and its applications, 27 (1987), 275–283.
[7] D. Cooper and D. Long, Virtually Haken Dehn filling, J. of Diff. Geom. 52 (1999), 173–187.
[8] T. Colding and W. Minnicozzi, Minimal surfaces, Courant Lecture Notes in Mathematics 4, 1999.
[9] H. Choi and R. Schoen, The space of embeddings of a minimal surface into a three dimensional manifold of positive Ricci curvature, Invent. Math. 85 (1985), 387–394.
[10] C. Epstein, The hyperbolic Gauss map and quasiconformal reflections, Journal für Mathematik, 372 (1986), 96–135.
[11] M. Freedman, J. Hass and P. Scott, Least area incompressible surfaces in 3-manifolds, Invent Math. 7 (1983), 609–642.
[12] B. Freedman and M.H. Freedman, Kneser-Haken finiteness for bounded 3-manifolds, locally free groups and cyclic covers, Topology 37 (1998), 133–147.
[13] D. Gabai, On the geometric and topological rigidity of hyperbolic 3-manifolds, Journal of the Amer. Math. Soc. 10 (1997), 37–74.
[14] D. Gabai, W. Meyerhoff and N. Thurston, Homotopy hyperbolic 3-manifolds are hyperbolic, Annals of Math. 157 (2003), 335–431.
[15] M. Gromov and W. Thurston, Pinching constants for hyperbolic 3-manifolds, Invent. Math. 89 (1987), 1–12.
[16] J. Hass, Minimal surfaces in Seifert fibred spaces, Topology Appl. 18 (1984), 145–151.
[17] J. Hass, J. Pitts and J.H. Rubinstein, Existence of unstable minimal surfaces in manifolds with homology and applications to triply periodic minimal surfaces, Proc. Symp. Pure App. Math. 54-I, Amer. Math. Soc., Providence, 1993, 147–162.
[18] J. Hass, J.H. Rubinstein and S. Wang, Boundary slopes of immersed surfaces in 3-manifolds, J. Diff. Geom. 52 (1999), 303–325.
[19] J. Hass and P. Scott, The existence of least area surfaces in 3-manifolds, Trans. Amer. Math. Soc. 310 (1988), 87–114.
[20] J. Hempel, 3-manifolds, Annals of Math. Studies 86, Princeton University Press, 1976.
[21] W. Jaco, Lectures on the topology of 3-manifolds, CBMS Conferences 43, Amer. Math. Soc. Providence R.I. 1977.
[22] W. Jaco and J.H. Rubinstein, Efficient triangulations and finiteness of Heegaard splittings of 3-manifolds, in preparation.
[23] H. Karcher, U. Pinkall, I. Sterling, New minimal surfaces in S^3, J. Diff. Geom. 28 (1988), 169–185.
[24] H. B. Lawson, The unknottedness of minimal embeddings, Invent Math. 11 (1970), 183–187.
[25] H. B. Lawson, Complete minimal surfaces in S^3, Annals of Math. 90 (1970), 335–374.
[26] H. B. Lawson, Lectures on minimal submanifolds, Publish or Perish, 1980.
[27] B. Mazur, A note on some contractible 4-manifolds, Annals of Math. 73 (1961), 221–228.
[28] W. Meeks, Lectures on Plateau's problem, IMPA, Rio de Janeiro, 1978.

[29] W. Meeks, The geometry, topology and existence of periodic minimal surfaces, Proc. Symp. Pure Appl. Math. 54-I, Amer. Math. Soc., Providence, 1993.
[30] W. Meeks, The theory of triply periodic minimal surfaces, Indiana Uni. Math. J. 39 (1990), 877–936.
[31] W. Meeks, L. Simon and S.-T. Yau, Embedded minimal surfaces, exotic spheres, and manifolds with positive Ricci curvature, Annals of Math. 116 (1982), 621–659.
[32] J. Milnor, Morse theory, Annals of Math. Studies, 51 Princeton University Press 1963.
[33] Y. Moriah and J.H. Rubinstein, Heegaard structures of negatively curved 3-manifolds, Comm. in Anal. and Geom. 5 (1997), 375–412.
[34] Y. Moriah and J. Schultens, Irreducible Heegaard splittings of Seifert fibred spaces are horizontal or vertical, Topology 37 (1998), 1089–1112.
[35] J. Pitts, Existence and regularity of minimal surfaces on riemannian manifolds, Lecture Notes, Princeton University Press, Princeton, 1981.
[36] J. Pitts and J.H. Rubinstein, Existence of minimal surfaces of bounded topological type in 3-manifolds, Proc. Centre Math. Applic. Australian National University 10, (1985), 163–176.
[37] J. Pitts and J.H. Rubinstein, Applications of minimax to minimal surfaces and the topology of 3-manifolds, Proc. Centre Math. Applic. Australian National University 12, (1987), 137–170.
[38] J. Pitts and J.H. Rubinstein, The topology of minimal surfaces in Seifert fibred spaces, Mich. Math. Journal 42 (1995), 525–535.
[39] J. Pitts and J.H. Rubinstein, Equivariant minimax and minimal surfaces in geometric 3-manifolds, Bulletin Amer. Math. Soc. 19 (1988), 303–309.
[40] J. Pitts and J.H. Rubinstein, Minimal surfaces and finiteness of Heegaard splittings of 3-manifolds, in preparation.
[41] J.H. Rubinstein, One-sided Heegaard splittings of 3-manifolds, Pacific J. of Math. 76 (1978), 185–200.
[42] J.H. Rubinstein, The polyhedral geometry of 3-manifolds, Lecture Notes in preparation for CBMS series.
[43] J.H. Rubinstein, Embedded minimal surfaces in 3-manifolds of positive scalar curvature, Proc. Amer. Math. Soc. 95 (1985), 458–462.
[44] J.H. Rubinstein, The topology of embedded minimal surfaces in hyperbolic 3-manifolds, in preparation.
[45] P. Scott, The geometries of 3-manifolds, Bull. London Math. Soc. 15 (1983), 401–487.
[46] M. Scharlemann and A. Thompson, Thin position for 3-manifolds, Geom. Top. Haifa 1992, Cont. Math. Series 164, Amer. Math. Soc. 1994, 231–238.
[47] F. Smith, On the existence of embedded minimal 2-spheres in the 3-sphere, endowed with an arbitrary Riemannian metric, PhD thesis, supervisor L. Simon, University of Melbourne 1982.
[48] R. Schoen, Curvature estimates for stable minimal surfaces in three-dimensional manifolds, Seminar on minimal submanifolds, Ann of Math. Studies 103, Princeton University Press, 1983, 111–126.
[49] R. Schoen and S.-T. Yau, Existence of incompressible surfaces and the topology of 3-dimensional manifolds with non-negative scalar curvature, Annals of Math. 119 (1979), 127–142.
[50] W. Thurston, The geometry and topology of 3-manifolds, lecture notes at Princeton University 1978.
[51] K. Uhlenbeck, Closed minimal surfaces in hyperbolic 3-manifolds, Seminar on minimal submanifolds, Ann of Math Studies 103, Princeton University Press, 1983, 147–168.
[52] B. White, Existence of least area mappings of N-dimensional domains, Annals of Math. 118 (1983), 179–185.
[53] B. White, The space of minimal submanifolds for varying Riemannian metrics, Indiana Math. Journal 40 (1991), 161–200.
[54] B. White, Curvature estimates and compactness theorems in 3-manifolds for surfaces that are stationary for parametric elliptic functionals, Invent. Math. 88 (1987), 243–256.
[55] S. T. Yau, Open problems in geometry, Proc. Symp. Pure and Appl Math. 54-I, Amer. Math. Soc., Providence, 1993.

Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria 3010, Australia
E-mail address: rubin@ms.unimelb.edu.au

Clay Mathematics Proceedings
Volume 2, 2004

Cousins of Constant Mean Curvature Surfaces

Karsten Große-Brauckmann

These notes are a detailed version of my two lectures at the Conference *The Global Theory of Minimal Surfaces* held at MSRI in Berkeley, June and July 2001. My first lecture is covered by Section 1 to 5 while Section 6 covers the second.

Based on the description of conjugate minimal surfaces, given in Section 1, conjugation for constant mean curvature surfaces is explained in Section 2. These surfaces are called *cousin* surfaces, following Bryant's terminology. Conjugation is presented in the language of differential forms, not in the classical formulation of the fundamental theorem of surfaces. Thus the standard integrability conditions of forms replace the Gauss and Codazzi equations. This leads to a simple algebraic description. It is complemented with a discussion of the geometry.

The examples of Section 4 indicate how existence proofs for sufficiently symmetric constant mean curvature surfaces can be given by constructing their conjugates. The discussion of these examples is based on the symmetry properties of conjugation, explained in Section 3. In addition, transversality properties of cousin surfaces are dealt with in Section 5. This is interesting in its own right, and will be useful in the next section. The goal of the first five sections is to give an overview on conjugate cousins; though much of the material is covered in [**GKS1**], the presentation here is broader. Let me add that the third section of [**Kr2**] contains the same approach using, however, a formulation in terms of classical surface theory.

Section 6 contains a discussion of the existence proof for constant mean curvature surfaces with genus zero and three ends, which are called triunduloids for short. This is work joint with R. Kusner and J. Sullivan [**GKS1**]. After explaining the result, I take these notes as an opportunity to explain the main geometric ideas of the somewhat involved proof in an intuitive way.

The computer graphics contained in the present paper was produced using software written by Bernd Oberknapp (based on a program by Konrad Polthier) for the graphics environment Grape, developed at the SFB 256 at Bonn University.

I thank the referee for the proposed corrections and helpful remarks. This place is a welcome opportunity to thank the organizers of the Clay Mathematics Institute Summer School, in particular David Hoffman, for putting together a conference in equal parts interesting and enjoyable.

©2004 Clay Mathematics Institute

1. Conjugate Minimal Surfaces

A good starting point for presenting the conjugate cousin construction is a discussion of the minimal surface case. As is well-known, each of the three coordinate functions of a conformally parametrized minimal surface represents the real part of a holomorphic function. The three imaginary parts are also minimal, and parametrize the conjugate surface. Let us recall the details in a way that will make our constant mean curvature presentation below an obvious generalization.

For $\Omega \subset \mathbb{R}^2$ we consider an immersion $f\colon \Omega \to \mathbb{R}^3$ which is minimal. We always assume the minimal surface is oriented by its normal or Gauss map $\nu : \Omega \to \mathbb{S}^2$, and f is orientation preserving. In general we avoid specializing to conformal parametrizations using the following Riemannian approach: We endow Ω with the pullback metric and let J be the oriented 90 degree rotation on Ω. Then the conjugate minimal surface can be described as follows:

THEOREM 1.1. *Let $f\colon \Omega \to \mathbb{R}^3$ be a minimal immersion, where $\Omega \subset \mathbb{R}^2$ is a simply connected domain. Then*

$$d\tilde{f} = df \circ J \tag{1–1}$$

has a solution $\tilde{f}\colon \Omega \to \mathbb{R}^3$, unique up to translation. Moreover, f, $\tilde{f}$ are isometric immersions with Gauss maps $\nu = \tilde{\nu}$, and $\tilde{f}$ is minimal.

The immersion $\tilde{f}$ is called the *minimal surface conjugate to f*.

We consider (1–1) an equality of 1-forms, stating how df and $d\tilde{f}$ act on tangent vectors. If f is conformal, then it satisfies $df \circ J = f_y\,dx - f_x\,dy$, as is seen by evaluation on the standard basis $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y} = J\frac{\partial}{\partial x}$. Thus for f conformal (1–1) becomes the Cauchy–Riemann system

$$\tilde{f}_x = f_y, \qquad \tilde{f}_y = -f_x.$$

A geometric interpretation of (1–1) is as follows (Figure 1). A tangent vector to the immersion f at $p \in \Omega$ is first rotated by 90 degrees, and then parallel translated from $T_{f(p)}\mathbb{R}^3$ to $T_{\tilde{f}(p)}\mathbb{R}^3$. The resulting vector is tangent to the conjugate minimal immersion $\tilde{f}$ at p.

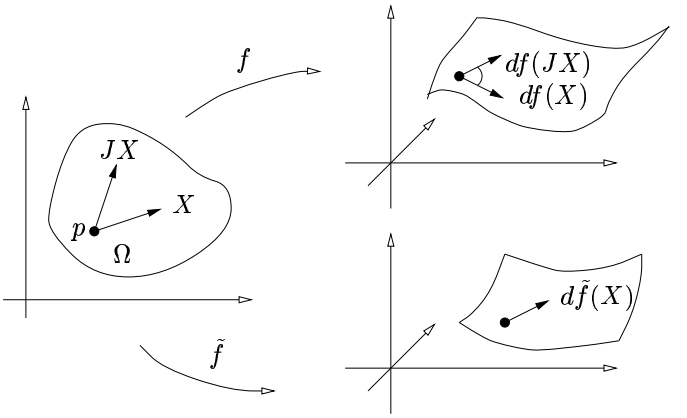


FIGURE 1. Conjugate minimal surfaces f and $\tilde{f}$.

PROOF. Before we show existence, let us show it implies the other properties claimed. Since J is an isometry of the pullback metric for f on Ω, the pullback metric for $\tilde{f}$ as in (1–1) is isometric; in particular $\tilde{f}$ must be an immersion. Moreover, since the tangent spaces at $p \in \Omega$ satisfy $T_p f = T_p \tilde{f}$ up to parallel translation, we have $\nu(p) = \tilde{\nu}(p)$ for the choice of normal consistent with the orientation on Ω.

The statement of the theorem remains invariant under orientation-preserving diffeomorphisms of Ω. Moreover, for any surface, there is a local parametrization in terms of conformal coordinates. Thus it suffices to prove the theorem under the assumption that f be conformal. For this case,

$$d(df \circ J) = f_{yy}\, dy \wedge dx - f_{xx}\, dx \wedge dy = -\Delta f\, dx \wedge dy. \tag{1–2}$$

But this vanishes since a conformal immersion f is minimal if and only if $\Delta f = 0$. Thus $df \circ J$ is closed. Since Ω is simply connected, $df \circ J$ is also exact, which means precisely that $\tilde{f}$ solving (1–1) exists.

It remains to show $\tilde{f}$ is minimal. We compose (1–1) with J to find $d\tilde{f} \circ J = df \circ J^2 = -df$, yielding

$$0 = d^2 f = -d(d\tilde{f} \circ J) \overset{(1-2)}{=} \Delta \tilde{f}\, dx \wedge dy.$$

It follows that $\Delta \tilde{f} = 0$. But as shown before $\tilde{f}$ is also conformal, and therefore minimal. □

The key point of our simple proof is (1–2), which we rephrase: The mean curvature equation for f happens to be the integrability condition for its conjugate $\tilde{f}$. In this form, the statement will continue to hold for the constant mean curvature case.

2. Cousins of Surfaces with Constant Mean Curvature 1

It is as interesting as useful to see that conjugation generalizes to surfaces with constant mean curvature 1, or CMC surfaces for short. However, the conjugate surface of a CMC surface in $\mathbb{R}^3$ is not CMC again, but turns out to be a minimal surface in $\mathbb{S}^3$. That is, we encounter a shift from mean curvature to ambient curvature. We will refer to a pair of such surfaces, defined precisely only below, as *cousins*.

Let us present the geometric picture before we give the algebraic description (see Figure 2). To relate a tangent vector of the CMC surface in $\mathbb{R}^3$ with the tangent vector of the cousin in $\mathbb{S}^3$ we depend on the Lie group structure of $\mathbb{S}^3$. Given the CMC surface $f\colon \Omega \to \mathbb{R}^3$, a tangent vector at the point $f(p)$ is first rotated by 90° in $T_p f$. By parallel translation, we then identify the ambient tangent space $T_{f(p)}\mathbb{R}^3$ with $\mathbb{R}^3$, which we in turn identify with $T_1\mathbb{S}^3$, the tangent space at unity $1 \in \mathbb{S}^3$. Left translation by $\tilde{f}(p)$ in $\mathbb{S}^3$ has a differential, mapping $T_1\mathbb{S}^3$ to $T_{\tilde{f}(p)}\mathbb{S}^3$. It finally maps our tangent vector in $T_1\mathbb{S}^3$ to a vector tangent to $\tilde{f}(p) \in \mathbb{S}^3$.

The formula describing this geometric recipe is

$$d\tilde{f} = dL_{\tilde{f}}(df \circ J). \tag{2–1}$$

Here, $L_q\colon \mathbb{S}^3 \to \mathbb{S}^3$ denotes left translation by q. At the point $1 \in \mathbb{S}^3$, it has the differential $dL_q\colon T_1\mathbb{S}^3 \to T_q\mathbb{S}^3$. When writing (2–1) we consider $\big(df \circ J\big)(X) \in T_{f(x)}\mathbb{R}^3 = T_1\mathbb{S}^3$ an identification. To show that (2–1) is solvable, that is, a cousin exists, it is convenient to reformulate the equation quaternionically.

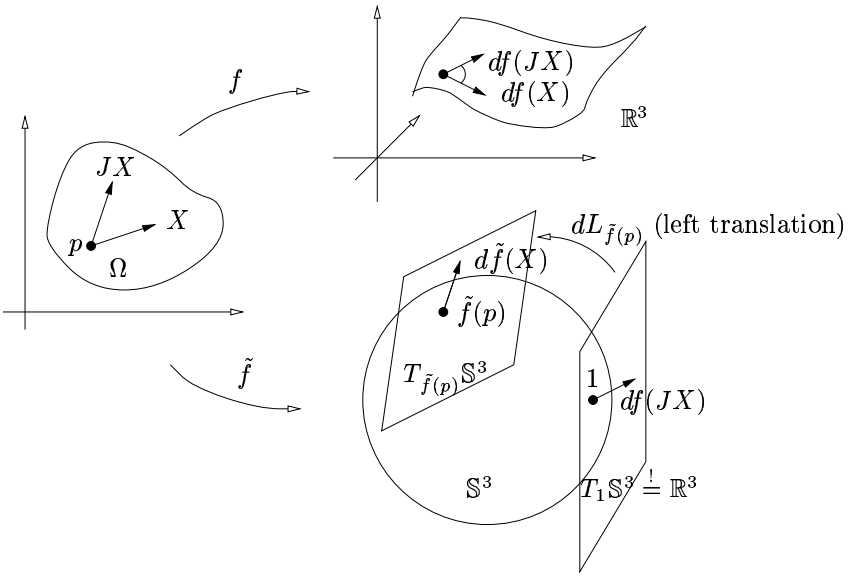


FIGURE 2. CMC surface f and its cousin $\tilde{f}$ schematically.

2.1. Quaternions. The quaternions $\mathbb{H}$ are the vector space $\mathbb{R}^4$ with basis $1, \mathbf{i}, \mathbf{j}, \mathbf{k}$, together with a bilinear product satisfying

$$1\mathbf{i} = \mathbf{i}, \quad \mathbf{i}^2 = -1, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \qquad \text{plus cyclic permutations.}$$

The Lie group $\mathbb{S}^3$ is the set of unit quaternions, with left translation becoming the quaternion product $L_p(q) := pq$ for $p, q \in \mathbb{S}^3$. Differentiation gives $dL_p(v) = pv$ where $v \in T_p\mathbb{S}^3$. Note that the right-hand side of the previous two quaternion formulas agree, although the first formula refers to the point $q \in \mathbb{S}^3$ while the second refers to the tangent vector $v \in T_p\mathbb{S}^3$. However unfamiliar this may appear in the case of the Lie group $\mathbb{S}^3$, it is nevertheless standard not to distinguish between points and tangent vectors for addition in Euclidean space.

Now we can write (2–1) as an equation of quaternion-valued one-forms:

$$d\tilde{f} = \tilde{f}(df \circ J)$$

As before, we have implicitly assumed an identification, namely that $df \circ J$ takes images in $T_1\mathbb{S}^3 = \operatorname{Im}\mathbb{H} := \operatorname{span}\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, the *imaginary quaternions*. This is achieved by either considering $\mathbb{R}^3 = \operatorname{Im}\mathbb{H}$ an identification or assuming directly that $f : \Omega \to \operatorname{Im}\mathbb{H} \subset \mathbb{H}$. In addition, we also find it convenient to identify $\operatorname{Re}\mathbb{H} := \mathbb{R} \subset \mathbb{H}$, similar to the common identification $\operatorname{Re}\mathbb{C} = \mathbb{R}$.

When we discuss integrability below, we will use the following algebraic rules. If $p \in \operatorname{Im}\mathbb{H}$, then $p^2 = -|p|^2 \in \operatorname{Re}\mathbb{H}$. More generally,

$$pq = -\langle p, q\rangle + p \times q = \operatorname{Re}(pq) + \operatorname{Im}(pq) \qquad \text{for } p, q \in \operatorname{Im}\mathbb{H}, \tag{2–2}$$

where we make use of the identification of $\mathbb{H}$ with $\mathbb{R} \oplus \operatorname{Im}\mathbb{H}$. So, geometrically, $p, q \in \operatorname{Im}\mathbb{H}$ commute if and only if they are parallel, and anticommute if and only if they are perpendicular.

2.2. The cousin theorem. Unlike for conjugate minimal surfaces, orientation matters when dealing with cousin surfaces. For this reason let us fix our orientation convention at this point.

