Proceedings of 12th Gökova
Geometry-Topology Conference
pp. 1 – 48

Published online at
GokovaGT.org

Seven short stories on blowups and resolutions

Herwig Hauser

To Raoul Bott – with great respect.

"At that time, blowups were the poor man's tool to resolve singularities." This phrase of the late 21st century mathematician J.H.Φ. Leicht could become correct. In our days, however, blowups are still the main device for resolution purposes (cf. fig. 1).

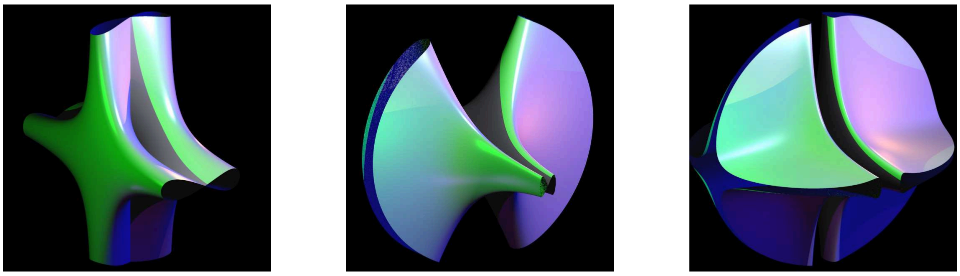


Figure 1: Resolution of the surface Helix: $x^2 - x^4 = y^2z^2$ by two blowups.

These notes shall give an informal introduction to the subject. They are complemented by the discussion of many special and less known features of blowups.

The lectures adress to students and geometers who are not experts in the field, but who need to use blowups occasionally or who just want to have a good comprehension of them. References are scattered in the literature and mostly concentrate on only part of the story. This text is neither complete, but hints at least at the variety of properties, results and techniques which are related to blowups and which make them so attractive. Actually, it may serve as the starting point to write a comprehensive treatise on blowups (which should in particular include the solutions to all exercises). The obvious objection from algebraic geometers to such a project will be that blowups are too simple to deserve a separate treatment. The many open and intricate questions listed in these notes may serve as a reply to this reproach.

The material stems from lectures held by the author at the Mathematical Sciences Research Institute (MSRI) at Berkeley in April and May 2004 and during the Conference on Geometry and Topology at Gökova, Turkey, in June 2005. Both classes were taught

Supported in part by FWF-project P 15551, and the bilateral exchange program Acciones Integradas.

ex tempore in order to react better to the doubts and questions of the audience. And the nicest compliment was the remark of a solid algebraic geometer saying "I thought to know already all about blowups, but I am no longer sure about this".

We are very grateful to Frank Sottile and Selman Akbulut for the kind invitations to Berkeley and Gökova, and to all the people from MSRI for their warm hospitality. The students of the classes interfered many times with interesting remarks and inquiries – this helped a lot to shape the material. As a pop star would formulate it: *"You are the best audience I ever had".*

Here is a list of reading literature. Eisenbud-Harris' book on the geometry of schemes is probably the source which gives the most inspiration on the topic [EiH]. Several phenomena were taken from there. Another excellent reading is the beginning of Hironaka's Annals article [Hi 1]. There are almost no proofs, but many of the main properties appear with precise details and in all generality. Abhyankar's papers can be very stimulating if one really succeeds to read the information hidden between the lines. In Hartshorne's book [Hs], blowups have a prominent place, but the discussion lacks the concrete calculations and the various aspects which can arise in examples. Let us also mention the articles of Encinas-Villamayor [Vi 1, Vi 2, EV 1, EV 2] and Bierstone-Milman [BM 1 to BM 6], the survey of Lipman [Lp 3], the notes of Cutkosky, Matsuki and Kollár [Cu 7, Ma, Ko]. Above all, we recommend reading Zariski, especially his paper on the resolution of threefolds [Za 4] and therein the section on resolution of surfaces.

The proof of resolution in characteristic zero which we develop in big lines towards the end of these notes stem's from the author's collaboration with Santiago Encinas [EH]. The enclosed pictures were produced by Sebastian Gann with the ray-tracing program POV-Ray.

1. Introduction and examples

By a *resolution* of an algebraic set (variety, scheme) X with singularities we understand a map $\pi : X' \to X$ which represents X as the image of a manifold X', i.e., which *parametrizes* X. It is by no means clear how to find such a parametrization.

Example 1: The first example of a resolution of a surface is the contraction of the cylinder X' given by $x^2 + y^2 = 1$ in $\mathbb{A}^3$ to the (double) cone $X : x^2 + y^2 - z^2 = 0$. The map is induced by $\pi : \mathbb{A}^3 \to \mathbb{A}^3$, $(x, y, z) \to (xz, yz, z)$. It collapses the xy-plane $z = 0$ to the origin of $\mathbb{A}^3$ so that the cone appears as the image of the cylinder contracted along a circle (cf. fig. 2). Despite its simplicity, the example is quite instructive, because it leads to the problem of reconstructing by a natural and general procedure the cylinder X' and the map π from the knowledge of X.

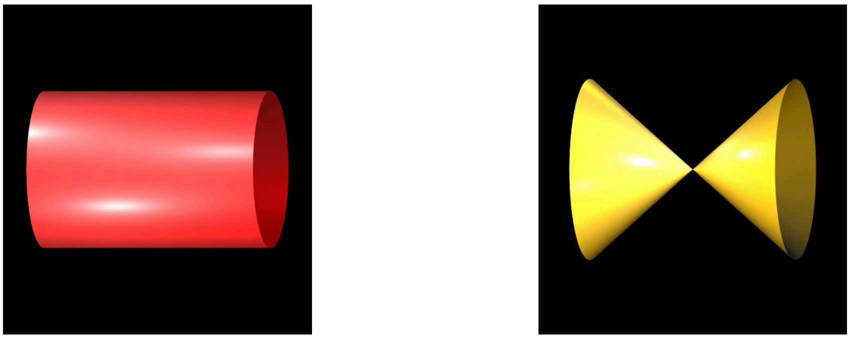


Figure 2: Cylinder and cone.

Example 1^a: To answer this question, we consider a yet simpler example, the crossing X of the two diagonals in $\mathbb{A}^2$, given by $x^2 - y^2 = 0$. The manifold X' should of course consist of two separate lines. Therefore, to separate the two branches of X, we lift the two diagonals to $\mathbb{A}^2 \times \mathbb{P}^1$ by associating to each point (x, y) the height z given by the slope of the line through (x, y) and $(0, 0)$, more precisely,

$$\sigma : \mathbb{A}^2 \setminus 0 \to \mathbb{A}^2 \times \mathbb{P}^1,$$
$$(x, y) \to ((x, y), (x : y)).$$

This map is defined outside the origin of $\mathbb{A}^2$, not only on X. Geometrically it is clear that the image of $X \setminus 0$ under σ consists of two "parallel" lines (each deprived of one point). The Zariski-closure X' of $\sigma(X \setminus 0)$ in $\mathbb{A}^2 \times \mathbb{P}^1$ consists of the entire lines and is thus a submanifold of $\mathbb{A}^2 \times \mathbb{P}^1$. The map $X' \to X$ is the restriction of the projection $\mathbb{A}^2 \times \mathbb{P}^1 \to \mathbb{A}^2$ onto the first two components. We leave it to the reader to compute the affine or projective equations of X' in $\mathbb{A}^2 \times \mathbb{P}^1$ as well as the chart expression of $\pi : X' \to X$.

The construction works fine because all points on one of the two lines of X have the same slope. What happens if this is not the case?

Example 1^b: Consider the variation of the preceding example, taking X defined in $\mathbb{A}^2$ by $(y - x^2)(y - x - x^3) = 0$, a parabola and a cubic, intersecting over the reals only at the origin. The two components meet transversally at 0. There are now two options how to separate the branches.

(i) Either we associate to points $(x, y) \in \mathbb{A}^2$ again the triple $((x, y), (x : y)) \in \mathbb{A}^2 \times \mathbb{P}^1$. Notice that for points (x, y) on the curve X, the projective point $(x : y)$ is the *secant line* through (x, y) and $(0, 0)$.

A computation shows that the Zariski-closure X' of the image of $X \setminus 0$ under the above map σ consists of two disjoint regular curves, and it comes with a surjective map $\pi : X' \to X$. This construction is called the *blowup of* $\mathbb{A}^2$ *with center the origin* (or the ideal (x, y) of

$K[x, y]$), and X' is the *strict* or *proper transform* of X. It's not a bad exercise to carry out this computation.

(ii) The second option is to associate to a point (x, y) on the curve X the triple $(x, y, t) \in \mathbb{A}^2 \times \mathbb{P}^1$, where t is now the *tangent line* to X in (x, y). This construction, of course, works for the moment only at (the regular) points of the curve, and not on whole $\mathbb{A}^2$. The resulting transformation of X is known as *Nash modification* (it corresponds to blowing up the jacobian ideal of the curve given by the partial derivatives of the defining equation, and is thus defined a posteriori again on whole $\mathbb{A}^2$). Though geometrically attractive, it has not been used and exploited systematically beyond curve and surface singularities.

Nota bene: Blowups whose centers are points or regular subvarieties of the ambient scheme represent the main modification of varieties to resolve their singularities. Their impact on the singularities, however, is very modest. They only succeed to improve them "picolissimo". Actually, it is very hard in general situations to show that an improvement actually occurs.

Therefore it would be nice to find substitutes which are more powerful and more refined, and Nash modifications are one possible option. You should always keep in mind to look out for possible new modifications to replace classical blowups.

After all, it can be shown by Hironaka's theorem that there does exist (at least in characteristic 0) a non reduced ideal structure on the singular locus of an arbitrarily difficult variety (i.e., an ideal with support the singular locus) such that blowing up the respective ideal resolves the singularities of X ad hoc in one capital stroke, and moreover preserves the regularity of the ambient scheme. Nobody has the slightest idea how this non-reduced structure should look like. Aside of non-significant examples nothing seems to be known. *End of N.b.*

Example 1^c: Consider the curve X defined in $\mathbb{A}^2$ by $(y-x^2)(y-x^k)$ with $k \geq 3$. The two components meet now tangentially at the origin. Taking the blowup with center $(0, 0)$, i.e., corresponding to secant lines, yields (in the relevant chart) the equation $(y-x)(y-x^{k-1})$. The two components now meet less tangentially, but this, obviously, has to be made precise and captured by an intrinsic measure of tangency.

Here is a first (and somewhat) preliminary definition of the blowup of affine space $\mathbb{A}^n$ with center an ideal P of $K[x_1, \ldots, x_n]$ (said differently, with center Z the subvariety $V(P)$ of $\mathbb{A}^n$): Choose generators $g_1, \ldots, g_k$ of P and consider the map

$$\mathbb{A}^n \setminus Z \to \mathbb{A}^n \times \mathbb{P}^{k-1},$$
$$x \to (x, g_1(x) : \ldots : g_k(x)).$$

This is a well defined injective morphism. The image is the graph $\Gamma(g)$ of $g : \mathbb{A}^n \setminus Z \to \mathbb{P}^{k-1}$, $x \to (g_1(x) : \ldots : g_k(x))$. The Zariski-closure $\widetilde{\mathbb{A}}^n$ of $\Gamma(g)$ in $\mathbb{A}^n \times \mathbb{P}^{k-1}$ is the *blowup of* $\mathbb{A}^n$ *with center* Z. It does not depend on the choice of the generators g_i up to isomorphism (show this). The restriction of the projection on the first factor $\pi : \widetilde{\mathbb{A}}^n \to \mathbb{A}^n$

is the associated *blowup map.* It is a birational morphism which is an isomorphism over the complement of Z, say $\pi : \widetilde{\mathbb{A}}^n \setminus \pi^{-1}(Z) \cong \mathbb{A}^n \setminus Z$. The subvariety $\pi^{-1}(Z)$ of $\widetilde{\mathbb{A}}^n$ which is contracted under π to Z is a hypersurface (show also this) and called the *exceptional divisor* E of the blowup.

From the definition, the algebraic equations of $\widetilde{\mathbb{A}}^n$ in $\mathbb{A}^n \times \mathbb{P}^{k-1}$ as well as the affine chart expressions of π can be directly deduced. The projective equations for $\widetilde{\mathbb{A}}^n$ in $\mathbb{A}^n \times \mathbb{P}^{k-1}$ are $u_i g_j(x) - u_j g_i(x) = 0$ for $1 \leq i, j \leq k$ and coordinates $u_1, \dots, u_k$ on $\mathbb{P}^{k-1}$.

For X a subvariety of $\mathbb{A}^n$ containing Z we obtain in a similar fashion the Zariski-closure $\widetilde{X}$ of the image of

$$\begin{aligned} X \setminus Z &\to X \times \mathbb{P}^{k-1}, \\ x &\to (x, g_1(x) : \ldots : g_k(x)), \end{aligned}$$

called the blowup of X with center Z. Again, $\pi : \widetilde{X} \to X$ is the associated blowup map.

It is not difficult to show that $\widetilde{\mathbb{A}}^n$ is regular if P is the reduced ideal of a regular subvariety Z of $\mathbb{A}^n$. For $Z = 0$ the origin in the real plane $\mathbb{R}^2$, we obtain (cum grano salis) for $\widetilde{\mathbb{A}}^2$ the Möbiusband in $\mathbb{R}^3$ (cf. fig. 3).

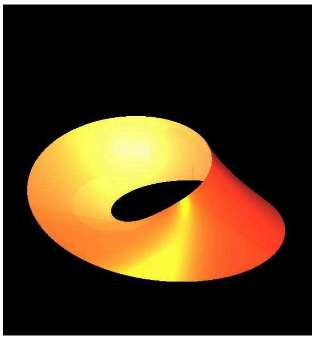


Figure 3: The Möbiusband, blowup of $\mathbb{A}^2$ with center a point.

Exercises. Writing down the above introductory examples with all details is almost mandatory for being able to enjoy the remaining material. There are also more complicated examples where everybody can try out its resolution instinct.

Exercise 2: To warm up, we start with some computations of blowups of affine space.

Exercise 2^a: Determine the affine chart expressions for the blowup map $\pi : \widetilde{\mathbb{A}}^n \to \mathbb{A}^n$ for $n = 2$ and $Z = 0$, $n = 3$ and $Z = 0$, $n = 3$ and $Z = V(x, y) = z$-axis. Show that in the last case, the blowup map π is the cartesian product of the blowup map in the first case with the identity map on the z-axis.

Exercise 2^b: Take various curves in $\mathbb{A}^2$ and $\mathbb{A}^3$ (e.g., $y^2 = x^3 - x$, $y^2 = x^2 - x^3$, $y^2 = x^3$ or the parametrized curves (t^2, t^3, t^4) or (t^3, t^4, t^5)) and compute their pullbacks under the previous blowups, both projectively and in the affine charts. Distinguish the locations of the intersections of the "pullbacks" of $xy = 0$, $x^2 - y^2 = 0$, $xy(x^2 - y^2) = 0$ with the exceptional divisor E (cf. fig. 4).

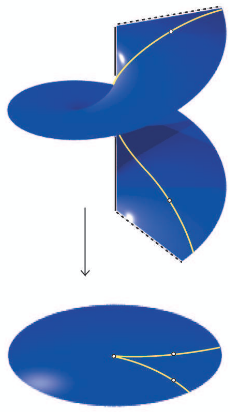


Figure 4: Blowup of $\mathbb{A}^2$ with center a point and transform of cuspidal curve.

Exercise 2^c: Show that the ideals (x, y), $(x, y)^2$ and more generally $(x, y)^k$ define the same blowup of $\mathbb{A}^2$ when taken as center.

Exercise 2^d: If the center is a hypersurface, the resulting blowup map is an isomorphism.

Exercise 2^e: Compute the point blowup of the ideal (x, y^2) in $\mathbb{A}^2$. Show that $\widetilde{\mathbb{A}}^2$ is singular. Resolve its singularities by a further blowup.

Exercise 2^f: Compute the blowup of the ideal $(x, y^2)(x, y)$ in $\mathbb{A}^2$. Show that $\widetilde{\mathbb{A}}^2$ is regular.

Exercise 2^g: Let $\widetilde{\mathbb{A}}^2$ be the blowup of $\mathbb{A}^2$ with center the origin. Let $a \in \widetilde{\mathbb{A}}^2$ be the point corresponding to the direction of the line $y = 0$ in $\mathbb{A}^2$ (make this precise). Blow it up. You get a composition of two blowups, say $\tau : \widehat{\mathbb{A}}^2 \to \mathbb{A}^2$. Show that it is the blowup of $\mathbb{A}^2$ with center an ideal (not reduced) supported by the origin. Determine this ideal. Show that it is not unique.

Exercise 2^h: Compute the blowup of $\mathbb{A}^3$ with center the circle $x^2 + y^2 = 1$, $z = 0$. Make sure to compute all affine charts. Show that $\tilde{\mathbb{A}}^3$ is regular.

Exercise 2^i: Show that the exceptional divisor E of a blowup is indeed a hypersurface. Compute its equation in various circumstances.

Exercise 2^j: Blow up the "arithmetic line" $\mathbb{A}^1_{\mathbb{Z}} = \mathrm{Spec}(\mathbb{Z}[x])$ in the ideal (x, p), $p \in \mathbb{Z}$ a prime. What could be the "affine" equations of $\widetilde{\mathbb{A}}^1_{\mathbb{Z}}$?

Exercise 3: We next try to resolve some singularities: The double-cone $x^2+y^2=z^2$ in $\mathbb{A}^3$ should not pose too many problems, as well as the Dingdong surface $x^2+y^2=z^2(1-z)$ (figure out first how this creature looks like?). More interesting is the Whitney-umbrella $x^2-y^2z=0$ in $\mathbb{A}^3$. The singular locus is the z-axis. Again, it is worth to try first a simpler example, namely

Exercise 3^a: The cylinder over the cusp x^2-y^3 in $\mathbb{A}^3$. The equation is $x^2-y^3=0$, but considered in three variables. What is the correct blowup to be applied? The question applies to any varieties which are cartesian products (cf. fig. 5).

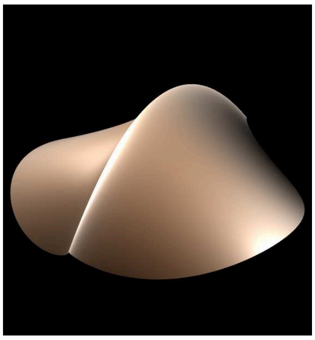


Figure 5: Plop, the cartesian product of a cusp with the parabola.

Exercise 3^b: A somewhat more delicate example is the Limão $x^2=y^3z^3$ in $\mathbb{A}^3$ (cf. fig. 6), which appears as a modification of the equation of the Whitney-umbrella $x^2=y^2z$ (cf. fig. 7). The singular locus consists of the y- and the z-axis; it has therefore a singular point at the origin. What is the correct choice of center of blowup? Can you preserve by your blowups the symmetry between y and z.

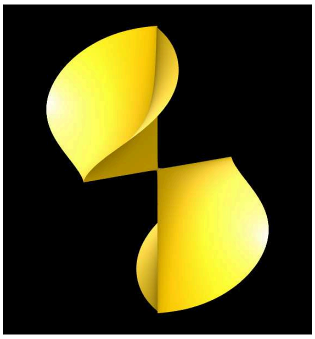


Figure 6: Limão, of equation $x^2-y^3z^3=0$.

Exercise 4: Let f be any polynomial in three variables, and denote by o its order of vanishing at 0, i.e., the order of the Taylor expansion at 0. Blow up the origin in $\mathbb{A}^3$ and consider the strict transform f' of f at a point a' of $\widetilde{\mathbb{A}}^3$ above a (i.e., take the pullback

f^* of f and factor from it the maximal power of the equation defining the exceptional divisor). Determine the cases where the order o' of f' at a' has remained constant.

2. Choosing the centers of blowup

We now turn to the problem of selecting for each singular variety a suitable center along which the ambient space and the variety shall be blown up. Recall that the blowup map is an isomorphism over the complement of the center, whereas it contracts the exceptional divisor, which is a hypersurface, to the center (which could be rather small). Therefore, our variety will be modified only along the center. And indeed, at its regular points we have no reason to change it, they cannot be improved any further. So will agree to choose our centers *inside the singular locus* of the variety.

This choice, however, is not a must, and there are several papers where the centers are chosen to stick out of the singular locus and even out of the variety. The defect is due to a lack of suitable stratifications of the variety, namely ones whose smallest stratum is regular and could thus be chosen as center. And in fact, allowing singular centers is quite a delicate job, see [Ha 2].

For the moment we will accept solely centers inside the singular locus of the variety. This requirement is known as the *economy of the resolution process*: Only points are modified where it is really necessary, because they are singular. Of course, in the resolution process, each singular point has to belong once in a while to a center of blowup, otherwise it would remain singular forever.

Let us then observe whether our singularities improve under the chosen blowups.

We shall concentrate on surfaces, because they represent a nice playground to develop some of the most prominent ideas. In contrast, the reader should know that dimension three is *the* critical dimension for bad phenomena to happen, especially but not exclusively in positive characteristic. Quasi all "counterexamples" to properties one could hope for are formulated for three-folds.

Let us start with isolated singularities. In this case there is no ambiguity how to choose the center. It should coincide with the finitely many singular points, so it suffices to consider one of these.

Example 4: The double-cone X given by $x^2 + y^2 - z^2 = 0$ in $\mathbb{A}^3$. The only singularity sits in the origin. Blowing it up gives three charts, with a symmetry between the x- and y-chart. In the x-chart we obtain $x^2(1 + y^2 - z^2) = 0$, in the z-chart $z^2(x^2 + y^2 - 1) = 0$. Again, we will omit the exceptional monomial factors x^2 and z^2, so that the interesting equations are $1 + y^2 - z^2 = 0$ and $x^2 + y^2 - 1 = 0$. All three charts of the blowup X' are regular.

The exceptional divisor E of $\pi : \tilde{\mathbb{A}}^3 \to \mathbb{A}3$ is the two-dimensional projective space $\mathbb{P}^2$, and X' meets E entirely in the z-chart (so we can discard the other charts). The intersection is the circle $x^2 + y^2 - 1 = 0$, which shows that X' is indeed the cylinder (because outside

0 we did not alter X). We recover our example from the beginning of the first lecture. Check that the blowup map $\pi : X' \to X$ is precisely the contraction described there.

Example 4^a: Let us now change coordinates. As a first variation of the preceding example, "rotate" the double-cone by taking for X the equation $x^2 - yz = 0$, with a symmetry between y and z. The equation is obtained from the previous example by substituting there $y - z$ by y and $y + z$ by $-z$. The equations for X' are now in the x-chart $1 - yz = 0$ and in the y-chart $x^2 - z = 0$. Therefore, the intersection of X' with E is no longer contained in one of the three affine charts of E. This becomes relevant when computing local invariants of X' at points of E. The fewer the charts are that have to be considered, the shorter are the computations.

As a general observation, the choice of suitable coordinates (suitable with respect to different intentions as e.g. achieving small complexity or defining coordinate independent data) is a very subtle point and marks the entire literature on the subject.

Example 4^b: What happens if we do not take the origin of the cone as center, but a whole line through the origin? For simplicity we take X given by $x^2 - yz = 0$. There are two types of lines through 0, those contained in X and those which are transversal to X (in an heuristic sense, meaning simply that they are not contained in X). Let first the center $Z \subseteq X$ be the z-axis of ideal (x, y). The transform X' of X has equation (replace (x, y) by (x, yx) respectively (xy, y) and leave z untouched) $x - yz = 0$, respectively $x^2 y - z = 0$. It is hence non-singular, but the global geometry seems to be somewhat twisted. We leave it to the ambitious (or curious) reader to figure out whether X' is again the cylinder.

(Glueing affine charts may be tedious. We will encounter this task at many more occasions, and a certain routine will be helpful.)

Let us now take as center $Z \not\subseteq X$ the "transversal line" of ideal (y, z), i.e., the x-axis. By symmetry, we only have to consider the y-chart. The transform X' is given there by $x^2 - y^2 z = 0$, which is the Whitney-umbrella mentioned at the end of the first lecture. It has a whole line as singular locus, and the origin is by all means the worst singularity on this surface. It would be difficult to qualify X' as being less singular than the double-cone: Our blowup did not improve the singularities – it made them *worse*.

This comes a bit as a surprise, but cannot be avoided: If the center leaves the variety, or if it remains inside but is too small, the singularities may become more involved (further examples still to come). Moreover, in our example, the singularities are no longer isolated. We will see later that for instance the normality of surfaces need not persist under point blowups. This has a simple meta-mathematic interpretation: Normal surface singularities are isolated, but the variety has really to squeeze around these points in an intricate way in order not to create singular curves. Now, when considering blowups of normal surfaces, we think of all secants going through regular points and the (isolated) singular point. And there may appear in the limit – when the regular points come close to the singular one – very complicated configurations of secants.

In particular, as a side-product, we see that the blowup of X in a center Z not contained in X does not coincide with its blowup with center the intersection $Z \cap X$ (neither for the set-theoretic nor for the scheme-theoretic intersection $Z \cap X$). There is no way to expect commutativity of blowups with respect to restricting the center to X.

Exercise 1: Prove this accurately. What happens if you take as center the scheme-theoretic intersection $Z \cap X$?

Fortunately, many other and very practical commutativity properties do hold, we will list and prove them in detail. The first one appears in the next example.

Example 5: Let us take for X two transversal planes in $\mathbb{A}^3$, say of equation $x^2 - y^2 = 0$, considered in three variables. So X is the cartesian product of (or the cylinder over) the union C of two transversal lines in $\mathbb{A}^2$ with regular factor the z-axis, say $X = C \times \mathbb{A}^1$. We have two options:

(i) Blow up the origin in $\mathbb{A}^2$, get $\pi : \widetilde{\mathbb{A}}^2 \to \mathbb{A}^2$ and then take $\pi \times \mathrm{id}_{\mathbb{A}^1} : \widetilde{\mathbb{A}}^2 \times \mathbb{A}^1 \to \mathbb{A}^2 \times \mathbb{A}^1 = \mathbb{A}^3$. This is certainly an honest attempt, because X is a cartesian product, a fact which should be reflected by the transformation we choose.

(ii) The second option is to blow up $\mathbb{A}^3$ along the singular locus of X, i.e., with center the z-axis. The ideal is (x, y).

It is easily checked (especially, if you have done some of the earlier exercises), that both options are valid and resolve X, yielding for X' two separate (parallel) planes in $\mathbb{A}^2 \times \mathbb{P}^1 \times \mathbb{A}^1$, respectively $\mathbb{A}^3 \times \mathbb{P}^1$. And as you probably suspected by what was said before, the two options coincide (not hard to be verified). Perfect! We state explicitly our first functorial property:

Blowups commute with taking cartesian products with regular factors.

More explicitly: The blowup of $\mathbb{A}^n$ along a coordinate subspace Z equals the cartesian product of the point-blowup in a transversal subspace V of $\mathbb{A}^n$ (of complementary dimension) with the identity on Z.

Exercise 2: Prove this in all generality without referring to affine charts but by recalling that the blowup is the Zariski-closure of a graph.

Exercise 3^a: This sounds nice, so will immediately test it in a concrete situation. Compute the blowup of $\mathbb{A}^3$ with center the circle Z of ideal $(x^2 + y^2 - 1, z)$. You should at least remember here that this coincides with an earlier exercise. If you don't remember, you can almost be sure that you did not do this exercise. What are the affine chart expressions of the blowup map? Next, define a surface whose singular locus is precisely this circle. What is its transform under the blowup?

Exercise 3^b: Blow up the plain circle Z from before inside the surface $Z \times \mathbb{A}^2$.

Exercise 4: Let us take for X the cylinder over the cusp, given by the equation $x^2 - y^3 = 0$ in $\mathbb{A}^3$. You should have no problems in defining the center and the induced resolution of

X by one blowup. But how does X' intersect the exceptional divisor $E \cong \mathbb{P}^1 \times \mathbb{A}^1$? Are you pleased how they meet? If yes, go back to the blowup of the curve $x^2 - y^3$ in $\mathbb{A}^2$ and contemplate your satisfaction. If not, what do you suggest to remedy your discomfort?

To relax briefly and to prepare the field for new exercises, another definition of the blowup of affine space $\mathbb{A}^n$ in an ideal $P = I_Z = (g_1, \ldots, g_k)$ of $K[x_1, \ldots, x_n] = K[x]$ will be given. This time, we indicate the affine coordinate rings of $\widetilde{\mathbb{A}}^n$ as subrings of the quotient field $K(x)$. The i-th one is given as

$$K[x_1, \ldots, x_n, g_1/g_i, \ldots, g_k/g_i].$$

Exercise 5: Determine how these rings have to be glued in order to define $\widetilde{\mathbb{A}}^n$.

Exercise 6: Computations are fun, so show with this definition that $P = (x, y^2)$ in $\mathbb{A}^2$ and $P = (x, yz)$ in $\mathbb{A}^3$ give singular blowups $\widetilde{\mathbb{A}}^2$ and $\widetilde{\mathbb{A}}^3$, whereas $P = (x, y^2)(x, y)$ in $\mathbb{A}^2$ and $P = (x, yz)(x, y)(x, z)$ in $\mathbb{A}^3$ give regular ones. You may consult [Ha 2].

Exercise 7: This is a nice little exercise$^+$ in discrete or toric geometry. Determine the monomial ideals in $K[x_1, \ldots, x_n]$ which produce a regular blowup of affine space.

Exercise 8: This is a possibly nasty exercise^{++} in discrete geometry. Determine for monomial ideals in $K[x_1, \ldots, x_n]$ a natural and simple multiplication procedure by other monomial ideals (with smaller exponents) such that the entire product (which may have many many generators) produces a regular blowup of affine space (not too hard for $n = 2$ or $n = 3$).

We are back to singularities.

Exercise 9: Here comes the first serious example, the Whitney-umbrella $x^2 - y^2 z = 0$ in $\mathbb{A}^3$. Though easy, it shows many of the decisive features, so we will study it with patience. You are invited to carry out all computations in detail.

This surface has non-isolated singularities along the z-axis. Locally in the euclidean topology, at points on the z-axis but off the origin, it looks like two transversal planes (this is not the case not in the Zariski-topology), which reminds us example 5. At the origin, the surface bends around, making 0 strikingly more singular than the other points on the z-axis.

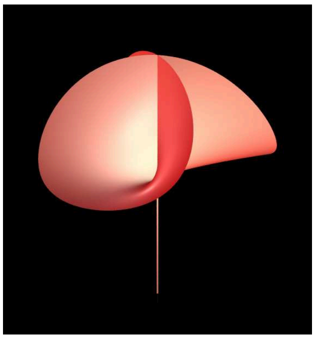


Figure 7: The Whitney-umbrella $x^2 = y^2 z$.

According to the philosophy that the worst singularities should be taken care of first, we blow up the origin with reduced ideal (x, y, z) in $\mathbb{A}^3$. We get three affine charts and a new surface X' of respective equations:

x-chart: $1 - xy^2z = 0$, no intersection with exceptional divisor E.

y-chart: $x^2 - yz = 0$, our old acquaintance the double-cone, which, for any point of view, should be simpler than the Whitney-umbrella. Note that the intersection of X' with E is the z-axis (with multiplicity 2).

z-chart: $x^2 - y^2z = 0$, and – surprise – the same singularity we had before blowing up pops up again at the origin of this chart. The singularity has survived our attack without damage (cf. fig. 8). The intersection of X' with E is now the y-axis (again with multiplicity 2).

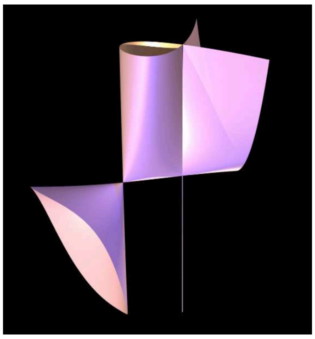


Figure 8: The Pirat, point blowup of the Whitney-umbrella.