1. When we say $f\colon \Omega \to \mathbb{R}^3$ is a CMC surface we mean that the mean curvature equals 1 with respect to a Gauss map ν (i.e., for $\mathbb{S}^2$ to be CMC we choose the inner normal for ν).
2. With respect to a given choice of Gauss map ν we define an orientation on Ω by requiring that J be chosen such that the frame $\big(df(X), df(JX), \nu\big)$ is positively oriented in $\mathbb{R}^3$ for X any nonzero tangent vector to Ω; equivalently $\nu = df(X) \times df(JX)$ whenever $|X| = 1$.
3. Let us now consider a conformally parametrized CMC surface f. Under the previous conventions the general curvature equation $\Delta f = 2Hf_x \times f_y$ becomes quaternionically $\Delta f = 2f_x f_y$; here we used that $f_x \perp f_y$ in (2–2).

THEOREM 2.1. *Let $\Omega \subset \mathbb{R}^2$ be a simply connected domain and $f\colon \Omega \to \mathbb{R}^3 = \operatorname{Im}\mathbb{H} \subset \mathbb{H}$ be a CMC immersion with Gauss map ν. Then*

$$d\tilde f = \tilde f\, df \circ J \tag{2–3}$$

has a solution $\tilde f\colon \Omega \to \mathbb{S}^3 \subset \mathbb{H}$, unique up to left translation. Moreover, $f, \tilde f$ are isometric immersions, with Gauss maps $\tilde\nu = \tilde f \nu$, and $\tilde f$ is minimal. Conversely, for each such minimal $\tilde f$ and choice of Gauss map $\tilde\nu$, there is a CMC immersion f with Gauss map $\nu = \tilde f^{-1}\tilde\nu$, solving (2–3), *unique up to translation.*

PROOF. We try to copy the proof step by step from Theorem 1.1, letting the calculations become quaternionic. As pointed out before, the key point is that the curvature equation of one surface happens to be the integrability equation for its cousin.

Both J and left multiplication with $\tilde f$ are isometries for the pullback metric. Therefore, f and $\tilde f$ solving (2–3) are isometric, and one of them is an immersion if and only if the other is.

Since the result is invariant under orientation preserving diffeomorphisms, we can assume that f (or $\tilde f$) is conformal. As pointed out above, in that case f is CMC if and only if $\Delta f = 2f_x f_y$; likewise, a conformal immersion $\tilde f$ into $\mathbb{S}^3$ is minimal if and only if $\tilde f$ is harmonic, that is $\Delta\tilde f = -\tilde f |d\tilde f|^2$.

Let us now check the integrability conditions. On a simply connected domain, closedness and exactness are equivalent. This time, let us start by assuming $\tilde f$ exists. Since $d\tilde f = \tilde f\, df \circ J$ is equivalent to $df = -\tilde f^{-1} d\tilde f \circ J$, to check for integrability of f amounts to verifying that

$$d(-\tilde f^{-1} d\tilde f \circ J) = -d(\tilde f^{-1}) \wedge d\tilde f \circ J - \tilde f^{-1} d(d\tilde f \circ J)$$

vanishes. Using $d(\tilde f^{-1}) = -\tilde f^{-1} d\tilde f\, \tilde f^{-1}$ we calculate the first term in conformal coordinates:

$$\begin{aligned}\tilde f^{-1}\, d\tilde f\, \tilde f^{-1} \wedge d\tilde f \circ J &= \tilde f^{-1}(\tilde f_x dx + \tilde f_y dy) \wedge \tilde f^{-1}(\tilde f_y dx - \tilde f_x dy)\\ &= -(\tilde f^{-1}\tilde f_x)^2\, dx \wedge dy + (\tilde f^{-1}\tilde f_y)^2\, dy \wedge dx = -\big((\tilde f^{-1}\tilde f_x)^2 + (\tilde f^{-1}\tilde f_y)^2\big) dx \wedge dy.\end{aligned}$$

Now $\tilde f^{-1}\tilde f_x$ is in the tangent space of unity, $T_1\mathbb{S}^3 = \operatorname{Im}\mathbb{H}$; indeed, since multiplying with $\tilde f$ is an isometry, $0 = \langle \tilde f, \tilde f_x\rangle = \langle 1, \tilde f^{-1}\tilde f_x\rangle$. It follows from (2–2) that

$$-\big(\tilde f^{-1}\tilde f_x\big)^2 - \big(\tilde f^{-1}\tilde f_y\big)^2 = |\tilde f^{-1}|^2\big(|\tilde f_x|^2 + |\tilde f_y|^2\big) = |d\tilde f|^2.$$

Combining these calculations with (1–2) we find

$$d(-\tilde{f}^{-1}d\tilde{f}\circ J) = \big(|d\tilde{f}|^2 + \tilde{f}^{-1}\Delta\tilde{f}\big)dx\wedge dy, \tag{2–4}$$

which indeed vanishes as $\tilde{f}$ is a harmonic map into $\mathbb{S}^3$. This establishes the existence of f, while (2–5) and (2–6) below prove f is CMC.

On the other hand, assume f is given and we want to solve (2–3) for $\tilde{f}$. Setting $\alpha := df\circ J$ this means we must solve

$$\tilde{f}^{-1}d\tilde{f} = \alpha;$$

but this time, the equation involves both $\tilde{f}$ and $d\tilde{f}$. It is known that for this case the Maurer–Cartan equation

$$d\alpha + \alpha\wedge\alpha = 0 \tag{2–5}$$

is the integrability condition, see Chapter 3 of Sharpe [**Sh**]. To verify (2–5), we use that for f conformal we have $\alpha = f_y dx - f_x dy$ and furthermore $f_y f_x = -f_x f_y$ by (2–2). This gives

$$\alpha\wedge\alpha = f_x f_y\, dx\wedge dy + f_y f_x\, dy\wedge dx = 2f_x f_y\, dx\wedge dy.$$

Together with (1–2) we obtain

$$\alpha + \alpha\wedge\alpha = \big(-\Delta f + 2f_x f_y\big)dx\wedge dy. \tag{2–6}$$

Thus the Maurer–Cartan equation holds precisely when f is CMC. Moreover, $\tilde{f}$ indeed takes values in $\mathbb{S}^3$, as $d\tilde{f}(X)$ is tangent to $\mathbb{S}^3$ for all $X\in T\Omega$; so an initial value in $\mathbb{S}^3$ confines the entire solution $\tilde{f}$ to $\mathbb{S}^3$. The fact that $\tilde{f}$ is minimal in $\mathbb{S}^3$ follows from $d^2 f = 0$ by (2–4). □

Let us note that there is also a second order description of cousins. It says that for cousins f and $\tilde{f}$ the respective shape operators S and $\tilde{S}$ satisfy

$$J\circ\tilde{S} = S - \mathrm{id}\,. \tag{2–7}$$

Hence $\tilde{S}$ is the $-90°$ rotation of the trace-free part of S; observe that it is necessary to rotate the trace-free part of S in order to obtain a symmetric cousin shape operator $\tilde{S}$. Applying the Gauss equation to (2–7) gives that for $\tilde{S}$ to be the shape operator of a surface $\tilde{f}$, the ambient curvature must be 1. Thus the "initial" mean curvature of f transforms to ambient sectional curvature of $\tilde{f}$.

The second order description of cousins is prior to the first order description which we presented above. Used by Lawson in 1970 [**Law**] it seems to go back to Calabi (see [**Kr2**]). On the other hand, the first order description was known to Karcher 1989 [**Kr1**] for the symmetry curves in cousins (see next section). In general, it appears to have been known quite some time before it was first referred to in the literature. As for written sources I am aware of an article by Konrad Polthier and myself (1996), a book by Frederic Hélein (1997), a diploma thesis by Bernd Oberknapp (Bonn, 1998), and [**GKS1**] (see there for references). I would like to add that Frank Pacard explained it to me in 1997.

With the term *cousin* we follow Bryant, who similarly related minimal surfaces in $\mathbb{R}^3$ with CMC surfaces in $\mathbb{H}^3$. In that case, however, a first order description is not known.

3. Symmetry Properties of Cousins

The main use of conjugate cousins has been to construct examples of CMC surfaces with sufficiently large symmetry. For this reason we discuss symmetry properties before we come to examples.

3.1. Hopf fibrations of $\mathbb{S}^3$. A particular circle fibration of $\mathbb{S}^3$, usually called *the* Hopf fibration, results from the decomposition $\mathbb{R}^4 = \mathbb{C} \times \mathbb{C}$. Here, we need an efficient description of the images of this fibration under the action of the group $\mathsf{SO}(4)$. Quaternions turn out to be a useful tool for doing this.

For each $u \in \mathbb{R}^3 = \operatorname{Im} \mathbb{H} = T_1\mathbb{S}^3$, the vector field $p \mapsto pu$ on $\mathbb{S}^3$ is left invariant. When $u \in \mathbb{S}^2$, the integral curve through $p \in \mathbb{S}^3$ is the great circle

$$t \mapsto \phi_t(p) := p(\cos t + u \sin t). \tag{3–1}$$

Indeed,

$$\phi_t'(p) = p(-\sin t + u \cos t) = p(u \sin t + \cos t)u = \phi_t(p)u,$$

using that $-u^2 = |u|^2 = 1$.

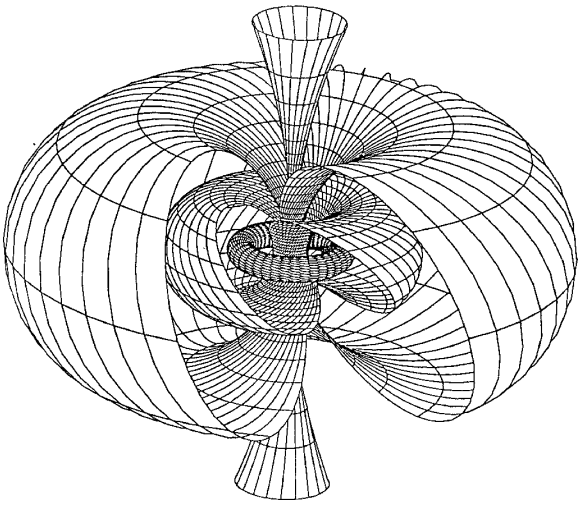


FIGURE 3. A Hopf fibration of $\mathbb{S}^3$, stereographically projected to $\mathbb{R}^3$. The distance tori in $\mathbb{S}^3$ about any u-Hopf circle are foliated with u-Hopf circles. For the two circles which project to the vertical line and the horizontal unit circle, the distance tori project to tori of revolution. Four of these tori are shown, with two of them cut along a pair of u-Hopf circles. (In addition to u-Hopf circles the depicted tori carry some horizontal circles which do not represent great circles of $\mathbb{S}^3$.) Figure courtesy of K. Polthier.

We will use the following terminology:

DEFINITION. Let $u \in \mathbb{S}^2 \subset \operatorname{Im} \mathbb{H}$.

- The *u-Hopf field* is the left invariant vector field $p \mapsto pu$ on $\mathbb{S}^3$.
- A *u-Hopf circle* is an integral curve of a u-Hopf field.
- The *u-Hopf projection* is the mapping $\Pi_u \colon \mathbb{S}^3 \to \mathbb{S}^2$, $\Pi_u(p) := pu\overline{p}$, where $\overline{p} := \operatorname{Re} p - \operatorname{Im} p$.

A quaternion v is in $\operatorname{Im}\mathbb{H}$ if and only if $\overline{v} = -v$. Moreover, $\overline{pq} = \overline{q}\,\overline{p}$ for all $p, q \in \mathbb{H}$. Hence

$$\overline{\Pi_u(p)} = \overline{pu\overline{p}} = \overline{\overline{p}}\,\overline{u}\,\overline{p} = -pu\overline{p} = -\Pi_u(p),$$

showing $\Pi_u(p) \in \operatorname{Im}\mathbb{H}$, while $|\Pi_u(p)| = 1$ is obvious; this gives that indeed $\Pi_u(p) \in \mathbb{S}^2$.

The u-Hopf circles u-Hopf project to points:

$$\begin{aligned}\Pi_u\big(p(\cos t + u\sin t)\big) &= p(\cos t + u\sin t)u(\cos t - u\sin t)\overline{p}\\ &= p(u\cos^2 t - u^3\sin^2 t)\overline{p} = \Pi_u(p).\end{aligned}$$

Conversely, it can be checked that the inverse image of a point in $\mathbb{S}^2$ is precisely a u-Hopf circle. Thus the u-Hopf projection is a circle fibration. It follows that each oriented great circle c in $\mathbb{S}^3$ is uniquely characterized by its Hopf field $u \in \mathbb{S}^2$ as well as its Hopf projection $\Pi_u(c) \in \mathbb{S}^2$. So in fact we have described a bijection of the Grassmannian of oriented 2-planes in $\mathbb{R}^4$ to $\mathbb{S}^2 \times \mathbb{S}^2$.

3.2. Symmetries of conjugate minimal surfaces. Consider a curve γ in Ω, parametrized with unit speed. Then

$$d\tilde{f}(\dot{\gamma}) \overset{(1\text{–}1)}{=} (df \circ J)(\dot{\gamma}) = df(J\dot{\gamma}). \tag{3–2}$$

Thus the vector tangent to the curve $\tilde{f} \circ \gamma$ is at the same time the conormal vector of the curve $f \circ \gamma$. Here, a vector $df(X)$ is called *conormal* if $X = J\dot{\gamma}(t)$, that is, $df(X)$ is a unit normal to the curve $f \circ \gamma$ taken within $T_{\gamma(t)}f$.

Suppose a minimal surface f contains a curve $f \circ \gamma$ with constant conormal $u \in \mathbb{S}^2$. Then not only is the curve contained in a plane $P \subset \mathbb{R}^3$ perpendicular to u, but the surface normal along $f \circ \gamma$ must be parallel to P. By the Schwarz reflection principle, reflection in P leaves the minimal surface invariant. Thus we call a curve with constant conormal a *curve of planar reflection* for the surface. >From (3–2) it follows that $f \circ \gamma$ is a curve of planar reflection, if and only if $\tilde{f} \circ \gamma$ has a constant tangent vector $u \in \mathbb{S}^2$; the latter is true precisely when $\tilde{f} \circ \gamma$ is a straight line.

This property is useful for constructing examples of symmetric minimal surfaces: When solving a Plateau problem for a spatial polygon of straight arcs, the conjugate minimal surface can be reflected across its planar boundary arcs. If the angles of the spatial polygon are integer fractions of π, reflection yields a surface in the sense of differential geometry. But note that instead of attaining a vertex angle $\frac{\pi}{k}$ a Plateau solution might also attain the complementary angle $2\pi - \frac{\pi}{k}$; so for the last statement to hold it is necessary to construct *barriers* (see [**Kr1**], [**G**]).

3.3. Symmetries of conjugate cousins. From the cousin theorem we derive:

COROLLARY 3.1. *$f \circ \gamma$ is a curve of planar reflection, for a plane perpendicular to $u \in \mathbb{S}^2$, if and only if $\tilde{f} \circ \gamma$ traces out a u-Hopf circle of $\mathbb{S}^3$.*

PROOF. If $f \circ \gamma$ is a curve of planar reflection it has the constant conormal $(df \circ J)(\gamma') = u \in \mathbb{S}^2$. Hence by (2–3),

$$(\tilde{f} \circ \gamma)' = d\tilde{f}(\gamma') = \tilde{f} \circ \gamma(df \circ J)(\gamma') = (\tilde{f} \circ \gamma)u,$$

meaning that the curve $\tilde{f} \circ \gamma$ is an integral curve of the u-Hopf field. Thus it is contained in a u-Hopf circle. The same proof shows the converse. □

4. Examples

We want to illustrate how conjugate cousins yield existence proofs by discussing two examples. We skip the simplest example, which is the sphere $\mathbb{S}^2 = \operatorname{Im}\mathbb{H} \cap \mathbb{S}^3 \subset \mathbb{H}$; it presents the only CMC surface which agrees with its cousin.

4.1. Lawson surfaces. Periodic CMC surfaces contained in a slab provide the simplest type of example. These surfaces also appear as interfaces in biological and chemical systems. They can be regarded as perforated double membranes, that is, they look like a double cover of the plane perforated periodically with small handles. There are three one-parameter families of such surfaces, having different symmetry groups [**G**] (see also [**Kr1**]). Two of these three families contain a surface first described by Lawson [**Law**].

For each surface, the desired fundamental domain is simply connected, namely a disk. Its boundary loop is contained in the boundary of a half-infinite prism over a right-angled triangle. To obtain embedded surfaces, it is necessary that this triangle tessellates the plane. More precisely, each of the four planar boundary components of the prism contains a boundary arc of planar reflection. By the Symmetry Corollary 3.1 the cousin has four great circle arcs as boundary, whose Hopf fields are given by the oriented normals to the containing symmetry planes. The lengths of these great circle arcs are not immediate from the symmetries.

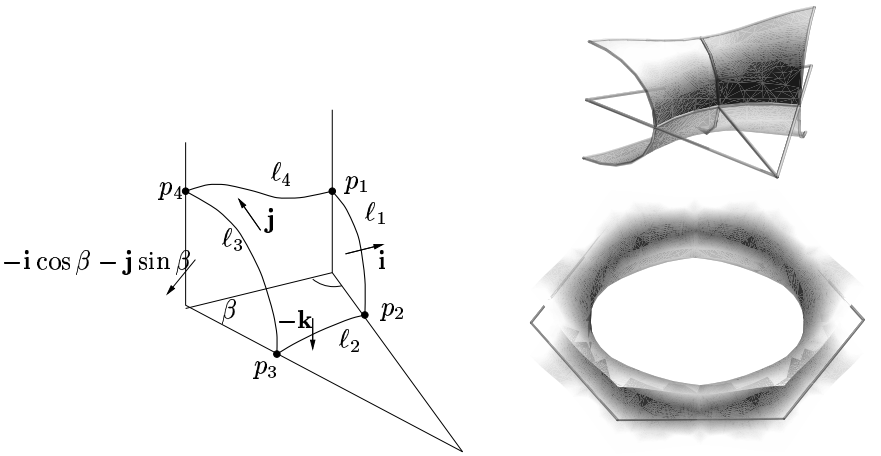


FIGURE 4. Left: Boundary polygon of a fundamental domain of a Lawson surface. Right: Four reflected copies of the fundamental domain (top), and 24 copies forming a translational fundamental domain (bottom); shown is the surface with $\beta = \frac{\pi}{3}$.

Choosing a labeling as in Figure 4 we can express the vertices $\tilde{p}_i$, $i = 1, \ldots, 4$ of the cousin polygon in terms of the as yet unknown lengths $\tilde{\ell}_i = \ell_i$ of the polygon. To remove the freedom of left translating the entire polygon we find it convenient to fix $\tilde{p}_1 := 1 \in \mathbb{S}^3$. Then

$$\begin{aligned} \tilde{p}_2 &= \cos\ell_1 + \mathbf{i}\sin\ell_1, \quad \tilde{p}_3 = \tilde{p}_2\big(\cos\ell_2 - \mathbf{k}\sin\ell_2\big), \\ \tilde{p}_4 &= \tilde{p}_3\big(\cos\ell_3 - (\mathbf{i}\cos\beta + \mathbf{j}\sin\beta)\sin\ell_3\big). \end{aligned} \tag{4–1}$$

Now the spherical quadrilateral is closed, that is,

$$\tilde{p}_4\big(\cos\ell_4 + \mathbf{j}\sin\ell_4\big) = \tilde{p}_1 = 1. \tag{4–2}$$

Relations satisfied by the lengths $\ell_1, \ell_2, \ell_3, \ell_4$ can be derived algebraically from (4–1) and (4–2). Indeed, spherical trigonometry gives the following statement for the quadrilateral (see [**G**]):

LEMMA 4.1. *For each $0 < \beta < \frac{\pi}{2}$ there is a one-parameter family of quadrilaterals $\Gamma(t,\beta) \subset \mathbb{S}^3$ parametrized with $0 < t < \frac{\pi}{2}$. More precisely, specifying $t := \ell_4(t) \in (0, \frac{\pi}{2})$, we can find distinct points $1 = \tilde{p}_1$, $\tilde{p}_2(t)$, $\tilde{p}_3(t)$, $\tilde{p}_4(t)$ in $\mathbb{S}^3$ and lengths $0 < \ell_1(t)$, $\ell_2(t)$, $\ell_3(t)$, $\ell_4(t) = t < \frac{\pi}{2}$ satisfying (4–1) (4–2).*

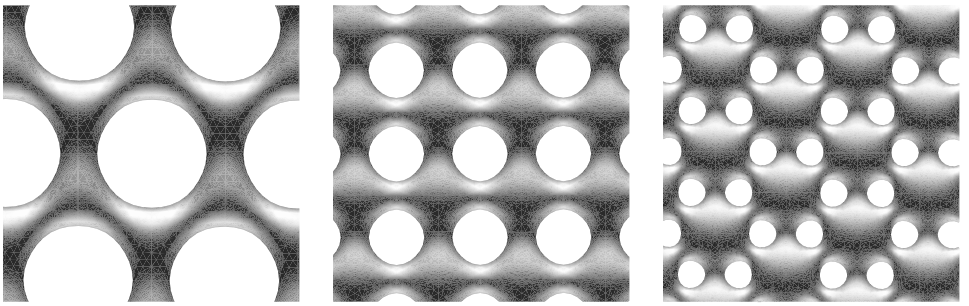


FIGURE 5. The two Lawson CMC surfaces (first two images) and a similar doubly periodic surface (third image). Translational fundamental domains are contained in a hexagonal, square, and rhombic prism, respectively. These surfaces are the cousins of the Lawson minimal surfaces $\xi_{2,2}, \xi_{3,3}, \xi_{5,5} \subset \mathbb{S}^3$ [**Law**]. The three images are shown at the same scale (H agrees).