Tautology : If the singularity remains the same, there is no way to declare it less singular than the original one.

Conclusion : Our choice of center was not appropriate (if we believe that resolutions exist). We should have probably better taken a larger center.

Parenthesis : We have cheated here a little bit. Even though the singularity was preserved in the z-chart, the situation has improved in a certain (and not yet revealed) sense. And, actually, all existing resolution algorithms and their implementations choose for the Whitney-umbrella the origin as the first center of blowup. The reason for this “misbehaviour” of the algorithms will be explained later.

Exercise 10: In the y-chart we have an isolated singularity at the origin (the one of the double-cone), in the z-chart we have a non-isolated singularity (the one of the Whitney-umbrella). Explain this strange occurence. Think again of the configuration of secant lines along the z-axis of X. How do the three charts patch?

Exercise 11: (mandatory) Show that blowing up the z-axis in $\mathbb{A}^3$ resolves the Whitney-umbrella in one step. Could this be expected geometrically before doing the calculation?

Exercise 12^a: Resolve Himmel & Hölle $x^2 - y^2z^2 = 0$ (cf. fig. 9).

Exercise 12^b: Resolve Limão $x^2 - y^3z^3 = 0$.

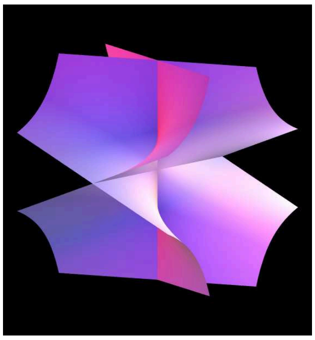


Figure 9: Himmel & Hölle, of equation $x^2 - y^2z^2 = 0$.

3. Transforms of ideals

Now that we have got an approximate idea how blowups are defined and how they may affect singularities, let us see more systematically how various types of objects transform under blowup. These objects can be polynomials or hypersurfaces, ideals or subvarieties, parametrizations of curves and surfaces, jacobian ideals of polynomials or singular loci of varieties, sum and product decompositions of ideals, coefficient ideals, coordinate systems, automorphisms, group actions, local flags, local invariants, vector fields, differential forms, etc.

Given is a blowup $\pi : W' \to W$ with center $Z \subset W$ of ideal P in $\mathcal{O}_W$ and exceptional divisor $E = \pi^{-1}Z$. We shall mostly assume that W is equal to affine space $\mathbb{A}^n$, so that P is an ideal in $K[x_1, \ldots, x_n]$, with $Z = V(P)$ regular. We wish to *lift* a given object $\mathcal{B}$ in W to an object $\mathcal{B}'$ in W' in a natural and significant way.

There are two options: Either we take the usual pullback $\mathcal{B}^*$ of $\mathcal{B}$ under π. Or, more subtly, we restrict $\mathcal{B}$ to $W \setminus Z$, get an object $\mathcal{C}$ on $W \setminus Z$, apply to the restriction $\mathcal{C}$ the inverse φ^{-1} of the isomorphism $\varphi : W' \setminus E \to W \setminus Z$ induced by π and get an object $\mathcal{C}'$ on $W' \setminus E$, which we can then try to extend to an object $\mathcal{B}'$ defined on whole W'. If the original object is defined locally on W at a point of Z, the lifted object may only exist locally on W' at certain points of E.

When realizing this procedure algebraically, we shall fix for convenience (local or affine) coordinates $x_1, \dots, x_n$ on W, and assume that Z is defined by the ideal $P = (x_1, \dots, x_k)$ for some $k \leq n$. Moreover, we will place ourselves w.l.o.g. in the k-th affine chart of W' so that the blowup map $W' \to W$ has chart expression $\pi : \mathbb{A}^n \to \mathbb{A}^n$ sending x_i to x_i if $i \geq k$ and to $x_i x_k$ if $i < k$. The exceptional divisor E has equation $x_k = 0$ in this chart. Of course, our setting specifies a point $a \in Z$ and a point a' above a in E by considering the respective origins of the charts. Note here that the location of a' on E depends on the choice of the coordinates $x_1, \dots, x_n$ at a.

Hint: It is much simpler to treat the chart expression of the blowup map as a map of affine space onto *itself* with the same coordinates on source and image rather than to consider two different affine spaces with different coordinate systems.

In any case, we have not specified yet what we mean by coordinates. For the moment it suffices to take any regular system of parameters of $K[x_1, \dots, x_n]$, later on we shall mostly work in the local context with regular parameters of the formal power series ring $K[[x_1, \dots, x_n]]$. In the second case, coordinate changes (i.e., K-algebra-automorphisms of $K[[x]]$) are easier to handle.

Our study of transforms begins geometrically. Take a subvariety X of W (no component of X should be contained in the center Z). How does it lift to W'? We follow the recipe described above: Consider $U = X \setminus Z$ with preimage $U' = \pi^{-1}(U) = \pi^{-1}(X \setminus Z)$ in $W' \setminus E$. Clearly, $\pi : U' \to U$ is an isomorphism. We are interested in extending U' over the exceptional divisor E. Notice here that the way how U' approaches E may be substantially different from the way how U approaches Z. By our assumption, U is Zariski-dense in X, i.e., the points of X inside Z are "limit" or "boundary" points of U.

It is then more than natural to define X' as the Zariski-closure of U' in W', i.e., to add to U' those points of E to which the points of U' tend in the limit when approaching the exceptional divisor. We call X' for the moment the *geometric transform* of X. It is contained in the *total transform* $X^* = \pi^{-1}(X)$ of X, which is just the whole inverse image of X under π and therefore contains the entire exceptional divisor E. So the irreducible components of X' will be precisely the components of X^* which are different from E. Clearly, X' is the geometrically interesting object, whereas X^* is easier to work with algebraically.

It will be useful to provide an algebraic description of X' in terms of the ideal defining X in W (the ideal of X' will be the *strict* or *proper transform* of the ideal of X, and X' will inherit this name).

So let I be the ideal defining X in W. Assume in a first instance that I is principal, i.e., that X is a hypersurface given by a polynomial equation $f = 0$. Let f^* be the pullback of f to W'. In the k-th chart, E is defined by $x_k = 0$, and f^* is again a polynomial. As E is an irreducible component of X^* (we shall always assume that Z is irreducible), f^* must factor into a power of x_k and another polynomial, say f', not divisible by x_k,

$$f^* = x_k^r \cdot f'.$$

So r is the maximal power to which x_k can be factored from f^*.

Exercise 1: Show that the exponent r of x_k does not depend on the chart. It is called the exceptional multiplicity of f^* or X^*. Determine its precise value in terms of f.

Exercise 2: Verify your formula for r in case $Z = V(x, y, z, u)$ in $\mathbb{A}^6$ and

$$f = x^{33} + y^{21}z^{13} + u^{35}v^{11}w^{11} + x^8y^8z^8u^8v^8w^8 \in K[x, y, z, u, v, w].$$

We can thus write globally $f^* = m_E^r \cdot f'$ where $m_E = 0$ is the reduced equation of E in W'. As we just stated, m_E is locally in the affine charts a monomial in one of the variables. Now it is (almost) evident that the polynomial f' defines X' in W'. Said differently, the geometric transform X' of X has equation $f' = m_E^{-r} \cdot f^* = 0$ in W'.

Exercise 3: Prove this with rigor over any ground field, including finite ones.

We call f' the *strict transform* of f. If X is not a hypersurface in W, things become more complicated. Let again I be the ideal of X in W.

Fact: The geometric transform X' of X under blowup of W with center the ideal P is defined by the ideal

$$I' = \bigcup_{s \in \mathbb{N}} m_E^{-s} \cdot (I \cap P^s)^*.$$

This formula breaks with the leisurely going style we have gotten used to. So let's have a closer look at it: The intersection $I \cap P^s$ simply means that you have to take all total transforms (i.e., pullbacks) f^* of elements f of I which vanish at least with order s along Z; then factor from each of them the s-th power of (the equation of) the exceptional divisor (you know that this is possible, because you have done exercise 1 above). Finally take the union over s. The resulting ideal defines X'. Not too difficult, isn't it.

Exercise 4: Prove the fact.

Everybody will agree that the above description of I' is not very handy for practical purposes. The curious reader may ask: "Does it suffice to consider here a generator system of I?" – and the answer will be an exercise.

Exercise 5: Let $I = (y^2 - xz, yz - x^3, z^2 - x^2y) = (f, g, h) \subset K[x, y, z]$ define the monomial curve X in $\mathbb{A}^3$ of parametrization (t^3, t^4, t^5). Blow up the origin and prove that I' is generated by the strict transforms f', g' and h' of f, g and h. What are the respective exceptional multiplicities?

Exercise 5^a: By the way, does this blowup resolve X? If not, resolve it by further blowups. Is the resolved curve transversal to the exceptional divisor? If not, achieve transversality by still further blowups.

Exercise 5^b, 5^c, 5^d: Compute the Nash modification of X, then the toric resolution and finally the normalization.

Exercise 5^e: Define directly the transform of a parametrized curve $\mathbb{A}^1 \to W$. Then compute it in case of (t^3, t^4, t^5).

Exercise 6: In exercise 5, we saw that the strict transforms of the three generators of the defining ideal suffice to describe the geometric transform of the curve. Try to prove this in all generality for the geometric transform of subvarieties X in $\mathbb{A}^n$ defined by ideals of $K[x_1, \ldots, x_n]$.

Exercise 7^a: Verify your proof in the case of the plane curve X in $\mathbb{A}^3$ defined by $x = 0$ and $y^2 - z^2 = 0$ (caution: the curve is embedded in three-space). What are the respective multiplicities?

Exercise 7^b: Verify your proof in the case of the curve X in $\mathbb{A}^3$ which is the intersection of the two surfaces $S_1 : x^2 - y^3 = 0$ and $S_2 : xy - z^3 = 0$. Be sure to understand completely how the geometric transform X' of X looks like, in particular, what are its irreducible components. Then do the same for the intersection of the geometric transforms of the two surfaces S_1 and S_2. How do S_1' and S_2' meet the exceptional divisor E? Finally, compute in all cases the respective exceptional multiplicities.

Exercise 8: Show now that your proof from exercise 6 was erroneous. Then try to fix it by modifying the original assertion on how to compute the strict transform of an ideal in terms of generators.

Here is the answer: Let I be an ideal in $K[x_1, \ldots, x_n]$, and let $f_1, \ldots, f_q$ be a (local) *Macaulay basis* of I, i.e., a generator system whose initial homogeneous forms (i.e., of minimal degree) generate the ideal of *all* initial forms of elements of I. Then the strict transform I' of I under the blowup of $\mathbb{A}^n$ in a regular center is generated by the strict transforms of $f_1, \ldots, f_q$:

$$\langle f_1, \ldots, f_q \rangle' = \langle f_1', \ldots, f_q' \rangle$$

with $f_i' = m_E^{-\mathrm{ord}_P f_i} \cdot f_i^*$. You may want to check this on your own or look it up in chapter III of Hironaka's Annals paper from 1964. There, the statement is proven in the local context for formal power series. Such generator system were originally called by Hironaka *standard bases*. Only later, the notion of nowadays standard bases (with respect to initial monomials selected by a monomial order on $\mathbb{N}^n$) was introduced.

Standard bases served Hironaka to compute the strict transform of ideals, but more decisively he used them to define his local resolution invariant $\nu_a(I)$ as the vector of orders of the Taylor expansions at the point a of a minimal standard basis. To compare the invariant before and after blowup, it was necessary to dispose also of a standard basis of the transformed ideal (which is not automatic) and to relate it to the one below. To ensure this, he introduced the concept of *reduced* standard basis (whose existence proof anticipated, at least implicitly, his famous division theorem). Nice thing: Reduced standard bases are persistent under blowup.

The definition of the invariant suggests an immediate refinement: Instead of the vector of the orders of the Taylor expansions of a reduced standard basis, take directly the vector given by the initial monomials of the initial ideal of I with respect to a chosen monomial order. The monomials will be listed increasingly. This vector can be made coordinate independent by taking either generic or most specific (formal) coordinates [Ha 5]. It represents a substitute for $\nu_a(I)$ and also for the more popular Hilbert-Samuel function. Advantage: It is much easier to work with. And, by choosing different monomial orders, it offers more flexibility and information.

Note: It seems to be an open problem how to describe the strict transform of a $\mathcal{D}$-module in terms of generators (where the strict transform is defined again by restricting first to $W \setminus Z$ and taking then an extension from $W' \setminus E$ to W').

4. Properties of blowups

Blowups can be introduced axiomatically by a universal property. This goes as follows. Let X be any scheme, and let E be a subscheme of X. Then E is a *Cartier divisor* in X if, locally on X, E is defined by a non-zero divisor f in the structure sheaf of X: For any $a \in X$ there is an affine neighborhood $U = \operatorname{Spec} R$ of a in X such that $E \cap U$ equals the subscheme $V(f)$ of U defined by a non-zero divisor $f \in R$.

If X is regular, a Cartier divisor is just a (closed) hypersurface of X. If X is singular, the notion is more subtle.

Exercise 1: Let E be one of the axes of the cross $X = V(xy)$ in $\mathbb{A}^2$. Is it a Cartier divisor? Same question for E the origin.

Exercise 2: Let E be the reduced origin of $X = V(xy, x^2)$ in $\mathbb{A}^2$. Show that E is not a Cartier divisor.

Exercise 3: Any line through 0 inside the cone $X = V(x^2 + y^2 - z^2)$ in $\mathbb{A}^3$ is not a Cartier divisor of X (though, locally at points different from the origin, it is).

Exercise 4: Are $E = V(x^2)$ in $\mathbb{A}^1$ and $E = V(x^2y)$ in $\mathbb{A}^2$ Cartier divisors?

It can be shown that a hypersurface E of X is a Cartier divisor of X if and only if it is rare in X, i.e., if and only if $X \setminus E$ is (scheme-theoretically) dense in X. The examples of the exercises show that this property depends in part on the location of E with respect to the (embedded) components of X.

Here is the universal property of blowups. The blowup of X in a closed subscheme Z is a morphism $\pi : X' \to X$ such that $E = \pi^{-1}(Z)$ is a Cartier divisor of X' and such that for any morphism $\tau : Y \to X$ with $\tau^{-1}(Z)$ a Cartier divisor in Y there exists a unique morphism $\rho : Y \to X'$ with $\tau = \pi \circ \rho$.

If it exists, the blowup is unique (up to unique isomorphism). The existence can be proven by different methods, constructing suitably the scheme X' and the morphism π. For affine schemes, our method with the closure of the graph of the map defined by the equations of Z does the job. There is a more general way to construct the blowup of X as the Proj of the Rees-Algebra $\oplus_{k=0}^{\infty} P^k$ where P denotes the ideal sheaf defining Z in X.

The blowup of a regular scheme X in a closed hypersurface Z is an isomorphism, by the universal property. This is no longer true if X is singular and/or Z is not a Cartier divisor.

Exercise 5: Show that the blowup of the cone $X = V(x^2 - yz)$ in $\mathbb{A}^3$ along the line $Z = V(x, y)$ is an isomorphism locally at all points outside 0, but not globally on X.

Exercise 6: Blow up the fat point $X = V(x^2)$ in $\mathbb{A}^2$ in the (reduced) origin $Z = 0$. What do you get?

Exercise 7: Blow up the subscheme $X = V(x^2, xy)$ of $\mathbb{A}^2$ in the reduced origin.

Exercise 8: Blow up the subscheme $X = V(xz, yz)$ of $\mathbb{A}^3$ first in the origin, then in the x-axis. At which points are the resulting morphisms local isomorphisms?

The universal property, as usual, is helpful when proving statements about blowups. The most important is the commutativity of blowups with base change. Many special features of blowups follow from this property.

Base change. Let $\pi : X' \to X$ be the blowup of X with center $Z \subset X$, let $\varphi : Y \to X$ be a morphism (the base change) and let $V = \varphi^{-1}(Z)$ be the preimage of Z in Y. Consider the fibre product $\widetilde{Y} = Y \times_X X'$ of Y and X' over X with induced diagram

$$\begin{array}{ccc} \widetilde{Y} & \to & X' \\ \downarrow & & \downarrow \\ Y & \to & X \end{array}$$

where $\widetilde{Y} \to Y$ and $\widetilde{Y} \to X'$ denote the canonical projections. Let Y' be the Zariski-closure of the preimage of $Y \setminus V$ in $\widetilde{Y}$. Then the induced morphism $\tau : Y' \to Y$ is the blowup of Y along V.

We prove the base change property after a few remarks and exercises.

Special cases: Taking in the above situation as base change the inclusion $Y \subset X$ of an open subscheme, we find that blowups can be constructed from local data by gluing blowups of open subsets: Given an open covering $\{U_i\}$ of X the blowups of U_i with center Z_i glue to a global blowup of X if the Z_i patch on the overlaps $U_i \cap U_j$ thus defining a global center Z in X.

Also, it follows that blowups are compatible with taking localizations at a point (or at a closed subscheme) and completions. For $a \in Z$ and a' a point of E above a, say $\pi(a') = a$, we call the induced local morphism $(X', a') \to (X, a)$ corresponding to the inclusion of the local rings $\mathcal{O}_{X,a} \to \mathcal{O}_{X',a'}$ or their completions the local blowup of X at a and a' with center Z.

Taking $Y \subset X$ a closed subscheme of a regular scheme X and a center Z inside Y, we recover the earlier notion of strict transform of Y: The strict transform of Y under the blowup of X in $Z \subset Y$ coincides with the blowup of Y in Z.

If Z is regular but not contained in Y, the intersection $V = Z \cap Y$ may be singular and non-reduced. It then yields the blowup of Y in a center which is no longer regular.

Exercise 9: Take the cone $Y : x^2 - yz = 0$ in $X = \mathbb{A}^3$ and as center the line $Z : y = z = 0$. The intersection $V = Z \cap Y$ is the origin with ideal $P = (\overline{y}, \overline{z})$ in $\mathcal{O}_Y$, the bars denoting residues. Compute the blowup of Y in V (compare with exercise 4^b of chapter II).

Exercise 10: Explain, why $\mathcal{O}_{X,a} \to \mathcal{O}_{X',a'}$ is an inclusion of rings. Then take X and Z regular, $a \in Z$, and express the local blowup with respect to (suitably chosen) regular systems of parameters of $\mathcal{O}_{X,a}$ and $\mathcal{O}_{X',a'}$.

We may also use the base change property of blowups for field extensions.

Exercise 11: Let Y be the real points of the complex surface $X : x^4 - y^4 - yz^2 = 0$ in $\mathbb{A}^3_{\mathbb{C}}$. Let $Z \subseteq X$ be the curve $x^2 + y^2 = z = 0$. Compare the blowups of X in Z and of Y in $Y \cap Z$. Then do the same for the center defined by the ideal $(x^2+y^2, z)(x^2+y^2, xz, yz, z^2)$.

Exercise 12: Consider $L = \mathbb{F}_2$ as a subfield of $K = \mathbb{F}_{32}$, and let X and Y be the surfaces defined by $x^2 + y^3 + z^4 = 0$ in $\mathbb{A}^3_K$, respectively $\mathbb{A}^3_L$. How many points have the blowups of X and Y with center the origin? Do these resolve the surfaces in one blowup?

Exercise 13: Let $K = \mathbb{C}((t))$ be the ring of formal Laurent series in one variable. Let $X : tx^2 - y^3 = 0$ be the "cusp" in $\mathbb{A}^2_K$. Blow up the origin and compute X'.

Proof. The proof of the base change property goes by diagram chasing. So let $\varphi : Y \to X$ be given and let $\tau : U \to Y$ be a morphism for which $\tau^{-1}(V)$ is a Cartier divisor in U. As $(\varphi\tau)^{-1}(Z) = \tau^{-1}(V)$, the universal property of the blowup $X' \to X$ implies that $\varphi\tau : U \to X$ factors through $X' \to X$ via some morphism $\rho : U \to X'$. By the universal property of fibre products, we get from τ and ρ a morphism $\sigma : U \to \widetilde{Y} = Y \times_X X'$. Its composition with the projection to Y equals τ. We have to show that σ maps U into Y'. By construction, σ maps $\tau^{-1}(Y \setminus V)$ into Y'. As $\tau^{-1}(Y \setminus V)$ is dense in U and Y' is closed in $\widetilde{Y}$, we conclude that σ maps U into Y', thus getting the required factorization of τ.

We constructed earlier, via graphs, the blowup $\pi : \widetilde{\mathbb{A}}^n \to \mathbb{A}^n$ of affine space $\mathbb{A}^n$ in a closed subscheme Z. It was done by choosing generators $g_1, \ldots, g_k$ of the ideal P defining Z in

$\mathbb{A}^n$, taking the morphism $g : \mathbb{A}^n \setminus Z \to \mathbb{P}^{k-1}$ with projective components $(g_1 : \ldots : g_k)$ and the Zariski-closure $\widetilde{\mathbb{A}}^n$ of its graph in $\mathbb{A}^n \times \mathbb{P}^{k-1}$. So let us establish the universal property for $\pi : \widetilde{\mathbb{A}}^n \to \mathbb{A}^n$. It will in particular show that, up to isomorphism, the construction does not depend on the choice of the generators $g_1, \ldots, g_k$.

We have already seen that $E = \pi^{-1}(Z)$ is a Cartier divisor, its affine equations were $g_i = 0$ (check this again accurately). Take a morphism $\tau : Y \to \mathbb{A}^n$ with $\tau^{-1}(Z)$ a Cartier divisor in Y. We wish to show that τ factors through $\pi : \widetilde{\mathbb{A}}^n \to \mathbb{A}^n$ via a morphism $\rho : Y \to \widetilde{\mathbb{A}}^n$. The morphism $\tau g : Y \setminus \tau^{-1}(Z) \to \mathbb{P}^{k-1}$ lifts to a morphism $\widetilde{\rho} : Y \setminus \tau^{-1}(Z) \to G$ into its graph $G \subset Y \times \mathbb{P}^{k-1}$, which, by construction, is contained in the graph of $g : \mathbb{A}^n \setminus Z \to \mathbb{P}^{k-1}$. Taking closures on both sides and using that $Y \setminus \tau^{-1}(Z)$ is dense in Y, we get the desired morphism $\rho : Y \to \widetilde{\mathbb{A}}^n$.

This proves the existence of the blowup of $\mathbb{A}^n$ in a closed center Z. From this, by the base change property for subschemes, we get the existence of blowups of arbitrary affine schemes with centers closed subschemes. The gluing property of blowups then implies the existence of blowups for arbitrary schemes.

Let us summarize the advantages and drawbacks of our five equivalent definitions of blowups.

Universal property: Very handy for proofs and general properties of blowups.

Rees-algebra and Proj: Most general explicit construction. Allows to read off global and geometric properties of the blowup as regularity, normality, Cohen-Macaulayness, etc. by applying methods of ring theory and commutative algebra. Very efficient if center is monomial ideal. Somewhat unpleasant for local computations.

Closure of graph: Works for affine schemes. Geometrically intuitive. Allows explicit computations for regular centers. Obsolete if the ideal of center has large number of (complicated) generators.

Affine charts expression of morphism: Best suited for local computations if center is regular. Allows often to choose privileged coordinates for which the local blowup is given by a monomial substitution of the variables.

Equations of blowup: Works for affine schemes $X \subseteq \mathbb{A}^n$ and their blowup $X' \subset \mathbb{A}^n \times \mathbb{P}^{k-1}$. Both, affine or projective coordinates and equations may be taken.

We now come to more geometric properties of schemes and how they behave under blowup.

Exercise 14: Let X be a regular scheme, and Z a regular closed subscheme. Show that the blowup X' of X along Z is again regular.

Blowups preserve also normal crossings singularities, provided the center is sufficiently transversal. We make this more precise. Two subschemes X and Z of a regular ambient scheme W are said to be *transversal* if the scheme defined by the product of the ideals of X and Z in W is a normal crossings scheme (we do not take the intersection of the ideals).

Exercise 15: Show that a regular subscheme Z of a regular scheme X meets X transversally.

Exercise 16: Let X be a surface in $\mathbb{A}^3$ consisting of two regular components X_1 and X_2 meeting transversally. Let Z be a regular surface in $\mathbb{A}^3$ which is transversal to X_1 and X_2. Show that Z need not be transversal to X.

Exercise 17: Let X be a normal crossings subscheme of $\mathbb{A}^n$, and let Z be a regular closed subscheme which is transversal to X. Show that the blowup X' of X along Z is again a normal crossings scheme (cf. fig. 10).

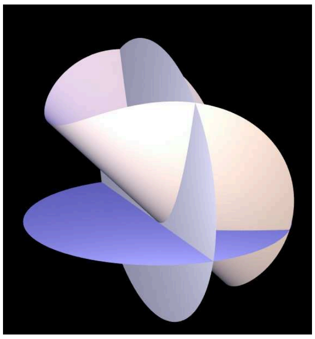


Figure 10: Tülle, non-normal crossings surface.

Exercise 18*: Let X be a *mikado scheme* in $\mathbb{A}^n$, i.e., a scheme which is a union of regular components all whose intersections are regular (scheme theoretically). Example: A finite union of *linear* subspaces V_i in $\mathbb{A}^n$ so that $V_i \setminus \bigcup_{j \neq i} V_j$ is dense in V_i for all i. Show that there is a sequence of blowups over $\mathbb{A}^n$ which transforms X in a normal crossings scheme (taking always total transforms). Write a program for this and determine its complexity. Compare with the papers of De Concini, Procesi [DCP] and Feichtner, Koslov [FK].

Exercise 19: Let $X' \to X$ be the blowup of any X in a chosen ideal P. Find three other ideals which yield the same blowup.

Exercise 20*: Let $\pi : (W', a') \to (W, a)$ be a local blowup of a regular scheme W along a regular center Z (you may choose first Z to be the point a). It should be clear what is meant by a local flag $\mathcal{F}$ of regular subschemes in W at a. Assume that Z is transversal to $\mathcal{F}$ (in the obvious sense). Show that $\mathcal{F}$ lifts in a natural way to a flag $\mathcal{F}'$ in W' at a'. (Hint: Do first the point blowups of $\mathbb{A}^2$ and $\mathbb{A}^3$ in the case where the flags are given by coordinate subspaces. You may wish to consult [Ha 5] for the general case.)

Exercise 21: Now choose local coordinates on W at a for which the local blowup from before is given by a monomial substitution of the variables (you may work in the completions of the local rings). Show that the coordinates can be chosen so that the flag in W is given by coordinate subspaces. Then determine the equations of the induced flag in W'.

There are two other important existential properties of blowups: The composition of blowups is again a blowup (in a more complicated center), and any birational projective morphism is a blowup. We refer to Bodnár's paper [Bo] for the first and to Hartshorne [Hs, p. 166] for the second.

Exercise 22: Blow up $\mathbb{A}^3$ in the normal crossings center Z of ideal $P = (xy, z)$, getting a scheme $\widetilde{\mathbb{A}}^3$ with an isolated singularity (locally, it is a three-dimensional cone). Show that the blowup of this point in $\widetilde{\mathbb{A}}^3$ yields a regular scheme. Then find an ideal in $\mathbb{A}^3$ which, when taken as center, defines the composition of the two blowups. (Hint: you may find an appropriate candidate for the ideal in one of the exercises of the other chapters.)

Exercise 23: Same as before with the irreducible center $P = (x^2 - y^2 - y^3, z)$.

Exercise 24^{**}: Let P be an ideal in $X = \mathbb{A}^n$, inducing the blowup $\pi : X' \to X$. Determine explicitly an ideal R in X such that the blowup $\tau : X'' \to X$ of X in the product ideal $P \cdot R$ yields a regular scheme X'' and such that τ factors through π, i.e., $\tau = \pi\rho$ with $\rho : X'' \to X'$. In particular, the morphism ρ will define a resolution of the singularities of X'. (A decent solution to exercise 24 in non-trivial cases is probably worth a publication in an equally decent journal.)

---

5. Improving singularities by blowups

One of the roles of blowups in algebraic geometry is to eliminate successively the singularities of a given variety. As this cannot be done by just one blowup (actually, it could be done, it is just that we don't know how to choose a suitable center), a whole series of different blowups is necessary. Some invariant shall document that the singularities become better with each blowup. This relies on a numerical measure of the complexity of a singularity. We therefore ask: Which number tells us the "distance" of a singular point from being a non-singular point? And: Does this number drop under blowup?

There are many proposals for such a measure, and each fullfils the expectation of certain objectives. In our context, we shall confine in a first instance to measures which, at least, do not deteriorate under blowup. This is not yet sufficient for the induction argument. Nevertheless, it is an important step towards it, because in most cases the measure will drop, and the remaining cases are so special that they can be treated by another type of induction.

A suitable numeric measure will be called a resolution invariant. By this we mean that the measure belongs to a well ordered set Γ, typically $\mathbb{N}^n$, that it is attached in an upper-semicontinuous way to each point a of a variety X, and that it does not depend on any choices: It is invariant under local isomorphisms of X at a. Said differently, it depends only on the local ring $\mathcal{O}_{X,a}$ or its completion. This is sometimes expressed by saying that the invariant is intrinsic. Morever, if the center of blowup is chosen suitably (this will have to be specified), it decreases with respect to the ordering on Γ.

We have already seen in the exercises that regular varieties remain regular under blowup if the center itself is regular and transversal to or included in the variety. Which number expresses regularity? Of course, this is a local notion. By the implicit function theorem, a point a on a hypersurface is regular, if and only if the order of vanishing of the defining polynomial f at a is 1. This order is just the order of the Taylor expansion of f at a, which we will denote by $\mathrm{ord}_a f$.

For non-hypersurfaces the situation is trickier: Measuring regularity involves the dimension of the vartiety as well as the jacobian matrix of the defining equations. Hironaka used his vector ν mentioned in the last chapter. Also the local multiplicity or the Hilbert-Samuel function can be considered as invariants. For the sake of simplicity, we shall restrict in the sequel to hypersurfaces. Non-hypersurfaces are technically a little bit more complicated, but all the interesting phenomena already appear for hypersurfaces. This justifies our restriction.

Observation: A regular hypersurface remains regular under blowup if we consider its strict transform. This does not tell us how the transform meets the exceptional divisor, in particular, if the intersection is transversal. We could instead consider the total transform of the hypersurface, because it contains also this piece of information. But: The total transform will never be a regular variety, even if the strict transform is regular. The best we can hope for is to achieve a normal crossings singularity.

Now, it's not difficult to prove that if we blow up a regular center which is (sufficiently) transversal to the variety, a normal crossings singularity has total transform which has again only normal crossings singularities (see exercise 17 in chapter IV). This suggests to measure at a singular point instead of the distance to regularity the distance to having a normal crossings point. Algebraically, this is the distance of an ideal of polynomials from being a monomial ideal, possibly in other (formal) coordinates.

Quandary: A significant and genuine invariant indicating the distance of a singularity from having normal crossings at a given point has not been discovered yet.

So there is homework to do. Be aware that it may not be easy: The invariant should be intrinsic, upper-semicontinuous, and it should not increase under blowup (if the center is chosen with care). A first candidate goes as follows: Factor from your polynomial the monomial of maximal degree, over all choices of coordinates. Then take the order of the remaining factor at the point in question.

This is what (almost) everybody does, though it is rather clumsy. Essentially, it boils down to considering the order of the strict transform of the hypersurface. Main drawback: It does not work well in positive characteristic, cf. [Ha 1, Ha 4]. *End of observation.*

By lack of any better, we shall stick to the order of a polynomial and of its strict transform as our main (though very rough) measure of singularity. And the question is: Does it never increase under blowup? Some examples will lift the fog.

Example 1: The plane curve $X : x^m - y^k = 0$ with $m < k$ of order m at 0. We blow up the origin. The tangent cone of X (given by the homogeneous form of lowest degree of the defining polynomial) at 0 is the y-axis $x = 0$ (would not hold for $m = k$). We are therefore led to consider in the exceptional divisor E only the y-chart, and there the point $x = 0$, i.e., the origin (prove this with care, cf. exercise 4 of chapter I). The strict transform X' has equation $x^m - y^{k-m} = 0$ there. Ah! If $m < 2k$, the order has dropped to $k - m < m$, and otherwise it has remained constant equal to m. Easy.

Example 2: The surface $X : x^m - y^k z^\ell = 0$ in $\mathbb{A}^3$ with $m \leq k$ and ℓ of order m at 0. Which blowup is suited? With center the origin? Or a line? Let us do both.

First the origin. The tangent cone of X at 0 is the yz-plane $x = 0$ (because of $m < k+\ell$), it hence suffices to consider the y- and the z-chart of the blowup and there only the points of E with x-coordinate equal to 0. In the x-chart, X' does not meet the exceptional divisor. And outside of E, the order must have remained the same since the blowup is an isomorphism there. By symmetry, we only consider the y-chart with strict transform X' given by $x^m - y^{k+\ell-m} z^\ell = 0$. As $k + 2\ell > 2m$, the order has remained constant equal to m at the origin of E. Observe that X' intersects E in the y-chart along the z-axis and has order m at each of its points. This shows that at no point of E the order has increased. Fine!

So let's blow up the line $x = y = 0$, i.e., the z-axis. In the y-chart, the equation of X' is $x^m - y^{k-m} z^\ell = 0$ since $m \leq k$. The order has remained constant at the origin and did not increase at the other points (check this carefully at all points of the chart and determine the respective orders). Fine again! What about the x-chart? It suffices to look there at the origin, the other points being covered already by the y-chart. The equation of X' is $1 - y^{k-m} z^\ell = 0$ and X' does not pass through the origin of this chart (the order is 0 there). Done!

Example 3: As the exercise-experienced reader already suspected, these examples are misleading. Take the preceding surface $X : x^m - y^k z^\ell = 0$ but with exponents $k < m \leq \ell$. We assume that $m \leq k + \ell$ so that X has order m at 0. We blow up the z-axis $x = y = 0$. In the x-chart we get X' defined by $x^{m-k} - y^k z^\ell = 0$ of order $\leq m$. In the y-chart the equation of X' is $x^m y^{m-k} - z^\ell = 0$. The order at the origin equals the minimum of $2m-k$ and ℓ. As both $2m - k$ and ℓ are larger than m, the order has increased. Bad news!

But why did this happen? There is no apparent reason for the increase. But our algebraic manipulations tend to hide the internal structure of the singularities.

Zariski was one of the first – if not the first – to notice and understand this phenomenon. Actually, it is not very complicated: In examples 2 and 3, compare the behaviour of X along the points of the z-axis Z, which was our chosen center. The relative size of the exponents k, ℓ and m comes into play.

If $m \leq k, \ell$, the order of X along Z is constant equal to m, because $f = x^m - y^k z^\ell$ belongs to the m-th power of the ideal (x, y) defining Z (it suffices to have $k \geq m$). If $k < m \leq \ell$, this is not the case. Indeed, the order at 0 is m but at points of Z outside the origin it is $k < m$. We have a drop of the order when leaving the origin.

We look closer at the total transform X^* of X at the origin of the y-chart. In both cases it is given by $f^* = x^m y^m - y^k z^\ell = 0$. The maximal power of the exceptional variable y which we can factor from f^* depends on the relative size of k and m. If $k < m$ we can factor only a power y^k whose exponent k is *smaller* than the order m of f at 0. Therefore, the strict transform f' has order which is *larger* than this order. If $k \geq m$, we can factor the whole power y^m.

You may notice that we saw this type of question already in exercise 1 of chapter III.

Exercise 1: Do example 3 with center the origin.

Fact: *The order of a variety does not increase at any point of its strict transform if we blow up a regular center along which the variety has constant order. In particular, regular points remain regular.*

We shall establish the fact for hypersurfaces only. Let $Z \subset X$ be the center, X defined in $W = \mathbb{A}^n$ by $f = 0$, and $\mathrm{ord}_a f = \mathrm{const}$ for $a \in Z$. Let $\pi : W' \to W$ be the induced blowup, and X' the strict transform of X in W'. Choose a point $a \in Z$ and a point a' above a, i.e., so that $\pi(a') = a$. Set $o = \mathrm{ord}_a f$ and $o' = \mathrm{ord}_{a'} f'$. We wish to prove that $o' \leq o$.

This is done by specifying very nice coordinates at a and a'. With these, the proof becomes almost automatic, so that the details can be grouped in a series of exercises. We may work in the completion of the local rings of W and W' at a and a', because completion does not alter the order. Local shall always mean formally local, i.e., working in $K[[x_1, \ldots, x_n]]$.

Exercise 2: Choose local coordinates $x_1, \ldots, x_n$ at a so that $a = 0$ and Z is defined by $x_1 = \ldots = x_k = 0$, where $k = \mathrm{codim}_W Z$.

Exercise 3: Permute the first k coordinates so that a' belongs to the x_k-chart of the blowup W'.

Exercise 4: Apply a lower triangular linear coordinate change in W (thus preserving the ideal defining Z in W) so that a' becomes the origin of the x_k-chart.

Exercise 5: Show that in these coordinates, the local chart expression of π at a' and a is given by a monomial substitution of the variables (the usual one, see chapter III).

Exercise 6: Show that in the coordinates chosen at a, the expansion of f has order $\geq o$ with respect to the variables $x_1, \ldots, x_k$ (i.e., belongs to the ideal $(x_1, \ldots, x_k)^o$).

Exercise 7: Use this to show that the exceptional monomial x_k^o can be factored from the total transform f^* of f.

Exercise 8: Now use the fact that $(x_1, \ldots, x_k)^o$ is the maximal power to which f belongs to show that f' has order at most o with respect to all variables, i.e., f' does not belong to $(x_1, \ldots, x_n)^{o+1}$.

Exercise 9: Conclude that $o' \leq o$, say $\mathrm{ord}_{a'} f' \leq \mathrm{ord}_a f$.

Question: This proof is very explicit, though not really conceptual. Where did you use that the order of f is constant along Z?

Exercise 10: For those familiar with the Rees algebra of blowups, find a less computational and more direct proof of the fact. Does it work also for singular centers?

Exercise 11: We have already seen that the total transform of a normal crossings singularity has again normal crossings, provided that the center is sufficiently transversal to the variety. Prove this again by exhibiting an invariant which measures the distance from being normal crossings and which does not increase under blowup.

Exercise 12: You may choose at your taste 27 different lines and 15 different planes in $\mathbb{A}^3$, all passing through the origin, no line contained in a plane. Then modify their union by blowups and taking total transforms until you get a variety with normal crossings. Intuitively, it is clear that this should work in finitely many steps. So write a computer program for it. What is the minimal number of blowups (with regular centers) needed?

6. The induction argument of resolution

Starting with this chapter, pace and level will go up. But before doing so, one and a half examples.

Example 1: Once again our lemon-tree $x^2 - y^3z^3 = 0$. The order at 0 is 2, and this is also the case at the points of the y- and z-axis. The locus of maximal order is therefore the cross formed by these two axes. Hesitating which axis would be better as our center of blowup we take, to be on the sure side, their intersection the origin 0. The point blowup gives in the y-chart the transform $x^2 - y^4z^3$ (the z-chart is symmetric, and the x-chart is irrelevant). The order has remained constant equal to 2 at the origin of the y-chart, not too bad, but the polynomial itself has become worse as the degree of the monomial in y and z has increased. How to build up an induction on this? Unclear.

Example $1\frac{1}{2}$: Hardly enchanted by the starter, we opt for an uncommon way to pursue: We complicate the example, taking instead the hypersurface X in $\mathbb{A}^6$ defined by $x^3 - y^\alpha z^\beta u^\gamma v^\delta w^\varepsilon = 0$. The exponents α, β, γ, δ and ε may be quite arbitrary and are only subject to $\alpha + \beta + \gamma + \delta + \varepsilon \geq 3$ so as to have a singularity of order 3 at 0 (the actual value 3 does not matter here). Let us not care about symmetry when choosing the center.

We just take any regular subvariety of the locus L of order 3 of f. But we require that it has maximal possible dimension among all regular subvarieties of L. By the shape of L in the present situation, Z will be a coordinate subspace. To fix ideas, assume that we choose for Z the vw-plane of equations $x = y = z = u = 0$.

There are four charts. The x-chart can be discarded by what we have already learned, and up to a permutation we may restrict our attention to the y-chart. There, the total transform X^* is defined by $x^3y^3 - y^{\alpha+\beta+\gamma}z^\beta u^\gamma v^\delta w^\varepsilon = 0$. The strict transform is given by factoring the maximal power of y, which is the minimum of 3 and $\alpha+\beta+\gamma$. As X has, by the very choice of the center, order 3 along Z, we have $\alpha+\beta+\gamma \geq 3$ and the minimum is 3. We get X' of equation $x^3 - y^{\alpha+\beta+\gamma-3}z^\beta u^\gamma v^\delta w^\varepsilon = 0$, which we write as

$$x^3 - y^{\alpha-(3-\beta-\gamma)}z^\beta u^\gamma v^\delta w^\varepsilon = 0.$$

Well, where is the point in all this? It is as follows, and this will be decisive: Up to now, we have not used the hypothesis that Z has *maximal possible dimension* in L. This assumption implies that the three-dimensional coordinate-subspace S defined by $x = y = z = 0$ is not contained in L. Hence f does not have order 3 along S, which, in turn, signifies that $\beta + \gamma < 3$. Therefore, in the equation of X', the degree of the monomial $y^{\alpha-(3-\beta-\gamma)}z^\beta u^\gamma v^\delta w^\varepsilon$ has *decreased*. Of course, this is a general phenomenon, which we now state explicitly.

Exercise 1: Let $f = x^m - y_1^{\alpha_1}\cdots y_n^{\alpha_n} = x^m - y^\alpha$ be a binomial with $|\alpha| = \alpha_1+\ldots+\alpha_n \geq m$. Then $\mathrm{ord}_0 f = m$. Let L be the locus of points a where $\mathrm{ord}_a f = m$, and let Z be a coordinate subspace of L of maximal possible dimension. Blow up Z in $W = \mathbb{A}^{1+n}$ with strict transform f'. Then at all points a' of the exceptional divisor E either the order of f' has dropped, or the affine chart expression of f' is again a binomial $f = x^m - y_1^{\alpha'_1}\cdots y_n^{\alpha'_n} = x^m - y^{\alpha'}$ with a monomial $y^{\alpha'}$ of total degree strictly smaller than the degree of y^α, say $|\alpha'| < |\alpha|$.

Exercise 2: Conclude from this that finitely many blowups in maximal dimensional centers succeed in decreasing the order of binomials f as above.

Neglecting in this argument symmetry considerations is harmless as long as we work locally at the point in question (though the equivariance of the resolution with respect to group actions, say symmetries, will be lost). Globally, the center Z you may have chosen may return after some promenade inside X to the respective point from a different direction and thus form a normal crossings center there (recall that the center must be closed, cf. figure 11). Therefore we would have a singular center. This problem can be circumvented by applying first auxiliary blowups which help to separate the local components of possible (big) centers Z. For the moment, we do not discuss this topic any further.

Exercise 3: Construct a surface X in $\mathbb{A}^3$ whose locus L of order 2 is the node Y : $x^2 - y^2 - y^3 = z = 0$. Show that blowing up $\mathbb{A}^3$ in $Z = Y$ produces a singular variety. Next, blow up the origin $Z = 0$ in $\mathbb{A}^3$, and show that the locus L' of the strict transform

X' has become either regular or a union of regular components intersecting transversally (produce examples of X for both occurences). In the second case, show that there exists a resolution of X (or, equivalently, of X') which preserves all local and global symmetries.

Exercise 4: Is it possible to equip Y with a non-reduced structure Y' so that the blowup of $\mathbb{A}^3$ with center $Z = Y'$ is regular?

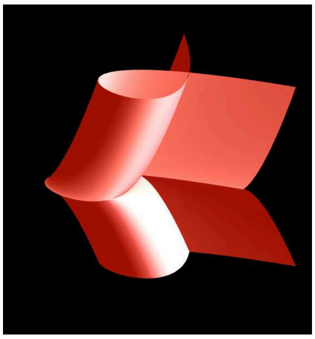


Figure 11: Quaste, cartesian product of cusp with node.

Let us look more closely at the order function: Given a variety X in a regular ambient space W, we stratify X according to the local order $o = \text{ord}_a I$ of the defining ideal $I = I_W(X)$ of X in W. Similarly as for hypersurfaces, $\text{ord}_a I$ denotes the maximal power of the maximal ideal of $\mathcal{O}_{W,a}$ at a which contains I. More generally, for S closed, we define $\text{ord}_S I$ as the maximal power of the ideal $I_W(S)$ of $\mathcal{O}_W$ defining S in W which contains I. It can be shown that $\text{ord}_S I = \min_{a \in S} \text{ord}_a I$.

Note that, locally at a point a, the variety may be singular though the order at a is 1. This happens when X is not minimally embedded locally at a. In particular, the order depends on the embedding of X in the ambient space W. (We are somewhat imprecise of what we understand by locally at a point: To avoid ambiguities and delicacies, you may always take the completed local rings of the variety and its ambient space. This is ok because the order is invariant under completion.)

As the order is an upper-semicontinuous function (you may convince yourself by some examples or try to prove it rigorously), the stratum $\text{top}(X)$ of points where the order attains its maximum is closed. We call it the *top locus* of X. Recall here that a function on a topological space with values in a totally ordered set is upper-semicontinuous if the subsets of points where the function takes values $\geq$ a given value are closed.

The top locus can be singular, so we cannot take it directly as center. Instead, we may choose any regular closed subvariety Z of $\text{top}(X)$, preferably of maximal possible dimension. We have seen before that this choice may be non-canonical, for instance in the case where X has a symmetry so that $\text{top}(X)$ has two regular components intersecting transversally. Then it will be up to us to choose one of these components. This problem of choice may appear locally as well as globally.

In the published proofs of resolution in characteristic 0, this ambiguity is resolved by prescribing ab initio a uniquely determined center, similarly as in exercise 3 from above.

So let the center Z inside $\mathrm{top}(X)$ be chosen, take the induced blowup $\pi : W' \to W$ of the ambient space and consider the total transform X^* of X in W'. Let a be a point of Z, and let a' be any point of $E = \pi^{-1}(Z)$ above a. For reasons to be explained later, we set $o = \mathrm{ord}_a X = \mathrm{ord}_Z X$ and define the *weak transform* $X^\curlyvee$ of X as the ideal $I^\curlyvee = I_E^{-o} \cdot I^*$. Here, I^* is the total transform (= pullback) of I in W' and I_E is the principal ideal defining E in W'. If you wonder whether the weak transform is well-defined (i.e., that the ideal I_E^o can indeed be factored from I^*), you missed to do one of the earlier exercises.

Exercise 5: Determine the geometric difference between weak and strict transform. Exhibit this difference in three significant examples. After that you should have recognized that working algebraically with the weak transform is more pleasant than with the strict transform.

Fact: *The order of the weak transform does not increase under blowup,*

$$\mathrm{ord}_{a'} I^\curlyvee \leq \mathrm{ord}_a I$$

for $a \in Z \subset \mathrm{top}(I)$ *and* $a' \in E$.

The proof is similar to the hypersurface case.

Observation 1: At points a' where $\mathrm{ord}_{a'} I^\curlyvee < \mathrm{ord}_a I$ induction applies to deduce that after finitely many steps the order of the weak transform has dropped to 0.

Exercise 6: Why 0 is better than 1?

Having order 0 implies that the total transform of X has become a normal crossings variety supported by the exceptional components, provided the centers have always been choosen transversal to the exceptional locus (so that the next exceptional locus is indeed a normal crossings divisor).

Observation 2: At points a' where $\mathrm{ord}_{a'} I^\curlyvee = \mathrm{ord}_a I$ induction does not apply. But: These *equiconstant* points are rare. In fact, they always lie inside a regular hypersurface $F = H' \cap E$ of E, where H' is a regular hypersurface in W' transversal to E (cf. exercise 4 of chapter I).

This can be seen by relating the equiconstant points to the tangent cone of X along Z. Actually, if τ is the minimal number of variables necessary to define the tangent cone of X at a, the equiconstant points lie inside a regular subvariety of codimension τ in the fiber $E_a \subseteq E$ of π over a.

Exercise 7: Prove this.

What does this help, if there do exist points in X' where the order did not decrease? How to apply any other type of induction?

To attack and overcome this intricacy, we consider two examples, a trivial and an almost trivial one.

Example 2: Let X be a fat point in $\mathbb{A}^1$, say the origin, defined by $x^m = 0$ (from now on, we will allow also non-reduced varieties). You would think there is not so much to do with this fat point geometrically, but algebraically it is of some interest. Let's blow up the (reduced) origin in $\mathbb{A}^1$. This is an isomorphism $\pi : \mathbb{A}^1 \to \mathbb{A}^1$, with $E = Z = 0$. Nevertheless, the weak transform $X^\curlyvee$ of X differs from X, because it is defined by the ideal $I_E^{-m} \cdot I^* = (x^{-m}x^m) = (1)$. So $X^\curlyvee$ is empty and the order has dropped to zero.

Conclusion: In ambient dimension 1, there arise no equiconstant points.

Example 3: What about dimension 2, say plane curves? We take again $x^m - y^k = 0$ with $k \geq m$, of order m at 0. We blow up the origin. If $k < 2m$, the order drops at all points of $E \cong \mathbb{P}^1$. If $k \geq 2m$ there is precisely one point in E where the order remains constant, namely the origin of the y-chart. There, the equation of the weak transform is $x^m - y^{k-m} = 0$. The exponent of the second monomial has dropped. How to profit of this drop?

The answer is to descend to the one-dimensional case from before. Intersect X in $\mathbb{A}^2$ and $X^\curlyvee$ in the y-chart of W' with the hyperplane $x = 0$. You get the ideals $I_1 = (y^k)$ respectively $I_1' = (y^{k-m})$ in one-dimensional affine space $\mathbb{A}^1$. There, the order has dropped. Nevertheless, this could just be by chance.

Now comes the clue: Possibly, I_1' is the transform of I_1 under the blowup of 0 in $\mathbb{A}^1$ as in example 1. Then the drop would have been forced, and hopefully lend itself to a similar argument in higher dimensions.

Obviously, I_1' is not the strict transform of I_1 (which is the ideal (1)) nor the total transform. But I_1' can be obtained from the total transform I_1^* of I_1 by factoring a *prescribed* power of the exceptional component. More precisely, $I_1' = I_E^{-m} \cdot I_1^*$ with $m = \mathrm{ord}_Z I$ and I_E the principal ideal of the exceptional divisor E. Observe that $m \leq \mathrm{ord}_0 I_1 = k$. This leads to defining the *controlled transform* $I^!$ of an ideal I with respect to $c \in \mathbb{N}$ under the blowup with center Z more generally as $I^! = I_E^{-c} \cdot I^*$ for any *control* $c \leq \mathrm{ord}_Z I$.

Let us conceptualize this very rough proposal in a diagram for ideals I in a regular ambient space W. Let $W' \to W$ be the blowup with center Z contained in $\mathrm{top}(I)$. Let H be a hypersurface of W containing Z, with induced blowup $H' \to H$.

Exercise 8: Show that H' is the strict transform of H under the blowup $W' \to W$. Moreover, it coincides with the weak transform of H.

Assume now that we dispose of a sophisticated construction which associates to I an ideal I_- in H, i.e., in one variable less. We will work here always locally at a given point

$a \in Z \subset H$. For plane curves this was just the passage from I to I_1. The commutative diagram

$$\begin{array}{ccc} W' & \supset & H' \\ \downarrow & & \downarrow \\ W & \supset & H \end{array}$$

gives rise to an incomplete diagram

$$\begin{array}{ccc} I^\curlyvee & \rightsquigarrow & I_?^? \\ \downarrow & & \downarrow \\ I & \rightsquigarrow & I_- \end{array}$$

Here, the lower curly arrow $\rightsquigarrow$ denotes the descent in dimension, associating to I the ideal I_-. In the upper horizontal row, we would like to replace $I_?^?$ in the upper right corner by the ideal $(I^\curlyvee)_-$ associated to $I^\curlyvee$. Similarly, in the right vertical column, we would like to replace $I_?^?$ in the upper right corner by the controlled transform $(I_-)^!$ of I_-. So we have two choices for $I_?^?$ to complete the diagram, and we don't know which one is better. Best would be if we had equality $(I^\curlyvee)_- = (I_-)^!$, because then the diagram would *commute*.

Unfortunately, equality does not hold. In general, the two ideals have nothing to do with each other, $(I^\curlyvee)_- \neq (I_-)^!$.

Exercise 10: Verify this by a concrete example.

Yet another clue: Recall that in E, we are only interested in equiconstant points a', say $\mathrm{ord}_{a'} I^\curlyvee = \mathrm{ord}_a I$. Locally at such points, there exists a hypersurface H' in W' which contains all equiconstant points. As H' is transversal to E, its image H in W under π is again regular, locally at a.

Fact: *Choosing such H and H', then, for any equiconstant point a' above a, we have the desired equality* $(I^\curlyvee)_- = (I_-)^!$.

Hence, at equiconstant points, we may write I' for $I^\curlyvee$ and I'_- for $(I^\curlyvee)_- = (I_-)^!$ and get the commutative diagram

$$\begin{array}{ccc} I' & \rightsquigarrow & I'_- \\ \downarrow & & \downarrow \\ I & \rightsquigarrow & I_- \end{array}$$

This is really very pleasant. Therefore we repeat it: If the order of the weak transform did not decrease, we may descent in dimension if we allow there the controlled transform instead of the weak transform. In H and H', we may apply induction on the dimension on any assertion we would like to prove – provided it holds in dimension 1 and is compatible with descent.

This principle is called – for obvious reasons – *cartesian induction*.

The descent in dimension can be performed in various ways and has been a common procedure in resolution of singularities since the work of Jung. Geometrically, one often takes a generic projection to a hypersurface and there the discriminant, see the survey of J. Lipman [Lp 3] for more details. Algebraically, Hironaka introduced the coefficient ideal as an ideal in one variable less. Although it is less suggestive than the discriminant it is easier to work with. All known proofs in characteristic 0 and for arbitrary dimension use one or the other form of coefficient ideals, cf. [EH] or [EV].

Encouraged by the perspective of nice commutative diagrams we apply induction on the dimension to show that we can resolve I_- inside H by a sequence of blowups with centers in H. If we are lucky, the centers are also contained in $\mathrm{top}(I)$ so as not to increase the order of I. Stop! The ideal I_- does not pass in this process to its strict or weak transform, but to its controlled transform, and, as we have seen many times, its order may *increase*.

This is not a big deal, in view of what we learned in the 1.5 introductory examples of this chapter. There, the ideal I_- corresponded to the monomial y^α, and we saw that once I_- is a monomial (if defined correctly), the drop of the order of I can be forced by choosing centers of maximal dimension. So, in the general situation, we will try to transform I_- into a monomial – instead of making the associated variety regular. With this new objective in mind, taking the controlled transform of I_- makes no harm: we just factor after each blowup the maximal exceptional power $I^r_{(E\cap H')}$ from $I^!_-$, so that the remaining factor is the weak transform of I_-. Now, induction applies: Its order will drop eventually to 0. Therefore, vice versa, the final controlled transform of I_- will have become a monomial.

In the course of this argument it is essential that the centers are chosen transversally to the already existing exceptional locus, otherwise the normal crossings property of the exceptional divisor is not ensured.

Pause: Before proceeding, please recapitulate what was said, read the last paragraphs again in order to digest well the various aspects of the reasoning.

Exercise 11: Check whether the reader followed the preceding advice.

Why do we need a break here? Simply, because there is another obstruction to come, the most serious one. If this (last) one would not occur, we would have outlined a proof of resolution of singularities in arbitrary characteristic, despite the fact that characteristic $p > 0$ is still wide open.

So where is the trap we have overseen? It lies in the choice of the local hypersurface H. We have found H by projecting the hypersurface H' of W' (which is not unique) down to W, getting a regular hypersurface H there. Then we take I_- in H and lift it to H', getting $I^!_-$.

Now, most probably, the order of I will not drop also in the subsequent blowup, say $W'' \to W'$. At least, this cannot be excluded and actually does happen. So the argument and the construction of the commutative diagram with the descent in dimension have to be repeated. For a'' an equiconstant point for I' in W'', choose H'' in W'' at a'' and project it to a regular hypersurface in W' containing the center Z' chosen there. This hypersurface of W' will in general be different from H', so that the right vertical columns of our diagrams do not match. In W' we will have to choose a *new* hypersurface $\widetilde{H}'$, and consequently we lose any control on I'_-, which may switch to anything under the change. That this really happens show examples in positive characteristic.

Lucky stroke: It is not completely clear who was the first to observe that in *characteristic zero* this change of hypersurfaces is superfluous: There exists a hypersurface H in W at a whose successive transforms H', H'', ... under blowup in centers $Z \subset H$, $Z' \subset H'$, ... contain *all* equiconstant points a', a'', ... above a. Wonderful!

Such hypersurfaces are called *hypersurfaces of maximal contact* by Hironaka, and *Tschirnhaus transformations* by Abhyankar. There is an explicit construction of them.

Abhyankar says that the concept appears first in a paper of his with Zariski, and that it has been discussed, investigated and exploited during a four day visit of Hironaka and Nagata to his house at the end of the fifties. Quote: "Hironaka did not stop asking questions, and if I possibly understood better the situation at the beginning, I am not sure if it was not him who did at the end."

That's it! The way was open to the proof of resolution of singularities in any dimension over fields of characteristic zero. The rest are "technicalities".

In the next and thus last lecture we will give the precise statement for the resolution of singularities in characteristic zero. The proof we shall present is based on the concept of *mobiles* and *setups* as developed by Encinas and the author in [EH]. These two data allow to define the resolution invariant elegantly and are then used to establish the required induction.

7. The resolution theorem and its proof

Let X be a reduced singular scheme. A *strong resolution of* X is, for every closed embedding of X into a regular ambient scheme W, a proper birational morphism $\varepsilon\delta\varepsilon\lambda\beta\varepsilon\iota\varsigma$ from a regular scheme W' onto W subject to the following conditions.

> *Explicitness.* $\mathcal{E}\delta\varepsilon\lambda\beta\varepsilon\iota\varsigma$ is a composition of blowups of W in regular closed centers Z transversal to the exceptional loci.
>
> *Embeddedness.* The strict transform X' of X is regular and has normal crossings with the exceptional locus in W'.
>
> *Excision.* The morphism $X' \to X$ does not depend on the embedding of X in W.

Equivariance. $\mathcal{E}\delta\varepsilon\lambda\beta\varepsilon\iota\varsigma$ commutes with smooth morphisms $W^- \to W$, embeddings $W \to W^+$, and separable field extensions.

Equivariance implies that $\mathcal{E}\delta\varepsilon\lambda\beta\varepsilon\iota\varsigma$ is an isomorphism outside the singular locus of X and commutes with group actions. The resolution commutes with open immersions, local and global isomorphisms and taking cartesian products with regular schemes. The smooth morphisms of equivariance need not be defined over the ground field. Passage to the completions implies resolution of formal schemes. One may add *effectiveness*: The centers are equal to the top locus of an upper-semicontinuous invariant $i_a(X)$ on W given by the local rings of X.

Theorem. (Hironaka 1964) *Reduced schemes of finite type over a field of characteristic zero admit a strong resolution.*

The existence of resolutions satisfying explicit- and embeddedness was established by Hironaka. His aproach was in part motivated by earlier work of Zariski and Abhyankar. Villamayor in 1989 and 1991 and Bierstone-Milman in 1997 described constructive resolution algorithms satisfying in addition excision, equivariance and effectiveness. The resolution process of Villamayor was implemented in Maple and Singular by Bodnár and Schicho in 2000. Włodarzcyk simplified in 2005 the descent in dimension by taking a different notion of coefficient ideal. Weak resolution theorems in characteristic zero have been established with different methods by Abramovich-de Jong, Abramovich-Wang and Bogomolov-Pantev.

In the following, we shall present an outline of the proof for the existence of strong resolution in characteristic 0 based on the concept of *mobiles*. This notion was introduced by Santiago Encinas and the author in [EH]. Mobiles allow to clarify substantially the structure of the induction argument of the various resolution proofs. Essentially all nowadays existing proofs (depending on the counting this is a theme with up to eleven variations) use explicit or implicitly mobiles – the only difference being that the transformation rules of the mobile under blowup may vary.

We wish to emphasize that all constructions and arguments below are very much inspired by the earlier proofs of Villamyor and Bierstone-Milman, which in turn, rely on Hironaka's work and ideas.

Throughout, we fix a regular ambient scheme W, always assumed to be of finite type over a field. By a *divisor* in W we shall mean an effective Weil divisor D. A closed subscheme D of W has *normal crossings* if it can be defined locally by a monomial ideal. The subscheme V meets D *transversally* if the product of the defining ideals of V and D defines a normal crossings scheme.

A *stratified ideal* in W is a collection of coherent ideal sheaves each of them defined on a stratum of a stratification of W by locally closed subschemes. A *stratified divisor* is defined by a stratified principal ideal. All ideals and divisors will be stratified without notice, except if said to be coherent.

A *shortcut* of a normal crossings divisor M in W is a divisor N obtained from M by deleting on each stratum of the underlying stratification of M some components of M. The divisor M is *labelled* if each shortcut N comes with a different non-negative integer N, its *label*. The empty shortcut has label 0.

A *handicap* on W is a sequence $D = (D_n, \ldots, D_1)$ of stratified normal crossings divisors D_i of W.

A *singular mobile* in W is a quadruple $\mathcal{M} = (\mathcal{J}, c, D, E)$ with $\mathcal{J}$ a coherent nowhere zero ideal sheaf on W, c a non-negative constant associated to W and D and E handicaps in W with D labelled and E reduced. We call c the *control* of $\mathcal{J}$, and D and E the *combinatorial* and *transversal handicap* of $\mathcal{M}$.

A *strong resolution of a mobile* $\mathcal{M} = (\mathcal{J}, c, D, E)$ in W with $\mathcal{J}$ a nowhere zero ideal in W is a sequence of blowups of W in regular closed and equivariant centers Z transversal to the exceptional loci such that the ideal $\mathcal{J}'$ of the final transform $\mathcal{M}' = (\mathcal{J}', c', D', E')$ of $\mathcal{M}$ in W' (as defined later) has become the locally principal monomial ideal defining D'_n in W'.

Theorem. (Encinas-Hauser 2002) *Singular mobiles defined over a field of characteristic zero admit a strong resolution.*

This result implies the result of Hironaka by taking for $\mathcal{J}$ the ideal defining X in W, as control c the supremum of the orders $\mathrm{ord}_a\mathcal{J}$ of $\mathcal{J}$ on W and empty handicaps D and E. Varying suitably the definition of the transform of a mobile, different resolution algorithms can be obtained.

The resolution of mobiles is found by associating to the mobile $\mathcal{M}$ in W and to each point $a \in W$ a local upper-semicontinuous invariant $i_a(\mathcal{M})$ in a well ordered set Γ. Its top locus $\mathrm{top}(\mathcal{M})$ is shown to be closed, regular and transversal to the divisors of E. Taking it as the center of blowup in W, one obtains the transform $\mathcal{M}'$ of $\mathcal{M}$ in W'. It is shown that for $a \in Z$ and $\mathcal{M}$ unresolved at a the invariant drops at any point $a' \in W'$ above a,

$$i_{a'}(\mathcal{M}') < i_a(\mathcal{M}).$$

It reaches therefore in finitely many steps its minimum. This, in turn, implies that the final transform of the mobile is resolved in the sense defined above.