The Plateau problem for $\Gamma(t,\beta)$ can be solved with the Morrey solution $\widetilde{M}(t,\beta)$; this surface does not have interior branch points. Its CMC cousin $M(t,\beta)$ has four arcs of planar reflection contained in the four faces of the triangular prism. Reflection in these faces leads to the desired complete and doubly periodic CMC surfaces when β equals $\frac{\pi}{6}, \frac{\pi}{4}$, or $\frac{\pi}{3}$ (Figure 5). To be exact, the result is a surface which has branch points possibly on the edges. To assert that no branching occurs, it is necessary to enclose the Plateau solution between barriers. Since all edges of the quadrilateral $\Gamma(t,\beta)$ have length at most $\frac{\pi}{2}$ we can use great spheres as barriers: The Plateau solution $\widetilde{M}(t,\beta)$ is contained in the intersection of four hemispheres of $\mathbb{S}^3$, each of which contains a different triple of the four vertices of $\Gamma(t,\beta)$.

We refer the reader to [**G**] for details. Let us mention that embeddedness for the surfaces in the three families $M(t,\frac{\pi}{6})$, $M(t,\frac{\pi}{4})$, $M(t,\frac{\pi}{3})$ has not yet been proven. Lawson claimed his two surfaces are embedded [**Law**, Thm.9], but did not supply a proof.

It is not likely that the symmetry displayed by the Lawson examples is necessary for embedded CMC surfaces to exist in a slab; however, no general method to produce similar unsymmetric examples is known. Let us add that any embedded CMC surface in a slab is expected to have a mirror symmetry plane. In the doubly periodic case it is straightforward to prove this fact using Alexandrov reflection; in general, however, the statement remains open.

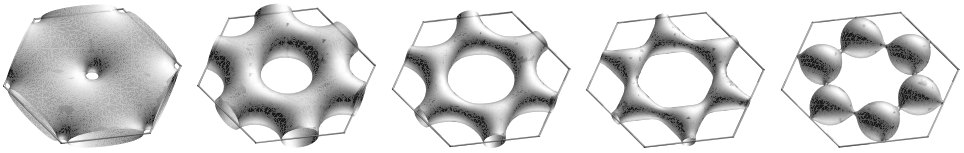


FIGURE 6. The surface family $M(t, \frac{\pi}{3})$, continued by reflection, is shown for five values of the parameter t. The family degenerates in a doubly covered plane (left), and in spheres (right); it contains the hexagonal Lawson surface (third image). The value of H is chosen such that the lattice of the surfaces agrees.

Quite similar is the existence proof for the families of CMC surfaces that contain the most well-known triply periodic minimal surfaces, namely the Schwarz P (primitive) and D (diamond) surface (see Figure 7).

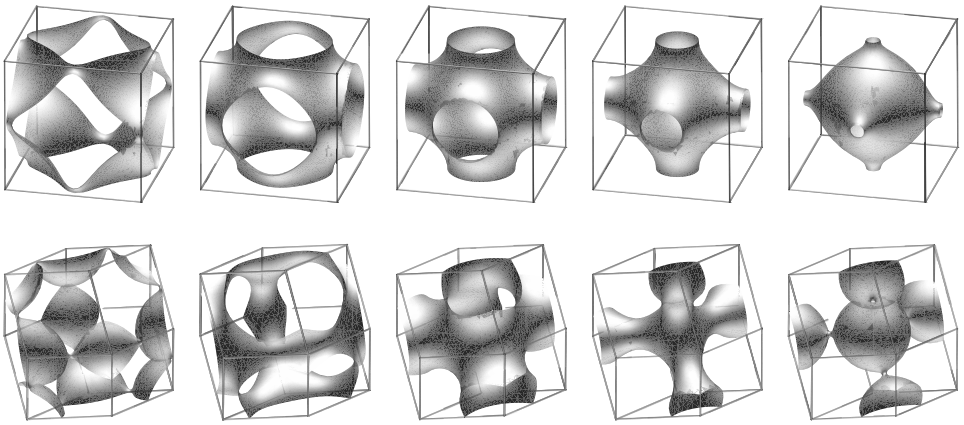


FIGURE 7. Top: The Schwarz P surface (middle) is contained in a family of CMC surfaces. Bottom: Similarly, the Schwarz D surface (middle) and its CMC family. For each surface family, the value of H is chosen such that the lattice agrees, that is, the depicted translational fundamental of a cube or rhombic dodecahedron appears at the same size. In each case, the faces of the containers contain curves of planar reflection of the surfaces, so that they can be continued periodically by taking mirror images.

4.2. Delaunay unduloids. In 1841 Delaunay determined the CMC surfaces of revolution [**D**]. Here we are interested in his one-parameter family of embedded surfaces, called *unduloids*, and will not consider the immersed *nodoids*. According to Delaunay's famous characterization, the meridian of an unduloid is the trace of the focus of an ellipse when the ellipse is rolled along a line. For H to be equal to 1, one semiaxis must have length $\frac{1}{2}$ while the other has length $0 < a \leq \frac{1}{2}$. Thus for $a = \frac{1}{2}$, the ellipse is a circle generating the CMC cylinder of radius $\frac{1}{2}$, while for $a \to 0$ it degenerates to a line of length 1 whose focus on an endpoint generates the meridian of a chain of spheres. The smallest circle on the CMC surface

has perimeter $n = 2\pi a$; we find it convenient to use this *necksize* $n \in (0, \pi]$ to parametrize the unduloid family.

FIGURE 8. Delaunay unduloids, the embedded CMC surfaces of revolution.

Each meridian of a surface of revolution is a curve of planar reflection. By the Symmetry Corollary 3.1 the cousin minimal surface must be ruled by great circles. (Similarly, the minimal surface of revolution is the catenoid, and its conjugate, the helicoid, is ruled by straight lines.)

We will now show how we can determine the parametrization $\tilde{f}(x, y)$ of the unduloid cousin in $\mathbb{S}^3$ from the symmetries of the CMC surface alone. Suppose $f(x, y)$ parametrizes a Delaunay unduloid, such that each x-curve $x \mapsto f(x, y)$ is a meridian and each y-curve $y \mapsto f(x, y)$ is a circle (see Figure 9). Moreover, for the x-meridians we require that they be parametrized with unit speed, while we parametrize each y-circle with constant speed by the interval $y \in (-\frac{1}{2}, \frac{1}{2}]$.

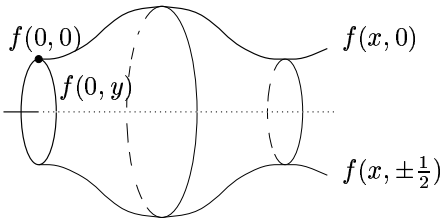


FIGURE 9. Parametrization of the unduloid.

We construct $\tilde{f}(x, y)$ in two steps. We assume the unduloid has the $\mathbf{i}$-line as its axis, and $f(0,0)$ is the top point (largest third coordinate) of a neck circle. Since the neck circle is perpendicular to an $\mathbf{i}$-plane and has perimeter n, it follows from Corollary 3.1 that $\tilde{f}(0, y)$ is the $\mathbf{i}$-Hopf circle

$$\tilde{f}(0, y) = \cos ny + \mathbf{i} \sin ny,$$

where we fixed the freedom to left translate by requiring $\tilde{f}(0,0) = 1 \in \mathbb{S}^3$.

Each meridian $x \mapsto f(x, y)$ is contained in a plane with normal $u(y)$ perpendicular to $\mathbf{i}$. To determine $u(y)$ explicitly, note that when y ranges from $-\frac{1}{2}$ to $\frac{1}{2}$, the normal $u(y)$ rotates once on a full circle. Moreover, since $u(0)$ contains a top point of a circle, $u(0) = \mathbf{j}$. This gives $u(y) = \mathbf{j} \cos 2\pi y + \mathbf{k} \sin 2\pi y$. By the corollary again, the cousin great circles of the meridians are the $u(y)$-Hopf circles through $\tilde{f}(0, y)$, that is, they are given by $y \mapsto \tilde{f}(0, y)\big(\cos x + u(y) \sin x\big)$. Consequently, we determined the unduloid cousins as

$$\tilde{f}(x, y) = \tilde{f}(0, y)\big(\cos x + u(y) \sin x\big).$$

Note that for each fixed y, the map $x \mapsto \tilde{f}(x, y)$ is indeed a great circle, so that $\tilde{f}$ is a ruled surface. For reasons explained in [**GKS1**], the surface $\tilde{f}$ can be considered a helicoid in $\mathbb{S}^3$.

It is Delaunay's result that f is CMC. Hence $\tilde{f}$ is minimal in $\mathbb{S}^3$ by the Cousin Theorem 2.1. On the other hand, it is not hard to check directly that $\tilde{f}$ is minimal in $\mathbb{S}^3$ (see [**G**]). Using this as a fact, the Cousin Theorem 2.1 can be used to establish the existence of the CMC surfaces of revolution.

5. Graph and Transversality Properties

A surface in $\mathbb{R}^3$ is a graph if it meets each vertical line in at most one point. While conjugation of minimal surfaces does not preserve the graph property, it preserves a weaker property: transversality to vertical lines. Indeed, if a surface is transversal to the vertical lines, then its Gauss map ν omits the horizontal direction. Since $\tilde{\nu} = \nu$ for conjugate minimal surfaces by Theorem 1.1, this property extends to the conjugate.

If a CMC surface is transversal to the vertical direction, then similarly we can make a transversality statement for its cousin:

LEMMA 5.1. *A CMC surface in $\mathbb{R}^3$ is transverse to the lines in direction $u \in \mathbb{S}^2$ if and only if its minimal cousin is transverse to the u-Hopf great circles. In the latter case, the u-Hopf projection of the surface is an immersion.*

PROOF. Left multiplication with $\tilde{f}$ gives an isometry of $T_x f \subset T_{f(x)}\mathbb{R}^3$ with $T_x\tilde{f} \subset T_{\tilde{f}(x)}\mathbb{S}^3$. Therefore

$$\langle \nu, u \rangle = \langle \tilde{f}\nu, \tilde{f}u \rangle \overset{\text{Thm.2.1}}{=} \langle \tilde{\nu}, \tilde{f}u \rangle,$$

and the lemma holds if this is nonzero everywhere. □

The upper half of a Delaunay surface U with $\mathbf{i}$-axis is a graph since its meridian curve is embedded. Let us choose a parametrization $f(x, y)$ of the Delaunay surface as in Section 4.2 and take $\Omega := \mathbb{R} \times (-\frac{1}{4}, \frac{1}{4})$ to parametrize its open upper half U^+. Since $f(\Omega)$ is a graph in the $\mathbf{k}$-direction of $\mathbb{R}^3$, the lemma gives that the cousin surface $\tilde{f}(\Omega)$ is transversal to the $\mathbf{k}$-Hopf circles of $\mathbb{S}^3$. In particular, $(\Pi_{\mathbf{k}} \circ \tilde{f})(\Omega)$ is an immersed disk in $\mathbb{S}^2$. Using symmetries and explicit calculation, we can characterize this disk geometrically (see [**GKS1**]):

LEMMA 5.2. *The Hopf projection of an unduloid half of necksize n, that is, the immersed disk $(\Pi_{\mathbf{k}} \circ \tilde{f})(\Omega)$, is isometric to the universal covering of $\mathbb{S}^2$ minus two points in spherical distance n.*

If we wish to consider a single bubble of U^+, enclosed between two neck semicircles, then its Hopf projection is a *slit sphere* (see Figure 10): It covers $\mathbb{S}^2$ once, with the boundary mapping to a geodesic arc (the slit) between two points in spherical distance n. Similarly the Hopf projection of two adjacent bubbles has degree 2; it is obtained by gluing two slit spheres with the same bounding slit together across the slit. Continuing, we see how the infinitely many bubbles in U^+ give the covering of infinite degree of the lemma; we consider it *a line of slit spheres*. Later, we will also use the notion of a *ray of slit spheres*. Removing a minimizing geodesic arc from a line of slit spheres decomposes it in an obvious way into rays of slit spheres.

6. Triunduloids

In recent work of the present author joint with Kusner and Sullivan [**GKS1**], the cousin construction was used to construct an entire moduli space of CMC surfaces.

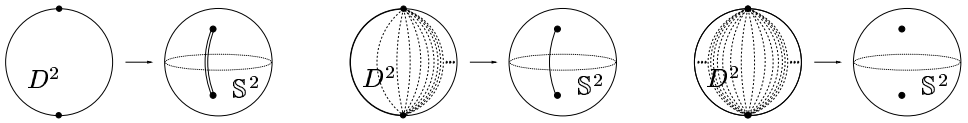


FIGURE 10. Left: A slit sphere is an immersion with degree 1 of the open disk to the complement of a minimizing geodesic arc in $\mathbb{S}^2$. Middle: A ray of slit spheres is an immersion with infinite degree; the preimages of the slit are indicated. Right: A line of slit spheres represents the universal covering of $\mathbb{S}^2$ minus two distinct points; preimages of the minimizing geodesic between these two points are indicated.

There is a short published announcement explaining the result of this work [**GKS2**]. In the following presentation we place more emphasis on the main ideas of the proof, in particular we will indicate how the cousin method is applied.

6.1. Examples of CMC surfaces which are Alexandrov embedded. The existence of a wealth of embedded CMC surfaces has only been established in the past two decades. With the unique compact example being $\mathbb{S}^2$ [**A1**], the task has been to find the proper CMC embeddings of the topological type of a compact surface Σ with genus g and k punctures. Properness means that a neighborhood of each of the k punctures in Σ maps to a neighborhood of infinity in $\mathbb{R}^3$; we call each image an *end* of the surface. No example with $k = 1$ is possible, and for $k = 2$ ends the Delaunay unduloids are the only surfaces, implying that $g = 0$ [**KKS**]. However, for all the cases $k \geq 3$, $g \geq 0$ Kapouleas found examples [**Kp**], and some more examples for all $k \geq 3$ with $g = 0$ were given by Mazzeo and Pacard [**MP**] as well as the present author [**G**].

Some of these examples are actually not embedded. Though we expect the remaining examples to be embedded, a proof of their embeddedness is typically tedious to give. Only for some Kapouleas surfaces with $g = 0$ and $k \leq 12$ it has been given. Nevertheless, most of the surfaces mentioned before are easily seen to be *Alexandrov embedded*, meaning roughly that they are immersions which extend as immersions to the 'interior' of the parametrizing surface Σ (see Figure 11, 12). The rigorous definition of Alexandrov embeddedness is as follows: We take Σ embedded in $\mathbb{R}^3$ and let Ω denote its interior with compact closure. Then we require there be a proper immersion of Ω extending smoothly to the given CMC immersion of $\Sigma = \partial\Omega$ minus the k punctures; moreover we demand that the mean curvature vector points into Ω.

Let us note for the interested reader that there are errors in the literature concerning Alexandrov embeddedness. In his original article [**A2**] of 1962, Alexandrov (or his translator?) requires the extending map only to be a *smooth mapping* instead of an *immersion* (see II_0 on page 303); moreover, in the proof of the reflection method, the maximum principle seems applied to two components of a CMC surface with opposite normal (p. 314). Spivak, in volume IV of his treatise [**S**], gives a wrong obstruction to Alexandrov embeddedness on page 511/512. It fails for the Alexandrov embedded circle with overlap shown in Figure 11 middle (or the similarly immersed sphere $\mathbb{S}^2$): If the overlap region contains the origin, the radial

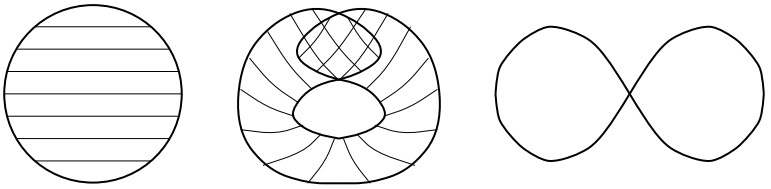


FIGURE 11. The first two immersions are Alexandrov embedded: These immersions of $\mathbb{S}^1$ extend to the disk D^2 as indicated. However, the figure-8 curve shown on the right is not Alexandrov embedded. The images are meant with multiplicity 1: their k-fold coverings would not be Alexandrov embedded when $k > 1$.

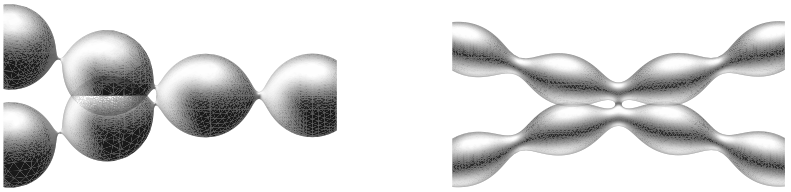


FIGURE 12. CMC immersions which are Alexandrov embedded, but not embedded.

projection to $\mathbb{S}^1$ (respectively, $\mathbb{S}^2$) has degree 2. However, Spivak asserts it must have degree 1 for an Alexandrov embedded surface.

Alexandrov embedded surfaces arose originally as the generalized class of surfaces to which the Alexandrov reflection technique applies. While Alexandrov himself observed this for the case of compact CMC surfaces, the following extension to non-compact CMC surfaces is due to Korevaar, Kusner, Solomon [**KKS**]:

THEOREM 6.1. *Let M be a proper CMC Alexandrov embedding of a compact surface Σ of genus g with k punctures.*
(i) If M is contained in a slab, there is a decomposition $M = M^+ \cup M^- \cup M_0$. Here, M_0 is the intersection of M with a mirror symmetry plane P, and $M^\pm$ are the two symmetric connected components of $M - P$. Moreover, the Gauss image of each symmetric half $M^\pm$ is contained in a hemisphere, $\nu(M^\pm) \subset \mathbb{S}^2_\pm$. Finally, $\nu(M_0) \subset P$, so that M_0 consists of curves of planar reflection.
(ii) Each end of M is asymptotic at infinity to a Delaunay unduloid of necksize $n \in (0, \pi]$.
(iii) In addition, if $k = 3$, then M is contained in a slab.

6.2. Existence and uniqueness for triunduloids. We call a proper constant mean curvature Alexandrov embedding of $\mathbb{S}^2$ minus three fixed punctures a *triunduloid*. We consider each parametrizing puncture a *labelling* of an end. We let $\mathcal{M}$ be the topological space of all triunduloids modulo those rigid motions of $\mathbb{R}^3$, that do not interchange the three punctures. Note that a similar *labeling* of the ends is familiar from the definition of Teichmüller space. The topology on $\mathcal{M}$ is induced by Hausdorff distance: A sequence in $\mathcal{M}$ converges when it can be represented with surfaces whose Hausdorff distance converges on each compact subset of $\mathbb{R}^3$.

The main result of [**KKS**] is the classification of $\mathcal{M}$ by an open 3-ball $\mathcal{T}$, namely the set of triples $(q_1, q_2, q_3) \in \mathbb{S}^2 \times \mathbb{S}^2 \times \mathbb{S}^2$ of pairwise distinct points in $\mathbb{S}^2$ modulo the action of the rotation group $\mathsf{SO}(3)$.

THEOREM 6.2. *There is a homeomorphism $\Psi\colon \mathcal{M} \to \mathcal{T}$, having the following property: If a representing triunduloid $M \in \mathcal{M}$ has necksizes n_1, n_2, n_3, then the triple $\Psi(M)$ has spherical distances n_1, n_2, n_3.*

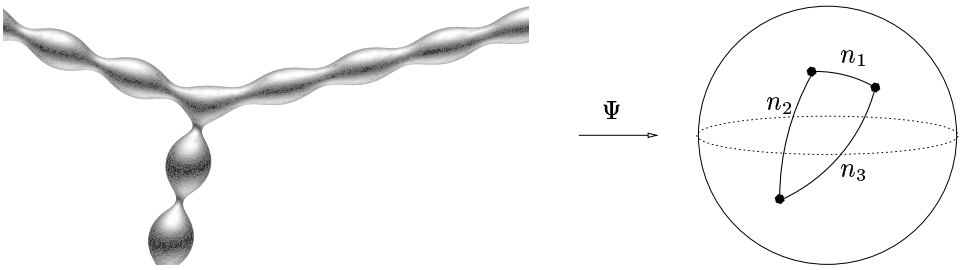


FIGURE 13. A triunduloid and its classifying triple in $\mathbb{S}^2$ (schematically).

This result at the same time establishes the existence of all triunduloids, as well as classifies the possible triunduloids by their asymptotic necksizes:

COROLLARY 6.3. *Let a triple $0 < n_1, n_2, n_3 \leq \pi$ be given.*
(i) For the triple to be the asymptotic necksizes of a triunduloid, there must hold

$$n_1 \leq n_2 + n_3, \quad n_2 \leq n_3 + n_1, \quad n_3 \leq n_1 + n_2, \qquad n_1 + n_2 + n_3 \leq 2\pi; \tag{6–1}$$

in particular at most one of the n_i can attain the value π corresponding to an end asymptotic to a CMC cylinder.
(ii) If all inequalities (6–1) are strict, there are exactly two different triunduloids in $\mathcal{M}$ with these necksizes; if one inequality holds with equality, there is just one such triunduloid.