The invariant $i_a(\mathcal{M})$ will be a vector of non-negative numbers defined through a local analysis of the mobile. We shall therefore work from now on with the stalks of the ideals at a given point $a \in W$, denoted by roman characters. For purposes of the induction we add subscripts indicating the dimension of the ambient space in which the objects are defined (with the exception of handicaps where the subscripts indicate the dimension where the divisor is used).

A *local flag* at a is a decreasing sequence $W_n \supset \ldots \supset W_1$ of closed i-dimensional regular subschemes W_i of a neighborhood $U = W_n$ of a in W. We shall define, for each mobile $\mathcal{M} = (\mathcal{J}, c, D, E)$ in W and each chosen local flag at a, several ideals J_i, I_i, P_i, Q_i, and

K_i, related to each other by various constructions. All together will form a *setup* of $\mathcal{M}$ at a.

Before giving the construction, we motivate the procedure by recalling the objectives of the resolution invariant $i_a(\mathcal{M})$. Its role is two-fold: First to define the center of blowup as the locus where it attains its maximum, second to decrease under blowup. Actually, these two tasks could also be fulfilled by two different invariants. Here, we will accomplish them simultaneously.

In the sequel, all constructions will be local in a sufficiently small neighborhood of a point a, respectively a point a' above a.

We start with J_n the stalk of $\mathcal{J}$ at a. Several blowups have already occurred, so J_n will factor into a product $J_n = M_n \cdot I_n$ with M_n a principal monomial ideal supported by the exceptional locus (which we denote by F) and an ideal I_n. We use here that F is a normal crossings divisor because we always chose the centers regular and transversal to the existing exceptional locus.

Our purpose is to simplify I_n by further blowups (until $I_n = 1$ and $J_n = M_n$ is a monomial ideal). Therefore our center Z should lie inside the top locus $\mathrm{top}(I)$ of I_n consisting of those points where I_n has maximal order. It is hence natural to define the first component of $i_a(\mathcal{M})$ as the order o_n of I_n at a and to consider the vector $i_a(\mathcal{M})$ with respect to the lexicographic order. In this way the top locus of $i_a(\mathcal{M})$ will be contained in $\mathrm{top}(I_n)$ locally at a.

Next, we wish to ensure that the center is transversal to the already existing exceptional locus F. If we neglected this requirement, we could complete by descending induction on the dimension the construction of the further components of the invariant (see below) and would get for the center its top locus $\tilde{Z}$ which, then, might not be transversal to F. Inspection shows that the locus $\tilde{Z}$ is, however, always transversal to some of the components of F, namely those which appeared in the last few blowups (which can be determined explicitly). The intersection with the other components of F is unknown and could be non-transversal. The first divisor E_n of the exceptional handicap E of $\mathcal{M}$ collects precisely these suspicious components.

As our final choice of the center Z has to be transversal to all exceptional components, the simplest though brute way is to require Z to lie locally at each point a inside the intersection of the components of E_n going through a. If we introduce the transversality ideal Q_n of $\mathcal{M}$ as the ideal defining E_n in W_n, its local top locus at a point is the intersection of the components of E_n there. Take then the order q_n of Q_n as the second component of the invariant

$$i_a(\mathcal{M}) = (o_n, q_n, \ldots).$$

With this agreement, the top locus of $i_a(\mathcal{M})$ is contained in $\mathrm{top}(I_n)$ and in all components of E_n it meets, locally at a.

It turns out that I_n passes to its weak transform under blowup, so its order does not increase. If its order remains constant, the determination of the new suspicious exceptional components E_n' shows that also Q_n passes to its weak transform under blowup. So its orders does not increase in this case. Put together, the first two components of $i_a(\mathcal{M})$ do not increase lexicographically,

$$(o_n', q_n') \leq_{lex} (o_n, q_n).$$

Up to now, neither the center is specified nor the defined portion of the invariant must drop under blowup. The next step will be to descend in dimension. Consider the product $K_n = I_n \cdot Q_n$ at a and its coefficient ideal J_{n-1} in any local hypersurface W_{n-1} of W_n through a. Later on, I_n will have to be replaced in the product by a slightly modified ideal P_n. We do not bother here how the coefficient ideal is constructed in reality, we shall only use its properties. For those who wish to see the details we refer to [Ha 3].

We have seen in the last chapter that the hypersurface W_{n-1} can be chosen so that its strict transform W_{n-1}' contains all equiconstant points of K_n, i.e., the points a' above a where the order of K_n' has remained constant. Moreover, we can achieve that it maximizes the order of J_{n-1} over all choices of hypersurfaces. And, finally, it is shown by a separate argument that then $\mathrm{top}(J_{n-1})$ is contained in $\mathrm{top}(K_n) = \mathrm{top}(I_n) \cap \mathrm{top}(Q_n)$.

Therefore, even though J_{n-1} depends on the choice of W_{n-1}, its order does not. As we have mentioned in the last chapter, coefficient ideals commute with blowup at equiconstant points, provided that we take their controlled transform. In the present case, the control for J_{n-1} will be $c_n = \mathrm{ord}_a K_n = o_n + q_n$.

As J_{n-1} passes under blowup to its controlled transform (at equiconstant points of K_n, the other points need not be considered since there (o_n, q_n) has already dropped), we will have again a factorization $J_{n-1} = M_{n-1} \cdot I_{n-1}$ (in W_{n-1} locally at a) with a monomial exceptional factor M_{n-1} and an ideal I_{n-1}. It is the second entry D_{n-1} of the combinatorial handicap D of $\mathcal{M}$ which prescribes the monomial factor, i.e., M_{n-1} is the ideal $I_{W_{n-1}}(D_{n-1} \cap W_{n-1})$ defining D_{n-1} in W_{n-1}. Of course, this is only a monomial ideal if D_{n-1} is transversal to W_{n-1}. The transversality will be ensured by the transformation rules for D_{n-1} and W_{n-1}. They are chosen so that the second factor I_{n-1} of J_{n-1} passes under blowup to its weak transform $I_{n-1}^\curlyvee$.

The story now repeats: The order of I_{n-1} will not increase at the points where (o_n, q_n) has remained constant, so $o_{n-1} = \mathrm{ord}_a I_{n-1}$ is the correct candidate for the third component of the invariant

$$i_a(\mathcal{M}) = (o_n, q_n, o_{n-1}, \ldots).$$

Taking this portion of the invariant, there will be again suspicious exceptional components to which the foreseen center of blowup $\tilde{Z}$ could be non-transversal. They are collected in the second entry E_{n-1} of the transversal handicap, yielding an ideal Q_{n-1} in W_{n-1}. Note here that E_{n-1} is defined globally in $W = W_n$ (though it is a stratified divisor), whereas

Q_{n-1} is defined only in the local hypersurface W_{n-1} at a. Setting $q_{n-1} = \mathrm{ord}_a Q_{n-1}$ we get

$$i_a(\mathcal{M}) = (o_n, q_n, o_{n-1}, q_{n-1}, \ldots),$$

form the product $K_{n-1} = I_{n-1} \cdot Q_{n-1}$ and take the coefficient ideal J_{n-2} of K_{n-1} in a local hypersurface W_{n-2} of W_{n-1} at a. This yields by induction on the dimension the complete invariant

$$i_a(\mathcal{M}) = (o_n, q_n, o_{n-1}, q_{n-1}, \ldots, o_1, q_1) \in \mathbb{N}^{2n}.$$

As all ideals I_i and Q_i pass under blowup to their weak transforms at points where the prior portion $(o_n, q_n, o_{n-1}, q_{n-1}, \ldots, o_{i+1}, q_{i+1})$ has not dropped, we can conclude that

$$i_{a'}(\mathcal{M}') \leq_{lex} i_a(\mathcal{M}).$$

So the invariant never increases lexicographically. Assume it remained constant. Then it remained constant also in dimension 1, $(o_1', q_1') = (o_1, q_1)$. But we saw in chapter V that the order of the weak transform of an ideal in one variable always drops to 0. So it could only remain constant if it was already equal to 0, say $I_1 = 1$ (Q_1 can be discarded because it equals always 1 – in one variable there is no transversality problem). Now, $I_1 = 1$ signifies that $J_1 = M_1$ is a monomial ideal.

Two cases: If $M_1 = 1$, then $J_1 = 1$ and K_2 is generated by a power of a variable, by the very definition of coefficient ideals. In this case, either $K_2 = 1$, $o_2 = q_2 = 0$ and we go one dimension higher as before, or not both o_2 and q_2 are simultaneously zero, with a forced drop of the order by the form of K_2 (to prove this could be an exercise if we wanted).

To conceptualize, let d be the maximal index so that $o_d = \ldots = o_1 = 0$. It can be shown that the transversality ideals $Q_d, \ldots, Q_1$ are all trivial and can therefore be discarded. As $o_d = 0$, the coefficient ideal $J_d = M_d$ of K_{d+1} is a principal monomial ideal. In this case we saw at the beginning of the last chapter (example $1\frac{1}{2}$) that if we choose the center of maximal possible dimension, we can make the *degree* of this monomial drop until the order of K_{d+1} has dropped. In this case, also our invariant will have dropped, so that induction applies.

In example $1\frac{1}{2}$ there occurred this ambiguity about symmetry. We now show how it is taken care of (again, not in the most economic way). Recall that $J_d = M_d$ is defined in W_d by restricting the combinatorial handicap D_d to W_d. As D_d is a collection of exceptional components, and as it is labelled, each of its shortcuts N_d comes with a different label (usually, the labels chosen for single components of D are just their birth date in the resolution process). Once the labels are chosen (and there is a prescibed rule how to do it when passing from D to D'), they are intrinsic information and respect any symmetry of the original mobile we wanted to resolve. Therefore, selecting the shortcut of D_d with maximal label among those whose order is $\geq$ the order of K_d (again, locally at a) and which do not have any proper shortcut of order $\mathrm{ord}_a K_d$ is a well defined process which is compatible with the requirements of *equivariance.* This *maximal tight* shortcut N_d of D_d has regular top locus $\mathrm{top}(N_d)$ which will form our desired center Z.

By the choice of the shortcut, Z has maximal dimension as a regular subscheme of $\mathrm{top}(K_{d+1})$, so the degree of M_d will drop after blowup, with an eventual drop of the order of K_{d+1} (if not one of the earlier components of $i_a(\mathcal{M})$ has already dropped). The center is transversal to the existing exceptional locus F, because it lies in $\mathrm{top}(Q_n)\cap\ldots\mathrm{top}(Q_{d+1})$ so that it is contained in all suspicious exceptional components. It lies in $\mathrm{top}(I_n)\cap\ldots\cap\mathrm{top}(I_{d+1})$ by the presence of the orders $o_n,\ldots,o_{d+1}$ in the invariant.

In all this we use the inclusions

$$\mathrm{top}(I_i)\supset\mathrm{top}(P_i),\ \mathrm{top}(E_i)=\mathrm{top}(Q_i)\supset\mathrm{top}(K_i)\qquad\text{and}$$
$$\ldots\supset\mathrm{top}(K_{i+1})\supset\mathrm{top}(J_i,c_{i+1})\supset\mathrm{top}(P_i)\supset\mathrm{top}(K_i)\supset\ldots\supset Z,$$

where $c_{i+1}=\mathrm{ord}_aK_i$ is the control for J_i and $\mathrm{top}(J_i,c_{i+1})$ denotes the locus of points where J_i has order at least c_{i+1}. The ideal P_i is the *companion ideal* of I_i. It is almost always equal to I_i, except in case when the order of I_i has become too small. Then it is introduced as a technical device to ensure the inclusion $\mathrm{top}(P_i)\subset\mathrm{top}(J_i,c_{i+1})$, and K_i is defined as $P_i\cdot Q_i$ instead of $K_i=I_i\cdot Q_i$. Its precise definition is

$$P_i=I_i+M^{\frac{o_i}{c_{i+1}-o_i}}\qquad\text{if }\ 0<o_i=\mathrm{ord}_aI_i<c_{i+1},$$
$$P_i=I_i\qquad\text{otherwise}$$

(don't worry about the rational exponent). The product K_i is called the *composition ideal*.

The invariant is intrinsic (i.e., does not depend on any choices), mainly, because the local flag $W_n\supset\ldots\supset W_1$ is chosen so as to maximize the orders of the induced coefficient ideals. We say that W_i has *weak maximal contact* with K_{i+1}.

Adding the pair m_d consisting of the order and the label of the maximal tight shortcut N_d of M_d to our invariant $i_a(\mathcal{M})$, we conclude that its top locus $Z=\mathrm{top}(i_a(\mathcal{M}))$ has all the required properties to be chosen as center. It is independent of any choices, regular, transversal to the exceptional locus, and satisfies the requirements of equivariants. So blow up Z in W, say $\pi:W'\to W$.

Taking the transformed mobile $\mathcal{M}'=(\mathcal{J}',c',D',E')$ of $\mathcal{M}=(\mathcal{J},c,D,E)$ in W' (where the ideal $\mathcal{J}$ passes to the controlled transform $\mathcal{J}^!$ of $\mathcal{J}$ with respect to c, c remains constant and D and E transform so as to capture the exceptional monomial factors of J_i respectively the suspicious exceptional components in each dimension) it follows from the various aspects described above that the invariant drops,

$$i_{a'}(\mathcal{M}')<_{lex}i_a(\mathcal{M}),$$

until it equals its minimal value $(0,\ldots,0)$. In this case the mobile has reached its final parking position and the seat-belt signs will be switched off, $J_n=M_n=I_W(D_n)$.

All this was for motivations. It remains to define the objects systematically. For further details, see the paper [EH] of Encinas and the author, or [Ha 3].

A *punctual setup* of $\mathcal{M}$ at a is a sequence $(J_n, \dots, J_1)$ of stalks of ideals J_i in a local flag $(W_n, \dots, W_1)$ of W at a satisfying for all $i \leq n$

(1) $J_i = M_i \cdot I_i$ with $M_i = I_{W_i}(D_i \cap W_i)$ and I_i an ideal in W_i at a.

(2) M_i defines a normal crossings divisor in W_i at a.

(3) W_{i-1} has weak maximal contact at a with the composition ideal $K_i = P_i \cdot Q_i$ in W_i of (J_i, c_{i+1}, D_i, E_i). Here, c_{i+1} is the control of J_i on W_i, given for $i < n$ as the order of K_{i+1} in W_{i+1} at a, and setting $c_{n+1} = c$.

(4) J_{i-1} is the coefficient ideal of K_i in W_{i-1}.

Setups depend on and are determined by the choice of the local flag subject to the above conditions. They commute with the operations described in *equivariance*.

Let $\mathcal{M} = (\mathcal{J}, c, D, E)$ be a mobile in W. Assume that $\mathcal{M}$ admits locally on W punctual setups $(J_n, \dots, J_1)$. Set

$$i_a(\mathcal{M}) = (t_n, \dots, t_1) \in \mathbb{N}^{4n}$$

with $t_i = (o_i, k_i, m_i)$ the *tag* of (J_i, c_{i+1}, D_i, E_i) at a, $o_i = \mathrm{ord}_a I_i$, $k_i = \mathrm{ord}_a K_i = \mathrm{ord}_a I_i + \mathrm{ord}_a Q_i$ and m_i the order and label of the maximal tight shortcut N_i of M_i of order $\geq c_{i+1}$ (m_i is set equal to $(0,0)$ if $i \neq d$ for d maximal with $o_d = 0$). Equipping $\mathbb{N}^{4n}$ with the lexicographic order this vector satisfies the following properties.

$i_a(\mathcal{M})$ does not depend on the chosen setup of $\mathcal{M}$ at a and commutes with the operations described in *equivariance*.

The map $a \to i_a(\mathcal{M})$ is upper-semicontinuous on V. The induced stratification of V refines the stratification underlying D and E.

The top locus Z of $i_a(\mathcal{M})$ is regular. Locally, Z lies in the top loci of all I_i, P_i, Q_i and K_i. It only depends on the restriction of $i_a(\mathcal{M})$ to the support of $\mathcal{J}$.

Z is transversal to all D_i and E_i.

We have seen above that then, for the transform $\mathcal{M}'$ of $\mathcal{M}$ under blowup, our resolution invariant drops

$$i_{a'}(\mathcal{M}) < i_a(\mathcal{M})$$

as long as $\mathcal{M}$ is not resolved. Done!

BIBLIOGRAPHY

[Ab 1] Abhyankar, S., *Local uniformization of algebraic surfaces over ground fields of characteristic* $p \neq 0$, Ann. Math. **63** (1956), 491-526.

[Ab 2] Abhyankar, S., *Resolution of singularities of arithmetical surfaces*, In: Arithmetical Algebraic Geometry, Harper and Row 1965, 111-152.

[Ab 3] Abhyankar, S., *Resolution of singularities of embedded algebraic surfaces*, Acad. Press 1966. 2nd edition, Springer 1998.

[Ab 4] Abhyankar, S., *An algorithm on polynomials in one determinate with coefficients in a two dimensional regular local domain*, Annali Mat. Pura Appl., Serie 4, **71** (1966), 25-60.

[Ab 5] Abhyankar, S., *Three dimensional embedded uniformization in characteristic p*, Lectures at Purdue Univeristy, Notes by M.F. Huang 1968.

[Ab 6] Abhyankar, S., *Good points of a hypersurface*, Adv. Math. **68** (1988), 87-256.

[Ab 7] Abhyankar, S., *Current status of the resolution problem*, Summer Institute on Algebraic Geometry 1964. Proc. Amer. Math. Soc.

[Ab 8] Abhyankar, S., *Desingularization of plane curves*, In: Summer Institute on Algebraic Geometry. Arcata 1981, Proc. Symp. Pure Appl. Math. **40** Amer. Math. Soc.

[Ab 9] Abhyankar, S., *Analytic desingularization in characteristic zero*, Appendix to the 2nd edition of [Ab 3].

[Ab 10] Abhyankar, S., *Algebraic Geometry for Scientists and Engineers*, Math. Surveys and Monographs **35**. Amer. Math. Soc. 1990.

[Ab 11] Abhyankar, S., *Ramification theoretic methods in algebraic geometry*, Princeton Univ. Press 1959.

[Ab 12] Abhyankar, S., *On the valuations centered in a local domain*, Amer. J. Math. **78** (1956), 321-348.

[Ab 13] Abhyankar, S., *Resolution of Singularities of Algebraic Surfaces*, In: Proceedings of the 1968 Bombay International Colloquium, Oxford University Press 1969, 1-11.

[AHV1] Aroca, J.-M., Hironaka, H., Vicente, J.-L., *The theory of the maximal contact*, Memorias Mat. Inst. Jorge Juan Madrid 29 (1975).

[AHV2] Aroca, J.-M., Hironaka, H., Vicente, J.-L., *Desingularization theorems*, Memorias Mat. Inst. Jorge Juan Madrid 29 (1975).

[AJ] Abramovich, D., de Jong, J., *Smoothness, semi-stability and toroidal geometry*, J. Alg. Geometry **6** (1997), 789-801.

[AM] Abhyankar, S., Moh, T.-T., *Newton-Puiseux expansion and generalized Tschirnhausen transformation I, II*, J. Reine Angew. Math. **260** (1973), 47-83, **261** (1973), 29-54.

[At] Artin, M., *Lipman's proof of resolution of singularities*, In: Arithmetic Geometry (eds. G. Cornell, J. H. Silverman), Springer 1986.

[Bb] Brandenberg, M., *Aufblasungen affiner Varietäten*, Thesis Zürich 1992.

[Be] Berthelot, P., *Altérations de variétés algébriques*, Séminaire Bourbaki 1995/96. Astérisque **241** (1997), 273-311.

[BE] Ban, C., MacEwan, L., *Canonical resolution of a quasi-ordinary surface singularity*, Preprint Ohio State University 1998.

[BEV] Bravo, A., Encinas, S., Villamayor, O., *A simplified proof of desingularization and applications*, Preprint.

[BK] Brieskorn, E., Knörrer, H., *Ebene algebraische Kurven*, Birkhäuser 1981. English translation: Plane algebraic curves. Birkhäuser 1986.

[BM 1] Bierstone, E., Milman, P., *Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant*, Invent. Math. 128 (1997), 207-302.

[BM 2] Bierstone, E., Milman, P., *Uniformization of analytic spaces*, J. Amer. Math. Soc. 2 (1989), 801-836.

[BM 3] Bierstone, E., Milman, P., *A simple constructive proof of canonical resolution of singularities*, In: Effective Methods in Algebraic Geometry (eds. T. Mora, C. Traverso). Progress in Math. 94, Birkhäuser 1991, 11-30.

[BM 4] Bierstone, E., Milman, P., *Semianalytic and subanalytic sets*, Publ. Math. IHES 67 (1988), 5-42.

[BM 5] Bierstone, E., Milman, P., *Resolution of singularities*, In: Several complex variables (Berkeley, 1995-1996), 43-78. Math. Sci. Res. Inst. Publ. 37, Cambridge Univ. Press, Cambridge, 1999.

[BM 6] Bierstone, E., Milman, P., *Desingularization algorithms. I. Role of exceptional divisors*, Mosc. Math. J. **3** (2003), 751-805.

[Bn] Bennett, B.-M., *On the characteristic function of a local ring*, Ann. Math. **91** (1970), 25-87.

[Bo] Bodnár, G., *Computation of blowing up centers*, J. Pure Appl. Algebra **179** (2003), 221-233.

[BP] Bogomolov, F., Pantev, T., *Weak Hironaka Theorem*, Math. Res. Letters **3** (1996), 299-307.

[Br 1] Brieskorn, E., *Über die Auflösung gewisser Singularitäten von holomorphen Abbildungen*, Math. Ann. **166** (1966), 76-102.

[Br 2] Brieskorn, E., *Die Auflösung der rationalen Singularitäten holomorpher Abbildungen*, Math. Ann. **178** (1968), 255 - 270.

[BS 1] Bodnár, G., Schicho, J., *Automated Resolution of Singularities for Hypersurfaces*, J. Symb. Comp. 30 (2000), 401-428.

[BS 2] Bodnár, G., Schicho, J., *A computer program for the resolution of singularities*, In: Resolution of singularities (ed. H. Hauser at al.), 231-238, Progr. in Math. 181, Birkhäuser, 2000.

[BS 3] Bodnár, G., Schicho, J., *Two computational techniques for singularity resolution*, Computer algebra and mechanized reasoning (St. Andrews, 2000). J. Symbolic Comput. **32** (2001), 39-54.

[BEV] Bravo A., Encinas S., Villamayor O., *A simplifed proof of desingularization and applications*, Rev. Mat. Iberoamericana **21** (2005), 349-458.

[BV] Bravo, A., Villamayor, O., *A strengthening of resolution of singularities in characteristic zero*, Proc. London Math. Soc. **86** (2003), 327-357.

[CGO] Cossart, V., Giraud, J., Orbanz, U., *Resolution of surface singularities*, Lecture Notes in Math. vol. 1101, Springer 1984.

[Co 1] Cossart, V., *Desingularization of embedded excellent surfaces*, Tôhoku Math. J. **33** (1981), 25-33.

[Co 2] Cossart, V., *Polyèdre caractéristique d'une singularité*, Thèse d'Etat, Orsay 1987.

[Co 3] Cossart, V., *Uniformisation et désingularisation des surfaces d'après Zariski*, In: Resolution of Singularities (ed.: H. Hauser et al.), Progress in Math. 181, Birkhäuser 2000.

[Co 4] Cossart, V., *Contact maximal en caractéristique positive et petite multiplicité*, Duke Math. J. **63** (1991), 57-64.

[Co 5] Cossart, V., *Desingularization in dimension two*, In: Resolution of surface singularities. Lecture Notes in Math. vol. 1101, Springer 1984.

[CP] Cano, F., Piedra, R., *Characteristic polygon of surface singularities*, In: Géométrie algébrique et applications II (eds. J.-M. Aroca, T. Sánchez-Giralda, J.-L. Vicente), Proc. of Conference on Singularities, La Rábida 1984. Hermann 1987.

[Cu 1] Cutkosky, D., *When are the blow ups of two ideals the same scheme?*, Comm. Algebra **19** (1991), 1119-1124.

[Cu 2] Cutkosky, D., *Local factorization of birational maps*, Adv. Math. **132** (1997), 167-315.

[Cu 3] Cutkosky, D., *Local factorization and monomialization of morphisms*, Astérisque 260 (1999).

[Cu 4] Cutkosky, D., *Monomialization of morphisms from 3-folds to surfaces*, Lecture Notes in Mathematics, 1786. Springer 2002.

[Cu 5] Cutkosky, D., *Resolution of morphisms*, Valuation theory and its applications, Vol. I (Saskatoon 1999), 117-140, Fields Inst. Commun., 32, Amer. Math. Soc. 2002.

[Cu6] Cutkosky, D., *Generically finite morphisms and simultaneous resolution of singularities*, In: Commutative algebra (Grenoble/Lyon, 2001), 75-99, Contemp. Math. 331, Amer. Math. Soc. 2003.

[Cu 7] Cutkosky, D., *Resolution of singularities*, Graduate Studies in Mathematics, 63. Amer. Math. Soc. 2004.

[DCP] De Concini, C., Procesi, C., *Wonderful models of subspace arrangements*, Selecta Math. **1** (1995), no. 3, 459-494.

[dJ] de Jong, A.J., *Smoothness, semi-stability and alterations*, Publ. Math. IHES **83** (1996), 51-93.

[EiH] Eisenbud, D., Harris, J., *The geometry of schemes*, Graduate Texts in Mathematics, 197. Springer 2000.

[EH] Encinas, S., Hauser, H., *Strong resolution of singularities in characteristic zero*, Comment. Math. Helv. **77** (2002), 821-845.

[En] Encinas, S., *Constructive resolution of singularities of families of schemes*, Thesis, Valladolid 1996.

[ENV] Encinas, S., Nobile, A., Villamayor, O., *On algorithmic equi-resolution and stratification of Hilbert schemes*, Proc. London Math. Soc. **86** (2003), 607-648.

[EV 1] Encinas, S., Villamayor, O., *Good points and constructive resolution of singularities*, Acta Math. **181** (1998), 109-158.

[EV 2] Encinas, S., Villamayor, O., *A course on constructive desingularization and equivariance*, In: Resolution of Singularities (ed.: H. Hauser et al.), Progress in Math. 181, Birkhäuser 2000.

[EV 3] Encinas, S., Villamayor, O., *A new proof of desingularization over fields of characteristic zero*, Rev. Mat. Iberoamericana **19** (2003), no. 2, 339-353.

[FK] Feichtner, E., Koslov, D., *Incidence combinatorics of resolutions*, Selecta Math. **10** (2004), 37-60.

[Gi 1] Giraud, J., *Etude locale des singularités*, Cours de 3e Cycle, Orsay 1971/72.

[Gi 2] Giraud, J., *Sur la théorie du contact maximal*, Math. Z. **137** (1974), 285-310.

[Gi 3] Giraud, J., *Contact maximal en caractéristique positive*, Ann. Scient. Ec. Norm. Sup. Paris **8** (1975), 201-234.

[Gi 4] Giraud, J., *Desingularization in low dimension. In: Resolution of surface singularities*, Lecture Notes in Math. vol. 1101, Springer 1984, 51-78.

[Gr] Grothendieck, A., *Traveaux de Heisouke Hironaka sur la résolution des singularités*, Actes Congrès International Math., Nice 1970.

[GT] Goldin, R., Teissier, B., *Resolving singularities of plane analytic branches with one toric morphism*, In: Resolution of Singularities, Progress in Math. 181, Birkhäuser 2000.

[Ha 1] Hauser, H., *Seventeen obstacles for resolution of singularities*, In: The Brieskorn Anniversary volume 1996. Progress in Math. 162, Birkhäuser 1997.

[Ha 2] Hauser, H., *Excellent surfaces and their taut resolution*, In: Resolution of Singularities, Progress in Math. 181, Birkhäuser 2000.

[Ha 3] Hauser, H., *The Hironaka Theorem on resolution of singularities*, Bull. Amer. Math. Soc. **40** (2003), 323-403.

[Ha 4] Hauser, H., *Why Hironaka's proof of resolution of singularities fails in positive characteristic*, Preprint University of Innsbruck, 51 pp.

[Ha 5] Hauser, H., *Three power series techniques*, Proc. London Math. Soc. **88** (2004), 1-24.

[Ha 6] Hauser, H., Regensburger, G., *Explizite Auflösung von ebenen Kurvensingularitäten in beliebiger Charakteristik*, L'Ens. Mathém. **50** (2004), 305-353.

[Ha 7] Hauser, H., *Desingularization of ideals and varieties*, Lecture notes of a course at CIRM, Luminy 2005, 18 pp.

[Hi 1] Hironaka, H., *Resolution of singularities of an algebraic variety over a field of characteristic zero*, Ann. Math. 79 (1964), 109-326.

[Hi 2] Hironaka, H., *Characteristic polyhedra of singularities*, J. Math. Kyoto Univ. **7** (1967), 251-293.

[Hi 3] Hironaka, H., *Certain numerical characters of singularities*, J. Math. Kyoto Univ. **10** (1970), 151-187.

[Hi 4] Hironaka, H., *Desingularization of excellent surfaces*, Notes by B. Bennett at the Conference on Algebraic Geometry, Bowdoin 1967. Reprinted in: Cossart, V., Giraud, J., Orbanz, U., *Resolution of surface singularities*, Lecture Notes in Math. 1101, Springer 1984.

[Hi 5] Hironaka, H., *Additive groups associated with points of a projective space*, Ann. Math. **92** (1970), 327-334.

[Hi 6] Hironaka, H., *Idealistic exponents of singularity*, In: Algebraic Geometry, the Johns Hopkins Centennial Lectures. Johns Hopkins University Press 1977.

[Hi 7] Hironaka, H., *Schemes etc*, In: 5th Nordic Summer School in Mathematics (ed. F. Oort), Oslo 1970, 291-313.

[Hi 8] Hironaka, H., *Desingularization of complex analytic varieties*, Actes Congrès International Math., Nice 1970, 627-631.

[Hi 9] Hironaka, H., *On the presentations of resolution data*, (Notes by T.T. Moh), In: Algebraic Analysis, Geometry and Number Theory 1988 (ed. J.I. Igusa), The Johns Hopkins University Press 1989.

[Hi 10] Hironaka, H., *Bimeromorphic smoothing of a complex analytic space*, Mimeographed notes, University of Warwick 1971.

[Hi 11] Hironaka, H., *Stratification and flatness*, In: Nordic Summer School, Oslo 1976 (ed. P. Holm), Sijthoff and Noordhoff 1977.

[Hi 12] Hironaka, H., Preface to *Zariski's collected papers I*, 307-312. MIT Press 1979.

[HIO] Herrmann, M., Ikeda, S., Orbanz, U., *Equimultiplicity and blowing up*, Springer 1988.

[HR] Hauser, H., Regensburger, G., *Explizite Auflösung von ebenen Kurvensingularitäten*, L'Enseign. Math. (2004)

[Ju] Jung, H., *Darstellung der Funktionen eines algebraischen Körpers zweier unabhängiger Veränderlicher x, y in der Umgebung einer Stelle $x = a$, $y = b$*, J. Reine Angew. Math. **133** (1908), 289-314.

[Ko] Kollár, J., *Lectures on resolution of singularities*, Preprint 2005, ArXiv: math.AG/0508332.

[KV] Kiyek, K.-H., Vicente, J. L., *Resolution of curve and surface singularities in characteristic zero*, Algebras and Applications, Kluwer 2004.

[Kz] Kunz, E., *Algebraische Geometrie IV*, Vorlesung Regensburg.

[LO] Lê, D.T., Oka, M., *On resolution complexity of plane curves*, Kodai Math. J. **18** (1995), 1-36.

[Lp 1] Lipman, J., *Desingularization of 2-dimensional schemes*, Ann. Math. **107** (1978), 151-207.

[Lp 2] Lipman, J., *Rational singularities, with applications to algebraic surfaces and unique factorization*, Publ. Math. IHES **36** (1969), 195-279.

[Lp 3] Lipman, J., *Introduction to resolution of singularities*, Proceedings Symp. Pure Appl. Math. **29** Amer. Math. Soc. 1975, 187-230.

[LT] Lejeune, M., Teissier, B., *Contribution à l'étude des singularités du point de vue du polygone de Newton*, Thèse d'Etat, Paris 1973.

[Ma] Matsuki, K., *Notes on the inductive algorithm of resolution of singularities by S. Encinas and O. Villamayor*, ArXiv: math.AG/0103120

[Mo 1] Moh, T.-T., *On a stability theorem for local uniformization in characteristic p*, Publ. RIMS Kyoto Univ. **23** (1987), 965-973.

[Mo 2] Moh, T.-T., *Canonical uniformization of hypersurface singularities of characteristic zero*, Comm. Alg. **20** (1992), 3207-3249.

[Mo 3] Moh, T.-T., *On a Newton polygon approach to the uniformization of singularities of characteristic p*, In: Algebraic Geometry and Singularities (eds. A. Campillo, L. Narváez). Proc. Conf. on Singularities La Rábida. Birkhäuser 1996.

[Mu] Mulay, S.B., *Equimultiplicity and hyperplanarity*, Proc. Amer. Math. Soc. **87** (1983), 407-413.

[MZ] Muhly, H.T., Zariski, O., *The resolution of singularities of an algebraic curve*, Amer. J. Math. **61** (1939), 107-114.

[Na 1] Narasimhan, R., *Monomial equimultiple curves in positve characteristic*, Proc. Amer. Math. Soc. **89** (1983), 402-413.

[Na 2] Narasimhan, R., *Hyperplanarity of the equimultiple locus*, Proc. Amer. Math. Soc. **87** (1983), 403-406.

[No] Noether, M., *Über die algebraischen Functionen einer und zweier Variabeln*, Gött. Nachr. (1871), 267-278.

[Od 1] Oda, T., *Infinitely very near singular points*, Adv. Studies Pure Math. **8** (1986), 363-404.

[Od 2] Oda, T., *Hironaka's additive group scheme*, In: Number Theory, Algebraic Geometry and Commutative Algebra, in honor of Y. Akizuki, Kinokuniya, Tokyo, 1973, 181-219.

[Od 3] Oda, T., *Hironaka group schemes and resolution of singularities*, In: Proc. Conf. on Algebraic Geometry, Tokyo and Kyoto 1982. Lecture Notes in Math. 1016, Springer 1983, 295-312.

[Ok] Oka, M., *Geometry of plane curves via toroidal resolution*, In: Algebraic Geometry and Singularities (eds. A. Campillo, L. Narváez). Proc. Conf. on Singularities La Rábida. Birkhäuser 1996.

[Or] Orbanz, U., *Embedded resolution of algebraic surfaces after Abhyankar (characteristic 0)*, In: Resolution of surface singularities. Lecture Notes in Math. vol. 1101, Springer 1984.

[Pf] Pfeifle, J., *Aufblasung monomialer Ideale*, Diplomarbeit Innsbruck 1996.

[Ro] Rosenberg, J., *Blowing up nonreduced toric subschemes of A^n*, Preprint 1998.

[RS] Rodríguez Sánchez, C., *Good points and local resolution of threefold singularities*, Thesis, Univ. Léon 1998.

[Si 1] Singh, B., *Effect of permissible blowing up on the local Hilbert function*, Invent. Math. **26** (1974), 201-212.

[Si 2] Singh, B., *Formal invariance of local characteristic functions*, In: Seminar D. Eisenbud, B. Singh, W. Vogel, vol. 1, 44-59. Teubner-Texte zur Math. 29, 1980.

[Sp 1] Spivakovsky, M., *A solution to Hironaka's Polyhedra Game*, In: Arithmetic and Geometry. Papers dedicated to I.R. Shafarevich on the occasion of his sixtieth birthday, vol II (eds.: M. Artin, J. Tate). Birkhäuser 1983, 419-432.

[Sp 2] Spivakovsky, M., *A counterexample to Hironaka's 'hard' polyhedra game*, Publ. RIMS, Kyoto University **18** (1983), 1009-1012.

[Sp 3] Spivakovsky, M., *A counterexample to the theorem of Beppo Levi in three dimensions*, Invent. Math. **96** (1989), 181-183.

[T] Teissier, B., *Valuations, deformations, and toric geometry*, In: Valuation theory and its applications, Vol. II (Saskatoon 1999), 361-459, Fields Inst. Commun. 33, Amer. Math. Soc. 2003.

[Ur] Urabe, T., *Resolution of singularities of germs in characteristic positive associated with valuation rings of iterated divisor type*, Preprint 1999.

[Va] Vasconcelos, W., *Arithmetic of blowup algebras*, London Mathematical Society Lecture Note Series, 195. Cambridge University Press 1994.

[Vi 1] Villamayor, O., *Constructiveness of Hironaka's resolution*, Ann. Scient. Ec. Norm. Sup. Paris **22** (1989), 1-32.

[Vi 2] Villamayor, O., *Patching local uniformizations*, Ann. Scient. Ec. Norm. Sup. Paris **25** (1992), 629-677.

[Vi 3] Villamayor, O., *An introduction to the algorithm of resolution*, In: Algebraic Geometry and Singularities (eds. A. Campillo, L. Narváez). Proc. Conf. on Singularities La Rábida. Birkhäuser 1996.

[Vi 4] Villamayor, O., *On equiresolution and a question of Zariski*, Acta Math. **185** (2000), 123-159.

[Vi 5] Villamayor, O., *An introduction to constructive desingularization*, ArXiv: math.AG/ 0507537.

[Wa 1] Walker, R.J., *Reduction of singularities of an algebraic surface*, Ann. Math. **36** (1935), 336-365.

[Wa 2] Walker, R.J., *Algebraic Curves*, Princeton Univ. Press 1950. Reprinted: Dover 1962.

[W] Włodarzcyk, H., *Simple Hironaka resolution in characteristic zero*, J. Amer. Math. Soc. **18** (2005), 779-822.

[Yo] Youssin, B., *Newton polyhedra without coordinates*, Memoirs Amer. Math. Soc. **433** (1990), 1-74, 75-99.

[Za 1] Zariski, O., *The reduction of singularities of an algebraic surface*, Ann. Math. **40** (1939), 639-689.

[Za 2] Zariski, O., *Local uniformization on algebraic varieties*, Ann. Math. **41** (1940), 852-896.

[Za 3] Zariski, O., *A simplified proof for resolution of singularities of an algebraic surface*, Ann. Math. **43** (1942), 583-593.

[Za 4] Zariski, O., *Reduction of singularities of algebraic three dimensional varieties*, Ann. Math. **45** (1944), 472-542.

[Za 5] Zariski, O., *The compactness of the Riemann manifold of an abstract field of algebraic functions*, Bull. Amer. Math. Soc. **45** (1945), 683-691.

[Za 6] Zariski, O., *Exceptional singularities of an algebraic surface and their reduction*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., serie VIII, **43** (1967), 135-196.

[Za 7] Zariski, O., *A new proof of the total embedded resolution theorem for algebraic surfaces*, Amer. J. Math. **100** (1978), 411-442.

[Za 8] Zariski, O., *Polynomial ideals defined by infinitely near base points*, Amer. J. Math. **60** (1938), 151-204.

[Za 9] Zariski, O., *Algebraic Surfaces*, Ergebnisse der Mathematik, vol 61. 2nd edition, Springer 1971.

[Za 10] Zariski, O., *The concept of a simple point of an abstract algebraic variety*, Trans. Amer. Math. Soc. **62** (1947), 1-52.

[Za 11] Zariski, O., *Normal varieties and birational correspondences*, Bull. Amer. Math. Soc. **48** (1942), 402-413.

[Za 12] Zariski, O., *Collected papers, vol. I. Foundations of algebraic geometry and resolution of singularities*, MIT Press 1979.

[Za 13] Zariski, O., *Collected papers, vol. IV. Equisingularity*, MIT Press 1979.

Institut für Mathematik, Universität Innsbruck, Austria, www.hh.hauser.cc
E-mail address: herwig.hauser@uibk.ac.at

Proceedings of 12^{th} Gökova
Geometry-Topology Conference
pp. 49 – 58

Published online at
GokovaGT.org

Real algebraic structures

Selman Akbulut

Dedicated to the memory of Raoul Bott.

Abstract. A brief survey of real algebraic structures on topological spaces is given.

0. Introduction

The question of when a manifold M is homeomorphic (or diffeomorphic) to a real algebraic set V is an old one. If we start with an imbedding $M \subset \mathbb{R}^n$ and insist on finding an algebraic subset V of $\mathbb{R}^n$ which is isotopic to M in $\mathbb{R}^n$, the problem encounters additional difficulties coming from complexification. Hence it is natural to break the question into two parts: (1) Stable: If M homeomorphic (or diffeomorphic) to some real algebraic set. (2) Ambient: If M isotopic to a real algebraic subset in $\mathbb{R}^n$. While the first problem has a complete solution, the second one encounters obstructions. Here we give a quick summary of some of the related results. This brief survey is by no means complete, and it is biased towards author's interest in the field. For the basics the reader can consult the book [AK3].

1. Stable results

By [N] and [T] every closed smooth manifold is diffeomorphic to a nonsingular real algebraic set, and by [AK5] every closed PL manifold is homeomorphic to real algebraic set. Also if $M \subset V$ is a closed smooth submanifold of a nonsingular variety V, we can ask whether M can be made algebraic in $V \times \mathbb{R}^k$ for some large k . The complete solution of this is also known. To explain this we first need some definitions: We call a homology cycle of V *algebraic* if it is represented by a real algebraic subset. For example, it is known that all the Steifel-Whitney and Pontryagin classes of V are represented by algebraic cycles [AK1],[AK8]. Let $H_*^A(V, \mathbb{Z}_2)$ be the subgroup of $H_*(V, \mathbb{Z}_2)$ generated by the real algebraic subsets. We call a real algebraic set *totally algebraic* if $H_*(V, \mathbb{Z}_2) = H_*^A(V, \mathbb{Z}_2)$. Clearly $\mathbb{R}^n$ is totally algebraic, and the (unoriented) Grassmannians $G(k, n)$ of k planes in $\mathbb{R}^n$ are totally algebraic [AK1] (because their homology is generated by the Schubert cycles which are algebraic subsets).

1991 *Mathematics Subject Classification.* 58D27, 58A05, 57R65.

Key words and phrases. real algebraic sets.
The author is partially supported by NSF grant DMS 9971440.

Theorem 1. ([AK1]) *Every closed smooth submanifold $f : M \hookrightarrow V$ of a real algebraic set V is ϵ-isotopic to a real algebraic subset in $V \times \mathbb{R}^k$ for some large k, if and only if V is totally algebraic.*

$$\begin{array}{ccc} & & V \times \mathbb{R}^k \\ & \nearrow & \downarrow \\ M & \overset{f}{\hookrightarrow} & V \end{array}$$

Remark 1. *It should be noted that not every closed smooth manifold V is diffeomorphic to a nonsingular totally algebraic set* [BD1], *but surprisingly every closed smooth V is homeomorphic to a totally algebraic set Z* [AK4]. *Evidently the singularities of Z is related to the obstructions arising from* [BD1].

2. Ambient results

Theorem 2. ([AK6]) *Every closed smooth submanifold $M \subset \mathbb{R}^n$ is ϵ-isotopic to nonsingular points of an algebraic subset $V \subset \mathbb{R}^n$. That is M is isotopic to a topological component V_0 of a real algebraic set V which is nonsingular, and the other components $V - V_0$ are smaller dimensional.*

Here ϵ-isotopy means arbitrarily small isotopy. In the proof of the above theorem one can arrange so that the extra components $V - V_0$ are arbitrarily far away from V_0. Also this theorem implies that any $M \subset \mathbb{R}^n$ is ϵ-isotopic to a nonsingular algebraic subset $Z \subset \mathbb{R}^{n+1}$. This is because if $V = f^{-1}(0)$ and $V - V_0 = g^{-1}(0)$ for some polynomials $f(x)$ and $g(x)$, then we can take

$$Z = \{(x, t) \mid f(x) = 0 \ , \ tg(x) = 1\}.$$

There is also a more sophisticated version of this theorem for immersed submanifolds, to explain it we need some definitions: Recall that the Thom construction associates an imbedding of any closed smooth manifold $f : M^m \hookrightarrow \mathbb{R}^{m+k}$ an element in the homotopy group of the Thom space $[f] \in \pi_{m+k}(MO_k)$, which classifies imbeddings up to cobordism in $\mathbb{R}^n \times [0,1]$. In the more general case of f is an immersion, Thom construction gives an element in the homotopy group of the iterated loop space suspension $[f] \in \pi_{m+k}(\Omega^\infty \Sigma^\infty MO_k)$; this is a group which classifies immersions up to immersed cobordisms in $\mathbb{R}^n \times [0,1]$. Let $\pi^{alg}_{m+k}(\Omega^\infty \Sigma^\infty MO_k)$ be the subgroup generated by the cobordism classes of 'almost nonsingular' algebraic subsets. We call an algebraic subset of $\mathbb{R}^n$ an *almost nonsingular algebraic subset* if it is an image of a smoothly immersed manifold $f : M \looparrowright \mathbb{R}^n$, and each sheet of the immersion is nonsingular. In particular if f is an imbedding, then the almost nonsingular algebraic set corresponding to f is a nonsingular algebraic subset. Let $dim(M) = m$, and $n = m + k$, then we can state:

Theorem 3. ([AK7]) *An immersed closed smooth manifold $f : M \looparrowright \mathbb{R}^n$ is ϵ-isotopic to an almost nonsingular algebraic subset of $\mathbb{R}^n$ if and only if $[f] \in \pi^{alg}_{m+k}(\Omega^\infty \Sigma^\infty MO_k)$.*

In particular, if an imbedding of a closed smooth manifold $f : M \hookrightarrow \mathbb{R}^n$ is cobordant through immersions to a closed smooth submanifold $N \subset \mathbb{R}^{n-1} \times \{0\}$, then it is isotopic to a nonsingular algebraic subset. This is because, by the remarks following Theorem 2, N is ϵ-isotopic to a nonsingular algebraic subset of $\mathbb{R}^n$, then by Theorem 3 we can isotope M to a nonsingular algebraic subset of $\mathbb{R}^n$. So the relevant topological question is whether we can cobord $M \subset \mathbb{R}^n$ into a lower dimensional subspace of $\mathbb{R}^n$:

(i) When the normal bundle of $M \subset \mathbb{R}^n$ splits a trivial line bundle?

(ii) When $[f]$ lies in the image of the suspension map Σ?

$$\pi_{m+k-1}(MO_{k-1}) \xrightarrow{\Sigma} \pi_{m+k}(MO_k)$$

Answering the first question would give sufficient conditions when M immerses into $\mathbb{R}^{n-1}$, answering the second second question would help us decide if M is cobordant to a submanifold of $\mathbb{R}^{n-1}$. Ideally one would hope to reduce the answers to the conditions on characteristic classes of the normal bundle of M, then (ii) would be sufficient conditions isotoping $M \subset \mathbb{R}^n$ to a nonsingular real algebraic subset.

3. Real algebraic characteristic numbers

Surprisingly, we still don't know whether a closed smooth submanifold $M \subset \mathbb{R}^n$ is isotopic to a (singular or nonsingular) real algebraic subset (though we came close to answering this in the affirmative in Section 2). So one can try to find obstructions. Currently this can only be achieved either by relaxing the condition of smoothness of M (this section), or by strengthening the notion of non-singularity (Section 4). Underlying topological space of every algebraic set is a stratified space (a polyhedron in particular), so it is natural to generalize the above questions from smooth manifolds to stratified subspaces of $\mathbb{R}^n$.

Theorem 4. *(*[CK]*,* [AK3], [AK10]*) There is a stratified space $Z^3 \subset \mathbb{R}^4$ which is homeomorphic to an algebraic set, but can not be isotopic to an algebraic subset of $\mathbb{R}^4$.*

To explain this, we need to review the general program of topologically characterizing real algebraic sets, in particular we need to recall some topologically defined structures ('topological resolution tower structures') on stratified spaces, which enable us to identify the obstructions of making stratified spaces homeomorphic to algebraic sets. These structures on stratified sets give a topological model for algebraic sets.

In [AK3] a topological characterization program for real algebraic sets is introduced, here is a brief summary: Just like an algebraic number such as $2+\sqrt{3}$ is determined by a pair of integers and a monomial $\{2, 3, x^2 = 3\}$ (integers are glued by a monomial), an algebraic set V is determined by a collection of nonsingular algebraic sets and a collection of "compatible" monomial maps between them $\mathcal{F} = \{V_i, p_{ij}\}$, and V is obtained by gluing these nonsingular V_i's by monomial p_{ij}'s. We denote this by $|\mathcal{F}|$

$$V = |\mathcal{F}| = \cup V_i / p_{ij}(x) \sim x \tag{1}$$

(e.g. Figure 2 is obtained by gluing an S^2 and two copies of S^1 together by a fold map) This result is obtained by resolving various strata of V. These objects $\mathcal{F} = \{V_i, p_{ij}\}$ are called *algebraic resolution towers*. By imitating this, we can define analogous objects in the topological category $\mathcal{F}_{top} = \{M_i, p_{ij}\}$, where M_i are smooth manifolds and p_{ij} are certain topological version of monomial maps between them. As in (1), by gluing the objects of $\mathcal{F}_{top}$ by its maps we obtain certain stratified spaces $X = |\mathcal{F}_{top}|$. We call $\mathcal{F}_{top}$ *topological resolution towers*. We have the following categories of sets:

$\mathcal{A} = \{\mathcal{F}\}$: Algebraic resolution towers
$\mathcal{T} = \{\mathcal{F}_{top}\}$: Topological resolution towers
$|\mathcal{A}| = \{ |\mathcal{F}| \}$: Realization of algebraic resolution towers
$|\mathcal{T}| = \{ |\mathcal{F}_{top}| \}$: Realization of topological resolution towers

We have the following maps (2), where the vertical arrows are gluings (and they are onto by construction), the horizontal right-pointing arrows are the forgetful maps, and the top left-pointing arrow is the important *algebraization map*: it is a generalized version of Theorem 1 (it turns a collection of smooth manifolds and compatible topological maps between them into nonsingular algebraic sets and compatible rational maps, in such a way that gluing them gives an algebraic set).

$$\begin{array}{ccc} \mathcal{A} & \overrightarrow{\leftarrow} & \mathcal{T} \\ \downarrow & & \downarrow \\ |\mathcal{A}| & \rightarrow & |\mathcal{T}| \end{array} \tag{2}$$

Therefore the stratified sets in $|\mathcal{T}|$ topologically characterize the real algebraic sets in $|\mathcal{A}|$. Also if $\mathcal{A}lg$ denotes the category of all real algebraic sets, by the above gluing process (1) we get a surjection $\gamma : \mathcal{A}lg \rightarrow |\mathcal{A}|$. This appears to give a complete topological characterization of all real algebraic sets. Unfortunately this is not quite so. For this we need that the elements $\mathcal{F} = \{V_i, p_{ij}\}$ in the image of γ to be *submersive*, that is we need the maps p_{ij} to be "submersive" on each strata. We may impose this property as the part of the definition of $\mathcal{T}$ (this makes the diagram (2) commutative). This property is known to exist for algebraic sets of dimension < 4, and it would hold in general provided that there is a certain map version of the "resolution of singularities theorem", which we don't know if it exists or whether it follows from a modification of the usual resolution of singularities theorem of Hironaka [AK3], [K].

There is an interesting subclass of stratified spaces in $|\mathcal{T}|$, called *A-spaces*, which behave nicely on PL manifolds. For example, the existence of this structure on PL manifolds can be reduced to an algebraic topology problem, i.e. a bundle lifting problem (fortunately with zero obstruction). Furthermore A- spaces are submersive elements of $\mathcal{T}$. This is why all PL manifolds are homeomorphic to real algebraic sets [AK2] [AK5], [AT].

Therefore any topological obstruction for a stratified space X to lie in $|\mathcal{T}|$ is an obstruction X to be isomorphic (as stratified space) to an algebraic set. It is already known by Sullivan, that every real algebraic set must be an *Euler space* (a stratified set such that the link of every point has even Euler characteristic). It turns out that in dimensions ≤ 2

this is also the sufficient condition for a stratified set to be homeomorphic to an algebraic set. This is proven by showing that every 2-dimensional Euler spaces lie in $|\mathcal{T}|$ ([AK1], [AK3], [BD2]). By studying the topology of $\mathcal{F}_{top} = \{M_i, p_{ij}\}$ carefully, one can start defining inductively a sequence of characteristic numbers on n-dimensional Euler spaces X_n (here $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$)

$$\beta = \beta(n) : X_n^{(0)} \to \mathbb{Z}_2^{d(n)}$$

whose vanishing is necessary and sufficient (or just necessary) for X_n to lie in $|\mathcal{T}|$, where $X_n^{(0)}$ is the 0-skeleton of X_n. In our context this means that the $n-1$ dimensional links of the vertices of X, which are already in $\mathcal{T}$, should bound in $\mathcal{T}$. Roughly $\beta(n)$ are the cobordism characteristic numbers of $n-1$ dimensional elements of $\mathcal{T}$.

In [AK3] this program was carried out in the first nontrivial case of $n = 3$. It turns out $d(3) = 4$. In fact it was shown that X_3 is homeomorphic to an algebraic set if and only if its characteristic numbers $\beta = 0$. It goes as follows: To every 1-dimensional Euler space X_1 we associate numbers $\alpha_j = \alpha_j(X_1)$ for $j = 0, .., 7$, which are numbers of vertices of X_1 whose links has j $(mod\ 8)$ number of points. Then to X_1 we associate the following (well defined after subdivision) 4-tuple number (e.g. Figure 1)

$$\epsilon(X_1) = (\alpha_0, \alpha_6, (\alpha_0 + \alpha_4)/2, (\alpha_2 + \alpha_6)/2) \in \mathbb{Z}_2^4$$

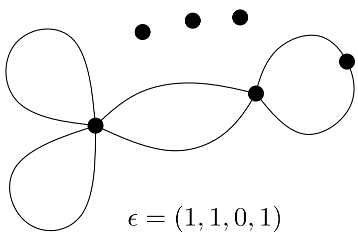


Figure 1.

Hence to the vertices of any 2-dimensional Euler space X_2 we can associated 4-tuple numbers (because their links are 1-dimensional Euler spaces). For the example of Figure 2 we calculated these numbers (and drew the links of its vertices)
Now for $\epsilon \in \mathbb{Z}_2^4$, we can associate numbers to the verticies of any 3-dimensional Euler space: $\beta_\epsilon : X_3^{(0)} \to \mathbb{Z}_2$ by $p \mapsto$*number of vertices (mod 2) in* $L := link(p)$*, with* $\epsilon(L) = \epsilon$.
Then the definition of $\beta : X_3^{(0)} \to \mathbb{Z}_2^4$ is given by the following expression

$$\beta = (\beta_{0100} + \beta_{0101},\ \beta_{1000} + \beta_{1001},\ \beta_{1100} + \beta_{1101},\ \beta_{1110} + \beta_{1111})$$

For example, if X_3 is the suspension of the 2-dimensional Euler space of Figure 2, it is not homeomorphic to an algebraic set; while the suspension Z of the Euler space of Figure 3 is homeomorphic to a real algebraic set. This Z satisfies Theorem 4.

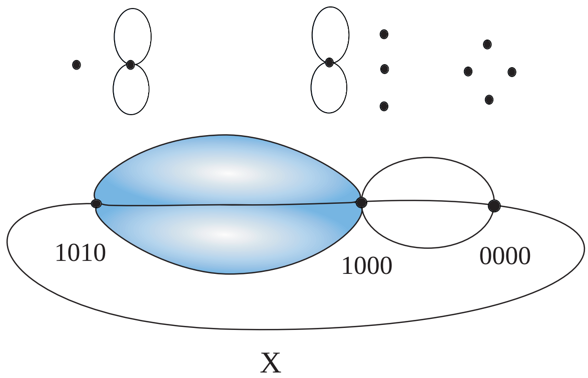


FIGURE 2.

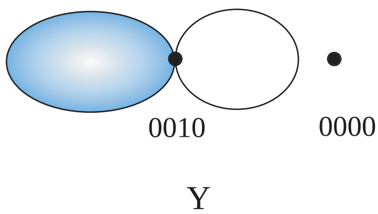


FIGURE 3.

More recently the "necessary" part of this program was generalized to dimension 4 by [P] and [MP2] (where [P] generalizes this approach, whereas [MP2] uses an alternative approach), and also [P] finds sufficient invariants (not all in $\mathbb{Z}_2$ but in a larger group). One difficulty with this is $d(n)$ grows too fast to calculate, for example [MP2] shows $d(4) \geq 2^{43} - 43$.

Note that in examples above the Euler characteristic *mod* 4 of the link of the 1-dimensional strata is generically constant. This is no coincidence, in fact it was observed in [CK] that general real algebraic varieties $V \subset X$ satisfy the property (*): The Euler characteristic $\chi(lk_x(V,X))$ of the link of V in X ($x \in V$) is generically constant *mod* 4 (also see [AK3], [MP1]). In [CK] this property is used to show that the set $Z \subset \mathbb{R}^4$ above satisfies the second claim of Theorem 4 as follows: Let $I \subset Z$ be the arc, which is the suspension of the isolated point of Y. By (*) in any algebraic model of Z, the Zariski closure of I is a 1-dimensional algebraic subset $V \subset Z$ such that $I' = V - I$ generically lies in the 3-manifold part of Z. Now if $Z \subset \mathbb{R}^4$ were an algebraic hyper-surface, we can choose a square free polynomial equation $f(x) = 0$ of Z. Say, f takes signs $\pm$ on the inside and outside regions $B_\pm \subset \mathbb{R}^4$ separated by $S^3 \subset Z$. Define

$$\widetilde{\mathbb{R}^4} := \{(x,t) \in \mathbb{R}^4 \times \mathbb{R} \mid t^2 = f(x)\} \supset Z \supset V$$

Then $\chi(lk_x(V,\widetilde{\mathbb{R}^4})) - \chi(lk_x(V,\mathbb{R}^4)) = \chi(lk_x(V,B_+)) - \chi(lk_x((V,B_-))$ mod 4 is not generically constant (it is 2 or 0 when x is in the interior of I or I'), this violates (*).

4. Transcendental manifolds

If we ask whether a smooth submanifold $M \subset \mathbb{R}^n$ is isotopic to a nonsingular real algebraic set V in a strong sense, then we can find genuine obstructions to doing this, even when M is already nonsingular algebraic set in $\mathbb{R}^n$. Here nonsingular in the "strong sense" means V is the real part of a nonsingular complex variety in $\mathbb{CP}^n$, where $\mathbb{R}^n \subset \mathbb{RP}^n$ is identified with one of the standard charts.

Theorem 5. *(*[AK9]*) There are smooth submanifolds of $M \subset \mathbb{RP}^n$ which are isotopic to nonsingular real algebraic subsets, but can not be isotopic to the real parts of nonsingular complex algebraic subsets of $\mathbb{CP}^n$.*

We will break the proof into elementary steps which are mostly special cases of more general results (e.g. [AK3]), some of which are already mentioned in Section 1. Since possible, we will outline the proofs from scratch for the benefit of non-specialist.

- **Step 1***: Grassmannians of k-planes in $\mathbb{R}^n$ is a nonsingular real algebraic set:*

$$G_k(\mathbb{R}^n) = \{A \in \mathbb{R}^{n(n+1)/2} \mid A^2 = A, trace(A) = k\ \}$$

where $\mathbb{R}^{n(n+1)/2}$ denotes the $n\times n$ symmetric matrices. If $V^{n-k} \subset \mathbb{R}^n$ is a compact nonsingular real algebraic set, the normal (or tangent) Gauss map $\gamma : V \to G_k(\mathbb{R}^n)$ is an entire rational map (non-zero denominator). Hence in particular the duals of Steifel-Whitney classes of V are represented by real algebraic subsets (since Steifel-Whitney classes of the universal bundle is represented by Shubert cycles in $G_k(\mathbb{R}^n)$).

Proof. We can identify $G_k(\mathbb{R}^n)$ with the algebraic set defined on the right via the map $L \mapsto$ projection matrix to L. Since nonsingular V can be covered by finitely many Zariski-open sets $V = \cup_{j=1}^N V_j$, each of which is described in $\mathbb{R}^n$ as the zeros of k polynomials $f_i(x) = 0$, $i = 1,..,k$, whose Jacobian $A(x) = (\partial f_i/\partial x_s)$ has rank k. Then by standard linear algebra $\gamma(x) = A(x)(A^t(x)A(x))^{-1}A^t(x)$. In particular this allows us to describe the restriction of the Gauss map $\gamma|_{V_j} = P_j(x)/q_j(x)$ as a matrix whose entries are entire rational maps. Then $\gamma = \sum_{j=1}^N P_j/q_j$ is the answer. □

- **Step 2***: Assume $V \subset \mathbb{RP}^N$ is a compact nonsingular real algebraic set with a nonsingular complexification $j : V \hookrightarrow V_\mathbb{C} \subset \mathbb{CP}^N$, and let $L \subset V$ be an algebraic subset with complexification $L_\mathbb{C} \subset V_\mathbb{C}$. Then the restriction of the Poincaré dual of the fundamental class of $L_\mathbb{C}$ is the cup square of the Poincaré dual of the fundamental class of L, i.e. as $\mathbb{Z}_2$ classes:*

$$j^*PD[L_\mathbb{C}] = PD[L]^2$$

Proof. Let $g_{\mathbb{C}} : \tilde{L}_{\mathbb{C}} \to V_{\mathbb{C}}$ be the resolution of singularities map followed by the inclusion $(\tilde{L}_{\mathbb{C}}, \tilde{L}) \to (L_{\mathbb{C}}, L) \hookrightarrow (V_{\mathbb{C}}, V)$. Isotop $g_{\mathbb{C}}$ to a map $g' : \tilde{L}_{\mathbb{C}} \to V_{\mathbb{C}}$ which is transverse to V, then $j^*[L_{\mathbb{C}}]$ is the Poincaré dual of the intersection $g'(\tilde{L}_{\mathbb{C}}) \cap V$. Call $g_{\mathbb{C}}|_{\tilde{L}} = g$. Near V the map $g_{\mathbb{C}}$ is modeled by the map $g \times g : (\tilde{L} \times \tilde{L}, \Delta) \to (V \times V, \Delta)$ where Δ are the diagonals, i.e. $(x, y) \mapsto (g(x), g(y))$. Therefore $j^*[L_{\mathbb{C}}]$ is represented by the Poincaré dual of the self intersection of the class $[L]$ □

• **Step 3**: *If $V \subset \mathbb{R}^n$ is an algebraic set given by polynomials with highest degree terms are $|x|^{2d}$, then V is projectively closed. That is if $\lambda : \mathbb{R}^n \hookrightarrow \mathbb{RP}^n$ is the imbedding $(x_1, .., x_n) \mapsto [1, x_1, .., x_n]$, then $\lambda(V)$ is a projective algebraic set in $\mathbb{RP}^n$.*

Proof. Assume $f(x) = 0$ is a defining polynomial equation of V (by taking the sum of the squares of the defining equations, every real algebraic set in $\mathbb{R}^n$ can be described by a single polynomial equation). Let $f_d(x)$ be the highest degree term of $f(x)$. Then clearly the equations of $\lambda(V)$ is $0 = x_0^d f(x/x_0) = f_d(x) + x_0 q(x)$, for some polynomial $q(x)$. Hence the zeros $\lambda(V)$ in $\mathbb{RP}^n$ coincides with the zeros of V, i.e. it doesn't have additional zeros at the infinity chart $x_0 = 0$. □

• **Step 4**: *If $M^m \subset Y^{m+1} \hookrightarrow \mathbb{R}^n$ are imbeddings of closed smooth manifolds, such that M is separates Y, then M is ϵ-isotopic to a projectively closed nonsingular algebraic subset of $\mathbb{R}^n$.*

Proof. Let $Z^{n-1} \subset \mathbb{R}^n$ be a codimension one smooth submanifold of which intersects Y^{n+1} transversally at M^m, e.g. we can take Z to be the hypersurface which is the boundary of the narrow sausage shaped region as shown in the Figure 4.