Since $\Psi(M)$ is a spherical triple for each $M \in \mathcal{M}$, spherical trigonometry proves the corollary.

An open question regarding this result is whether all L^2-Jacobi fields vanish on each $M \in \mathcal{M}$, that is, whether the linearized mean curvature equation $\Delta_M u + |A|^2 u = 0$ has no nontrivial solutions $u \in L^2$ (here A is the shape operator of M). Surfaces with this property are called *nondegenerate*. Mazzeo and Pollack establish the existence of certain nondegenerate triunduloids with small necksizes [**MP**]. Ratzkin proved the following [**R**, Thm. 11]: Two triunduloids which are nondegenerate and have a common necksize can be glued end-to-end; the connecting pieces must be suitably long. (I believe that in Theorem 11 it must be assumed in addition that the common necksize be different from π; a similar condition is indeed included in Ratzkin's Theorem 1.) Since the resulting surfaces are nondegenerate, Ratzkin's construction can be iterated as to yield CMC surfaces with genus 0 and k ends. Provided our triunduloids are nondegenerate, their use in Ratzkin's construction would give CMC surfaces with genus 0 and $k > 3$ whose asymptotic necksizes are large (unlike those of [**Kp**]).

I would like to mention some related work. Schmitt [**KMS**], as well as Dorfmeister and Wu (work in progress), use a Weierstrass-type representation formula to represent triunduloids. For hyperbolic space, CMC surfaces with three ends were studied and classified by Umehara and Yamada [**UY**] as well as Collin and Rosenberg (unpublished).

6.3. The proof. Formally, the proof has four steps: 1. Definition of Ψ, 2. Ψ is injective, 3. Ψ is proper, 4. Ψ is surjective and homeomorphism. We want to explain the main geometric idea of each step.

6.3.1. *Definition of* Ψ. To define Ψ, we use Theorem 6.1. With part (iii) applying to any triunduloid $M \in \mathcal{M}$, part (i) yields a symmetric decomposition $M = M^+ \cup M^- \cup M_0$. Moreover, since the genus of M is 0, the sets $M^\pm$ are disks. Thus we can apply Theorem 2.1 to find a minimal cousin $\widetilde{M^+} \subset \mathbb{S}^3$.

As asserted by Theorem 6.1(ii), the three curves of the intersection M_0 of M with the symmetry plane P are curves of planar reflection of M. We find it convenient to assume from now on that the symmetry plane is horizontal, so that $\mathbf{k}$ is the normal of P. By the Symmetry Corollary 3.1 it follows that $\widetilde{M^+}$ is bounded by three $\mathbf{k}$-Hopf circles. Then we can simply use the $\mathbf{k}$-Hopf projection of $\widetilde{M_0} = \partial\widetilde{M^+}$ to define a map of $\mathcal{M}$ to $\mathcal{T}$ by

$$\Psi(M) := \Pi_{\mathbf{k}}(\partial\widetilde{M^+}).$$

Using the asymptotics result Theorem 6.1(ii) and passing to the cousin, it can be seen that indeed the spherical distances of $\Psi(M)$ are the necksizes of M.

6.3.2. Ψ *is injective.* Let M and N be triunduloids. We want to show that $\Psi(M) = \Psi(N)$ implies $M = N$. We will do this by invoking the cousins. On the one hand, by taking the inverse image of the common triple $\Psi(M) = \Psi(N)$ under $\Pi_{\mathbf{k}}$, the assumption implies that the cousins have the same great circle boundary $\partial\widetilde{M^+} = \partial\widetilde{N^+}$. Note that these great circles are covered infinitely often, since they have infinite length, just as the three curves of planar reflection in the (isometric) CMC surfaces have infinite length. On the other hand, it suffices to establish $\widetilde{M^+} = \widetilde{N^+}$ as this implies $M = N$ by passing to the cousin, followed by reflection. In conclusion, to prove the injectivity of Ψ is equivalent to showing the following: If the spherical boundaries $\partial\widetilde{M^+} = \partial\widetilde{N^+}$ agree, the minimal surfaces $\widetilde{M^+} = \widetilde{N^+}$ coincide.

We prove this claim by applying the maximum principle. Unfortunately, this is not straightforward: Our surfaces $\widetilde{M^+}, \widetilde{N^+} \subset \mathbb{S}^3$ are neither compact nor embedded! Consequently, it is not clear if the two surfaces with common boundary can be moved so as to touch one another tangentially with one-sided contact.

Let us consider an analogous situation in $\mathbb{R}^3$. The doubly periodic Scherk surface S has a translational fundamental domain over the unit square $Q := [0,1] \times [0,1]$. Its boundary consists of the four vertical lines over the vertices of Q. Jenkins and Serrin prove a maximum principle saying that any minimal graph in $Q \times \mathbb{R}$ with the same boundary and the same asymptotics is a vertical translate of the Scherk surface.

To carry the analogy a step further, let us consider a quotient in the z-direction: The circle bundle over the square, $T := Q \times \mathbb{S}^1$, contains the quotient Scherk surface S_T as an immersed minimal surface. The boundary of S_T is the four geodesic fibers over the vertices of Q; they are covered infinitely often. Assuming we are

given $S_T \subset T$, we could reconstruct the original embedded Scherk surface S from the simply connected surface S_T as a lift in the universal covering $Q \times \mathbb{R}$ of T. Since we can pass to the covering, the Jenkins–Serrin maximum principle allows us to formulate again a maximum principle, saying that two immersed minimal surfaces in T with the same boundary and asymptotics agree.

Like $S_T \subset T$, our minimal surfaces $\widetilde{M^+}, \widetilde{N^+} \subset \mathbb{S}^3$ cover their three bounding great circles infinitely often. Moreover, by Theorem 6.1(ii) their asymptotics agree: Each end of M, N is asymptotic to some unduloid and so (2–3) gives that each end of our minimal disks $\widetilde{M^+}, \widetilde{N^+} \subset \mathbb{S}^3$ is asymptotic to some unduloid cousin half. In order to apply the maximum principle, we must specify an ambient space E which contains an embedded lift of the cousin disks $\widetilde{M^+}, \widetilde{N^+}$.

The isometric flow in the z-direction, which leaves the bounding circles of the Scherk surface S_T invariant, corresponds to the isometries ϕ_t of (3–1), leaving the great circle boundaries of the cousins invariant. Moreover, as S_T is transversal to the z-circles, so are the disks $\widetilde{M^+}, \widetilde{N^+}$ transversal to the $\mathbf{k}$-Hopf circles by the Transversality Lemma 5.1. Therefore, we must take the $\mathbf{k}$-Hopf circles as the fibers of our desired bundle, and take the $\mathbf{k}$-Hopf projection of $\widetilde{M^+}$ to define the base manifold of the bundle. The sphere $\mathbb{S}^3$ is the Hopf circle bundle over its Hopf projection $\mathbb{S}^2$; unlike the circle bundle $T = Q \times \mathbb{S}^1$ this bundle does not simplify to a product. So our ambient space E must be the universal covering of the $\mathbf{k}$-Hopf circle bundle over the $\mathbf{k}$-Hopf projection of $\widetilde{M^+}$. The group of isometries ϕ_t induces a group of isometries on E, which makes E a principal bundle.

There is a further complication of the cousin surfaces compared to our Scherk example: While the z-projection embeds S_T into the square Q, the $\mathbf{k}$-Hopf projection of $\widetilde{M^+}, \widetilde{N^+}$ is only an immersion to $\mathbb{S}^2$ (by the Transversality Lemma 5.1). In order to construct an embedded minimal lift of $\widetilde{M^+}, \widetilde{N^+}$ in the bundle, we need to pass to the appropriate 'covering' of our base manifold $\mathbb{S}^2$; to be precise, we take the base disk as the Hopf projection of $\widetilde{M^+}$ in the sense of an immersion. The observed difference could be interpreted as to say that the cousins also have 'horizontal' self-intersections in addition to the 'vertical' self-intersections already encountered with the quotient Scherk surface $S_T \subset T$.

Let us now give a geometric description of the base disk of the bundle, which is defined to be the Hopf projection of $\widetilde{M^+}$; it turns out to agree with the Hopf projection of $\widetilde{N^+}$. Let us recall the notion of a ray of slit spheres from Lemma 5.2 and Figure 10. A slit sphere represents the Hopf projection of the cousin of an unduloid end. In fact, it also represents the Hopf projection of a triunduloid end, cut at an appropriate position (roughly at a neck). Let us now join the vertices of the triple $\Psi(M)$ by a minimizing triangle in $\mathbb{S}^2$, as indicated in Figure 13. We can glue three rays of spheres, with the appropriate slit lengths, across the three minimizing geodesics. This defines an immersed disk in $\mathbb{S}^2$ with infinite degree. It happens to be the isometric description of $\Pi_{\mathbf{k}}(\widetilde{M^+})$: The central triangle is the Hopf projection of the central triple junction in the triunduloid, and each bubble on a triunduloid end contributes a slit sphere in the appropriate ray of spheres.

Now we can formally define the bundle. We consider an $\mathbb{S}^1$-bundle over the immersed disk $\Pi_{\mathbf{k}}(\widetilde{M^+})$. We endow it with a Riemannian metric locally isometric to $\mathbb{S}^3$, by pulling back the metric from the $\mathbf{k}$-Hopf bundle over $\widetilde{M^+} \subset \mathbb{S}^3$. This

way $\widetilde{M^+}$ becomes a section of a Riemannian circle bundle. Since the base of this bundle is a disk the fiberwise covering of this bundle gives a Riemannian line bundle E. Fiberwise translation, induced by ϕ_t defines a group of isometries of E (E is principal). The maximum principle is applied in E as follows:

- We foliate E with translates of $\widetilde{M^+}$. The surface $\widetilde{N^+}$ lifts to a section of E.
- Due to the asymptotics result, Theorem 6.1(ii), the section $\widetilde{N^+}$ is tangent to a leaf of the foliation over each end.
- Now we use the period condition: In $\mathbb{R}^3$, the three bounding curves of the CMC surface M^+ have the same height; likewise for N^+. We can conclude that the section $\widetilde{N^+}$ is tangent to the same leaf of the foliation at all three ends.
- From the maximum principle we conclude that $\widetilde{N^+}$ is identical to a leaf of the foliation.

This proves $\widetilde{M^+} = \widetilde{N^+}$ and therefore $M = N$ as desired.

6.3.3. *Ψ is continuous and proper.* The remaining two steps can be considered a continuity method. Using injectivity, we can identify $\Psi(\mathcal{M})$ with a subset of $\mathcal{T}$. To show that $\Psi(\mathcal{M})$ is the entire connected space $\mathcal{T}$, it would be sufficient to show $\Psi(\mathcal{M})$ is non-empty and $\Psi(\mathcal{M})$ is both closed and open.

As $\mathcal{M}$ is non-compact, instead of closedness of $\Psi(\mathcal{M})$ we show properness of the map Ψ. To show that any sequence in $\Psi^{-1}(\mathcal{K})$ has a convergent subsequence, we need estimates, uniform on any compact subset $\mathcal{K} \subset\subset \mathcal{T}$. Observe that triples in $\mathcal{K}$ attain a minimal pairwise distance. Hence the three necksizes of $\Psi^{-1}(\mathcal{K})$ are bounded away from zero, $n_1, n_2, n_3 > \epsilon$ for some $\epsilon > 0$. Using an extension of the curvature estimates of Korevaar and Kusner [**KK**] to the case of Alexandrov embedded surfaces, we establish a uniform curvature bound on $\Psi^{-1}(\mathcal{K})$. Moreover, the same paper gives uniform area bounds: The amount of area enclosed in an ambient ball is uniformly bounded, independently of the choice of $\mathcal{K}$. This establishes properness, and also the continuity of Ψ.

6.3.4. *Ψ is a homeomorphism.* Again, we illustrate the main idea only. If $\mathcal{M}$ were a manifold, $\Psi(\mathcal{M})$ would certainly be open due to the injectivity of Ψ. Though our main result in fact establishes that $\mathcal{M}$ is a topological manifold, this does not follow from general theory.

However, a result of Kusner, Mazzeo, Pollack [**KMP**] gives for our setting:
(i) $\mathcal{M}$ locally is a real analytic variety of finite dimension,
(ii) $\mathcal{M}$ is a manifold of dimension 3 in a neighborhood of those $M \in \mathcal{M}$ which do not admit L^2-Jacobi fields.

From a result of Montiel and Ros in conjunction with work of Mazzeo and Pacard [**MP**], we establish the existence of a triunduloid $M \in \mathcal{M}$ for which (ii) holds. Therefore, the dimension of $\mathcal{M}$ (which is, the highest dimension of a stratum in $\mathcal{M}$), is at least 3. On the other hand, since Ψ maps $\mathcal{M}$ injectively and continuously to the 3-manifold $\mathcal{T}$ the dimension cannot be any larger than 3. We conclude it must equal 3.

There is a structure theory for real analytic manifolds, going back to D. Sullivan, Burghelea/Verona, and Hardt. We use it to derive the following general statement which proves Theorem 6.2 when applied to $\Psi : \mathcal{M} \to \mathcal{T}$:

PROPOSITION 6.4. *Let f be a continuous, proper, and injective map from a real analytic variety of dimension d to a connected d-manifold. Then f is surjective and a homeomorphism.*

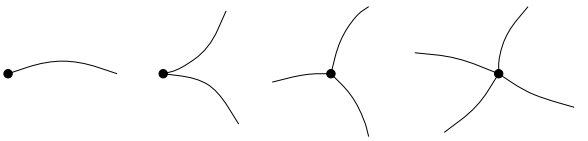


FIGURE 14. The first and third picture represent impossible local structures of real analytic varieties of dimension 1.

To explain this result, it is instructive to consider the simplest case of dimension $d = 1$. Then structure theory says that a real-analytic 1-manifold consists of smooth open arcs and singular points with the following property: If an arc has a boundary point, then it must be a singular point, and each singular point is the common boundary point of an even number m of curves (see Figure 14). In addition, $m = 4, 6, 8, \ldots$ is impossible due to the existence of an injective map to a 1-manifold. Thus all singular points have $m = 2$, which means that topologically, the singularity can be removed. But a continuous, proper and injective map between manifolds of the same dimension is a homeomorphism.

References

[A1] A. D. Alexandrov, Uniqueness theorems for surfaces in the large, V, Vestnik Leningrad Univ. 19:13 (1958), 5–8; Amer. Math. Soc. Transl. (Ser. 2) 21 (1962), 412–416.

[A2] ———, A characteristic property of spheres, Ann. Mat. Pura Appl. 58 (1962), 303–315.

[D] C. Delaunay, Sur la surface de révolution, dont la courbure moyenne est constante, Journal de mathématiques 6 (1841), 309–320.

[G] K. Große-Brauckmann, New surfaces of constant mean curvature, Math. Zeit. 214 (1993), 527–565.

[GKS1] K. Große-Brauckmann, R. Kusner and J. M. Sullivan, Classification of embedded constant mean curvature surfaces with genus zero and three ends, preprint, Bonn SFB 256, 1997.

[GKS2] ———, Constant mean curvature surfaces with three ends, Proc. Natl. Acad. Sci. USA 97:26 (2000), 14067–14068; math.DG/9903101.

[KK] N. Korevaar and R. Kusner, The global structure of constant mean curvature surfaces, Invent. Math. 114 (1993), 311–332.

[KKS] N. Korevaar, R. Kusner and B. Solomon, The structure of complete embedded surfaces with constant mean curvature, J. Diff. Geom. 30 (1989), 465–503.

[KMP] R. Kusner, R. Mazzeo and D. Pollack, The moduli space of complete embedded constant mean curvature surfaces, Geom. Funct. Anal. 6 (1996), 120–137.

[KMS] M. Kilian, I. McIntosh and N. Schmitt, New constant mean curvature surfaces, Experimental Math. 9:4 (2000), 595–611.

[Kp] N. Kapouleas, Complete constant mean curvature surfaces in Euclidean three-space, Annals of Math. 131 (1990), 239–330.

[Kr1] H. Karcher, The triply periodic minimal surfaces of A. Schoen and their constant mean curvature companions, Manus. Math. 64 (1989), 291–357.

[Kr2] ———, Introduction to conjugate Plateau constructions, manuscript, October 2001.

[Law] H. B. Lawson, Complete minimal surfaces in $\mathbb{S}^3$, Annals of Math. 92 (1970), 335–374.

[MP] R. Mazzeo and F. Pacard, Constant mean curvature surfaces with Delaunay ends, preprint, 1998.

[R] J. Ratzkin, An end-to-end gluing construction for surfaces of constant mean curvature, PhD thesis, University of Washington, 2001.

[S] M. Spivak, A comprehensive introduction to differential geometry, Publish or Perish, 1979.

[Sh] R. W. Sharpe, Differential geometry, Springer, 1997.
[UY] M. Umehara and K. Yamada, Metrics of constant curvature 1 with three conical singularities on the 2-sphere, Illinois J. Math. 44 (2000), 72–94.

Universität Bonn, Mathematisches Institut, Beringstr. 1, 53115 Bonn, Germany
E-mail address: kgb@math.uni-bonn.de

Clay Mathematics Proceedings
Volume 2, 2004

An Approach to the Willmore Conjecture

Peter Topping

Abstract. We highlight one of the methods developed in [**6**] to give partial answers to the Willmore conjecture, and discuss how it might be used to prove the complete conjecture.

The Willmore conjecture, dating from around 1965, asserts that the integral of the square of the mean curvature H over a torus immersed in $\mathbb{R}^3$ (with the induced metric) should always be at least $2\pi^2$:

$$W := \int_{T^2} H^2 \geq 2\pi^2.$$

A weaker lower bound of 4π for this 'Willmore energy' is rather easy to obtain (see [**6**]) but this bound is known not to be sharp (see [**5**]) unless we change the problem to allow integrals over spheres instead of tori.

It is well known that this problem has an equivalent formulation in S^3. Indeed, if we transport the torus from $\mathbb{R}^3$ to S^3 via inverse stereographic projection, the Willmore energy may be written in terms of the new induced metric on the torus, and the new mean curvature, as

$$W = \int_{T^2} (1 + H^2).$$

This energy is preserved if we move the torus via conformal transformations of the ambient S^3 (see [**6**]).

Once formulated within S^3, we see that the conjecture asserts that the Clifford torus, which has area $2\pi^2$ and satisfies $H = 0$, should be a minimiser for W. Of course, the Clifford torus is the case $r = \frac{1}{\sqrt{2}}$ of the family of flat tori

$$\mathbb{T}_r := \{(z_1, z_2) \in \mathbb{C}^2 \ : \ |z_1| = r, \ |z_2| = \sqrt{1 - r^2}\} \hookrightarrow S^3 \hookrightarrow \mathbb{C}^2$$

which, as r varies within $(0, 1)$, foliate S^3 minus the great circles $\mathbb{T}_0$ and $\mathbb{T}_1$.

Given a general torus $\mathcal{M} \hookrightarrow S^3$, we can use the degenerate foliation $\{\mathbb{T}_r\}$ to define a 'tangency map'

$$\Phi : G_{2,4} \to \mathbb{N} \cup \{0, \infty\}$$

from the Grassmannian of oriented 2-planes in $\mathbb{R}^4 \hookleftarrow S^3$. More precisely, given a plane $p \in G_{2,4}$, we may rotate the degenerate foliation $\{\mathbb{T}_r\}$ within S^3 so that the great circle $\mathbb{T}_0$ lies within the plane p, and then define $\Phi(p)$ to count the number of points of tangency between this rotated foliation and the original torus $\mathcal{M}$.

©2004 Clay Mathematics Institute

It is possible to compute (see the proofs in [**6**]) that in the case that $\mathcal{M}$ is the Clifford torus, we have $\Phi(p) \equiv 8$ for almost all planes $p \in G_{2,4}$. When $\mathcal{M}$ is a general torus, a Morse theory argument, explained in [**6**], provides the lower bound $\Phi \geq 4$ almost everywhere. This information about Φ should be compared to the main result we wish to discuss in this article (taken from [**6**, Theorem 4]):

THEOREM 1. *Given an immersed torus in S^3, its Willmore energy is bound by*

$$W \geq \frac{\pi^2}{4} \fint_{G_{2,4}} \Phi.$$

We see that the general Morse theoretic bound $\Phi \geq 4$ a.e. coupled with Theorem 1 is only enough to establish that $W \geq \pi^2$ which is already implied by the trivial bound $W \geq 4\pi$ mentioned earlier. However, if our torus $\mathcal{M}$ is invariant under the antipodal map in S^3, then we showed in [**6**] that by performing Morse theory on the quotient of $\mathcal{M}$ under the antipodal map, and then lifting back to the covering $\mathcal{M}$, we have the stronger bound $\Phi \geq 8$, and hence

COROLLARY 1. (Taken from [**6**, Corollary 5]. Proved with different methods by Ros [**4**].)
For any torus $\mathcal{M}$ immersed in S^3 which is invariant under the antipodal map, the Willmore conjecture $W \geq 2\pi^2$ holds.

Combining with a result of Kitagawa [**2**] which states that any flat torus in S^3 is invariant under the antipodal map, we recover the following well-known result of Chen [**1**] (which has extensions given in [**3**] and [**6**]).

COROLLARY 2. *For any flat torus $\mathcal{M}$ immersed in S^3, the Willmore conjecture $W \geq 2\pi^2$ holds.*

We would also like to give another corollary of Theorem 1 which is much more general than the two given above. Before doing so, we make the following observation of the consequences of antipodal symmetry.

PROPOSITION 1. *If $\mathcal{M}$ is a torus immersed in S^3 which is invariant under the antipodal map, then:*
(P1) *Given any great circle C in S^3 disjoint from $\mathcal{M}$, the distance function $\mathcal{M} \to \mathbb{R}$ given by $x \to dist_{S^3}(x, C)$ has at least two local minima.*
(P2) *Given any great circle C intersecting $\mathcal{M}$, if all of the intersections are transverse, then there must be at least four of them.*

PROOF. To establish property P1, take any point $x \in \mathcal{M} \hookrightarrow S^3 \hookrightarrow \mathbb{R}^4$ where the given distance function is minimised. Then $-x$ is also a minimum.

Property P2 follows from the fact that $\mathcal{M}$ separates S^3 into two components Σ_1 and Σ_2, and the fact, proved by Ros in [**4**], that the antipodal map *preserves* each of these components, rather than switches them, since $\mathcal{M}$ has odd genus. (Note, of course, that if $x \in \mathcal{M} \cap C$ then $-x \in \mathcal{M} \cap C$.) □

This proposition shows that the following corollary of Theorem 1 is a strict generalisation of the other two.

COROLLARY 3. *For any torus $\mathcal{M}$ immersed in S^3 which satisfies both properties P1 and P2, the Willmore conjecture $W \geq 2\pi^2$ holds.*

PROOF. The proof of this final corollary follows along the same lines as the proof of the Morse theoretic bound $\Phi \geq 4$ mentioned above (see [**6**]). By Theorem 1, all we must show is that $\Phi(p) \geq 8$ for almost all planes $p \in G_{2,4}$. Given a plane $p \in G_{2,4}$ spanned by orthonormal vectors e_1 and e_2 in $\mathbb{R}^4$, consider the function $f : \mathcal{M} \to [0,1]$ defined by

$$f(x) = \langle x, e_1 \rangle^2 + \langle x, e_2 \rangle^2,$$

and representing the square of the length of x once projected in $\mathbb{R}^4$ onto the plane p. We propose to apply Morse theory to f, for p within the set of full measure for which f is a Morse function *and* for which any intersection between either great circle $p \cap S^3$ or $p^\perp \cap S^3$ and $\mathcal{M}$ occurs transversely.

The critical points of f with critical value within $(0,1)$ (that is, neither 0 nor 1) are then the points of tangency counted by $\Phi(p)$. Therefore, all we must do is to use properties P1 and P2 to establish the existence of at least eight of these.

After reflecting on the similarity between f and the distance function appearing in property P1, we make the following observations.

First, if $p^\perp \cap \mathcal{M} = \varnothing$, we may apply P1 with $C = p^\perp \cap S^3$ to find that f has at least two local minima with critical value greater than 0 (that is, both of these critical points are counted by $\Phi(p)$).

Second, if $p \cap \mathcal{M} = \varnothing$, we may apply P1 with $C = p \cap S^3$ to find that f has at least two local maxima with critical value less than 1 (that is, both of these critical points are counted by $\Phi(p)$).

Third, if $p^\perp \cap \mathcal{M} \neq \varnothing$, we may apply P2 with $C = p^\perp \cap S^3$ to find that f has at least four local minima with a critical value of zero (that is, none of these critical points are counted by $\Phi(p)$).

Fourth, if $p \cap \mathcal{M} \neq \varnothing$, we may apply P2 with $C = p \cap S^3$ to find that f has at least four local maxima with a critical value of 1 (that is, none of these critical points are counted by $\Phi(p)$).

Finally, since Morse theory implies that the number of saddle points of f is at least the total number of maxima and minima (since $\mathcal{M}$ is a torus! — see [**6**, Section 4.3]) the observations above provide us with at least eight critical points with critical value within $(0,1)$. □

With a view to proving the entire Willmore conjecture, we are then led to ask:

QUESTION 1. Can every torus $\mathcal{M} \hookrightarrow S^3$ be moved by a conformal transformation of S^3 so that properties P1 and P2 hold?

STRATEGY OF THE PROOF OF THEOREM 1. A general principle of which several incarnations may be found in [**6**] is that the Willmore energy is related to the harmonic map energy of various forms of the Gauss map. In this proof, we consider the usual Gauss map $\mathcal{G} : \mathcal{M} \to G_{2,4}$ which assigns, to each point on the torus $\mathcal{M} \hookrightarrow S^3 \hookrightarrow \mathbb{R}^4$, the oriented tangent plane to the surface. It is well known, by representing elements of $G_{2,4}$ by unit simple 2-vectors in $\bigwedge^2(\mathbb{R}^4)$ and projecting onto components of equal length in the self-dual and antiself-dual subspaces $\bigwedge^2_+(\mathbb{R}^4)$ and $\bigwedge^2_-(\mathbb{R}^4)$, that $G_{2,4}$ may be identified with $S^2 \times S^2$ (see [**6**]). This enables us to see the Gauss map as a map $\sigma : \mathcal{M} \to S^2 \times S^2$, and decompose into components $\sigma = (a, b)$ with $a : \mathcal{M} \to S^2$ and $b : \mathcal{M} \to S^2$, and our strategy is to perform an integral geometric construction on the target $S^2 \times S^2$.

In order to do this, we assign to each point $(r,s) \in S^2 \times S^2$, the associated torus of points in $S^2 \times S^2$ defined by

$$T_{r,s} := \{(k,l) \in S^2 \times S^2 \hookrightarrow \mathbb{R}^3 \times \mathbb{R}^3 \mid \langle r,k\rangle = \langle s,l\rangle = 0\}.$$

For example, when r and s are the north poles in their respective spheres, $T_{r,s}$ is the product of the equators.

Our construction is then to assign to each point $x \in \mathcal{M}$ the torus $T_{a(x),b(x)} \hookrightarrow S^2 \times S^2$. As we move x around the two-dimensional surface $\mathcal{M}$, the two-dimensional torus $T_{a(x),b(x)}$ then sweeps out *four*-dimensional volume in $S^2 \times S^2$. We are then able to relate this total swept volume V both to the Willmore energy of $\mathcal{M}$, and the average of the tangency map Φ.

Indeed, local considerations suggest that if $\mathcal{M}$ has large curvature in some region, then a and b will move substantially within that region, and thus so will $T_{a(x),b(x)}$, sweeping out a large volume in $S^2 \times S^2$. A calculation along these lines (see [**6**]) tells us that

$$W \geq \tfrac{1}{64} V. \tag{0–1}$$

Meanwhile we can count the number of times the sweeping tori $T_{a(x),b(x)}$ hit an arbitrary point $(k,l) \in S^2 \times S^2$. First we observe the pretty duality, typical of integral geometry, that $(k,l) \in T_{a,b}$ is exactly equivalent to $(a,b) \in T_{k,l}$. A calculation (see [**6**]) then shows that the Gauss map condition $\sigma(x) = (a(x), b(x)) \in T_{k,l}$ is precisely equivalent to the condition that the surface $\mathcal{M}$ at x is tangent to a torus in the rotated family $\{\mathbb{T}_r\}$ for which $\mathbb{T}_0$ lies in the plane corresponding to (k,l) under the identification $G_{2,4} \simeq S^2 \times S^2$. By definition of Φ, we then have that (k,l) is hit by $T_{a(x),b(x)}$ a total of $\Phi(k,l)$ times, and hence that

$$V = \int_{S^2 \times S^2} \Phi = 16\pi^2 \fint_{G_{2,4}} \Phi. \tag{0–2}$$

Combining (0–1) and (0–2) we then have the desired result. □

References

[1] B. Y. Chen, *On the total curvature of immersed manifolds V. C-surfaces in Euclidean m-space*, Bull. Inst. Math., Acad. Sin. **9** (1981), 509–516.

[2] Y. Kitagawa, *Embedded flat tori in the unit 3-sphere*, J. Math. Soc. Japan **47** (1995), 275–296.

[3] P. Li and S.-T. Yau, *A new conformal invariant and its applications to the Willmore conjecture and the first eigenvalue of compact surfaces*, Invent. Math. **69** (1982), 269–291.

[4] A. Ros, *The Willmore conjecture in the real projective space*, Math. Res. Lett. **6** (1999), 487–493.

[5] L. M. Simon, *Existence of surfaces minimizing the Willmore functional*, Comm. Anal. Geom. **1**:2 (1993), 281–326.

[6] P. M. Topping, *Towards the Willmore conjecture*, Calc. Var. **11** (2000), 361–393.

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom

E-mail address: see `www.maths.warwick.ac.uk/~topping/`

Clay Mathematics Proceedings
Volume 2, 2004

Minimal Surfaces and Harmonic Maps into Singular Geometry

Chikako Mese

Abstract. The purpose of this paper is, first, to summarize our investigation of harmonic maps into singular spaces; in particular, we consider minimal surface theory as an extension of the harmonic map theory in this general setting. Second, as an application of these generalized harmonic maps and minimal surfaces, we will outline a completion of the Gerstenhaber–Rauch program of constructing the Teichmüller map between two compact Riemann surfaces. The use of the variational characterization of the Teichmüller map, suggested by Gerstenhaber and Rauch in the 1940s, has attracted the attention of many researchers.

The study of harmonic maps into singular targets have been a subject of far-reaching research. As a generalization of the classical harmonic map theory, this investigation has explicitly revealed the role that curvature plays in the analysis of solutions to geometric variational problems. From the classic paper of Eells and Sampson [**ES**], we know that the key geometric ingredient for harmonic map theory is the curvature of the target space. Recent developments extend much of the existing harmonic map theory to singular target spaces for which the curvature is bounded from above in the sense of the triangle comparison hypothesis. Many remarkable results have been obtained by Gromov and Schoen [**GS**], Korevaar and Schoen [**KS1**] [**KS2**], Jost [**Jo1**] [**Jo2**] and others. There are also many reasons to study this problem beside its own intrinsic interest as harmonic map theory has applications in other fields. For example, [**GS**] provides a new approach in the study of p-adic representations of lattices in noncompact semisimple Lie groups. It has also been used by Hardt and Lin [**HL**] to study nematic liquid crystal droplets. Wolf [**Wo1**] [**Wo2**] [**Wo3**], Fischer and Tromba [**FT**], and Yamada [**Ya**] and others have used it to understand Teichmüller spaces.

This paper is divided into two parts. Section 1 sketches the harmonic map and minimal surface theories in metric spaces of curvature bounded from above. Section 2 outlines the construction of the Teichmüller map between compact Riemann surfaces using these theories.

1. Harmonic Maps and Minimal Surfaces

1.1. Energy of maps into metric space (X, d). We first summarize the notion of the Korevaar–Schoen [**KS1**] energy for a map into a metric space. Let

©2004 Clay Mathematics Institute

$f : \Omega \to X$ be a map from an m-dimensional compact Riemannian domain (Ω, h) into space X. If X is an n-dimensional Riemannian manifold with metric g, then the energy ${}^g E^f$ of the map f with respect to the metric g is

$$ {}^g E^f = \int_\Omega |\nabla f|^2 d\mu $$

where

$$ |\nabla f|^2(x) = \sum_{i,j=1}^{n} \sum_{\alpha,\beta=1}^{m} g_{ij}(f(x)) h^{\alpha\beta}(x) \frac{\partial f^i}{\partial x^\alpha} \frac{\partial f^j}{\partial x^\beta} $$

is the energy density function of the map f. Physically, $|\nabla f|^2(x)$ measures (a constant multiple of) the average stretch of an infinitesimal sphere about x by the map f. To define energy when X is only equipped with a distance function, we first consider (a constant multiple of) the average stretch of $S_{\epsilon,x}$, a sphere of radius ϵ centered at x. This is the ϵ-approximate energy density function,

$$ e^f_\epsilon(x) = \int_{S_{\epsilon,x}} \frac{d^2(f(x), f(\xi))}{\epsilon^2} \frac{d\sigma_\epsilon(\xi)}{\epsilon^{n-1}}, $$

where $d\sigma_\epsilon$ is the normalized measure on $S_{\epsilon,x}$.

Let Ω_ϵ be the set of points in Ω with distance from the boundary greater than ϵ. The function e^f_ϵ defines a linear functional $E^f_\epsilon : C^\infty_c(\Omega_\epsilon) \to \mathbb{R}$ by

$$ E^f_\epsilon(\varphi) = \int_{\Omega_\epsilon} \varphi e^f_\epsilon d\mu, \qquad \varphi \in C^\infty_c(\Omega_\epsilon). $$

The energy of the map f is then defined by

$$ E^f = \sup_{\varphi \in C^\infty_c,\ 0 \le \varphi \le 1} \limsup_{\epsilon \to 0} E^f_\epsilon(\varphi). $$

It turns out that $e^f_\epsilon(x) d\mu$ converge as measures to a measure that is absolutely continuous with respect to the Lebesgue measure, which we write $|\nabla f|^2(x) d\mu$. The function $|\nabla f|^2(x)$ is the *generalized* energy density function and this notion is consistent with the usual definition when X is a Riemannian manifold. Similarly, $|f_*(Z)|^2(x)$ makes sense for a tangent vector field $Z \in \Gamma(T\bar{\Omega})$, and is consistent with the norm squared of the directional derivative when X has a smooth Riemannian metric. In order to generalize various results from classical harmonic map theory, we need an upper curvature bound on our metric space X.

1.2. Curvature of metric spaces. We assume that our metric spaces are length spaces, i.e., for each P, $Q \in X$, there exists a curve γ_{PQ} such that the length of γ_{PQ} is exactly $d(P, Q)$ (which we will sometimes write as d_{PQ}). We call γ_{PQ} a geodesic between P and Q. We say that a space X has curvature bounded from above by κ if small geodesic triangles in X are thinner than a comparison triangle in S_κ, a surface of constant curvature κ. More precisely, (X, d) is an NPC (non-positively curved) space if every point of X is contained in a neighborhood U so that for every $P, Q, R \in U$,

$$ d^2_{PQ_\tau} \le (1-\tau) d^2_{PQ} + \tau d^2_{PR} - \tau(1-\tau) d^2_{QR}, \tag{1–1} $$

where Q_τ is the point on γ_{QR} so that $d_{QQ_\tau} = \tau d_{QR}$. Note that equality is achieved for every triplet $P, Q, R \in \mathbb{R}^2$. Similarly, (X, d) is said to have curvature bounded

from above by κ if

$$\cosh d_{PQ_\tau} \leq \frac{\sinh(1-\tau)\kappa d_{QR}}{\sinh \kappa d_{QR}} \cosh d_{PQ} + \frac{\sinh \tau\kappa d_{QR}}{\sinh \kappa d_{QR}} \cosh d_{PR} \tag{1–2}$$

for $\kappa < 0$ and

$$\cos d_{PQ_\tau} \geq \frac{\sin(1-\tau)\kappa d_{QR}}{\sin \kappa d_{QR}} \cos d_{PQ} + \frac{\sin \tau\kappa d_{QR}}{\sin \kappa d_{QR}} \cos d_{PR} \tag{1–3}$$

for $\kappa > 0$. Note that if (X, d) is S_κ, then the equality is achieved for (1–2) when $\kappa > 0$ and for (1–3) when $\kappa < 0$ for every triplet $P, Q, R \in S_\kappa$

Basic results in classical harmonic map theory, such as the existence of solutions for the Dirichlet problem and homotopy problem as well as the regularity of these solutions, hold true when we impose this curvature bound on the target metric space (see [**KS1**], [**Se1**], [**Se2**]). We call f a harmonic map if it is locally energy minimizing.

For any finite energy map $f : \Omega \to (X, d)$, we can generalize the notion of the pull-back metric if (X, d) has an upper curvature bound; that is, there exists an inner product structure

$$\pi : \Gamma(T\bar{\Omega}) \times \Gamma(T\bar{\Omega}) \to L^1(\bar{\Omega})$$

defined by

$$\pi(Z, W) = \frac{1}{4}|f_*(Z + W)|^2 - \frac{1}{4}|f_*(Z - W)|^2.$$

Again, this definition is consistent with the usual pull-back metric when X is a Riemannian manifold.

Recall that harmonic maps are crucial in the theory of parameterized minimal surfaces. For example, harmonic maps are the main ingredient of the Douglas solution to the Plateau Problem. We have enough tools of harmonic map theory to study (parameterized) minimal surfaces in the more general setting of metric spaces.

1.3. Minimal surfaces. Consider a map $u : D \to X$ from a disk D in the complex plane. With the use of the generalized pull-back metric, it is natural to define the area functional as

$$\mathcal{A}(u) = \int_D \sqrt{\det(\pi_{ij})}\, dx,$$

where ∂_1 and ∂_2 are the standard coordinate basis vectors of D and $\pi_{ij} = \pi(\partial_i, \partial_j)$. With this, we can formulate:

The Plateau Problem. *Let Γ be a closed Jordan curve in X and let*

$$C_\Gamma = \{u \in W^{1,2}(D, X) : u|_{\partial D} \text{ parameterizes } \Gamma \text{ monotonically}\}.$$

There exists $u \in C_\Gamma$ so that $\mathcal{A}(u) = \inf\{\mathcal{A}(u) : v \in C_\Gamma\}$. Moreover, u is weakly conformal, i.e., $\pi_{11} = \pi_{22}$ and $\pi_{12} = 0$, and Lipschitz in the interior of D and Hölder continuous in $\bar{D}$.

For a locally compact metric space of curvature bounded from above or for a non-positively curved metric space without an assumption of local compactness, this can be solved by a straightforward extension of Morrey's [**Mo**] solution of the Plateau Problem in Riemannian manifolds.

Let Σ be a Riemann surface with local coordinate $w = u + iv$. From the above, it is natural to define $u : \Sigma \to X$ to be a (parameterized) minimal surface if u is a conformal, harmonic map. Thus, if u is a minimal surface, we have an induced metric on Σ which we can write $\lambda|dw|^2$ with $\lambda = \pi_{11}$. We will call λ the conformal factor of u. Note that we cannot hope that the metric $\lambda|dw|^2$ is smooth or non-degenerate. On the other hand, λ satisfies the following weak differential inequality:

THEOREM 1 (Mese [**Me1**]). *Let (X, d) be a metric space of curvature bounded from above and $u : D \to X$ be a minimal surface. The functions λ and $\log \lambda$ are locally integrable in D and $\frac{1}{2}\triangle \log \lambda \geq -\kappa\lambda$ weakly. In other words,*

$$\int_D (\triangle\varphi) \log \lambda \, dx \geq -2\kappa \int_D \varphi\lambda \, dx$$

for any $\varphi \in C_c^\infty(D)$.

The above analytical result can be interpreted in the following geometric way: Recall that in a smooth Riemannian manifold (M, g), the Euler–Lagrange equation for the area functional is $b_{11} + b_{22} = 0$ where b_{ij}'s are the components of the vector valued second fundamental form. The Gauss equation then gives,

$$K_\Sigma = K_M + b_{11}b_{22} - b_{12}^2 = K_M - \frac{1}{2}\sum_{i,j=1}^{2} b_{ij}^2$$

where K_Σ is the Gaussian curvature of the surface and K_M is the sectional curvature of the tangent plane to the surface in the manifold. This shows that the curvature of a minimal surface is less than or equal to that of the ambient space. Furthermore, when a smooth surface Σ has a conformal metric with conformal factor λ, it is well known that the Gaussian curvature K_Σ is given by the formula,

$$K_\Sigma = -\frac{1}{2\lambda}\triangle \log \lambda.$$

Hence, our result says that the curvature inequality that one obtains for minimal surfaces in a smooth Riemannian manifold also holds for minimal surfaces in our general setting. In fact, this geometric interpretation of the analytical result makes sense in that there is a distance function d_λ induced by the metric $\lambda\,|dw|^2$ and:

THEOREM 2 (Mese [**Me1**]). *The metric space (D, d_λ) defines an Alexandrov space of curvature bounded from above by κ.*

This gives a pleasant result that the curvature of the minimal surface is bounded in the same sense (i.e., curvature bound is given by way of the triangle comparison) as that of the ambient space.

The inequality of Theorem 1, used in conjunction with the theory of subharmonic functions, yields several nice results. One example is the following:

COROLLARY 2.1 (Mese [**Me1**]). *The branch points, i.e., the set of zeroes of the function λ, are of Hausdorff dimension 0.*

The theory of branch points in Euclidean space and Riemannian manifolds has a rich history. The fact that the branch points are isolated in the Euclidean case and the Riemannian manifold case were proven by Osserman and Gulliver, respectively.

Let Σ be a Riemann surface with boundary $\partial\Sigma$. When (X, d) is a non-positively curved metric space, we have found that a minimal surface $u : \Sigma \to X$ has many of the same properties as classical minimal surfaces.

THEOREM 3 (Mese [**Me2**]). *The boundary map $u|_{\partial\Sigma}$ is a monotone parameterization of $u(\partial\Sigma)$. The set $u(\Sigma)$ is contained in $\mathcal{C}(u(\partial\Sigma))$, the smallest geodesically convex set containing $u(\partial\Sigma)$. The image of $u(\Sigma)$ satisfies an isoperimetric inequality; that is there exists C so that $L^2 \geq CA$ where L is the length of $u(\partial\Sigma)$ and A is the area of $u(\Sigma)$.*

These results lead us to believe that our definition is a very natural way to define minimal surfaces in this setting.

As an application of these results for minimal surfaces, we are able to obtain an isothermal coordinate system for a surface with a distance function. This will be important in Section 2.