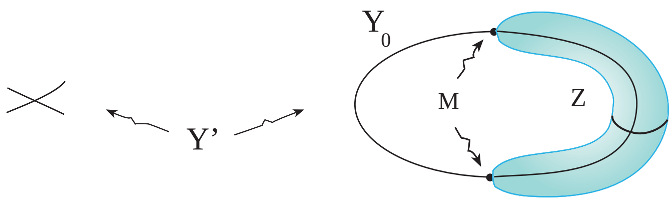


Figure 4.

By Theorem 2 we can approximate Y by a nonsingular component Y_0' of an algebraic set Y'. We claim that Z^{n-1} can be approximated by a projectively closed nonsingular algebraic subset $Z' \subset \mathbb{R}^n$. This is because we can write $Z = f^{-1}(0)$ for some proper smooth function $f : \mathbb{R}^n \to \mathbb{R}$ with 0 as a regular value (since Z is codimension one), and then approximate f by a polynomial of the form $F = g + \epsilon |x|^{2k}$ where $k >> 1$ and take

$Z' = F^{-1}(0)$. Therefore by Step 3 the set $Y' \cap Z'$ is a projectively closed nonsingular algebraic set isotopic to M. □

Proof of Theorem 5. We can construct examples from imbeddings $\mathbb{RP}^m \subset \mathbb{R}^{2m-s}$. For a nonorientable example we need m even and $s \geq 3$, and for an orientable example we need $m = 4k+1$ and $s \geq 5$. For example [Nu] gives $\mathbb{RP}^{10} \subset \mathbb{RP}^{17}$, and [Th] gives $\mathbb{RP}^{13} \subset \mathbb{R}^{20}$. Let us do the nonorientable case: Take the obvious imbeddings

$$M^{11} = \mathbb{RP}^{10} \times S^1 \subset \mathbb{R}^{17} \times \mathbb{R}^3 = \mathbb{R}^{20} \subset \mathbb{RP}^{20}$$

Let $Y^{12} = \mathbb{RP}^{10} \times S^2$ in $\mathbb{RP}^{20}$. Here $S^1 \subset S^2$ are the standard spheres. By Step 4 we can isotope M^{11} to a nonsingular projective algebraic subset $V^{11} \subset \mathbb{RP}^{20}$. We claim that V^{11} can not be the real part of a nonsingular complex algebraic subset $V_{\mathbb{C}}$ of $\mathbb{CP}^{20}$ (defined over $\mathbb{R}$). Suppose such a $V_{\mathbb{C}}$ exists. By the Lefschetz hyperplane theorem (e.g. [H]), for $i \leq 2(11) - 20 = 2$ the restriction induces an isomorphism

$$H^i(\mathbb{CP}^{20};\mathbb{Z}) \xrightarrow{\cong} H^i(V_{\mathbb{C}};\mathbb{Z})$$

By Step 1 the Poincaré dual of the Steifel-Whitney class $w_1(V) = \alpha \times 1$ is represented by an algebraic subset $L \subset V$; here $\alpha \in H^1(\mathbb{RP}^{10};\mathbb{Z}_2)$ is the generator. But since V lies in a chart $\mathbb{R}^{20}$ of $\mathbb{RP}^{20}$, the right vertical (restriction) map in the following commuting diagram is the zero map. This is a contradiction to Step 2.

$$\begin{array}{ccccc} PD[L_{\mathbb{C}}] \in & H^2(V_{\mathbb{C}};\mathbb{Z}_2) & \xrightarrow{j^*} & H^2(V;\mathbb{Z}_2) & \ni PD[L]^2 = \alpha^2 \times 1 \neq 0 \\ & \cong \uparrow & & \uparrow \text{zero} & \\ & H^2(\mathbb{CP}^{20};\mathbb{Z}_2) & \xrightarrow{j^*} & H^2(\mathbb{RP}^{20};\mathbb{Z}_2) & \end{array}$$

For an orientable example we start with $\mathbb{RP}^{13} \subset \mathbb{R}^{20}$ and use $w_2(\mathbb{RP}^{13}) \neq 0$ □

Remark 2. *Though the likelihood is slim, if one can demonstrate an imbedding of a smooth $M \subset \mathbb{R}^n$ such that one of its Steifel-Whitney or Pontryagin classes can only be represented by a bad singular space $X \subset M$ with the property of Theorem 4, then by Step 1 M can not be isotopic to a nonsingular algebraic subset of $\mathbb{R}^n$.*

Remark 3. *Recall in gauge theory (e.g.* [DK], [AMR]*) the space of connections on an $SU(2)$-bundle $P \to M^4$ modulo the gauge group of P, is identified by a component of the mapping space $\mathcal{B}(P) = Map(M, BSO(3))$. Prescribing a metric g on M allows us to consider the space self-dual connections $\mathcal{M}_g \subset \mathcal{B}(P)$ which is finite dimensional. Then evaluating various cohomology classes of $\mathcal{B}(P)$ on $\mathcal{M}_g$ gives Donaldson invariants. So just as using metrics on M^4 as auxiliary objects to define natural finite dimensional subspaces $\mathcal{M}_g$, one can try to use real algebraic structures on a smooth submanifold $M^{n-k} \subset \mathbb{R}^n$ as auxiliary objects to define finite dimensional subsets $\mathcal{Z}_d \subset Map(M, G_k(\mathbb{R}^n))$, such as $\mathcal{Z}_d = \{$entire rational maps of degree $\leq d\}$, and imitate Donaldson invariants (here compactness is the main problem to be faced).*

References

[AK1] S. Akbulut, H. King, *The topology of real algebraic sets with isolated singularities,* Ann. of Math. 113 (1981), 425-446.

[AK2] S. Akbulut, H. King, *Topology of real algebraic sets,* L'Enseignement Math, 29 (1983) 221-261

[AK3] S. Akbulut, H. King, *Topology of real algebraic sets,* MSRI book series no. 25, Springer-Verlag (1992), ISBN: 0-387-97744-9.

[AK4] S. Akbulut, H. King, *All compact manifolds are homeomorphic to totally algebraic real algebraic sets,* Comm. Math. Helv. 66, (1991), 139-149.

[AK5] S. Akbulut, H. King, *Real algebraic structures on topological spaces,* Publ. Math. Inst. Hautes Étud. Sci. 53 (1981) 79-162.

[AK6] S. Akbulut, H. King, *On isotoping submanifolds of* $\mathbb{R}^n$ *to real algebraic subsets and a solution to the Nash conjecture,* Invent.Math. 107 (1992) 87-98

[AK7] S. Akbulut, H. King, *Algebraicity of immersions,* Topology 31 (4) (1992) 701-712.

[AK8] S. Akbulut, H. King, *Transcendental submanifolds of* $\mathbb{R}^n$*,* Comment. Math. Helvetici 68 (1993) 308-318.

[AK9] S. Akbulut, H. King, *Transcendental submanifolds of* $\mathbb{RP}^n$*,* Comment. Math. Helvetici 80 (2005), 427-432.

[AK10] S. Akbulut, H. King, *Constructing strange real algebraic sets,* Real alg. Geom. Proceedins, Rennes, Lec Notes of Math 1524 Springer-Verlag (1991) 120-127.

[AT] S. Akbulut, L. Taylor, *A topological resolution theorem,* Publ. Math. Inst. Hautes Étud. Sci. 53 (1981) 163-195.

[AMR] S. Akbulut, T. Mrowka, Y. Ruan, *Torsion class and universal invariants for odd manifolds*, TAMS, vol 347, no.1 (1995), 63-76.

[BD1] R. Benedetti, M. Dedo, *Counterexamples to representing homology classes by real algebraic subvarieties up to homeomorphism,* Composito Math. 53 (1984), 143-151.

[BD2] R. Benedetti, M. Dedo, *The topology of two-dimensional real algebraic varieties,* Annali di Math pura ed applicata 127 (1981) 141-171

[CK] M. Coste, K. Kurdyka, *On the link of a stratum in a real algebraic set,* Topology 31, no 2 (1992) 323-336.

[DK] S. Donaldson, P. Kronheimer, *The geometry of four-manifolds,* Oxford mathematical monographs (1990) ISBN: 0-19-853553-8

[H] R. Hartshorne, *Equivalence relations on algebraic cycles and subvarieties of small codimension*, Proc. of Symposia in Pure Math. vol 29, 129-164 (1975)

[K] H. King, *Resolving singularities of maps,* Real algebraic geometry and topology, AMS, Contemp. Math. 182 (1993), 135-154.

[MP1] C. McCrory and A. Parusiński, *Complex monodromy and the topology of real algebraic sets*, arXiv:alg-geom/9407011 (1994)

[MP2] C. McCrory and A. Parusiński, *Algebraically Constructible Functions: real algebraic topology*, math.AG/0202086 (2002)

[N] J. Nash, *Real algebraic manifolds,* Ann. of Math. 56 (1952), 405-421.

[Nu] F. Nusbaum, *Obstruction theory of possibly nonorientable fibrations,* Northwestern Univ. PhD thesis (1970)

[P] M. Previte, *The topology of four-dimensional real algebraic sets,* U.Maryland PhD thesis(2000).

[Th] E. Thomas *Embedding manifolds in Euclidean space,* Osaka J. Math. 13 (1976) 163-186

[T] A. Tognoli, *Su una congettura di Nash,* Annali della Scuola Normale Superiore di Pisa 27 (1973), 167-185.

Department of Mathematics, Michigan State University, MI, 48824
E-mail address: akbulut@math.msu.edu

Proceedings of 12^{th} Gökova
Geometry-Topology Conference
pp. 59 – 74

Published online at
GokovaGT.org

Lefschetz decomposition and the *cd*-index of fans

Kalle Karu

Dedicated to the memory of Raoul Bott.

Abstract. The goal of this article is to give a Lefschetz type decomposition for the *cd*-index of a complete fan.

To a complete simplicial fan one can associate a toric variety X, the even Betti numbers h_i of X and the numbers $g_i = h_i - h_{i-1}$. If the fan is projective, then non-negativity of g_i follows from the Lefschetz decomposition of the cohomology.

In the case of a nonsimplicial complete fan, one can analogously compute the flag h-numbers h_S and, by a change of variable formula, the cd-index. We give an analogue of the Lefschetz operation for the cd-index. This gives another proof of the non-negativity of the cd-index for complete fans.

1. Introduction

Let Δ be a complete simplicial n-dimensional fan. Let f_i be the number of i-dimensional cones in Δ and let h_k be defined by the formula

$$\sum_i f_{n-i}(t-1)^i = \sum_k h_{n-k}t^k.$$

The numbers h_k for $k = 0, \ldots, n$ are the even Betti numbers $h_k = \dim H^{2k}(X_\Delta, \mathbb{C})$ of a toric variety X_Δ if the fan Δ is rational. If Δ is also projective, then there exists a Lefschetz operation:

$$L : H^{2k}(X_\Delta, \mathbb{C}) \to H^{2k+2}(X_\Delta, \mathbb{C}), \qquad L^k : H^{n-k}(X_\Delta, \mathbb{C}) \stackrel{\simeq}{\to} H^{n+k}(X_\Delta, \mathbb{C}),$$

giving rise to the Lefschetz decomposition of the cohomology. The existence of a Lefschetz operation implies that the numbers $g_k = h_k - h_{k-1}$ are non-negative for $0 \leq k \leq n/2$.

For a complete but not necessarily simplicial fan one can consider the barycentric subdivision $B\Delta$ of Δ, construct cohomology spaces $H^S(B\Delta)$ of dimension h_S for $S \in \mathbb{N}^n$, and from the numbers h_S compute the *cd*-index $\Psi_\Delta(c, d)$ of Δ (see below). Our goal is to find linear maps on $H^S(\Delta)$, the analogs of the Lefschetz operation, that guarantee non-negativity of the *cd*-index. Unlike the simplicial case, it is not clear how such maps should be defined. We will give in Definition 1.1 a rather weak notion of a Lefschetz operation which, nevertheless, is sufficient to imply non-negativity of the *cd*-index. We also

The author was supported by NSERC.

conjecture a stronger version in which the maps are defined by conewise linear functions on the fan, just as in the simplicial case. The rest of the introduction is spent on constructing the *cd*-index and motivating the definition of a Lefschetz operation.The precise statements and proofs are given in Sections 2 and 3 below.

Let us start by recalling the construction of the cohomology $H^{2*}(X_\Delta, \mathbb{C})$ (which we will denote simply $H^*(\Delta)$) in the case of a simplicial fan Δ (the references [1, 3] contain more details and generalization to the intersection cohomology). Let $\mathcal{A}(\Delta)$ be the vector space of complex-valued conewise polynomial functions on the fan Δ. In other words, an element of $\mathcal{A}(\Delta)$ is a continuous function on the support of the fan Δ that restricts to a polynomial on each cone $\sigma \in \Delta$. We can multiply a conewise polynomial function with a globally polynomial function. In fact, this makes the space $\mathcal{A}(\Delta)$ into a free module under the action of the ring $A = \mathbb{C}[x_1, \ldots, x_n]$ of global polynomial functions, graded by degree. The graded vector space $\mathcal{A}(\Delta)/m\mathcal{A}(\Delta)$, where $m \subset A$ is the maximal homogeneous ideal, is the cohomology space $H^*(\Delta)$ with Poincaré polynomial

$$P_\Delta(t) = \sum_k h_k t^k, \qquad h_k = \dim H^k(\Delta).$$

The fan Δ is projective if and only if there exists a strictly convex conewise linear function $L \in \mathcal{A}(\Delta)$. Multiplication with L induces a Lefschetz operation in cohomology.

In case when the fan Δ is complete, but not necessarily simplicial, we proceed as follows (see [6] or Section 2 below for more details). Let $B\Delta$ be a first barycentric subdivision of Δ and consider the space $\mathcal{A}(B\Delta)$ of conewise polynomial functions on this subdivision, which again is a free A-module. It is possible to modify the A-module structure, so that $\mathcal{A}(B\Delta)$ has a grading by $\mathbb{N}^n$ and the module structure is compatible with this grading (where $A = \mathbb{C}[x_1, \ldots, x_n]$ has the standard grading by $\mathbb{N}^n$). To do this, note that the generating rays of a maximal cone $\sigma \in B\Delta$ are labeled by $1, \ldots, n$: the i'th ray is the barycenter of a cone of dimension i. Let's map the cone σ linearly onto the positive orthant of $\mathbb{R}^n$ so that the i'th ray goes to the i'th coordinate axis. These maps can be chosen compatibly for all maximal cones, defining a piecewise linear map from $B\Delta$ to $\mathbb{R}^n$ that "folds" the fan onto the positive orthant. The A-module structure on $\mathcal{A}(B\Delta)$ is defined via pullback by this map. Since this map identifies polynomial functions on a maximal cone with polynomials on the positive orthant, i.e., with A, we get a grading by $\mathbb{N}^n$ on polynomials on each cone σ, hence on $\mathcal{A}(B\Delta)$.

With the module structure and grading on $\mathcal{A}(B\Delta)$ defined above, the quotient $H^*(B\Delta) := \mathcal{A}(B\Delta)/m\mathcal{A}(B\Delta)$ inherits a similar grading. Consider the corresponding Poincaré polynomial

$$P_{B\Delta}(t_1, \ldots, t_n) = \sum_{S \in \mathbb{N}^n} h_S t_1^{S_1} \cdots t_n^{S_n}, \qquad h_S = \dim H^S(B\Delta).$$

The cohomology $H^*(B\Delta)$ satisfies Poincaré duality $h_S = h_{(1,\ldots,1)-S}$. In fact, this duality is defined by a non-degenerate Poincaré pairing (see Section 2.4). This in particular implies that the nonzero coefficients in the Poincaré polynomial can be indexed by subsets $S \subset \{1, \ldots, n\}$. The numbers h_S are called the flag h-numbers of the fan Δ.

Let $\mathbb{Q}\langle c, d\rangle$ be the polynomial ring in non-commuting variables c and d of degree 1 and 2, respectively. There is an embedding of vector spaces

$$\phi : \mathbb{Q}\langle c, d\rangle \hookrightarrow \mathbb{Q}[t_1, t_2, \ldots],$$

defined as follows. ϕ maps constants to constants and if $f(c,d)c+g(c,d)d$ is a homogeneous cd-polynomial of degree $m > 0$, define inductively

$$\phi(f(c,d)c + g(c,d)d) = \phi(f(c,d))(t_m + 1) + \phi(g(c,d))(t_{m-1} + t_m).$$

For example, there are 3 cd-monomials of degree 3:

$$c^3 = (t_1 + 1)(t_2 + 1)(t_3 + 1),$$
$$cd = (t_1 + 1)(t_2 + t_3),$$
$$dc = (t_1 + t_2)(t_3 + 1).$$

It is shown in [2] that the Poincaré polynomial $P_{B\Delta}(t_1, \ldots, t_n)$ of a complete fan Δ can be expressed as a homogeneous cd-polynomial of degree n, called the cd-index $\Psi_\Delta(c, d)$ of Δ. The coefficients of the polynomial are integers [2] and non-negative [7, 6].

One approach to proving non-negativity of the coefficients of the cd-index is to decompose the cohomology $H^*(B\Delta)$ into summands corresponding to different cd-monomials, so that the coefficients of $\Psi_\Delta(c, d)$ are the dimensions of the corresponding components. If we know the non-negativity of the cd-index, then the existence of such a decomposition follows trivially. Figure 1 shows the dimensions of the pieces corresponding to different cd-monomials in the 3-dimensional case. The bold dots indicate the t_i-monomial being a summand of the cd-monomial.

In analogy with the Lefschetz decomposition in the singly-graded case, we expect the decomposition to be defined by linear maps. More precisely, we look for endomorphisms $L_i : H^*(B\Delta) \to H^*(B\Delta)$ of degree $e_i = (0, \ldots, 1, \ldots, 0)$. If a cd-monomial m can be written as $m = \ldots (t_i + 1) \ldots$, then L_i should map in the corresponding piece H^*_m of the cohomology decomposition:

$$L_i : H_m^{(*,\ldots,*,0,*,\ldots,*)} \xrightarrow{\simeq} H_m^{(*,\ldots,*,1,*,\ldots,*)}.$$

For example, L_1 should define an isomorphism from the back face of the cube to the front face in Figure 1 for the monomials c^3 and cd; the component corresponding to the monomial dc should lie in the kernel of L_1.

1.1. The main construction

The definition of a Lefschetz operation is given inductively using a construction that we call "the main construction". It essentially describes the action of L_1 on the A-module $\mathcal{A}(B\Delta)$ as described in the previous paragraph.

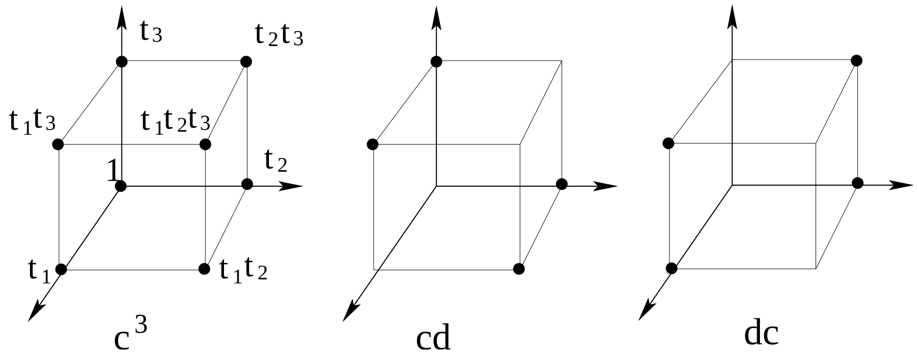


FIGURE 1. cd-monomials in terms of t_i-monomials.

Let $A_{l,m}$ be the polynomial ring $\mathbb{C}[x_l, \ldots, x_m]$, graded by $\mathbb{N}^{m-l}$, with x_i having degree e_i. Let the dualizing module of $A_{l,m}$ be $\omega_{l,m}$, the principal ideal in $A_{l,m}$ generated by $x_l \cdots x_m$.

Let M be a finitely generated free graded $A_{l,m}$-module. A Poincaré pairing on M is an $A_{l,m}$-bilinear symmetric map

$$< \cdot, \cdot >: M \times M \to \omega_{l,m},$$

inducing a non-degenerate pairing on $\overline{M} = M/(x_l, \ldots, x_m)M$. We always assume that M is graded in non-negative degrees. Then the existence of a Poincaré pairing implies that $\overline{M}$ is graded by subsets of $\{l, \ldots, m\}$.

Let M be a free $A_{l,m}$-module with a Poincaré pairing and let $L : M \to M$ be an endomorphism of degree e_l which is self-adjoint with respect to the pairing:

$$< Lm_1, m_2 >=< m_1, Lm_2 > .$$

We can write

$$M/(x_l)M = M^0 + M^1,$$

where M^i consists of elements of degree $(i, *, \ldots, *)$. Then L induces a map $M^0 \to M^1$.

Assume that the map $L : M^0 \to M^1$ is injective and the quotient is annihilated by x_{l+1}:

$$0 \to M^0 \xrightarrow{L} M^1 \to Q \to 0, \qquad x_{l+1}Q = 0. \tag{1}$$

It is elementary (see Lemma 3.1 in Section 3) that Q is a free $A_{l+2,m} = \mathbb{C}[x_{l+2}, \ldots, x_m]$-module and we get a long-exact Tor sequence:

$$0 \to Q[e_l - e_{l+1}] \to M^0/(x_{l+1})M^0 \xrightarrow{L} M^1/(x_{l+1})M^1 \to Q \to 0. \tag{2}$$

Let C be the cokernel of the embedding $Q[e_l - e_{l+1}] \to M^0/(x_{l+1})M^0$:

$$0 \to Q[e_l - e_{l+1}] \to M^0/(x_{l+1})M^0 \to C \to 0. \tag{3}$$

Then C is also a free $A_{l+2,m}$-module (see Section 3 below). We will show that Q and C both inherit Poincaré pairings from M. The construction of Q and C from M and L is what we call the main construction.

Let us explain how the construction of C and Q corresponds to the *cd*-variables c and d, respectively. First, since Q lies in degrees $(1, 0, *, \ldots, *)$, let us replace it with the shifted module $Q' = Q[e_l]$, which is a free $A_{l+2,m}$-module in degrees $(0, 0, *, \ldots, *)$. Also, since c and d have degrees 1 and 2, respectively, we replace C with the free $A_{l+1,m}$-module $C' = C \otimes_{A_{l+2,m}} A_{l+1,m}$. Now let $P_M(t_1, \ldots, t_n)$ (respectively $P_{Q'}(t_{l+2}, \ldots, t_m)$, $P_{C'}(t_{l+1}, \ldots, t_m)$) be the Hilbert polynomial of $\overline{M}$ (respectively $\overline{Q'}$, $\overline{C'}$). From the exact sequences (2) and (3), we get

$$P_M = (1 + t_l)P_{C'} + (t_l + t_{l+1})P_{Q'} = cP_{C'} + dP_{Q'}. \tag{4}$$

Thus, if $P_{C'}$ and $P_{Q'}$ are both *cd*-polynomials with non-negative coefficients, then the same is true for P_M. (Here we changed slightly the map ϕ on $\mathbb{Q}\langle c, d\rangle$ by letting its image be in $\mathbb{Q}[t_l, \ldots]$ for P_M, $\mathbb{Q}[t_{l+1}, \ldots]$ for $P_{C'}$, and similarly for $P_{Q'}$.) From this formula, we see that the main construction of C' and Q' from M corresponds to contracting the polynomial P_M from the left with c and d, respectively.

Definition 1.1. Let M be a finitely generated free $A_{l,m}$-module with a Poincaré pairing. We say that M has a *Lefschetz operation* if there exists an endomorphism $L : M \to M$ of degree e_l, satisfying the assumptions of the main construction, such that the modules C' and Q' also have Lefschetz operations. More precisely:

- L is self-adjoint with respect to the pairing on M.
- $L : M^0 \to M^1$ is injective with cokernel annihilated by x_{l+1}.
- Inductively, the $A_{l+1,m}$-module $C' = C \otimes_{A_{l+2,m}} A_{l+1,m}$ and the $A_{l+2,m}$-module $Q' = Q[e_l]$ have Lefschetz operations.

To start the induction, if $l = m + 1$ and M is a finite dimensional vector space, then it trivially has a Lefschetz operation.

From the computation (4) above, it is clear that if M has a Lefschetz operation, then the Hilbert function of $\overline{M}$ can be written as a homogeneous *cd*-polynomial of degree $m - l$ with non-negative integer coefficients.

The main result of this article is:

Theorem 1.2. *Let Δ be a complete fan of dimension n. Then the $A_{1,n}$-module $\mathcal{A}(B\Delta)$ has a Lefschetz operation. In particular, the cd-index of Δ has non-negative integer coefficients.*

Recall that $\mathcal{A}(B\Delta)$ is a ring. If $L_i \in \mathcal{A}(B\Delta)$ is an element of degree e_i, then multiplication with L_i defines an endomorphism of $\mathcal{A}(B\Delta)$ of degree e_i, self-adjoint with respect to the natural Poincaré pairing. Thus, L_1 is a good candidate for the Lefschetz operation on $\mathcal{A}(B\Delta)$, and inductively, L_i for $i > 1$ could be used to define the endomorphisms of Q and C.

Conjecture 1.3. *Let $L_i \in \mathcal{A}(B\Delta)$ be a general element of degree e_i for $i = 1, \ldots, n$. Then L_i define a Lefschetz operation on $\mathcal{A}(B\Delta)$.*

We remark that a Lefschetz operation on M does not define a canonical decomposition of $\overline{M}$ into components corresponding to the *cd*-monomials. To decompose $\overline{M}$, we need to choose a splitting of the sequence (3), so that

$$\overline{M} \simeq \overline{C}' \oplus \overline{C}'[-e_l] \oplus \overline{Q}'[-e_l] \oplus \overline{Q}'[-e_{l+1}],$$

corresponding to the formula (4). Inductive decomposition of $\overline{C}'$ and $\overline{Q}'$ then give a complete decomposition of $\overline{M}$.

To prove Theorem 1.2, we express $\mathcal{A}(B\Delta)$ as the space of global sections of a sheaf $\mathcal{L}$ on Δ. The main construction can be sheafified, i.e., performed on the stalks of the sheaf $\mathcal{L}$ simultaneously. We show that a Lefschetz operation on the space of global sections comes from a sheaf homomorphism.

We also consider Conjecture 1.3 in the context of sheaves and reduce it to a Kleiman-Bertini type problem of torus actions on a vector space. Let an algebraic torus T act on a finite dimensional vector space V with possibly infinitely many orbits. When does the general translate of a subspace $K \subset V$ intersect another subspace transversely? Conjecture 3.13 claims sufficient conditions for this, implying Conjecture 1.3.

Theorem 1.2 gives another proof of non-negativity of the *cd*-index for a complete fan. In [6] non-negativity was proved more generally for Gorenstein* posets. The current proof does not extend to that more general situation. The two proofs are based on the same idea. However, the proof we give here is simpler because we work with modules only, avoiding derived categories.

2. Sheaves on fans

All our vector spaces are over the field of complex numbers $\mathbb{C}$. Let $A_{l,m} = \mathbb{C}[x_l, x_{l+1}, \ldots, x_m]$, graded so that x_i has degree e_i. For a graded $A_{l,m}$-module M we denote the shift in grading by $M[\cdot]$. We also write $\overline{M} = M/(x_l, \ldots, x_m)M$.

For a graded set, the superscript refers to degree. If Δ is a fan, then $\Delta^{\geq m}$ consists of all cones of dimension at least m. Similarly, $\Delta^{[l,m]}$ is the subset of cones of dimension $d \in [l, m]$.

2.1. Fan spaces

Let us recall the notion of sheaves on fans. The main reference for the general theory is [1, 3] and for the specific sheaves used here [6].

We fix a complete n-dimensional fan Δ (see [5] for terminology). Consider Δ as a finite partially ordered set of cones, graded in degrees $0, \ldots, n$. It is sometimes convenient to add a maximal element $\hat{1}$ of degree $n+1$ to Δ.

The fan Δ is given the topology in which open sets are the (closed) subfans of Δ. Then a sheaf F of vector spaces on Δ consists of the data:

- A vector space F_σ for each $\sigma \in \Delta$.
- Linear maps $res^\sigma_\tau : F_\sigma \to F_\tau$ for $\sigma > \tau$, satisfying the compatibility condition $res^\tau_\rho \circ res^\sigma_\tau = res^\sigma_\rho$ for $\sigma > \tau > \rho$.

On sheaves we can perform the usual sheaf operations. For example, a global section $f \in \Gamma(F, \Delta)$ consists of the data $f_\sigma \in F_\sigma$ for each $\sigma \in \Delta$, such that $res^\sigma_\tau f_\sigma = f_\tau$. Equivalently, we only need to give $f_\sigma \in F_\sigma$ for maximal cones σ, such that their restrictions to smaller dimensional cones agree.

Define a sheaf of rings $\mathcal{A}$ on Δ as follows:

- $\mathcal{A}_\sigma = A_{1,d} = \mathbb{C}[x_1, \ldots, x_d]$ if $\dim \sigma = d$.
- $res^\sigma_\tau : \mathbb{C}[x_1, \ldots, x_d] \to \mathbb{C}[x_1, \ldots, x_l]$ is the standard projection $x_i \mapsto x_i$ for $1 \leq i \leq l$ and $x_i \mapsto 0$ for $i > l$.

Given the sheaf of rings $\mathcal{A}$ on Δ, we consider sheaves of $\mathcal{A}$-modules $\mathcal{F}$. This means that the stalks $\mathcal{F}_\sigma$ are $\mathcal{A}_\sigma$-modules and the restriction maps are module homomorphisms. Note that the sheaf $\mathcal{A}$ is graded by $\mathbb{N}^n$. We assume that all sheaves of $\mathcal{A}$-modules are similarly graded.

There exists an indecomposable sheaf $\mathcal{L}$ of $\mathcal{A}$-modules satisfying the following conditions:

- Locally free: $\mathcal{L}_\sigma$ is a graded free $\mathcal{A}_\sigma$-module.
- Minimally flabby: for $\sigma = 0$, $\mathcal{L}_0 = \mathbb{C}$ in degree 0; for $\sigma > 0$, the restriction maps induce an isomorphism
$$\overline{\mathcal{L}}_\sigma \to \overline{\Gamma(\mathcal{L}, \partial\sigma)},$$
where $\partial\sigma$ is the boundary fan of σ.

These two conditions define $\mathcal{L}$ up to an isomorphism. In fact, $\Gamma(\mathcal{L}, \partial\sigma)$ is a free $A_{1,d-1}$-module if $\dim \sigma = d$, and we can inductively define
$$\mathcal{L}_\sigma = \Gamma(\mathcal{L}, \partial\sigma) \otimes_{A_{1,d-1}} A_{1,d}.$$

2.2. Barycentric subdivisions

Let $B\Delta$ be a barycentric subdivision of Δ. As a poset, it consists of chains $x = (0 < \sigma_1 < \ldots < \sigma_m)$ in Δ. Define a sheaf of rings $\mathcal{B}$ on $B\Delta$ as follows:

- $\mathcal{B}_x = \mathbb{C}[x_i]_{i \in S}$, where $x = (0 < \sigma_1 < \ldots < \sigma_m)$, $S = \{\dim \sigma_1, \ldots, \dim \sigma_m\}$.
- res^x_y is the standard projection.