1.4. Isothermal coordinates. Theorem 1 leads us to the existence of an isothermal coordinate system for a surface (S, d) of curvature bounded from above. In the case S has a smooth Riemannian metric, the standard approach in obtaining isothermal coordinates is to solve a system of PDE, namely the Beltrami equations. Here, the lack of smoothness prohibits us from using this approach and we apply a variational approach instead. For a simple curve $\Gamma \subset S$, let $u : D \to S$ be the Plateau solution with $u(\partial D) = \Gamma$. Using the fact the set of zeroes of the pull-back metric is a very small set (Corollary 2.1), we are able to show that u is actually a homeomorphism. We obtain a new proof of the following theorem of Reshetnyak's:

THEOREM 4 (Reshetnyak [**Re**], Mese [**Me3**]). *Let S be a surface with a distance function d which makes (S, d) into a metric space of curvature bounded from above. Assume that the metric topology is equivalent to the surface topology. Then for every $P \in S$, there is a neighborhood U of P and a conformal homeomorphism $u : D \to U$ from the unit disk.*

In [**Me3**], Theorem 4 is only proved when (S, d) is non-positively curved, but by using results of [**Me4**], it is not hard to extend the proof to cover the case when (S, d) has curvature bounded from above by κ. Theorem 4 is a special case of a more general theorem:

THEOREM 5 (Mese [**Me3**]). *Let (X, d) be a locally compact metric space of non-positive curvature and Hausdorff dimension 2. Let $P \in X$. Assume that (1) there exists $\delta > 0$ so that every geodesic emanating from P can be extended to length greater than δ and (2) there exists a simple closed curve Γ which is non-trivial in $X - \{P\}$. Then there exists a conformal homeomorphism u from the unit disk D into X so that $P \in u(D)$.*

2. The Existence of Teichmüller Maps: A Variational Approach

Gerstenhaber and Rauch [**GR**] proposed the problem of constructing the Teichmüller map by a maximum-minimum approach involving harmonic maps. This idea has been studied by many researchers including Reich [**R**], Reich–Strebel [**RS**], Miyahara [**Mi**], Leite [**Le**], and Kuwert [**Ku**]. This section summarizes the completion of the Gerstenhaber–Rauch program for maps between compact Riemann surfaces. For more details, see [**Me6**]. The key idea is to consider a class of metrics

on the target which include singular metrics and use the harmonic map theory in this setting.

Let Σ_1 and Σ_2 be compact Riemann surfaces of the same genus ≥ 2 and $f : \Sigma_1 \to \Sigma_2$ be an orientation preserving homeomorphism. For a sufficiently smooth f, we measure the deviation of the map from conformality at each point $z \in \Sigma_1$ by the dilatation $K_f(z)$, defined by the ratio of the axes of the infinitesimal ellipse into which f takes an infinitesimal circle around z. (See [**Ah1**] for more details.) Let $K[f] = \sup_{z\in\Sigma_1} K_f(z)$.

Given an orientation preserving homeomorphism $h : \Sigma_1 \to \Sigma_2$, let K^* be the infimum of $K[f]$ amongst all quasiconformal maps homotopic to h. Teichmüller's Theorem asserts the existence of a unique map f_0 with the property that $K_{f_0}(z) \equiv K^*$ everywhere except at isolated points. The extremal map f_0 can be described analytically by two holomorphic quadratic differentials Φ and Ψ defined on Σ_1 and Σ_2 respectively; f_0 can be written $u = Kx$ and $v = y$ where $z = x + iy$ and $w = u + iv$ are local parameters so that $\Phi = dz^2$ and $\Psi = dw^2$. The Teichmüller distance between Σ_1 and Σ_2, relative to the homotopy class of h, is defined by $\log K^*$. This distance function makes Teichmüller space (an equivalence class of conformal structures on a compact surface where two conformal structures are considered to be equivalent if there exists a conformal diffeomorphism between them which is homotopic to the identity map) into a metric space. The Riemann–Roch theorem and the fundamental relation between the Teichmüller space and the space of holomorphic quadratic differentials show that the Teichmüller space is homeomorphic to the Euclidean space of dimension $6g-6$. The topological structure of the Teichmüller space was already known [**FK**] since the early 20th century, but Teichmüller realized its connection with holomorphic quadratic differentials in the 1940s. Complete details of Teichmüller's claims were worked out by L. V. Ahlfors [**Ah2**] and L. Bers [**Be**] in the 1950s and 60s.

In their 1964 paper [**GR**], Gerstenhaber and Rauch proposed an alternative approach; they attempted to characterize the Teichmüller map via a variational characterization using harmonic maps. They consider the following: Let $g = \lambda|dw|^2$ be a conformal metric on Σ_2 so that

$$\int_{\Sigma_2} \lambda(w)\, du\, dv = 1 \qquad (w = u + iv).$$

Let $\mathcal{M}$ be a family of such metrics and $\mathcal{F}_h$ be a family of maps from Σ_1 to Σ_2 homotopic to a given homeomorphism h. We assume that $g \in \mathcal{M}$ and $f \in \mathcal{F}_h$ are sufficiently nice so that the Dirichlet energy of f with respect to $g = \lambda|dw|^2$,

$${}^gE^f = \int_{\Sigma_1} \lambda(f(z))(|f_z|^2 + |f_{\bar{z}}|^2)\, dx\, dy \qquad (z = x + iy),$$

makes sense. Gerstenhaber and Rauch conjectured that

$$\sup_{g\in\mathcal{M}} \inf_{f\in\mathcal{F}_h} {}^gE^f = \frac{1}{2}\left(K^* + \frac{1}{K^*}\right) \tag{2–1}$$

and proposed constructing the Teichmüller map via this variational characterization. The above equality was later proved by E. Kuwert [**Ku**] assuming the existence of the Teichmüller map.

To construct the Teichmüller map using this approach, we enlarge the class of target metrics by allowing singular surfaces. To do this, we define conformal

classes for distance functions on the target. Let S be a surface and d a distance function so that (S, d) has curvature bounded from above. By Theorem 4 and the uniformization theorem, there is a Riemann surface Σ and a conformal homeomorphism $\iota : \Sigma \to (S, d)$. Recall that a point in the Teichmüller space with base surface S is represented by a pair (Σ, ϕ) where Σ is a Riemann surface and $\phi : S \to \Sigma$ is a homeomorphism. Two pairs (Σ, ϕ) and (Σ', ϕ') represent the same point if and only if $\phi' \circ \phi^{-1}$ is homotopic to a holomorphic map.

DEFINITION 5.1. *Let S be a compact surface, d be a distance function on S, Σ a Riemann surface with local coordinate $w = u + iv$, $\phi : S \to \Sigma$ a homeomorphism and (Σ, ϕ) be a point in Teichmüller space with base surface S. Suppose that (S, d) is a metric space of curvature bounded from above by κ and the metric topology of d is equivalent to the surface topology. Then d is said to be in the Teichmüller class of (Σ, ϕ) if there exists a conformal homeomorphism $\iota : \Sigma \to (S, d)$ homotopic to ϕ^{-1}. The area of S with respect to d is*

$$A(d) = \int_\Sigma \lambda du\, dv,$$

where λ is the conformal factor of the map ι.

Let d be a distance function on a surface S, $w = u + iv$ a local coordinate on Σ_2, and $\phi : S \to \Sigma_2$ a homeomorphism. If d is in the Teichmüller class of (Σ_2, ϕ), then (S, d) can be identified with (Σ_2, g) where $g = \lambda |dw|^2$ and λ is the conformal factor of ι. By Theorem 1, λ satisfies $\triangle \log \lambda \geq -2\kappa\lambda$ weakly.

Now let $\mathcal{D}_{a,\kappa}$ be the space of distance functions d in the Teichmüller class of (Σ_2, ϕ) with curvature bounded from above by κ and with area equal a, and let $\mathcal{M}_{a,\kappa}$ be the space of metrics $g = \lambda\, |dw|^2$ on Σ_2 with λ satisfying $\triangle \log \lambda \geq -2\kappa\lambda$ weakly and $\int_{\Sigma_2} \lambda\, dx\, dy = a$. By the discussion above, the spaces $\mathcal{D}_{a,\kappa}$ and $\mathcal{M}_{a,\kappa}$ are equivalent.

We seek to solve the variational problem described by the left-hand side of equation (2–1) by replacing $\mathcal{M}$ by $\mathcal{M}_{a,\kappa}$ (or equivalently by $\mathcal{D}_{a,\kappa}$):

$$\sup_{g \in \mathcal{M}_{a,\kappa}} \inf_{f \in \mathcal{F}_h} {}^g E^f = \frac{1}{2}\Big(K^* + \frac{1}{K^*}\Big) a.$$

A key to any variational construction is a compactness property; more precisely, a limit of a maximizing or a minimizing sequence of a given functional in a chosen set must also belong to that set. We consider the functional $\mathcal{E} : \mathcal{M}_{a,\kappa} \to \mathbb{R}$ (or equivalently $\mathcal{E} : \mathcal{D}_{a,\kappa} \to \mathbb{R}$) defined by

$$\mathcal{E}(g) = \inf_{f \in \mathcal{F}_h} {}^g E^f.$$

We use the following compactness theorem for distance functions and energy minimizing maps:

THEOREM 6 (Mese [**Me5**]). *Let $\{d_i\}$ be a sequence in $\mathcal{D}_{\kappa,a}$. There exists a subsequence $\{d_{i'}\} \subset \{d_i\}$ so that $d_{i'}$ converge uniformly to $d_0 \in \mathcal{D}_{\kappa,a}$.*

THEOREM 7 (Korevaar–Schoen [**KS2**], Mese [**Me4**]). *Let $\{d_i\}$ be a sequence of distance functions on S with curvature bounded from above by κ so that d_i converge uniformly to a distance function d_0. Assume S is compact with respect to the metric topology induced by d_i. Let $h : \Sigma_1 \to S$ be a continuous map and let $f_i : \Sigma_1 \to (S, d_i)$ be energy minimizing maps in the homotopy class of h so that the energy of f_i is*

bounded from above by K for each i. Then there exists a subsequence $\{f_{i'}\} \subset \{f_i\}$ and an energy minimizing map f_0 with respect to d_0 so that $d_{i'}(f_{i'}(\cdot), f_{i'}(\cdot))$ converge uniformly to $d_0(f_0(\cdot), f_0(\cdot))$ and the energies of $f_{i'}$ converge to that f_0.

By taking a maximizing sequence $\{g_i\}$ of the functional $\mathcal{E}$, we can obtain a metric g_0 (by Theorem 6) and a map f_0 (by Theorem 7) which satisfy

$$ {}^{g_0}E^{f_0} = \lim_{i' \to \infty} {}^{g_{i'}}E^{f_{i'}} = \sup_{g \in \mathcal{M}_{a,\kappa}} \inf_{f \in \mathcal{F}_h} {}^{g}E^{f}. $$

Finally, we show that the map f_0 is a Teichmüller map if $\kappa > 0$. To accomplish this, we show that a first variation type argument remains valid in this singular setting. We outline this argument below: If $K_{f_0}(z_0) > K^*$ for some $z_0 \in \Sigma_1$, then $K_{f_0} > K^*$ in a neighborhood of z_0 by the upper semicontinuity of the dilatation function. We bump-up the metric $g_0 = \lambda_0 |dw|^2$ in this neighborhood to construct a one-parameter family of metrics $g_t = \lambda_t |dw|^2$ with $\lambda_t > \lambda_0$ near z_0 for $t > 0$. In doing so, we are careful to control the lower curvature bound $\kappa(t)$ of g_t. Let $a(t)$ be the area of Σ_2 with respect to g_t. Next, we check that there is a sequence $t_i \to 0$ so that if f_{t_i} is the energy minimizing map with respect to the metric g_{t_i}, then $K_{f_{t_i}}$ converge almost everywhere to K_{f_0}. This is the key to showing the first variation type inequality;

$$ \liminf_{t_i \to \infty} \frac{{}^{g_{t_i}}E^{f_{t_i}} - {}^{g_0}E^{f_0}}{t_i} > \frac{1}{2}\Big(K^* + \frac{1}{K^*}\Big) a'(0). $$

Let

$$ \mathcal{E}(a, \kappa) = \sup_{g \in \mathcal{M}_{a,\kappa}} \inf_{f \in \mathcal{F}_h} {}^{g}E^{f}. $$

By definition of $\mathcal{E}(a, \kappa)$,

$$ \frac{d}{dt}\mathcal{E}(a(t), \kappa(t))|_{t=0} \geq \liminf_{t_i \to \infty} \frac{{}^{g_{t_i}}E^{f_{t_i}} - {}^{g_0}E^{f_0}}{t_i}. $$

Hence, by the chain rule,

$$ \frac{\partial \mathcal{E}}{\partial a}(a(0), \kappa(0)) \cdot a'(0) + \frac{\partial \mathcal{E}}{\partial \kappa}(a(0), \kappa(0)) \cdot \kappa'(0) > \frac{1}{2}\Big(K^* + \frac{1}{K^*}\Big) a'(0). \tag{2–2} $$

Recall that the curvature bound can be adjusted by rescaling the metric. Rescaling the metric just changes the area of the metric, and we see that

$$ \frac{\partial \mathcal{E}}{\partial \kappa} \leq \frac{1}{\kappa} \frac{\partial \mathcal{E}}{\partial a}. \tag{2–3} $$

Because we can control the curvature bound, $\kappa'(0)$ can be made sufficiently small and inequality (2–2) and (2–3) then implies

$$ \frac{\partial \mathcal{E}}{\partial a} > \frac{1}{2}\Big(K^* + \frac{1}{K^*}\Big). $$

Now, note that for any fixed κ, $a \mapsto \mathcal{E}(a, \kappa)$ is a linear map with $\lim_{a \to 0} \mathcal{E}(a, \kappa) = 0$. Hence,

$$ \mathcal{E}(a, \kappa) = \int_0^a \frac{\partial \mathcal{E}}{\partial a} da > \frac{1}{2}\Big(K^* + \frac{1}{K^*}\Big) a. $$

On the other hand, if f^* is an extremal map (i.e., $K[f^*] = K^*$), then

$$ \mathcal{E}(a, \kappa) = \sup_{g \in \mathcal{M}_{a,\kappa}} \inf_{f \in \mathcal{F}_h} {}^{g}E^{f} \leq \sup_{g \in \mathcal{M}_{a,\kappa}} {}^{g}E^{f^*} \leq \Big(K^* + \frac{1}{K^*}\Big) a. $$

This contradiction shows that $K_{f_0} \leq K^*$ for all $z \in \Sigma_1$. A similar argument shows that $\mathcal{E}(a, \kappa) = (K^* + 1/K^*)\, a$, which in turn implies that $K_{f_0} \equiv K^*$ almost everywhere. This is enough to prove that f_0 is the Teichmüller map.

References

[Ah1] L. V. Ahlfors. *Lectures on quasiconformal mappings.* Van Nostrand, Princeton, 1966.

[Ah2] L. V. Alhfors. *On quasi-conformal mappings.* Jour. d'Anal. Math. 4 (1954), 1–58.

[Be] L. Bers. *Quasiconformal mappings and Teichmüller's theorem.* Seminar on analytic functions. Princeton University Press, 1960, 89–119.

[ES] J. Eells and J. H. Sampson. *Harmonic mappings of Riemannian manifolds.* Amer. J. Math. 86 (1964), 109–160.

[FT] A. E. Fischer and A. J. Tromba. *A new proof that Teichmüller space is a cell.* Trans. Amer. Math. Soc. 303 (1987), 357–262.

[FK] R. Fricke and F. Kleine. *Vorlesungen über die Theorie der automorphen Funktionen.* Teubner, Leipzig, 1926.

[GR] M. Gerstenhaber and H. E. Rauch. *On extremal quasiconformal mapppings I, II.* Proc. N. A.S. 40 (1954), 808–812, 991–994.

[GS] M. Gromov and R. Schoen. *Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one.* IHES Publ. Math. 76 (1992), 165–246.

[HL] R. Hardt and F. H. Lin. *Harmonic maps into round cones and singularities of nematic liquid crystals.* Math. Z. 213 (1993), 575–593.

[HK] W. K. Hayman and P. B. Kennedy. *Subharmonic functions, vol. 1.* Academic Press, London, 1976.

[Hu] A. Huber. *Zum potentialtheoretischen Aspekt der Alexandrowschen Flächentheorie.* Comment. Math. Helv. 34 (1960), 99–126.

[Jo1] J. Jost. *Equilibrium maps between metric spaces.* Calc. Var. 2 (1994), 173–204.

[Jo2] J. Jost. *Generalized harmonic maps between metric spaces.* Geom Anal. and Calc. Var. (1996), 143–174.

[KS1] N. Korevaar and R. Schoen. *Sobolev spaces and harmonic maps for metric space targets.* Comm. Anal. Geom. 1 (1993), 561–659.

[KS2] N. Korevaar and R. Schoen. *Global existence theorem for harmonic maps to non-locally compact spaces.* Comm. Anal. Geom. 5 (1997), 333–387.

[Ku] E. Kuwert. *Harmonic maps between flat surfaces with conical singularities.* Math. Z. 221 (1996), 421–430.

[Le] M. Leite. *Harmonic mappings of surfaces with degenerate metrics.* Amer. J. Math. 110 (1988), 399–412.

[Me1] C. Mese. *Curvature of minimal surfaces in singular spaces.* Comm. Anal. Geom. 9 (2001), 3–34.

[Me2] C. Mese. *Some properties of minimal surfaces in singular spaces.* Trans. Amer. Math. Soc. 352 (2000), 547–580.

[Me3] C. Mese. *The structure of singular spaces of dimension 2.* Manu. Math. 100 (1999), 375–389.

[Me4] C. Mese. *Harmonic maps into spaces with an upper curvature bound in the sense of Alexandrov.* to appear Math. Z.

[Me5] C. Mese. *Harmonic maps between surfaces and Teichmüller spaces.* to appear Amer. J. Math.

[Me6] C. Mese. *A variational construction of the Teichmüller map.* preprint.

[Mi] Y. Miyahara. *On some properties of a Teichmüller mapping.* TRU Math. 4 (1968), 36–43.

[Mo] C. B. Morrey. *The problem of Plateau on a Riemannian manifold.* Ann. of Math. 49 (1948), 807–851.

[R] E. Reich. *On the variational principle of Gerstenhaber and Rauch.* Ann. Acad. Sci. Fenn. 10 (1985), 469–475.

[RS] E. Reich and K. Strebel. *On the Gerstenhaber–Rauch Principle.* Israel J. Math. 57 (1987), 89–100.

[Re] Y. G. Reshetnyak. *Isothermal coordinates in manifolds of bounded curvature.* Dokl. Aka. Nauk SSSR 94 (1954), 191–264.

[Se1] T. Serbinowski. *Harmonic maps into metric spaces with curvature bounded above.* Ph.D. thesis, University of Utah, 1995.

[Se2] T. Serbinowski. *Boundary regularity of harmonic maps to nonpositively curved metric spaces.* Comm. Anal. Geom. 2 (1994), 139–154.

[Sch] R. Schoen. *Analytic aspects of the harmonic map problem.* In *Seminar in non-linear partial differential equations*, edited by S. S. Chern, MSRI publications vol. 2. Springer, New York, 1985.

[Te1] O. Teichmüller. *Extremale quasikonforme Abbildungen und quadratische differentiale.* Abh. Preuss. Akad. 22 (1940), 1–197. Reprinted as pp. 335–531 in his *Collected Papers*, Springer, Berlin, 1982.

[Te2] O. Teichmüller. *Bestimmung der extremalen quasikonformen Addildungen bei geschlossenen orientierten Riemannschen Flächen.* Abh. Preuss. Akad. 4 (1943), 1–42. Reprinted as pp. 635–676 in his *Collected Papers*, Springer, Berlin, 1982.

[Wo1] M. Wolf. *The Teichmüller theory of harmonic maps.* J. Diff. Geom. 29 (1989) 449–470.

[Wo2] M. Wolf. *High energy degeneration of harmonic maps between surfaces and rays in Teichmüller space.* Topology 30 (1991) 517–540.

[Wo3] M. Wolf. *Harmonic maps from surfaces to $\mathbb{R}$-trees.* Math. Z. 218 (1995) 577–593.

[Ya] S. Yamada. *Weil–Peterson convexity of the energy functional on classical and universal Teichmüller spaces.* J. Diff. Geom. 51 (1999).

Connecticut College, 270 Mohegan Ave., New London, CT 06320
E-mail address: cmes@conncoll.edu

Clay Mathematics Proceedings
Volume 2, 2004

Shortest Networks in 2 and 3 Dimensions

J. Hyam Rubinstein

Abstract. We give a survey of recent results on Steiner trees in the plane and in 3-dimensional space. Included are the solution of Graham's problem on the shortest network for points on a circle, a polynomial time algorithm for the Steiner tree connecting up terminals lying on a fixed set of smooth curves of finite length and a determination of the shortest networks for points lying in a regular array or lattice in the plane. In three dimensions, a generalized version of Melzak's construction is discussed, by finding the furthest point on an immersed high dimensional torus. Recent applications to underground mining optimization are also described briefly.