One can construct as above a sheaf $\mathcal{L}$ with respect to $\mathcal{B}$, but this sheaf is isomorphic to $\mathcal{B}$.

Lemma 2.1 ([6]). *We have*

$$\pi_* \mathcal{B} \simeq \mathcal{L},$$

where $\pi : B\Delta \to \Delta$ *is the subdivision map sending* $x = (0 < \sigma_1 < \ldots < \sigma_m)$ *to* σ_m. □

It is often more convenient to work with the sheaf $\mathcal{B}$ because it is a sheaf of rings. The space of global sections $\Gamma(\mathcal{B}, B\Delta)$ (which is isomorphic to $\Gamma(\mathcal{L}, \Delta)$ by the previous lemma) is what we called $\mathcal{A}(B\Delta)$ in the introduction. Since $\mathcal{B}$ and $\mathcal{L}$ are sheaves of $A_{1,n}$-modules, so are the spaces of global sections.

2.3. The cellular complex

Let us fix an orientation for each cone $\sigma \in \Delta$ and for $\sigma > \tau$, $\dim \sigma = \dim \tau + 1$, let

$$or_\tau^\sigma = \pm 1$$

depending on whether the orientations of σ and τ agree or not.

The cellular complex of a sheaf F on Δ is

$$C_n^\bullet(F, \Delta) = 0 \to C^0 \to C^1 \to \ldots \to C^n \to 0,$$

where

$$C^i = \bigoplus_{\dim \sigma = n-i} F_\sigma,$$

and the differentials are defined as sums of $or_\tau^\sigma res_\tau^\sigma : F_\sigma \to F_\tau$.

For a complete fan Δ, the cellular complex $C_n^\bullet(F, \Delta)$ computes the cohomology of F. Applying this to the flabby sheaf $\mathcal{L}$, we get

$$H^i(C_n^\bullet(\mathcal{L}, \Delta)) = \begin{cases} \Gamma(\mathcal{L}, \Delta) & \text{if } i = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $\Gamma(\mathcal{L}, \Delta)$ is a graded free $A_{1,n}$-module.

If $\sigma \in \Delta$, $\dim \sigma = d$, then $\partial\sigma$ is combinatorially equivalent to a complete fan of dimension $d-1$, hence we may use $C_{d-1}^\bullet(\mathcal{L}, \partial\sigma)$ to compute $\Gamma(\mathcal{L}, \partial\sigma)$. This gives an exact sequence

$$0 \to \mathcal{L}_\sigma / x_d \mathcal{L}_\sigma \to \bigoplus_{\tau < \sigma, \dim \tau = d-1} \mathcal{L}_\tau \to \bigoplus_{\rho < \sigma, \dim \rho = d-2} \mathcal{L}_\rho \to \ldots \to \mathcal{L}_0 \to 0. \tag{5}$$

2.4. Poincaré pairing

Define the dualizing module $\omega_{1,n} = (x_1 \cdots x_n) A_{1,n}$. I.e., $\omega_{1,n}$ is the principal ideal generated by $x_1 \cdots x_n$. There exists an $A_{1,n}$-bilinear non-degenerate pairing

$$\Gamma(\mathcal{L}, \Delta) \times \Gamma(\mathcal{L}, \Delta) \to \omega_{1,n}.$$

The pairing is best constructed using the isomorphism $\Gamma(\mathcal{L},\Delta) \simeq \Gamma(\mathcal{B}, B\Delta)$. On $\Gamma(\mathcal{B}, B\Delta)$ the pairing is defined by multiplication ($\mathcal{B}$ is a sheaf of rings), followed by an evaluation map into $\omega_{1,n}$.

One can give a simple description of the evaluation map as in [4], depending on the orientations or^σ_τ. For $x = (0 < \sigma_1 < \ldots < \sigma_n)$ a maximal element of $B\Delta$ of dimension n, define

$$\varepsilon_x = or^{\hat{1}}_{\sigma_n} or^{\sigma_n}_{\sigma_{n-1}} \cdots or^{\sigma_1}_0 = \pm 1.$$

Now if $f \in \Gamma(\mathcal{B}, B\Delta)$, then it can be shown that

$$\sum_{\dim x = n} \varepsilon_x f_x$$

is an element of $A_{1,n}$ that is divisible by $x_1 x_2 \cdots x_n$, hence lies in $\omega_{1,n}$. This defines the $A_{1,n}$-linear evaluation map $\Gamma(\mathcal{B}, B\Delta) \to \omega_{1,n}$ and the Poincaré pairing on $\Gamma(\mathcal{B}, B\Delta)$.

If $\sigma \in \Delta$ is a d-dimensional cone, then $\partial\sigma$ is combinatorially equivalent to a complete fan of dimension $d-1$. By the same construction as above we get a pairing on $\Gamma(\mathcal{L}, \partial\sigma) \simeq \mathcal{L}_\sigma / x_d \mathcal{L}_\sigma$.

In summary, for each cone $\sigma \in \Delta$, $\dim\sigma = d$, we have a non-degenerate symmetric bilinear pairing

$$< \cdot, \cdot >_\sigma : \mathcal{L}_\sigma / x_d \mathcal{L}_\sigma \times \mathcal{L}_\sigma / x_d \mathcal{L}_\sigma \to \omega_{1,d-1}.$$

These pairings are related as follows. For $f, g \in \mathcal{L}_\sigma / x_d \mathcal{L}_\sigma$,

$$< f, g >_\sigma = \sum_{\dim \tau = d-1} or^\sigma_\tau < f_\tau, g_\tau >_\tau,$$

where f_τ and g_τ are the restrictions of f and g to τ and the pairing on the right hand side is the $A_{1,d-1}$-bilinear extension of the $A_{1,d-2}$-bilinear pairing $< \cdot, \cdot >_\tau$.

3. The main construction on sheaves

Let us return to the situation of Section 1.1 and prove the claims made there.

We have a finitely generated free $A_{l,m}$-module M with Poincaré pairing

$$< \cdot, \cdot >_M : M \times M \to \omega_{l,m}.$$

Write

$$M / x_l M = M^0 \oplus M^1,$$

where M^i consists of elements of degree $(i, *, \ldots, *)$. Assume that $L : M^0 \to M^1$ is a $A_{l+1,m}$-module homomorphism of degree e_l, self-adjoint with respect to the pairing, and such that L is injective with quotient Q annihilated by x_{l+1}:

$$0 \to M^0 \xrightarrow{L} M^1 \to Q \to 0, \qquad x_{l+1} Q = 0.$$

Lemma 3.1. *Q is a free $A_{l+2,m}$-module.*

Proof. Since M^0 and M^1 are free $A_{l+1,m}$-modules, we get from the exact sequence above that

$$Tor_i^{A_{l+1,m}}(Q,\mathbb{C}) = 0, \qquad i \geq 2.$$

Because Q is a $A_{l+2,m}$-module, annihilated by x_{l+1}, this implies that

$$Tor_1^{A_{l+2,m}}(Q,\mathbb{C}) = 0,$$

hence Q is free. □

Now assuming that Q is free, we get an exact sequence

$$0 \to Q[e_l - e_{l+1}] \to M^0/(x_{l+1})M^0 \xrightarrow{L} M^1/(x_{l+1})M^1 \to Q \to 0,$$

where all terms are free $A_{l+2,m}$-modules. Define C by the exact sequence

$$0 \to Q[e_l - e_{l+1}] \to M^0/(x_{l+1})M^0 \to C \to 0.$$

Then $Tor_1^{A_{l+2,m}}(C,\mathbb{C}) = 0$ and C is also a free $A_{l+2,m}$-module.

Let us construct bilinear pairings on C and Q. On C the pairing is

$$< x, y >_C = < x, Ly >_M .$$

This is well-defined and gives an $A_{l+2,m}$-linear map of degree e_l

$$C \otimes_{A_{l+2,m}} C \to \omega_{l,m} \otimes_{A_{l,m}} A_{l+2,m}.$$

Dividing by x_l we get a degree 0 map into $\omega_{l+1,m} \otimes_{A_{l+1,m}} A_{l+2,m}$. Finally, replacing C by $C' = C \otimes_{A_{l+2,m}} A_{l+1,m}$ and extending the pairing linearly, we have a $A_{l+1,m}$-bilinear map

$$< \cdot, \cdot >_{C'} : C' \times C' \to \omega_{l+1,m}.$$

To define the pairing on Q, let α be the composition

$$\alpha : Q \xrightarrow{\cong} Q[e_l - e_{l+1}] \hookrightarrow M^0/x_{l+1}M^0.$$

On the elements $[q] \in Q$ this map is given by

$$\alpha([q]) = L^{-1}(x_{l+1}q).$$

Now define the pairing

$$< x, y >_Q = < \alpha(x), y >_M .$$

One can check that this pairing is well-defined. Taking into account that α has degree $e_{l+1} - e_l$, we get a degree 0 $A_{l+2,m}$-bilinear map on $Q' = Q[e_l]$

$$< \cdot, \cdot >_{Q'} : Q' \times Q' \to \omega_{l+2,m}.$$

It is easy to see that the bilinear maps on C' and Q' are symmetric.

Lemma 3.2. *The pairings $< \cdot, \cdot >_{Q'}$ and $< \cdot, \cdot >_{C'}$ are non-degenerate.*

Proof. One checks the non-degeneracy of the pairing on C using the definition and self-adjointness of L. Then it follows that the pairing between Q and $Q[e_l - e_{l+1}]$ is non-degenerate. □

We next want to sheafify the main construction. Recall that $\mathcal{L}$ is a sheaf on Δ with stalks $\mathcal{L}_\sigma$ free $\mathcal{A}_\sigma$-modules with Poincaré pairings. To perform the main construction simultaneously on all stalks of $\mathcal{L}$, the first step is to split

$$\mathcal{L}/x_1\mathcal{L} = \mathcal{L}^0 \oplus \mathcal{L}^1,$$

and then find a map of sheaves of degree e_1

$$L : \mathcal{L}^0 \to \mathcal{L}^1.$$

If one looks at the stalks, it becomes clear that $\mathcal{L}^i$ should be considered as sheaves on $\Delta^{\geq 2}$ (i.e., on the poset of cones of dimension at least 2), and the cokernel Q of the map L should be a sheaf on $\Delta^{\geq 3}$. Therefore we will consider sheaves on $\Delta^{\geq m}$ for $m \geq 1$.

3.1. Sheaves on $\Delta^{\geq m}$

We let $\Delta^{\geq m}$ have the the topology induced from Δ. To give a sheaf on $\Delta^{\geq m}$ is equivalent to giving a sheaf on Δ with all stalks zero on cones of dimension less than m.

Define the structure sheaf $\mathcal{A}$ on $\Delta^{\geq m}$ as follows. For $\sigma \in \Delta$, $\dim \sigma = d \geq m$, let

$$\mathcal{A}_\sigma = A_{m,d} = \mathbb{C}[x_m, \ldots, x_d],$$

with restriction maps res^σ_τ the standard projections.

Definition 3.3. Let $\mathcal{F}$ be a locally free sheaf of $\mathcal{A}$-modules on $\Delta^{\geq m}$. We say that $\mathcal{F}$ is *minimally flabby* if all the restriction maps res^α_β are surjective and for every $\sigma \in \Delta$, $\dim \sigma = d \geq m$, we have an exact sequence, the "augmented cellular complex" (compare with (5))

$$0 \to \mathcal{F}_\sigma/x_d\mathcal{F}_\sigma \to \bigoplus_{\tau<\sigma, \dim\tau=d-1} \mathcal{F}_\tau \to \ldots \to \bigoplus_{\rho<\sigma, \dim\rho=m} \mathcal{F}_\rho \to G_\sigma \to 0, \qquad (6)$$

where

- The augmentation G_σ is a vector space (i.e., an $A_{1,n}$-module annihilated by $x_1, \ldots, x_n$).
- All differentials, except the maps to G_σ, are defined by $or^\alpha_\beta res^\alpha_\beta$ as in the usual cellular complex.

Remark 3.4.
(1) It should be noted that a minimally flabby sheaf is not flabby in the topology of $\Delta^{\geq m}$.
(2) We do not need the surjectivity of the restriction maps res^α_β for the proof of Theorem 1.2. These conditions are only necessary to state Conjectures 1.3 and 3.13. However, surjectivity of the restriction maps follows easily for all sheaves we consider.

Example 3.5. (1) Let $\mathcal{L}$ be the indecomposable sheaf on Δ. Then $\mathcal{L}|_{\Delta^{\geq 1}}$ is a minimally flabby sheaf on $\Delta^{\geq 1}$. In this case we have $G_\sigma = \mathcal{L}_0 = \mathbb{C}$ for all σ.

(2) In general, the vector spaces G_σ depend on the cone σ. Let $\pi_1, \pi_2 \in \Delta$ be two cones of dimension $m-1$, and let $\mathcal{L}^{\pi_i}$ be the indecomposable sheaf constructed on the poset $\operatorname{Star} \pi_i$. Then $\mathcal{F} = \mathcal{L}^{\pi_1} \oplus \mathcal{L}^{\pi_2}|_{\Delta^{\geq m}}$ is a minimally flabby sheaf and we have

$$G_\sigma = \begin{cases} \mathbb{C} \oplus \mathbb{C} & \text{if } \pi_1, \pi_2 < \sigma \\ \mathbb{C} & \text{if } \pi_1 < \sigma \text{ or } \pi_2 < \sigma \text{, but not both} \\ 0 & \text{otherwise.} \end{cases}$$

Note that a minimally flabby sheaf on $\Delta^{\geq m}$ is determined by its restriction to $\Delta^{[m,m+1]}$. Indeed, the exact sequence (6) can be used to recover $\mathcal{F}_\sigma$ for $\dim \sigma > m+1$. Similarly, given two minimally flabby shaves $\mathcal{F}$ and $\mathcal{E}$, a morphism defined between the restrictions of these sheaves to $\Delta^{[m,m+1]}$ can be lifted to a morphism on $\Delta^{\geq m}$.

Lemma 3.6. *Let $\mathcal{E}$ and $\mathcal{F}$ be minimally flabby sheaves on $\Delta^{\geq m}$, and $L : \mathcal{E} \to \mathcal{F}$ a homomorphism of $\mathcal{A}$-modules.*

(1) *If L is injective on cones $\sigma \in \Delta$, $\dim \sigma = m$, then L is injective on all cones.*
(2) *If L is an isomorphism on cones $\sigma \in \Delta$, $\dim \sigma = m$, then the cokernel $\mathcal{Q}$ of L:*

$$0 \to \mathcal{E} \to \mathcal{F} \to \mathcal{Q} \to 0$$

is a minimally flabby sheaf on $\Delta^{\geq m+1}$.

Proof. The first statement follows by induction on $\dim \sigma$ from the exact sequence (6).

To prove the second statement, first note that the surjectivity of the restriction maps res^α_β for $\mathcal{Q}$ is clear. The morphism L defines a map between the augmented cellular complexes of $\mathcal{E}$ and $\mathcal{F}$ which is injective except possibly in the G_σ terms. The quotient gives the cellular complex for Q. By induction on $\dim \sigma$ it follows that Q_σ is annihilated by x_m, hence is a free $A_{m+1,d}$-module by Lemma 3.1. We get the augmentation for $\mathcal{Q}$ by removing tha augmentations of $\mathcal{E}$ and $\mathcal{F}$ and considering the long-exact cohomology sequence of the short-exact sequence of complexes. □

Definition 3.7. Let $\mathcal{F}$ be a minimally flabby sheaf on $\Delta^{\geq m}$. We say that $\mathcal{F}$ is a Poincaré sheaf if for every $\sigma \in \Delta$, $\dim \sigma = d \geq m$, we have an $A_{m,d-1}$-bilinear non-degenerate symmetric pairing

$$< \cdot, \cdot >_\sigma : F_\sigma / x_d F_\sigma \times F_\sigma / x_d F_\sigma \to \omega_{m,d-1},$$

satisfying the compatibility condition:

$$< f, g >_\sigma = \sum_{\tau < \sigma} or^\sigma_\tau < \operatorname{res}^\sigma_\tau f, \operatorname{res}^\sigma_\tau g >_\tau, \quad f, g \in F_\sigma / x_d F_\sigma. \tag{7}$$

Here on the right hand side $< \cdot, \cdot >_\tau$ denotes the $A_{m,d-1}$-bilinear extension of the $A_{m,d-2}$-bilinear pairing $< \cdot, \cdot >_\tau$.

Example 3.8. The sheaf $\mathcal{L}|_{\Delta^{\geq 1}}$ is a Poincaré sheaf on $\Delta^{\geq 1}$.

Let $\mathcal{F}$ be a Poincaré sheaf on $\Delta^{\geq m}$. Then $\overline{\mathcal{F}}_\sigma$ for $\dim\sigma = d \geq m$ is a vector space graded by subsets of $\{m, \ldots, d-1\}$. Write $\mathcal{F}/x_m\mathcal{F}$ for the sheaf with stalks

$$(\mathcal{F}/x_m\mathcal{F})_\sigma = \mathcal{F}_\sigma/x_m\mathcal{F}_\sigma.$$

This is a locally free sheaf on $\Delta^{\geq m+1}$, and we can split it as

$$\mathcal{F}/x_m\mathcal{F} = \mathcal{F}^0 \oplus \mathcal{F}^1,$$

where $\mathcal{F}^i_\sigma$ consists of elements of degree $(i, *, \ldots, *)$.

Lemma 3.9. *Let $\mathcal{F}$ be a Poincaré sheaf on $\Delta^{\geq m}$. Then $\mathcal{F}^0$ and $\mathcal{F}^1$ are minimally flabby sheaves on $\Delta^{\geq m+1}$.*

Proof. Let us cut the sequence (6) into two exact sequences

$$0 \to \mathcal{F}_\sigma/x_d\mathcal{F}_\sigma \to \bigoplus_{\tau<\sigma, \dim\tau=d-1} \mathcal{F}_\tau \to \ldots \to S \to 0,$$

$$0 \to S \to \bigoplus_{\rho<\sigma, \dim\rho=m} \mathcal{F}_\rho \to G_\sigma \to 0.$$

From the second sequence we get that S is a free $\mathbb{C}[x_m]$-module, hence the first sequence remains exact after taking quotient by the ideal (x_m) and splitting into two according to degree. The two sequences are the augmented cellular complexes for $\mathcal{F}^0$ and $\mathcal{F}^1$. □

Now we are ready to define the sheafified version of the main construction. Let $\mathcal{F}$ be a Poincaré sheaf on $\Delta^{\geq m}$ and $L : \mathcal{F} \to \mathcal{F}$ an endomorphism of $\mathcal{A}$-modules of degree e_m, such that $L_\sigma : \mathcal{F}_\sigma \to \mathcal{F}_\sigma$ is self-adjoint with respect to the pairing for each σ. (More precisely, $L_\sigma : \mathcal{F}_\sigma \to \mathcal{F}_\sigma$ has to be self-adjoint with respect to the $A_{m,d}$-linear extension of the pairing $< \cdot, \cdot >_\sigma$.) Assume that the induced morphism $L : \mathcal{F}^0 \to \mathcal{F}^1$ is injective on cones $\sigma \in \Delta$, $\dim\sigma = m+1$; then it is an isomorphism on these cones by Poincaré duality. Lemma 3.6 gives an exact sequence

$$0 \to \mathcal{F}^0 \to \mathcal{F}^1 \to \mathcal{Q} \to 0,$$

where $\mathcal{Q}$ is a minimally flabby sheaf on $\Delta^{\geq m+2}$. In order to have $\mathcal{Q}$ in correct degrees, we have to replace it with $\mathcal{Q}' = \mathcal{Q}[e_m]$.

We also construct the sheaf $\mathcal{C}$ as follows. First, we have an exact sequence of minimally flabby sheaves on $\Delta^{\geq m+2}$:

$$0 \to \mathcal{Q}[e_m - e_{m+1}] \to \mathcal{F}^0/x_{m+1}\mathcal{F}^0 \to \mathcal{F}^1/x_{m+1}\mathcal{F}^1 \to \mathcal{Q} \to 0.$$

Define $\mathcal{C}$ by the exact sequence

$$0 \to \mathcal{Q}[e_m - e_{m+1}] \to \mathcal{F}^0/x_{m+1}\mathcal{F}^0 \to \mathcal{C} \to 0.$$

Then one easily sees that $\mathcal{C}$ is also minimally flabby on $\Delta^{\geq m+2}$ (to get the augmented cellular complex for $\mathcal{C}$, it is more convenient to consider the short exact sequence

$$0 \to \mathcal{C} \xrightarrow{L} \mathcal{F}^1/x_{m+1}\mathcal{F}^1 \to \mathcal{Q} \to 0).$$

We should again replace $\mathcal{C}$ with an almost flabby sheaf $\mathcal{C}'$ on $\Delta^{\geq m+1}$, such that $\mathcal{C} = \mathcal{C}'/x_{m+1}\mathcal{C}'$. We will not do this because inductively, the next step to construct a Lefschetz operation is to go from $\mathcal{C}'$ to $\mathcal{C}$ and split it according to degree. The fact that we don't have $\mathcal{C}'$ that induces $\mathcal{C}$ will cause us some trouble later when we look for an endomorphism of $\mathcal{C}$.

Summarizing, we have defined the sheafified version of the main construction. Starting with a Poincaré sheaf $\mathcal{F}$ on $\Delta^{\geq m}$ and a morphism L, we constructed minimally flabby sheaves $\mathcal{Q}$ and $\mathcal{C}$ on $\Delta^{\geq m+2}$. The construction on stalks agrees with the main construction on modules. The stalks of the shaves $\mathcal{Q}$ and $\mathcal{C}$ inherit Poincaré pairings from the pairing on $\mathcal{F}$, which is clearly compatible with the restriction morphisms. Hence the two new sheaves are also Poincaré sheaves.

It remains to see when can we find an appropriate endomorphism L of $\mathcal{F}$.

Lemma 3.10. *Let $\mathcal{F}$ be a Poincaré sheaf on $\Delta^{\geq m}$ and $L : \mathcal{F} \to \mathcal{F}$ a homomorphism of degree e_m. Then L is self-adjoint with respect to the pairings on $\sigma \in \Delta^{\geq m}$ if and only if it is self-adjoint on cones ρ of dimension m.*

Proof. This follows by induction on the dimension of a cone from the formula (7). □

Lemma 3.11. *Let $\mathcal{F}$ be a Poincaré sheaf on $\Delta^{\geq m}$. Then there exists a homomorphism $L : \mathcal{F} \to \mathcal{F}$ of degree e_m that is self-adjoint with respect to the pairings on the stalks $\mathcal{F}_\sigma$ and such that the induced homomorphism $L : \mathcal{F}^0 \to \mathcal{F}^1$ is injective.*

Proof. For $\dim \rho = m$, let $L_\rho : \mathcal{F}_\rho \to \mathcal{F}_\rho$ be a self-adjoint homomorphism of degree e_m. (Note that $\mathcal{F}_\rho \simeq \mathbb{C}[x_m]^{\oplus a_\rho}$ for some $a_\rho \geq 0$.) We claim that a suitable collection of L_ρ induces the required L. For this we need to check that L_ρ can be extended to cones τ of dimension $m+1$ (hence can be extended to all cones), and that on such τ it defines an injection $\mathcal{F}^0_\tau \to \mathcal{F}^1_\tau$.

Let $\dim \tau = m+1$ and consider the augmented cellular complex of τ:

$$0 \to \mathcal{F}_\tau/x_{m+1}\mathcal{F}_\tau \to \bigoplus_{\rho<\tau} \mathcal{F}_\rho \to G_\tau \to 0.$$

Here $G_\tau \simeq \mathbb{C}^a$ for some $a \geq 0$, $\bigoplus_{\rho<\tau} \mathcal{F}_\rho \simeq \mathbb{C}[x_m]^{\oplus 2a}$ and $\mathcal{F}_\tau/x_{m+1}\mathcal{F}_\tau \simeq \mathbb{C}[x_m]^{\oplus a} \oplus \mathbb{C}[x_m][-e_m]^{\oplus a}$.

The maps L_ρ are compatible with the zero map $G_\tau \to G_\tau$ of the augmentation. It follows that L_ρ induce a map $L_\tau : \mathcal{F}_\tau \to \mathcal{F}_\tau$, compatible with restriction maps, hence there is an extension to a morphism $L : \mathcal{F} \to \mathcal{F}$.

Let $V = \bigoplus_{\rho<\tau} \overline{\mathcal{F}}_\rho \simeq \mathbb{C}^{2a}$ and let $K \simeq \mathbb{C}^a$ be the kernel of $V \to G_\tau$. Then $K = \mathcal{F}^0_\tau$. The map L_ρ comes from a linear map $\overline{L}_\rho = \frac{1}{x} L_\rho : \overline{\mathcal{F}}_\rho \to \overline{\mathcal{F}}_\rho$. The maps $\overline{L}_\rho$ together define a linear map $L_V : V \to V$. Now the condition that L is injective is equivalent to $L_\tau : \mathcal{F}^0_\tau \to \mathcal{F}^1_\tau$ being injective, which is equivalent to the condition that the intersection of K and $L_V(K)$ is zero.

Let us also bring the Poincaré pairing into the picture. We have a non-degenerate symmetric pairing on each $\overline{\mathcal{F}}_\rho$, combined to a pairing on V. The pairing on $\mathcal{F}_\tau$ induces a non-degenerate pairing between $\mathcal{F}_\tau^0$ and $\mathcal{F}_\tau^1$, which restricts to the zero pairing on $\mathcal{F}_\tau^0$, hence the compatibility condition implies that the pairing on V restricted to K is zero. In other words, $K = K^\perp$. The proof that a suitable set of L_ρ gives a required L is given in the lemma below.

Finally, let us consider the case when L is defined by a multiplication with an element in $L \in \Gamma(\mathcal{A}, \Delta)$ of degree e_m. In this case the linear maps $\overline{L}_\rho$ are given my multiplication with a constant c_ρ (where $L|_\rho = c_\rho x_m$). Note also that since the restriction maps res_ρ^τ are surjective, the projection $V \to \overline{\mathcal{F}}_\rho$ maps K onto $\overline{\mathcal{F}}_\rho$. Thus, if the conjecture below is true then L defines an injective morphism. □

Lemma 3.12. *Let $V = \oplus V_i$ be a finite dimensional vector space. Suppose that each V_i has a non-degenerate symmetric bilinear pairing, giving a pairing on V. Let $K \subset V$ be a subspace such that $K \subset K^\perp$. Then there exist self-adjoint linear maps $L_i : V_i \to V_i$, combined to $L : V \to V$, satisfying $K^\perp \cap L(K) = 0$.*

Proof. Let $v_1, \dots, v_{2a}$ be an orthogonal basis of V consisting of elements from V_i and let $y_1, \dots, y_{2a}$ be the dual basis giving coordinates on V. Let T be the algebraic torus of dimension $\dim V$ acting on V by:

$$(t_1, \dots, t_{2a}) \cdot (y_1, \dots, y_{2a}) = (t_1 y_1, \dots, t_{2a} y_{2a}).$$

An element $t \in T$ defines a linear map $V \to V$ of the required type. We claim that for a general t we have $K^\perp \cap t(K) = 0$.

Now V has finitely many T-orbits. By Kleiman-Bertini theorem, for a general t, the restrictions of $K^\perp$ and K to any orbit O intersect transversely. Thus, it suffices to show that the expected dimension of this intersection is zero.

Let $W \subset V$ be a subspace spanned by a subset of the v_j. Then the pairing on V restricts to a non-degenerate pairing on W. Since $K \subset K^\perp$, it follows that

$$\dim(K^\perp \cap W) + \dim(K \cap W) \leq \dim(W). \quad \square$$

Conjecture 3.13. *Let the notation be as in the previous lemma. Additionally assume that the projections $V \to V_i$ map K onto V_i for each i.Then the statement of the lemma remains true if we let L_i be multiplication by some constant c_i.*

Remark 3.14. Starting with a Poincaré sheaf $\mathcal{F}$ on $\Delta^{\geq m}$, we apply the previous lemmas to perform the main construction on $\mathcal{F}$ and produce new sheaves $\mathcal{Q}$ and $\mathcal{C}$. Then inductively we apply the same construction on $\mathcal{C}$ and $\mathcal{Q}$. As explained above, we should consider $\mathcal{C}$ as coming from a sheaf $\mathcal{C}'$ on $\Delta^{\geq m+1}$, so that the main construction should be applied to $\mathcal{C}'$ rather than $\mathcal{C}$. Let us show that we don't need $\mathcal{C}'$ for the existence of the required $L : \mathcal{C} \to \mathcal{C}$.

Recall that $\mathcal{C}$ was defined by the exact sequence of minimally flabby sheaves on $\Delta^{\geq m+2}$:

$$0 \to \mathcal{Q}[e_m - e_{m+1}] \to \mathcal{F}^0 / x_{m+1} \mathcal{F}^0 \to \mathcal{C} \to 0.$$

On the sheaf $\mathcal{F}^0$ we can define a bilinear pairing by the same formula as on $\mathcal{C}$. This pairing is degenerate, but it induces the pairing on $\mathcal{C}$. Now as in Lemma 3.11 we construct a homomorphism $L : \mathcal{F}^0 \to \mathcal{F}^0$ of degree e_{m+1}. We claim that this homomorphism induces the injective homomorphism $\mathcal{C}^0 \to \mathcal{C}^1$. Indeed, we are reduced to the same Lemma 3.12. The difference now is that we may have a strict inclusion $K \subset K^\perp$, while the two spaces were equal in the proof of Lemma 3.11.

Let us now put everything together and finish the proof of Theorem 1.2. We start with the Poincaré sheaf $\mathcal{L}|_{\Delta^{\geq 1}}$ and apply the main construction to produce new Poincaré sheaves $\mathcal{C}$ and $\mathcal{Q}$. Then inductively we apply the main construction to $\mathcal{C}$ and $\mathcal{Q}$. These constructions give a Lefschetz operation on each stalk $\mathcal{L}_\sigma / x_d \mathcal{L}_\sigma$, $\dim \sigma = d$. Considering

$$\mathcal{L}_{\hat{1}} / x_{n+1} \mathcal{L}_{\hat{1}} \simeq \Gamma(\mathcal{L}, \Delta),$$

we get a Lefschetz operation on $\Gamma(\mathcal{L}, \Delta)$ as stated in the theorem.

References

[1] G. Barthel, J.-P. Brasselet, K.-H. Fieseler and L. Kaup, *Combinatorial Intersection cohomology for Fans*, Tôhoku Math. J. 54 (2002) 1-41.

[2] M.M. Bayer and A. Klapper, *A new index for polytopes*, Discrete Comput. Geom. 6 (1991), 33-47.

[3] P. Bressler and V. A. Lunts, *Intersection cohomology on nonrational polytopes*, Compositio Math. 135 (2003), no. 3, 245-278.

[4] M. Brion, *The Structure of the Polytope Algebra*, Tôhoku Math. J. 49, 1997, 1-32.

[5] W. Fulton, *Introduction to toric varieties*, Princeton University Press, (1993).

[6] K. Karu *The cd-index of fans and lattices*, preprint math.AG/0410513.

[7] R.P. Stanley, *Flag f-vectors and the cd-index*, Math. Z. 216 (1994), 483-499.

Mathematics Department, University of British Columbia, 1984 Mathematics Road, Vancouver, B.C. Canada V6T 1Z2

E-mail address: karu@math.ubc.ca

Proceedings of 12th Gökova
Geometry-Topology Conference
pp. 75 – 109

Published online at
GokovaGT.org

Some remarks on G_2-structures

Robert L. Bryant

Dedicated to the memory of Raoul Bott.