1. Introduction

Shortest networks have a long and interesting history. The 'roadway problem' is to find the system of roads of shortest total length, connecting a given set of cities in the plane. In the simplest version of the problem, there are no obstacles and the earth is completely flat. This is often called the Steiner problem and was popularized by Courant and Robbins in their book, What is Mathematics? In the 1960s and 1970s, mathematicians at Bell labs (E. Gilbert, H. Pollak, F. Chung, M. Garey, R. Graham, F. Hwang, D. Johnson) became interested in the question, as it was relevant to costing for private phone lines used by corporations. A famous example mentioned by Pollak was Delta Airlines, which had office centers at New York, Chicago and Atlanta, forming the vertices of a triangle, which was close to being equilateral. It was realized that putting an additional fictitious 'office' in the middle of the triangle would save about 14% of the costs of the line rental, which was based on a minimal spanning tree distance between the centers.

The Steiner tree is the shortest possible network, which is obviously a tree. A minimal spanning tree is obtained by taking a collection of shortest edges connecting the given terminals, with no extra vertices allowed. For the Steiner tree, additional vertices are called Steiner points. It is easy to show that all the angles at a Steiner point must be 120^o. There are elegant algorithms to find both minimal

1991 *Mathematics Subject Classification.* Primary 05C05, 90B85, 68R10.

Key words and phrases. Steiner trees, minimal spanning trees, variational method, polynomial algorithms, combinatorial optimization.

M. Brazil, D. Lee, D. Thomas, N. Wormald, and J. Weng (in various combinations) are joint authors for the material for these notes.

©2004 Clay Mathematics Institute

spanning trees and Steiner trees. Notice that a Steiner tree can be viewed as a one-dimensional soap film, and one can find such trees by joining two parallel flat sheets of plastic by rods at the positions of the terminals and dipping it into soap solution. We will see later that simple ideas from the calculus of variations are very useful.

An important area of applications is to VLSI chip design. Distances in the plane are measured by the l_1 or Manhattan metric and the corresponding shortest Steiner trees are often called rectilinear. We will not deal with such networks here, but the techniques used are very similar.

2. Algorithms for Spanning Trees and Steiner Trees

For minimal spanning trees, Kruskal [**14**] gave a greedy algorithm which is linear time in the number n of terminals. Assume that all the edge lengths between pairs of terminals are known. Then keep choosing the shortest edge, with the constraint that a cycle is never produced in the graph.

Melzak [**16**] came up with a beautiful geometric method to build a Steiner tree, knowing its 'topology', that is, the way the terminals are connected to the extra Steiner points. A 'cherry' of the tree consists of two terminals V_1, V_2 connected to the same Steiner point S_1. Draw an equilateral triangle with two vertices as V_1, V_2, oriented so that the cherry is on the opposite side of the edge V_1V_2 to the triangle. The third vertex M is often called a *Melzak point.* Now elementary Euclidean geometry tells us that the line through M and S_1 gives 120^o angles with the edges S_1V_i for $i = 1, 2$ and the lengths are related by $|MS_1| = |S_1V_1| + |S_1V_2|$. So we can replace the two terminals V_1, V_2 by the Melzak point M and have now to find a Steiner tree on one fewer terminals. By induction, any Steiner tree *with a particular topology* can be found in linear time.

There is one subtlety here — note that initially we are only given the terminals and the topology. So a choice has to be made of which side to place the Melzak point, that is, the position of the equilateral triangle depends on knowing the position of the cherry. This problem was solved by Hwang [**12**] — one can predict which is the correct choice, avoiding an exponential number (2^n) of possibilities.

However Garey, Graham and Johnson [**9**] showed that finding the shortest Steiner tree is in the class NP, that is, one does not expect any polynomial time algorithm. The problem is that the number of topologies grows exponentially with n. Winter has written a program called GEOSTEINER to implement Melzak's algorithm, using sophisticated 'pruning methods' to rule out many topologies which do not give plausible shortest trees. His algorithm can quickly find shortest Steiner trees for $n \leq 100$.

Note that the topology of a Steiner tree is called *full* if there are the maximal number $n-2$ of Steiner vertices introduced. So all terminals occur at vertices of the tree of degree 1 and all other vertices are Steiner of degree 3. A nonfull tree splits into full components. It has some terminals which are of degree 2 or very rarely of degree 3. By decomposing along all such terminals of degree > 1, one obtains the full components of the Steiner tree. Many problems are tackled by reducing to the study of full trees.

Approximations, special cases and heuristic algorithms have been major topics of investigation. A good general reference is the book 'Steiner tree problems' by Hwang, Richards, Winter [**13**].

3. The Steiner Ratio Conjecture and the Variational Method

Gilbert and Pollak [**11**] had computed the ratio ρ between the lengths of a shortest Steiner tree L_S and the minimal spanning tree L_T for small examples. They came up with the Steiner ratio conjecture

$$\rho = \frac{L_S}{L_T} \geq \frac{\sqrt{3}}{2.}$$

This bound is achieved for equilateral triangles and shows that the computationally difficult problem of finding a shortest Steiner tree S saves at most 14% of the length compared to the easy algorithm of a minimal spanning tree T. Simple approximation algorithms have been developed by starting with T and trying to shrink the length by putting in extra (Steiner) vertices at terminals with angles $< 120^o$.

Doreen Thomas and I introduced the variational method to study the ratio conjecture [**17**]. A good case to consider is for three terminals X, Y, Z. We show that if XYZ is not equilateral, then by moving the terminals, the ratio ρ can be reduced. Think of ρ as a piecewise differentiable function on the configuration space, consisting of all triangles. By dividing out by translations, rotations and rescaling, one can parametrize triangles by the lengths $x = |XS_1|, y = |YS_1|, z = |ZS_1|$, where $x + y + z = 1, x, y, z \geq 0$, where S_1 is the Steiner vertex. (Note it is easy to deal with the case that there is no Steiner vertex, since then $S = T$. It is easy to verify this is precisely the case when one angle of the triangle is at least 120^o.) We call $\{(x, y, z) : x + y + z = 1, x, y, z \geq 0\}$ the *configuration space*. Generally for n terminals, this is a simplex of dimension $2n - 4$.

We compute the first derivative

$$\rho' = \frac{L_T{}'}{L_T}\left(\frac{L_S{}'}{L_T{}'} - \frac{L_S}{L_T}\right).$$

Notice that if L_T is decreasing, then ρ also decreases when the expression in brackets is positive. So we require that the ratio of derivatives of L_S and L_T be larger than ρ to show that the configuration of X, Y, Z does *not* achieve the minimum ratio. The difficulty is that there may be several different topologies for T, so we are actually dealing with several functions ρ. We need to show *each one* decreases for a given perturbation. It suffices to consider only one topology for S though.

If XYZ is neither isosceles nor equilateral, then there is a unique choice for T. Also we may assume that $T = XZ \cup YZ$ and $x > y > z$ without loss of generality. Now move X towards S_1 at unit speed. Then $L_T{}' = -1$ and $L_S{}' = -cos\theta$, where θ is the angle between XZ and XS_1. We see that the expression in brackets for $\dot{\rho}$ is $\frac{1}{cos\theta} - \rho$ which is clearly positive. (Note we are just differentiating a dot product to get $|XZ|$.)

If XYZ is isosceles, then we can suppose that $x > y = z$ and do the same argument. Notice that here there are two choices for T, namely $XZ \cup YZ$ or $XY \cup YZ$. However these both decrease at the same rate under the deformation above. So the proof is complete; we have shown that the equilateral triangle is the unique minimum for ρ.

This variational approach turns out to be useful for many questions about Steiner trees. A Morse theory of Min-type functions (that is, either maxima or minima of finite families of smooth functions) appears in [**10**]. Also we used the

Meeks-Yau exchange and round off trick from minimal surface theory [**15**], to show that all minimal spanning trees are "compatible", that is, do not cross each other.

In 1992, Du and Hwang [**8**] completely solved the ratio conjecture. They analyzed ρ on the same configuration space, but simplified the problem by observing that since one can put $L_S \equiv 1$ using rescaling, $\rho = \frac{1}{L_T}$ is concave for each choice of minimal spanning tree topology. They also used a wonderful trick of forming a 'spiral staircase' surface (very much like a Riemann surface which is a branched covering) over the plane, to hold enough spanning tree topologies simultaneously, so that all the terminals were on the boundary of the region. In this way, they could overcome the difficulty of seeing how *all* the lengths of the spanning trees change as the terminals are perturbed.

4. Special Cases

Around 1970, Ron Graham had asked if a collection of terminals on a unit circle is the shortest Steiner tree given by a minimal spanning tree following around the circle, assuming that there is at most one distance between adjacent terminals ≥ 1? It is easy to see that if there are several 'gaps' larger than distance 1, then a shorter network can be obtained by using Steiner vertices near the center of the circle. D. Thomas and I were able to solve this problem in [**18**]. The key was to study the Steiner ratio ρ between the length of some potential Steiner tree and the obvious minimal spanning trees. Negativity of the second derivative of ρ has a nice interpretation, as the terminals move around the circle. If a critical point of ρ occurs, then there are restrictions on the angles between the Steiner tree edges and the spanning tree edges, enabling one to estimate the second derivative. The conclusion was that $\rho \geq 1$ with Graham's condition. A *reversible variation* for the configuration is one for which $D\rho(V) = -D\rho(-V)$, where V denotes the direction vectors for perturbing the configuration of terminals and D is the directional derivative. If the configuration is a critical point for ρ, it is easy to see that any reversible variation has to have value 0. For such a variation, the second variation has a very useful form

$$D^2\rho(V) = \frac{L_S{}'' - \rho L_T{}''}{L_T}.$$

So if a configuration is a minimum for ρ, then any suitable reversible variation must satisfy $L_S{}'' - \rho L_T{}'' \geq 0$.

An interesting special class of general problems is where the terminals lie in a disjoint collection of continuously differentiable compact embedded curves $\mathcal{C}$ ([**20**])). A simple version of the monotonicity formula from geometric measure theory, gives control on the growth in length of a shortest network in a region of the plane which does not contain any terminals. Consequently there is a bounded number of Steiner points in the outside of a small collar region about $\mathcal{C}$. We used this as a starting point to show that there is a polynomial time algorithm to find the shortest Steiner tree, with the above assumptions. The complexity of the polynomial depends on the curvature of the curves and how close are different branches.

A major difficulty is that there can be arbitrarily long 'tails' of such shortest Steiner trees, but the lengths of the edges have to decrease at an exponential rate. A tail consists of a sequence of terminals on a branch B of a curve, which are exponentially converging and are joined to a collection of Steiner points by edges which are nearly parallel to B. The end of the tail is the final unique cherry.

We were able to show that locating the ends of such tails could be achieved in polynomial time and so the rest of the tail can be filled in readily. It turns out that this is the only way that Steiner points can accumulate close to $\mathcal{C}$.

Moreover, if a smooth graph (with intersecting branches) or some curves of infinite length were used, instead of a collection of embedded curves of finite length, then the problem becomes NP hard. Two instances are to find shortest networks for terminals on two parallel lines, or two straight line segments of finite length meeting at an angle of $< 120^o$. It is interesting to see this 'boundary' between polynomial time and NP problems. The basic reason is that there are many choices of where to put the cherries of the tree, by slightly adjusting positions of the terminals. So one can show that the complexity of choosing the topology is equivalent to other problems in the class NP. We have recently proved a similar result for the traveling salesman problem.

A final class of special cases is to determine the shortest Steiner trees on terminals forming a regular subset of a lattice in the plane. Both square or rectangular sets of points had been considered in articles by F. Chung, M. Gardner and R. Graham [**7**], [**6**]. They gave evidence based on extensive computations, of how the shortest networks should be constructed. In [**1**], [**3**],[**4**] all the problems were solved and in one case, a better answer was achieved than the computer had predicted!

The main idea (due to Nick Wormald) was to study the 'excess' of such networks. If a square of side length 2^k is chosen, then a very special shortest network can be found, consisting entirely of $X's$. For a network in a general lattice, the excess measures how much longer is the tree, than a theoretical collection of such $X's$. It turns out that the shortest pieces, needed to be combined to give the general solution, all have very small excess. The variational method is very useful again in estimating lengths to give these shortest pieces.

5. 3-Dimensional Shortest Networks

J. Smith and W. Smith [**22**], [**23**] studied the geometry of shortest networks in 3-dimensional space. Melzak's algorithm cannot be extended to this case, although it is still true that additional Steiner vertices are of degree 3 and the edges have angles of 120^o there. The problem is that the plane containing the three edges may have to twist at an adjacent Steiner vertex.

They also investigated the Steiner ratio conjecture in three dimensions and came up with the 'sausage' conjecture. For four points, the smallest ratio is obtained by terminals lying at the vertices of a regular tetrahedron (Rubinstein, Thomas, Weng [**19**]). Smith and Smith conjectured that for n points, the smallest ratio is obtained by forming a chain or sausage of $n-3$ regular tetrahedra glued together in a line. The terminals then lie on a triple helix. J. Smith has also asked if there is some type of minimum energy associated with this problem.

Our interest in this area has been stimulated by applications in design of underground mines. The two most common designs are via a shaft or a decline. In the former case, there is a large cost associated with the shaft and lifting equipment. Typically, drives and ramps are then run from the base of the shaft to the ore bodies. These are both to give access for blasting (so the drives may run to the top of the deposit) and for haulage (from the bottom). A key constraint is that the drives may not have gradient more than about $1:10$, since otherwise very expensive four-wheel drive vehicles are needed. There is often a constraint on the radius

of turning circles and separation of tunnels and avoidance issues (such as the ore body itself or unstable regions) are significant. However, the overall optimization problem is very similar to the Steiner tree problem. Here the edges of the tree have different weights, depending on whether they correspond to the shaft, haulage or access drives.

A decline is used if the ore body is reasonably close to the surface and the geology is suitable. A decline is just a long sloping ramp from the surface (can be of the order of a mile long), often in a helical shape. Again the drives and ramps are run off the decline and the optimization problems are similar to the case of a shaft.

In both cases, it is a very interesting mathematical problem to find the best solution *with a given topology* ([**5**]). Then simulated annealing or genetic programming is used to run through all the possible topologies. Our algorithm to do this allows for haulage costs over the life time of the mine, that is, we seek to minimize total capital and running costs. We are currently writing up the solution but have actually been working on a computer program over the last three years to achieve this in a practical way. A number of model design projects have been done with mining companies in Australia testing out the method ([**2**]). David Lee was the director of research at the University of South Australia and has been the key person introducing us to this area and formulating the mathematical models from the mining engineers.

To finish, I will describe briefly how to modify Melzak's algorithm in the 3-dimensional classical case. The key idea is that the Melzak point M now is free to rotate about a circle in space, since we do not know the plane through the first Steiner point S_1 connected to a cherry V_1, V_2. Notice that for four points, if we construct two Melzak circles C_1, C_2 from the two terminals, then the correct choices of Melzak points M_1, M_2 are the farthest apart points. It turns out that generally, one can consider replacing V_1, V_2 by M_1 moving around on C_1. Next for each choice of M_1, we can form a Melzak circle C_2 so have a circle worth of circles, that is, a (singular) torus. For n terminals, we need to do this $n-2$ times, i.e., have an immersed $(n-2)$-torus of choices. The true direction and length of the Steiner tree is given by connecting the last terminal to the furthest point on this torus. Alternatively we could expand the tree from two ends and finish with a pair of tori of dimensions $d, n-d-2$ and seek the pair of points which are furthest apart on these tori. (See [**21**] for a discussion of this construction and results about useful approximations of the true Steiner tree).

One can write down an explicit collection of simple equations to describe the tori. Another interesting method is the following ([**19**]). Consider how to determine the Steiner points for four points X, Y, Z, W in space, with a given full topology. So we will suppose that the Steiner points S_1, S_2 are adjacent to X, Y and Z, W respectively. Now it is easy to see that the extension of the line S_1S_2 meets the edges XY and ZW in points P, Q respectively. Now we can write down equations for the positions of P, Q using the following picture. Note that the correct position of the Melzak points M, M' for the equilateral triangles MXY and $M'ZW$ satisfies that M, P, Q, M is a straight line. Denote the angles XPQ and WQP by θ and ϕ respectively. Also denote the ratio of distances XP/XY by t and WQ/WZ by s.

Now the sine rule in the triangles MPX and $M'WQ$ implies that

$$t = \frac{1}{2} - \frac{\sqrt{3}}{2} cot\theta, \quad s = \frac{1}{2} - \frac{\sqrt{3}}{2} cot\phi.$$

Denote the vectors XY, YZ, ZW by $\mathbf{a}, \mathbf{b}, \mathbf{c}$ respectively. Elementary vector geometry gives the equations

$$cos\phi = \frac{\mathbf{u} \cdot \mathbf{c}}{|\mathbf{u}||\mathbf{c}|}, \quad cos\theta = \frac{\mathbf{u} \cdot \mathbf{a}}{|\mathbf{u}||\mathbf{a}|}, \quad \mathbf{u} = (1-t)\mathbf{a} + \mathbf{b} + (1-s)\mathbf{c}.$$

So combining these equations to eliminate the angles θ, ϕ gives two linked quadratic equations in the variables s, t. These equations can be viewed as the analogues of the Melzak method in the plane — the same method works to determine the position of the Steiner points for any given topology in $\mathbb{R}^3$.

References

[1] M. Brazil, T. Cole, J. H. Rubinstein, D. Thomas, J. Weng, N. Wormald, Minimal Steiner trees for $2^k \times 2^k$ Square Lattices, J. Comb. Theory, Series A, 73 (1996), 91-110.

[2] M. Brazil, D Lee, J. H. Rubinstein, D. Thomas, J. Weng, N. Wormald, Network optimisation of underground mine design, Australasian Institute of Mining and Metallurgy, 305 (2000), 57-65

[3] M. Brazil, J. H. Rubinstein, D. Thomas, J. Weng, N. Wormald, Minimal Steiner trees for Rectangular Arrays of Lattice Points, J. Comb. Theory, Series A, 79 (1997), 181-208.

[4] M. Brazil, J. H. Rubinstein, D. Thomas, J. Weng, N. Wormald, Full minimal Steiner trees on lattice sets, J. Comb Theory, Series A, 78 (1997), 51-91.

[5] M. Brazil, J. H. Rubinstein, D. Thomas, J. Weng, N. Wormald, Gradient-constrained minimum networks, I. Fundamentals, Journal of global optimisation, 21 (2001), 139-155.

[6] F. Chung and R. Graham, Steiner trees for ladders, Ann. Discr. Math. 2 (1978), 173-200.

[7] F. Chung, M. Gardner and R. Graham, Steinre trees on a chequerboard, Math. Magazine, 62 (1989), 83-96.

[8] D. Z. Du and F. Hwang, A proof of the Gilbert-Pollak Conjecture on the Steiner ratio, Algorithmica, 7 (1992), 121-135.

[9] M. Garey, R. Graham and D. Johnson, The complexity of computing Steiner minimal trees, SIAM J. Appl. Math. 32 (1977), 835-859.

[10] V. Gershkovich and J. H. Rubinstein, Morse theory for Min-type functions, Asian J. of Math. 1 (1997), 696-715.

[11] E. Gilbert and H. Pollak, Steiner minimal trees, SIAM J. Appl. Math. 16 (1968), 1-29.

[12] F. Hwang, A linear time algorithm for full Steiner trees, Oper. Res. Lett. 5 (1986), 235-237.

[13] F. Hwang, D. Richards and P. Winter, The Steiner Tree Problem, Annals of Discrete Mathematics, 53, North Holland, Amsterdam, 1992

[14] J. Kruskal, On the shortest spanning subtree of a graph and the travelling salesman problem, Proc. Amer. Math. Soc. 7 (1956), 48-50.

[15] W. Meeks and S. T. Yau, Topology of three-dimensional manifolds and the embedding problems in minimal surface theory, Annals of Math 112 (1980), 441-484.

[16] Z. Melzak, On the problem of Steiner, Canad. Math. Bull. 4 (1961), 143-148.

[17] J. H. Rubinstein and D. Thomas, A variational approach to the Steiner network problem, Ann. Oper. Res. 33 (1991), 481-499.

[18] J. H. Rubinstein and D. Thomas, Graham's problem on shortest networks for points on a circle, Algorithmica, 7 (1992), 193-218.

[19] J. H. Rubinstein, D. Thomas and J. Weng, Minimum networks for four points in space, Geom. Dedicata, 93, (2002), 57-70.

[20] J. H. Rubinstein, D. Thomas and N. Wormald, Steiner trees for terminals constrained to curves, SIAM J. Discr. Math. 10 (1997), 1-17.

[21] J. H. Rubinstein, J. Weng and N. Wormald, in preparation

[22] W. Smith and J. MacGregor Smith, On the Steiner ratio in 3-space, J. Comb. Theory, Series A, 69(1995), 301-332.

[23] J. MacGregor Smith and B. Toppur, Euclidean Steiner minimal trees, minimum energy configurations and the embedding problem of weighted graphs in E^3, Discrete Appl. Math. 71 (1996), 187-215.

Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria 3010, Australia

E-mail address: rubin@ms.unimelb.edu.au

Agol, Ian
MSCS
University of Illinois, Chicago
m/c 249, 322 SEO, 851 S. Morgan St.
Chicago, Il 60607
agol@math.uic.edu

Alencar, Hilario
Departamento de Matematica
Universidade Federal de Alagoas Maceio,
Alagoas 57072-900 Brazil
hilario@mat.ufal.br

Alvarez, Miguel Carrion
Department of Mathematics
University of California, Riverside
Riverside, CA 92521
miguel@math.ucr.edu

Auth, Matthew
Department of Mathematics & Statistics
University of Massachusetts, Amherst
710 N Pleasant Street
Amherst, MA 01003
auth@math.umass.edu

Batista, Valerio
IMECC-UNICAMP CP 6065
Campinas, SP 13083-970 Brazil
valerio@ime.unicamp.br

Beeson, Michael
Department of Mathematics
San Jose State University
San Jose, CA 95192-0103
beeson@cruzio.com

Beheshti, Shabnam
Department of Mathematics & Statistics
University of Massachusetts, Amherst
710 N. Pleasant St.,
Lederle Graduate Research Tower
Amherst, MA 01003
beheshti@math.umass.edu

Bobenko, Alexander
Institut fuer Mathematik
Technische Universitët, Berlin
Strasse des 17. Juni 136
Berlin, 10623 Germany
bobenko@math.tu-berlin.de

Bozin, Vladimir
Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue, Room 2-088
Cambridge, MA 02139
bozin@math.mit.edu

Cao, Xiaodong
Department of Mathematics
Columbia University
2990 Broadway New York, NY 10027
cao@math.columbia.edu

Chen, Xiaojun
Department of Mathematics
SUNY, Stony Brook
Stony Brook, NY 11794
chen@math.sunnysb.edu

Cheng, Xu
Inst. de Matematical Pura e Aplicada
Est. Dona Castorina 110,
Jardim Botanico
Rio de Janeiro, RJ 22460-320 Brazil
xcheng@impa.br

Choe, Jaegyoung
Department of Mathematics
Seoul National University
San 56-1, Shilim-dong, Kwanak-gu
Seoul, 151-742 South Korea
choe@math.snu.ac.kr

Collin, Pascal
Laboratoire Émile Picard
Université de Toulouse 3 (Paul Sabatier)
118 route de Narbonne
Toulouse, F-31062 France
collin@picard.ups-tlse.fr

Coskunuzer, Baris
Department of Mathematics,
Princeton University
Princeton, NJ 08540
baris@princeton.edu

Daily, Marilyn
Department of Mathematics
North Carolina State University
Box 8205
Raleigh, NC 27695-8205
medaily@math.ncsu.edu

Daniel, Benoit
École Normale Supérieure
45, rue d'Ulm
Paris, F-75005 France
daniel@clipper.ens.fr

Dean, Brian
Department of Mathematics
Johns Hopkins University
3400 N. Charles St., 404 Krieger Hall
Baltimore, MD 21218
bdean@math.jhu.edu

DeCoste, Rachelle
Department of Mathematics
University of North Carolina,
Chapel Hill Chapel Hill, NC 27599-3250
rdecoste@math.unc.edu

do Espirito-Santo, Nedir
Universidade Federal do Rio de Janeiro
Rua Assis Brasil 143 bloco 1 apt. 207
Rio de Janeiro, RJ 22030-010 Brazil
nedir@impa.br

Dorff, Michael
Department of Mathematics
Brigham Young University
281 TMCB Provo, UT 84602
mdorff@math.byu.edu

El Soufi, Ahmad
Université de Tours
Parc Grandmont
Tours, F-37200 France
elsoufi@univ-tours.fr

Fang, Yi
Centre for Math. and its Applications
Australian National University
Canberra, ACT 200 Australia
yi@maths.anu.edu.au

Fang, Yong
Department of Mathematics
University of Paris
61. Boulvard Jourdan
Paris, F-75014 France
fangyong1@yahoo.fr

Fleming, Thomas
Department of Mathematics
University of California, San Diego
9500 Gilman Dr La Jolla, CA 92093
tfleming@math.ucsd.edu

Fraser, Ailana
Department of Mathematics
University of British Columbia
121 - 1984 Mathematics Road
Vancouver, BC V6T 1Z2 Canada
afraser@math.ubc.ca

Freire de Lima, Ronaldo
U. Federal do Rio Grande do Norte
Av. Jaguarari, 1915/202 - Lagoa Nova
Natal, Rio Grande do Norte 59054-150
Brazil
ronaldo@ccet.ufrn.br

Futer, David
Department of Mathematics
Stanford University
450 Serra Mall, Building 380
Stanford, CA 94305
dfuter@stanford.edu

Gabai, David
Department of Mathematics
Princeton University
Washington Road, Fine Hall
Princeton, NJ 08544
gabai@princeton.edu

Gage, Michael
Department of Mathematics
University of Rochester
915 Hylan Building
Rochester, NY 14627
gage@math.rochester.edu

Ge, Yuxin
Departement de Mathematiques
Université de Paris
12 (val de Marne)
61, Avenue du General de Gaulle
Creteil, F-94010 France
ge@univ-paris12.fr

Georgieva, Bogdana
Department of Mathematics
Oregon State University
368 Kidder Hall
Corvallis, OR 97331
bogdana@math.orst.edu

Ghomi, Mohammad
Department of Mathematics
University of South Carolina
Columbia, SC 29208
ghomi@math.sc.edu

Glickenstein, David
Mathematics University of Arizona
P.O. Box 210089
Tucson, AZ 85721
glickenstein@math.arizona.edu

Groisman, Pavel
Department of Mathematics
U. of Illinois, Urbana-Champaign
1409 W. Green Street
Urbana, IL 61801
groisman@math.uiuc.edu

Grosse-Brauckmann, Karsten
Mathematisches Institut
Universitët Bonn
Beringstr. 1
Bonn, D-53115 Germany
kgb@mathematik.tu-darmstadt.de

Guan, Bo
Department of Mathematics
University of Tennessee
Knoxville, TN 37996
guan@math.utk.edu

Gulliver, Robert
School of Mathematics, 127 Vincent Hall
University of Minnesota
206 Church St. SE
Minneapolis, MN 55455
gulliver@math.umn.edu

Halverson, Denise
Department of Mathematics
Brigham Young University
Provo, UT 84602-0002
deniseh@math.byu.edu

Hardt, Robert
Department of Mathematics
Rice University
PO Box 1892 Houston, TX 77251-1892
hardt@rice.edu

Hass, Joel
Department of Mathematics
University of California, Davis
1 Shields Ave
Davis, CA 95616
hass@math.ucdavis.edu

Hauswirth, Laurent
Analyse et mathematiques appliquees.
Univ. de Marne-la-Vallée (Copernic)
5 bd Descartes
Champs-sur-Marne, F-77454 France
hauswirth@math-univ-mlv.fr

Hoffman, David
Math. Sciences Research Institute
17 Gauss Way
Berkeley, CA 94720-5070
david@msri.org

Hofmann, Jon
Department of Mathematics
University of California,
Santa Cruz, CA 95064
hofmann@math.ucsc.edu

Holt, Paul
Department of Mathematics & Statistics
Williams College
Williamstown, MA 01267
pholt@wso.williams.edu

Huang, Zheng
Department of Mathematics
Rice University
6100 Main Street, MS136
Houston, TX 77005
huangz@math.rice.edu

Hutchings, Michael
Department of Mathematics
University of California, Berkeley
923 Evans Hall
Berkeley, CA 94720-3840
hutching@math.berkeley.edu

Ilmanen, Tom
Math Department
Columbia University
2990 Broadway
New York, NY 10027
nonce1024@hotmail.com

Ionel, Marianty
Department of Mathematics & Statistics
McMaster University
Hamilton, ON L8S 4K1 Canada
ionelm@math.mcmaster.ca

Jaffe, Arthur
Department of Mathematics
Harvard University
1 Oxford Street
Cambridge, MA 02138
jaffe@math.harvard.edu

Jenkins, Paul
Department of Mathematics
Brigham Young University
Provo, UT 84602-0002
pmj5@email.byu.edu

Jin, Sun
School of Mathematical Science
Seoul National University
San 56-1, Shilim-dong, Kwanak-gu
Seoul, 151-742 South Korea
ssjin87@yahoo.co.kr

Joyce, Dominic
Mathematical Institute
Oxford University
24-29 St. Giles
Oxford, OX1 3LB U. K.
dominic.joyce@lincoln.ox.ac.uk

Kapouleas, Nicolaos
Brown University
234 President Avenue, Apt # 5
Providence, RI 02906
nicos@math.brown.edu

Karcher, Hermann
Department of Mathematics
Universitët Bonn
Beringstrasse 1
Bonn, D-53115 Germany
unm416@uni-bonn.de

Korevaar, Nick
Department of Mathematics
University of Utah
155 S 1400 E
Salt Lake City, UT 84112
korevaar@math.utah.edu

Korolev, Alexander
Department of Mathematics
Brigham Young University
Provo, UT 84602-0002
akor@math.byu.edu

Koshkin, Sergiy
Department of Mathematics
Kansas State University
138 Cardwell Hall
Manhattan, KS 66505-2602
koshkin@math.ksu.edu

Kudzin, Matthew
Department of Mathematics
SUNY, Stony Brook
Stony Brook, New York 11777
mkudzin@math.sunysb.edu

Kusner, Robert
Math. Sciences Research Institute
17 Gauss Way
Berkeley, CA 94720-5070
kusner@msri.org

Langevin, Remi
Université de Bourgogne
18 Bd de Magenta
Dijon, F-75010 France
langevin@u-bourgogne.fr

Lawlor, Gary
Department of Mathematics Education
Brigham Young University
Provo, UT 84602-0002
lawlor@mathed.byu.edu

Lee, Jaejeong
School of Mathematical Sciences
Seoul National University
San 56-1, Shilim-dong, Kwanak-gu
Seoul, 151-742 South Korea
izzy@math.snu.ac.kr

Leger, Nicholas
Department of Mathematics
University of Texas, Austin
Austin, TX 78712-1082
nickleger@mail.utexas.edu

Liu, Lixin
Zhongshan University
Xingangxilu Rd. 135
Guangzhou, Guangdong 510275-186
China
mcsllx@zsu.edu.cn

Lo, Yiu
Department of Mathematics
University of California, Irvine
Irvine, CA 92697
ylo@math.uci.edu

Lock, Anthony
Department of Mathematics
Imperial College, London
Huxley Building, 180 Queen's Gate
London, England SW7 2BZ U. K.
anthony_lock@yahoo.com

López, Francisco J.
Departamento Geometria y Topologia
Universidad de Granada
Fuentenueva s/n
Granada, E-18071 Spain
fjlopez@goliat.ugr.es

Ma, Renyi
Department of Mathematics
Tsinghua University
Beijing, 100084 China
rma@math.tsinghua.edu.cn

Martín, Francisco
Departamento de Geometria y Topologia
Universidad de Granada
Campus de Fuentenueva, s/n
Granada, E-18071 Spain
fmartin@ugr.es

Mazzeo, Rafe
Department of Mathematics
Stanford University
Bldg. 380 – Serra St.
Stanford, CA 94305
mazzeo@math.stanford.edu

McCuan, John
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332
mccuan@math.gatech.edu

McCune, Catherine
Department of Mathematics
University of Connecticut
Storrs, CT 06269
mccune@math.uconn.edu

McLelland, Matt
Department of Mathematics
Rice University
PO Box 1892
Houston, TX 77251-1892
mattpi@swbell.net

Meadows, Alexander
Department of Mathematics
Stanford University
450 Serra Mall, Building 380
Stanford, CA 94305
meadows@math.stanford.edu

Meeks, William, III
Department of Mathematics & Statistics
University of Massachusetts, Amherst
710 N Pleasant Street
Amherst, MA 01003
bill@gang.umass.edu

Mese, Chikako
Department of Mathematics
Connecticut College
270 Mohegan Ave.
New London, CT 06320
cmes@conncoll.edu

Minicozzi, Bill
Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218
minicozz@math.jhu.edu

Moore, Helen
American Institute of Mathematics
360 Portage Avenue
Palo Alto, CA 94306
moore@aimath.org

Morgan, Frank
Department of Mathematics
Williams College
Williamstown, MA 01267
frank.morgan@williams.edu

Nelli, Barbara
Dipartimento di Matematica
Universita' di L'Aquila
via Vetoio -Loc. Coppito
L'Aquila, I-67100 Italy
nelli@univaq.it

Neumann-Coto, Max
Instituto de Matematicas
U. Nacional Autonoma de Mexico
Facultad de Ciencias
Mexico City, DF 4510 Mexico
max@math.unam.mx

Noronha, Helena
Department of Mathematics
California State University
Northridge, CA 91330-8313
maria.noronha@csun.edu

Oliveira, Krerley
Departamento de Matematica
UFAL-Brazil
Maceio, Alagoas 57080000 Brazil
krerley@mat.ufal.br

Pacard, Frank
Département de Mathématiques
Université de Paris 12 (val de Marne)
61 Avenue du General de Gaulle
Cretiel, CEDEX F-94010 France
pacard@univ-paris12.fr

Park, Sungho
School of Mathematics
Georgia Institute of Technology
686 Cherry Street
Atlanta, GA 30332-0160
subuti00@yahoo.co.kr

Parwani, Kamlesh
Department of Mathematics
Northwestern University
2033 Sheridan Evanston, IL 60208
forty2@math.northwestern.edu

Pasko, Brian
Department of Mathematics
Kansas State University
138 Cardwell Hall
Manhattan, KS 66505-2602
pasko@ksu.edu

Pedit, Franz
GANG
University of Massachusetts, Amherst
Amherst, MA 01003
franz@gang.umass.edu

Pérez, Joaquín
Departamento de Geometria y Topologia
Universidad de Granada
Fuentenueva, S/N
Granada, Spain
jperez@ugr.es

Pirola, Gian
Dip. di Matematica,
Universita' di Pavia
via Ferrata, 1
Pavia, I-27100 Italy
pirola@dimat.unipv.it

Pollack, Daniel
Department of Mathematics
University of Washington
Box 354350
Seattle, WA 98195-4350
pollack@math.washington.edu

Polthier, Konrad
Technische Universitët, Berlin
Str. des 17. Juni 136
Berlin, D-10623 Germany
polthier@zib.de

Preuss, Eike
Department of Mathematics
Technische Universitët, Berlin
Föhrer Str. 12
Berlin, D-13353 Germany
eike@sfb288.math.tu-berlin.de

Qiu, Weiyang
Department of Mathematics
Stanford University
450 Serra Mall, Building 380
Stanford, CA 94305
qiu@math.stanford.edu

Ratzkin, Jesse
Department of Mathematics
University of Utah
155 South 1400 East
Salt Lake City, UT 84112
ratzkin@math.utah.edu

Ries, Richard
Department of Mathematics
University of California, Riverside
Riverside, CA 92521
ries@math.ucr.edu

Ritoré, Manuel
Departamento de geometría
Universidad de Granada
Avda. Fuentenueva
Granada, E-18071 Spain
ritore@ugr.es

Robadey, Anne
Ecole Normale Supérieure de Montrouge
1, rue Maurice Arnoux
Montrouge, F-92120 France
robadey@clipper.ens.fr

Rodriguez, Lucio
7427 S. Waterway Drive
Miami, FL 33155-2707
lucio@impa.br

Romon, Pascal
Département de Mathématiques
Univ. de Marne-la-Vallée (Copernic)
5 bd Descartes
Champs-sur-Marne, F-77454 France
romon@math.univ-mvl.fr

Ros, Antonio
Facultad de Ciencias
Universidad de Granada
Departamento de Geometria y Topologia
Granada, E-18071 Spain
aros@ugr.es

Rosenberg, Harold
13 Faubourg Montmartre
Paris, F-75009 France
rosen@math.jussieu.fr

Rossman, Wayne
Mathematics, Faculty of Science
Kobe University
1-1 Rokkodaicho, Nadaku
Kobe, 657-8501 Japan
wayne@math.kobe-u.ac.jp

Rubinstein, Joachim
Department of Mathematics & Statistics
University of Melbourne Parkville
Melbourne, Victoria 3010 Australia
rubin@ms.unimelb.edu.au

Rudd, Matthew
Department of Mathematics
University of Utah
155 S 1400 E
Salt Lake City, UT 84112
rudd@math.utah.edu

Sa Earp, Ricardo
Departamento de Matemática
Pontificia U. Catolica do Rio de Janeiro
Rua Marquês de São Vicente, 225, Gávea
Rio De Janeiro, RJ 22453-900 Brazil
earp@mat.puc-rio.br

Santos, Walcy
Universidade Federal do Rio de Janeiro
Rua Morais e Silva, 51 - APTO 1001 - Bloco 3
Rio de Janeiro, RJ 20271-030 Brazil
deloera@math.ucdavis.edu

Schoen, Richard
Department of Mathematics
Stanford University
450 Serra Mall, Building 383-Q
Stanford, CA 94305
schoen@math.stanford.edu

Schoenfeld, Eric
Williams College
Williamstown, MA 01267
03ems@williams.edu

Schwartz, Fernando
Department of Mathematics
Cornell University
310 Malott Hall
Ithaca, NY 14853
fschwart@math.cornell.edu

Seo, Keomkyo
School of Mathematical Sciences
Seoul National University
San 56-1, Shilim-dong, Kwanak-gu
Seoul, 151-742 South Korea
heretin@hanmail.net

Smale, Nathan
Department of Mathematics
University of Utah
155 S 1400 E
Salt Lake City, UT 84112
smale@math.utah.edu

Soret, Marc
17 rue Jobbe-Duval
Paris, F-75015 France
marc@gargan.math.univ-tours.fr

Spruck, Joel
Department of Mathematics
Johns Hopkins University
3400 N. Charles St., 404 Krieger Hall
Baltimore, MD 21218
js@math.jhu.edu

Steiner, Jean
Department of Mathematics
University of California, San Diego
Mail Code 0112, 9500 Gilman Dr.
La Jolla, CA 92093-0112
steiner@math.ucsd.edu

Sullivan, John
Department of Mathematics
TU Berlin, MA 3-2
Str. Des 17. Juni 136
Berlin, 10623 Germany
Sullivan@Math.TU-Berlin.de

Szopos, Marcela
Laboratoire Jacques-Louis Lions
Université de Paris 6
(Pierre et Marie Curie)
BC 187, 4 Place Jussieu
Paris, Cedex 05 F-75252 France
szopos@dptmaths.ens-cachan.fr

Taylor, Yuka
Department of Mathematics
Rutgers University, New Brunswick
110 Frelinghuysen Road
Piscataway, NJ 08854
yukatylr@math.rutgers.edu

Tenenblat, Keti
Departamento de Matematica
Universidade de Brasília
Brasília, DF 70.910-900 Brazil
keti@mat.unb.br

Topping, Peter
Department of Mathematics
University of Warwick
Coventry, CV4 7AL U. K.
topping@maths.warwick.ac.uk

Traizet, Martin
Université de Tours
Tours, F-37200 France
martin@gargan.math.univ-tours.fr

Tschirschwitz, Boris
Department of Mathematics
University of British Columbia
Vancouver, BC V6T 1Z2 Canada
boris@math.ubc.ca

Umehara, Masaaki
Hiroshima University
1-3-1 Kagamiyama
Higashi-Hiroshima,
Hiroshima 739-8526 Japan
umehara@math.sci.hiroshima-u.ac.jp

Walsh, Genevieve
Department of Mathematics
University of California, Davis
1 Shields Avenue
Davis, CA 95616
gwalsh@math.ucdavis.edu

Wang, Sung
Department of Mathematics
Pohang Inst. of Sci & Tech
Pohang, South Korea
wang@math.mcgill.ca

Weber, Matthias
Indiana University
107 S. Indiana Ave., Rawles Hall
Bloomington, IN 47405
matweber@indiana.edu

Wellin, Paul
Wolfram Research, Inc.
13361 Sheep Hallow Creek
Chico, CA 95973
wellin@wolfram.com

Wente, Henry
Department of Mathematics
University of Toledo
2801 West Bancroft Street
Toledo, OH 43606-3390
hwente@math.utoledo.edu

White, Brian
Department of Mathematics
Stanford University
450 Serra Mall, Building 380
Stanford, CA 94305
white@math.stanford.edu

Williams, Anca
Department of Mathematics & Statistics
Portland State University
PO Box 751
Portland, OR 97207-0751
ancajim@onemain.com

Wilson, Robin
Department of Mathematics
University of California, Davis
1 Shields Avenue
Davis, CA 95616
rtwilson@ucdavis.edu

Wolf, Michael
Department of Mathematics
Rice University
PO Box 1892
Houston, TX 77251-1892
mwolf@math.rice.edu

Wong, Chi
Department of Mathematics
University of Hong Kong
Pokfulam Road , Hong Kong, SAR
cwwong@submaths.hku.hk

Xu, Dezhen
Department of Mathematics
SUNY, Stony Brook
Stony Brook, NY 11794-3651
dezhen@math.sunysb.edu

Yamada, Kotaro
Graduate School of Mathematics
Kyushu University
6-10-1 Hakozaki, Higashi-ku
Fukuoka, 812-8581 Japan
kotaro@math.kyushu-u.ac.jp

Yamada, Sumio
Department of Mathematics
University of Alabama at Birmingham
1300 Univ. Blvd.
Birmingham, AL 35294
yamada@math.uab.edu

Yan, Yu
Department of Mathematics
Stanford University
450 Serra Mall, Building 380
Stanford, CA 94305
yyan@math.stanford.edu

Yang, Seong-Deog
Department of Mathematics
Brown University
Providence, RI 02912
sdyang@math.brown.edu

Yaskin, Vladyslav
Department of Mathematics
University of Missouri, Columbia
Mathematical Sciences Bldg.
Columbia, MO 65211
yaskinv@math.missouri.edu

Zhang, Sirong
Department of Mathematics
Johns Hopkins University
3400 N. Charles St., 404 Krieger Hall
Baltimore, MD 21218
szhang@math.jhu.edu