Abstract. This article consists of loosely related remarks about the geometry of G_2-structures on 7-manifolds, some of which are based on unpublished joint work with two other people: F. Reese Harvey and Steven Altschuler.

After some preliminary background information about the group G_2 and its representation theory, a set of techniques is introduced for calculating the differential invariants of G_2-structures and the rest of the article is applications of these results. Some of the results that may be of interest are as follows:

First, a formula is derived for the scalar curvature and Ricci curvature of a G_2-structure in terms of its torsion and covariant derivatives with respect to the 'natural connection' (as opposed to the Levi-Civita connection) associated to a G_2-structure. When the fundamental 3-form of the G_2-structure is closed, this formula implies, in particular, that the scalar curvature of the underlying metric is nonpositive and vanishes if and only if the structure is torsion-free. These formulae are also used to generalize a recent result of Cleyton and Ivanov [3] about the nonexistence of closed Einstein G_2-structures (other than the Ricci-flat ones) on compact 7-manifolds to a nonexistence result for closed G_2-structures whose Ricci tensor is too tightly pinched.

Second, some discussion is given of the geometry of the first and second order invariants of G_2-structures in terms of the representation theory of G_2.

Third, some formulae are derived for closed solutions of the Laplacian flow that specify how various related quantities, such as the torsion and the metric, evolve with the flow. These may be useful in studying convergence or long-time existence for given initial data.

Some of this work was subsumed in the work of Hitchin [12] and Joyce [14]. I am making it available now mainly because of interest expressed by others in seeing these results written up since they do not seem to have all made it into the literature.

Received by the editors Februay 01, 2005.

1991 *Mathematics Subject Classification.* 53C10, 53C29.

Key words and phrases. exceptional holonomy, Laplacian flows.

Thanks to Rice University and Duke University for their support via research grants and to the National Science Foundation for its support via DMS-8352009, DMS-8905207, and, most recently in DMS-0103884. Thanks also to the organizers of the April-May 2003 IPAM conference "Geometry and Physics of G_2 manifolds" for their kind support and for the interest expressed there in making these notes available. Finally, thanks to the organizers of the 2004 conference on topology and geometry at Gokova, Turkey for accepting this manuscript to appear in their proceedings and to their referee for pointing out several typos and mistakes.

Version 1.0 (`math.DG/0305124`) was posted to the arXiv on 8 May 2003. This is Version 4.1.

Contents

1. Introduction

This brief article consists of a collection of remarks on the geometry of G_2-structures on 7-manifolds, some of which are based on old unpublished joint work carried out on separate occasions with two other people: F. Reese Harvey and Steven Altschuler.

The work with Reese Harvey (recounted in §5) concerned techniques for calculating various quantities associated to a G_2-structure, possibly with torsion, and was carried out intermittently during the period 1988 through 1994.

The work with Steven Altschuler (recounted in §6) concerned the geometry of a natural Laplacian flow for G_2-structures and was carried out in 1992.

The main reason for making these remarks available now is that some of these formulae and results do not seem to have appeared yet in the literature and some people have expressed an interest in learning about them.

2. Algebra

This section will collect the main results about the group G_2 that will be needed. The reader may consult [2], [14], or [15] for details concerning the properties of the group G_2 that are not proved here. In general, the notation is chosen to agree with the notation in [2].

2.1. The group G_2

Let $e_1, e_2, \ldots, e_7$ denote the standard basis of $\mathbb{R}^7$ (whose elements will be referred to as column vectors of height 7) and let $e^1, e^2, \ldots, e^7 : \mathbb{R}^7 \to \mathbb{R}$ denote the corresponding dual basis.

For notational simplicity, write e^{ijk} for the wedge product $e^i{\wedge}e^j{\wedge}e^k$ in $\Lambda^3\big((\mathbb{R}^7)^*\big)$. Define

$$\phi = e^{123} + e^{145} + e^{167} + e^{246} - e^{257} - e^{347} - e^{356}. \tag{2.1}$$

By a theorem of Schouten [16] (see [2] for a proof), the subgroup of $GL(7,\mathbb{R})$ that fixes ϕ is a compact, connected, simple Lie group of type G_2. In this article, this result will be used to justify the following definition:

Definition 1 (The group G_2).

$$G_2 = \{\ g \in GL(7,\mathbb{R}) \mid g^*(\phi) = \phi\ \}\,. \tag{2.2}$$

2.2. Associated structures

A few properties of G_2 will be needed in this article. The reader may consult [2] for proofs.

The group G_2 acts irreducibly on $\mathbb{R}^7$ and preserves the metric and orientation for which the basis $e_1, e_2, \ldots, e_7$ is an oriented orthonormal basis. The notations g_ϕ and $\langle,\rangle_\phi$ will be used to refer to the metric. The Hodge star operator determined by this metric and orientation will be denoted $*_\phi$. Note, in particular, that G_2 also fixes the 4-form

$$*_\phi\phi = e^{4567} + e^{2367} + e^{2345} + e^{1357} - e^{1346} - e^{1256} - e^{1247}. \tag{2.3}$$

2.3. Some G_2 actions

The group G_2 acts transitively on the unit sphere $S^6 \subset \mathbb{R}^7$. The stabilizer subgroup of any non-zero vector in $\mathbb{R}^7$ is isomorphic to $SU(3) \subset SO(6)$, so that $S^6 = G_2/\,SU(3)$. Since $SU(3)$ acts transitively on $S^5 \subset \mathbb{R}^6$, it follows that G_2 acts transitively on the set of orthonormal pairs of vectors in $\mathbb{R}^7$.

However, G_2 does not act transitively on the set of orthonormal triples of vectors in $\mathbb{R}^7$ since it preserves the 3-form ϕ.

2.4. The ε-notation

It will be convenient to use an ε-notation that will now be introduced. This is the unique symbol that is skew-symmetric in either three or four indices and satisfies

$$\phi = \tfrac{1}{6}\varepsilon_{ijk}\, e^i \wedge e^j \wedge e^k \tag{2.4}$$

$$*_\phi\phi = \tfrac{1}{24}\varepsilon_{ijkl}\, e^i \wedge e^j \wedge e^k \wedge e^l\,. \tag{2.5}$$

Thus, for example, $\varepsilon_{123} = 1$ and $\varepsilon_{4567} = 1$, while $\varepsilon_{124} = \varepsilon_{3456} = 0$. Another way to think of this symbol is via the cross product: $e_i \times e_j = \varepsilon_{ijk} e_k$.

The symbol ε satisfies various useful identities. For example (using the summation convention),

$$\varepsilon_{ijk}\,\varepsilon_{ijl} = 6\delta_{kl} \tag{2.6}$$

$$\varepsilon_{ijq}\,\varepsilon_{ijkl} = 4\varepsilon_{qkl} \tag{2.7}$$

$$\varepsilon_{ipq}\,\varepsilon_{ijk} = \varepsilon_{pqjk} + \delta_{pj}\delta_{qk} - \delta_{pk}\delta_{qj} \tag{2.8}$$

$$\varepsilon_{ipq}\,\varepsilon_{ijkl} = \delta_{pj}\varepsilon_{qkl} - \delta_{jq}\varepsilon_{pkl} + \delta_{pk}\varepsilon_{jql} - \delta_{kq}\varepsilon_{jpl} + \delta_{pl}\varepsilon_{jkq} - \delta_{lq}\varepsilon_{jkp}\,. \tag{2.9}$$

These identities are actually quite easy to prove using the fact that G_2 acts transitively on orthonormal pairs. For example, identity (2.8) can be reduced to the case where $p = 1$ and $q = 2$. Then the only non-zero term on the left hand side is $\varepsilon_{312}\varepsilon_{3jk}$. By the definitions of ϕ and $*_\phi\phi$, both sides of the equation vanish unless $\{j, k\}$ is one of the subsets $\{1,2\}$, $\{4,7\}$, or $\{5,6\}$, and the identity clearly holds in those cases. The other identities can be proved similarly.

2.5. Matrix and vector representations

The ε-symbol can be used to describe the algebra $\mathfrak{g}_2$ as a subalgebra of $\mathfrak{so}(7)$, the space of skew-symmetric 7-by-7 matrices. A skew-symmetric matrix $a = (a_{ij})$ lies in $\mathfrak{g}_2$ if and only if $\varepsilon_{ijk}a_{jk} = 0$ for all i.

For any vector $v = v_i e_i \in \mathbb{R}^7$, define $[v] = (v_{ij}) \in \mathfrak{so}(7)$ by the formula $v_{ij} = \varepsilon_{ijk}v_k$. It then follows that

$$\mathfrak{so}(7) = \mathfrak{g}_2 \oplus \left[\mathbb{R}^7\right],$$

which is the G_2-invariant irreducible decomposition of $\mathfrak{so}(7)$. Note that $[v]$ is the matrix that represents the linear transformation of $\mathbb{R}^7$ induced by cross-product with $v \in \mathbb{R}^7$.

Conversely, define the map $\langle\cdot\rangle\colon \mathfrak{gl}(7) \to \mathbb{R}^7$ by $\langle\,(a_{ij})\,\rangle = (\varepsilon_{ijk}a_{jk})$. The kernel of this mapping intersected with $\mathfrak{so}(7)$ is $\mathfrak{g}_2$ and the ε-identities imply that, for all $a, b \in \mathbb{R}^7$,

$$\langle\,[a]\,\rangle = 6a \tag{2.10}$$

$$\langle\,[a][b]\,\rangle = 3[b]a = -3[a]b. \tag{2.11}$$

2.6. The G_2-type decomposition of exterior forms

To avoid writing $(\mathbb{R}^7)^*$ many times, I will, for the rest of this section, use V as an abbreviation for the vector space $\mathbb{R}^7$.

Although G_2 acts irreducibly on V and hence on $\Lambda^1(V^*)$ and $\Lambda^6(V^*)$, it does not act irreducibly on $\Lambda^p(V^*)$ for $2 \le p \le 5$. In order to understand the irreducible decomposition of $\Lambda^p(V^*)$ for p in this range, it suffices to understand the cases $p = 2$ and $p = 3$, since the operator $*_\phi$ induces an isomorphism of G_2-modules $\Lambda^p(V^*) = \Lambda^{7-p}(V^*)$.

In [2], it is shown that there are irreducible G_2-module decompositions

$$\Lambda^2(V^*) = \Lambda^2_{14}(V^*) \oplus \Lambda^2_7(V^*) \tag{2.12}$$

$$\Lambda^3(V^*) = \Lambda^3_{27}(V^*) \oplus \Lambda^3_7(V^*) \oplus \Lambda^3_1(V^*) \tag{2.13}$$

where $\Lambda^p_d(V^*)$ denotes an irreducible G_2-module of dimension d. For $p = 4$ or 5, adopt the convention that $\Lambda^p_d(V^*) = *_\phi(\Lambda^{7-p}_d(V^*))$.

These summands can be characterized as follows:

$$\begin{aligned}
\Lambda^2_7(V^*) &= \left\{\, *_\phi(\alpha \wedge *_\phi\phi) \mid \alpha \in \Lambda^1(V^*) \,\right\} \\
&= \left\{\, \alpha \in \Lambda^2(V^*) \mid \alpha \wedge \phi = 2{*_\phi}\alpha \,\right\} \\
\Lambda^2_{14}(V^*) &= \left\{\, \alpha \in \Lambda^2(V^*) \mid \alpha \wedge \phi = -{*_\phi}\alpha \,\right\} = \mathfrak{g}_2^\flat \\
\Lambda^3_1(V^*) &= \{\, r\phi \mid r \in \mathbb{R} \,\} \\
\Lambda^3_7(V^*) &= \left\{\, *_\phi(\alpha \wedge \phi) \mid \alpha \in \Lambda^1(V^*) \,\right\} \\
\Lambda^3_{27}(V^*) &= \left\{\, \alpha \in \Lambda^3(V^*) \mid \alpha \wedge \phi = 0 \text{ and } \alpha \wedge *_\phi\phi = 0 \,\right\} = \mathsf{i}_\phi\big(S^2_0(V^*)\big).
\end{aligned} \tag{2.14}$$

The notations $\mathfrak{g}_2^\flat$ and $\mathsf{i}_\phi\big(S^2_0(V^*)\big)$ used in (2.14) need some explanation.

First, $\mathfrak{g}_2^\flat$: Under the "musical isomorphism" $\flat\colon V \to V^*$ induced by the G_2-invariant inner product $\langle,\rangle_\phi$, the Lie algebra of G_2, namely $\mathfrak{g}_2 \subset V \otimes V^*$, is identified with $\mathfrak{g}_2^\flat = (\flat \otimes 1)(\mathfrak{g}_2) \subset \Lambda^2(V^*) \subset V^* \otimes V^*$. This subspace is an irreducible G_2-module since G_2 is simple.

Second, $\mathsf{i}_\phi\big(S^2_0(V^*)\big)$: Consider the linear mapping $\mathsf{i}_\phi\colon S^2(V^*) \to \Lambda^3(V^*)$, defined on decomposable elements by

$$\mathsf{i}_\phi(\alpha \circ \beta) = \alpha \wedge *_\phi(\beta \wedge *_\phi \phi) + \beta \wedge *_\phi(\alpha \wedge *_\phi \phi). \tag{2.15}$$

The mapping i_ϕ is G_2-invariant and one can show that $S^2(V^*) = \mathbb{R}g_\phi \oplus S^2_0(V^*)$ is a decomposition of $S^2(V^*)$ into G_2-irreducible summands. Evidently, i_ϕ is nonzero on each summand and is therefore injective. Hence, the image $\mathsf{i}_\phi\left(S^2_0(V^*)\right) \subset \Lambda^3(V^*)$ is 27-dimensional and irreducible. The equation

$$\Lambda^3_{27}(V^*) = \left\{\ \alpha \in \Lambda^3(V^*) \mid \alpha \wedge \phi = 0 \text{ and } \alpha \wedge *_\phi \phi = 0\ \right\} \tag{2.16}$$

defines $\Lambda^3_{27}(V^*)$ as a G_2-invariant, 27-dimensional subspace of $\Lambda^3(V^*)$. By dimension count, it must intersect $\mathsf{i}_\phi\left(S^2_0(V^*)\right)$ nontrivially. Since this intersection is also G_2-invariant and since $\mathsf{i}_\phi\left(S^2_0(V^*)\right)$ is G_2-irreducible, $\mathsf{i}_\phi\left(S^2_0(V^*)\right) = \Lambda^3_{27}(V^*)$.

Using the ε-notation, one can express the map i_ϕ in indices as

$$\mathsf{i}_\phi(h_{ij}e^ie^j) = \varepsilon_{ikl}\, h_{ij}\, e^j \wedge e^k \wedge e^l, \tag{2.17}$$

making it evident that $\mathsf{i}_\phi(g_\phi) = 6\phi$.

It will be useful to have a way to invert the map i_ϕ. Define $\mathsf{j}_\phi : \Lambda^3(V^*) \to S^2(V^*)$ by the formula

$$\mathsf{j}_\phi(\gamma)(v,w) = *_\phi\big((v \lrcorner \phi) \wedge (w \lrcorner \phi) \wedge \gamma\big). \tag{2.18}$$

for $\gamma \in \Lambda^3(V^*)$ and $v, w \in V$. It is not difficult to verify that

$$\mathsf{j}_\phi\big(\mathsf{i}_\phi(h)\big) = 8h + 4\big(\mathrm{tr}_{g_\phi}(h)\big)\, g_\phi \tag{2.19}$$

for all $h \in S^2(V^*)$. Note also that $\mathsf{j}_\phi(\phi) = 6g_\phi$, while $\mathsf{j}_\phi\big(\Lambda^3_7(V^*)\big) = 0$.

Note that i_ϕ and j_ϕ are not isometries when $S^2_0(V^*)$ and $\Lambda^3_{27}(V^*)$ are given their natural metrics.[1] Instead, $\gamma \in \Lambda^3_{27}(V^*)$ satisfies $|\mathsf{j}_\phi(\gamma)|^2 = 8\,|\gamma|^2$ while $h \in S^2_0(V^*)$ satisfies $|\mathsf{i}_\phi(h)|^2 = 8\,|h|^2$.

2.7. More G_2 representation theory

It will, from time to time, be useful to have some deeper knowledge of the representation theory of G_2, so some of these facts will be collected here. For details, consult [13].

Since G_2 is a simple Lie group of rank 2, its irreducible representations can be indexed by a pair of integers (p, q) that represent the highest weight of the representation with respect to a fixed maximal torus in G_2 endowed with fixed base for its root system. The irreducible representation of highest weight (p, q) will be denoted $\mathsf{V}_{p,q}$.

[1] The usual inner product on exterior forms is meant here, while, when $h = h_{ij}\, e^ie^j$ with (e^i) being a g-orthonormal coframe of V, one sets $|h|^2 = h_{ij}h_{ij}$.

2.7.1. The standard representation

The fundamental representation $\mathsf{V}_{1,0} \simeq \mathbb{R}^7$ is the 'standard' representation in which G_2 has been defined in this article.

The representation $\mathsf{V}_{p,0}$ for $p \geq 0$ is isomorphic to $\mathsf{S}^p_0(\mathbb{R}^7)$, i.e., the symmetric, trace-free polynomials of degree p in seven variables. (It is somewhat remarkable that these irreducible representations of SO(7) remain irreducible when thought of as representations of G_2.) In this article, the only representations $\mathsf{V}_{p,0}$ in this series that will be important are those for $p = 0, 1, 2$.

2.7.2. The adjoint representation

The other fundamental representation, $\mathsf{V}_{0,1} \simeq \mathbb{R}^{14}$ is isomorphic to $\mathfrak{g}_2$, i.e., is the adjoint representation of G_2. The representation $\mathsf{V}_{0,p}$ for $p \geq 0$ is then the irreducible constituent of highest weight in $\mathsf{S}^p(\mathfrak{g}_2)$.

In this article, only $\mathsf{V}_{0,1} \simeq \mathfrak{g}_2$ and $\mathsf{V}_{0,2} \simeq \mathbb{R}^{77}$ from this series will be important. (This latter one will be important because it is the space of curvature tensors of G_2-metrics.) The reader must be careful not to confuse the representation $\mathsf{V}_{0,2}$ with $\mathsf{V}_{3,0}$, which also happens to have dimension 77.

A few more facts about this representation will be needed: The group G_2 has rank 2 and a maximal torus for G_2 can be obtained by simply taking a maximal torus in the subgroup SU(3). Moreover, every element in $\mathfrak{g}_2$ is $\mathrm{Ad}(\mathrm{G}_2)$-conjugate to an element in such a maximal torus. Consequently, every element in $\Lambda^2_{14}(\mathbb{R}^7) = \mathfrak{g}_2^\flat$ is conjugate to an element of the form

$$\alpha = \lambda_1\, e^{23} + \lambda_2\, e^{45} - (\lambda_1{+}\lambda_2)\, e^{67} \tag{2.20}$$

since these span $\mathfrak{t}^\flat \subset \mathfrak{g}_2^\flat$, where $\mathfrak{t} \subset \mathfrak{g}_2$ is a Cartan subalgebra. Moreover, it is well-known that the ring of $\mathrm{Ad}(\mathrm{G}_2)$-invariant polynomials on $\mathfrak{g}_2$ is a free polynomial ring on two generators, one of degree 2 and one of degree 6. One sees from the above normal form that these two generators can be taken to be $|\alpha|^2$ and $|\alpha^3|^2$. Thus, two elements α and β in $\Lambda^2_{14}(\mathbb{R}^7)$ are conjugate under the action of G_2 if and only if they satisfy $|\alpha|^2 = |\beta|^2$ and $|\alpha^3|^2 = |\beta^3|^2$. In particular, the normal form (2.20) can be made unique by requiring that $0 \leq \lambda_1 \leq \lambda_2$.

In particular, one obtains, for all $\alpha \in \Lambda^2_{14}(V^*)$, the useful identity

$$|\,\alpha^2\,|^2 = |\alpha|^4 \tag{2.21}$$

and inequality

$$|\,\alpha^3\,|^2 \leq \tfrac{2}{3}\,|\alpha|^6, \tag{2.22}$$

which are easily verified by checking them on elements of the form (2.20).

In fact, using the normal form (2.20), one can prove other useful exterior algebra identities. One that will be needed later is

$$\alpha \wedge *_\phi(\alpha\wedge\alpha) = |\alpha|^2 *_\phi\alpha - \tfrac{1}{3} *_\phi(\alpha^3) \wedge *_\phi\phi \qquad \text{for } \alpha \in \Lambda^2_{14}(V^*). \tag{2.23}$$

2.7.3. Other representations

Of the representations $\mathsf{V}_{p,q}$ with p and q positive, only $\mathsf{V}_{1,1} \simeq \mathbb{R}^{64}$ will play any significant role in this article (and mainly as a nuisance at that). In fact, each of the other representations $\mathsf{V}_{p,q}$ with both p and q positive has dimension at least 189, so these can easily be ruled out for dimension reasons in the calculations to follow.

The following tensor product and Schur functor decompositions will be useful:

$$\begin{aligned}
\mathsf{S}^2\big(V_{1,0}\big) &\simeq \mathsf{V}_{0,0} \oplus \mathsf{V}_{2,0} \\
\Lambda^2\big(V_{1,0}\big) &\simeq \mathsf{V}_{1,0} \oplus \mathsf{V}_{0,1} \\
\mathsf{V}_{1,0} \otimes \mathsf{V}_{0,1} &\simeq \mathsf{V}_{1,0} \oplus \mathsf{V}_{2,0} \oplus \mathsf{V}_{1,1} \\
\mathsf{S}^2\big(V_{0,1}\big) &\simeq \mathsf{V}_{0,0} \oplus \mathsf{V}_{2,0} \oplus \mathsf{V}_{0,2} \\
\Lambda^2\big(V_{0,1}\big) &\simeq \mathsf{V}_{0,1} \oplus \mathsf{V}_{3,0}
\end{aligned} \tag{2.24}$$

2.7.4. An example of G_2-type decomposition

As an application of these formulae that will be used below, consider the problem of decomposing $\beta{\wedge}\beta \in \Lambda^4(V^*)$ into its G_2-types where β lies in $\Lambda^2_{14}(V^*) \simeq \mathsf{V}_{0,1}$. Since

$$\Lambda^4(V^*) \simeq \Lambda^4_1(V^*) \oplus \Lambda^4_7(V^*) \oplus \Lambda^4_{27}(V^*) \simeq \mathsf{V}_{0,0} \oplus \mathsf{V}_{1,0} \oplus \mathsf{V}_{0,2} \tag{2.25}$$

and since, by (2.24), we have $\mathsf{S}^2\big(\mathsf{V}_{0,1}\big) \simeq \mathsf{V}_{0,0} \oplus \mathsf{V}_{2,0} \oplus \mathsf{V}_{0,2}$, it follows that $\beta{\wedge}\beta$ can have no component in $\Lambda^4_7(V^*) \simeq \mathsf{V}_{1,0}$. Moreover, since there is, up to multiples, only one G_2-invariant quadratic form on $\mathsf{V}_{0,1}$ and since $*_\phi\phi$ spans $\Lambda^4_1(V^*) \simeq \mathsf{V}_{0,0}$, it follows that there is a constant λ such that

$$\beta \wedge \beta = \lambda\,|\beta|^2 *_\phi \phi + \big(\beta \wedge \beta - \lambda\,|\beta|^2 *_\phi \phi\big) \tag{2.26}$$

where the first term on the right lies in $\Lambda^4_1(V^*)$ while the second term (in parentheses) lies in $\Lambda^4_{27}(V^*)$.

The constant λ is determined as follows: Wedging both sides with ϕ and using the fact that $\beta{\wedge}\phi = -{*_\phi}\beta$ while $\gamma{\wedge}\phi = 0$ for $\gamma \in \Lambda^4_{27}(V^*)$ yields

$$-\,|\beta|^2 *_\phi 1 = \beta \wedge \beta \wedge \phi = \big(\lambda\,|\beta|^2 *_\phi \phi\big) \wedge \phi = 7\lambda\,|\beta|^2 *_\phi 1, \tag{2.27}$$

showing that $\lambda = -\frac{1}{7}$. Thus, the G_2-type decomposition is given by

$$\beta \wedge \beta = -\tfrac{1}{7}\,|\beta|^2 *_\phi \phi + \big(\beta \wedge \beta + \tfrac{1}{7}\,|\beta|^2 *_\phi \phi\big). \tag{2.28}$$

for $\beta \in \Lambda^2_{14}(V^*)$. Of course, this decomposition is orthogonal, so, using the identity (2.21), one can take the square norms of both sides, yielding

$$|\beta|^4 = |\beta \wedge \beta|^2 = \tfrac{1}{7}|\beta|^4 + \big|\beta \wedge \beta + \tfrac{1}{7}\,|\beta|^2 *_\phi \phi\big|^2. \tag{2.29}$$

Consequently, for $\beta \in \Lambda^2_{14}(V^*)$, one has

$$\big|\,\beta \wedge \beta + \tfrac{1}{7}\,|\beta|^2 *_\phi \phi\,\big|^2 = \tfrac{6}{7}|\beta|^4, \tag{2.30}$$

an identity that will be used below. (Note that (2.30) implies, in particular, that the $\Lambda^4_{27}(V^*)$-piece of $\beta\wedge\beta$ cannot vanish unless β itself vanishes, a result equivalent to Lemma 5.8 of [3].)

Similar sorts of calculations can be used to establish the (sharp) inequalities for quadratic forms

$$-\,2\,|\beta|^2\,g \;\le\; \mathsf{j}\big(*_\phi(\beta\wedge\beta)\big) \;\le\; \tfrac{2}{3}\,|\beta|^2\,g. \tag{2.31}$$

Details are left to the reader.

2.8. Definite forms

The dimension of G_2 is 14 and so, by dimension count, the $GL(V)$-orbit of ϕ in $\Lambda^3(V^*)$ is open. Denote this orbit by $\Lambda^3_+(V^*)$ and speak of the elements of $\Lambda^3_+(V^*)$ as *definite* 3-forms on V. Note that $\Lambda^3_+(V)$ has two components, since $GL(V)$ does and since G_2 is connected. Each component is the negative of the other. It is known [10] that $SO(7)/G_2 \simeq \mathbb{RP}^7$, so that each component of $\Lambda^3_+(V)$ is diffeomorphic to $\mathbb{RP}^7\times\mathbb{R}^{28}$.

2.8.1. On general 7-dimensional vector spaces

If W is any 7-dimensional vector space, an isomorphism $u\colon W \xrightarrow{\sim} V$ induces an isomorphism $u^*\colon \Lambda^3(V^*) \xrightarrow{\sim} \Lambda^3(W^*)$. Denote by $\Lambda^3_+(W^*)$ the open subset $u^*\big(\Lambda^3_+(V^*)\big) \subset \Lambda^3(W^*)$. Since $\Lambda^3_+(V^*)$ consists of a single $GL(V)$-orbit, this set does not depend on the choice of u.

2.8.2. Associated algebraic structures

Each $\varphi\in\Lambda^3_+(W^*)$ has a stabilizer in $GL(W)$ that is isomorphic to G_2 and hence defines a canonical inner product $\langle,\rangle_\varphi$ (with associated quadratic form g_φ) and orientation (Hodge star) $*_\varphi : \Lambda^p(W^*) \to \Lambda^{7-p}(W^*)$.

Similarly, using φ in the place of ϕ in the formulae (2.15) and (2.18), one defines mappings $\mathsf{i}_\varphi : S^2(W^*) \to \Lambda^3(W^*)$ and $\mathsf{j}_\varphi : \Lambda^3(W^*) \to S^2(W^*)$. These maps are frequently useful in formulae.

For example, let $\mathsf{G} : \Lambda^3_+(W^*) \to S^2_+(W^*)$ be the nonlinear $GL(W)$-equivariant mapping that satisfies $\mathsf{G}(\varphi) = g_\varphi$. It is not difficult to show that G is smooth and satisfies

$$\mathsf{G}'(\varphi)(\psi) = \tfrac{1}{2}\,\mathsf{j}_\varphi(\psi) - \tfrac{1}{3}\,*_\varphi(\psi\wedge *_\varphi\varphi)\,g_\varphi\,. \tag{2.32}$$

There is also an associated *vector cross product* $\times_\varphi : W\times W \to W$ defined by the condition

$$\langle w_1 \times_\varphi w_2\,, w_3\rangle_\varphi = \varphi(w_1, w_2, w_3). \tag{2.33}$$

Remark 1 (The vector cross product definition of G_2). Given a vector space V over $\mathbb{R}$ endowed with a positive definite inner product $\langle,\rangle : V\times V\to\mathbb{R}$, a (2-fold) *vector cross product* on $\big(V,\langle,\rangle\big)$ is a skew-symmetric bilinear pairing $\times : V\times V \to V$ that satisfies

$$\langle v_1\times v_2, v_1\rangle = 0 \qquad\text{and}\qquad |v_1\times v_2|^2 = |v_1|^2\,|v_2|^2 - \langle v_1, v_2\rangle^2 \tag{2.34}$$

for all $v_1, v_2\in V$.

It can be shown that the $GL(7,\mathbb{R})$-stabilizer of the vector cross product $\times_\phi$ is equal to G_2. Hence one could take $\times_\phi : \mathbb{R}^7 \times \mathbb{R}^7 \to \mathbb{R}^7$ as the algebraic structure defining G_2. In fact, this is what Gray did in his work on G_2-structures. However, I find that the 3-form formulation is more congenial for computations, so vector cross products will not play any significant role in this article.

2.8.3. Definite 4-forms

The canonical mapping $\mathsf{S} : \Lambda^3_+(W^*) \to \Lambda^4(W^*)$ defined by $\mathsf{S}(\varphi) = *_\varphi \varphi$ is a double covering onto an open set $\Lambda^4_+(W^*)$ in $\Lambda^4(W^*)$, which will be referred to as the space of 'definite' 4-forms on W.

The $GL(W)$-stabilizer of an element $\psi \in \Lambda^4_+(W^*)$ is then isomorphic to $\pm G_2 = G_2 \cup \big(G_2 \cdot (-\mathrm{id}_W)\big)$. Thus, a definite 4-form on W defines an inner product on W, but not an orientation.

3. G_2-structures

3.1. Definite forms on manifolds

Let M be a smooth manifold of dimension 7. The union of the subspaces $\Lambda^3_+(T^*_xM)$ is an open subbundle $\Lambda^3_+(T^*M) \subset \Lambda^3(T^*M)$ of the bundle of 3-forms on M.

Definition 2 (Definite 3-forms on manifolds). A 3-form σ on M that takes values in $\Lambda^3_+(T^*M)$ will be said to be a *definite* 3-form on M. The set of definite 3-forms on M will be denoted $\Omega^3_+(M)$.

3.1.1. G_2-structures and definite 3-forms

Each definite 3-form on M defines a G_2-structure on M in the following way:

Let $\mathcal{F}$ denote the principal right $GL(V)$-bundle over M consisting of V-coframes $u\colon T_xM \xrightarrow{\sim} V$. Given any $\sigma \in \Omega^3_+(M)$, define a G_2-bundle

$$F_\sigma = \{u \in \mathrm{Hom}(T_xM, V) \ \mid \ x \in M \text{ and } u^*(\phi) = \sigma_x\}\,. \tag{3.1}$$

Every G_2-reduction of $\mathcal{F}$ (i.e., G_2-structure on M in the usual sense) is of the form F_σ for some unique $\sigma \in \Omega^3_+(M)$. For this reason, a 3-form $\sigma \in \Omega^3_+(M)$ will usually, by abuse of language, be called a G_2-structure in this article.

Remark 2 (Alternative terminologies). Some authors use 'almost G_2-structure' to refer to what I am calling a G_2-structure in this article. Apparently, this practice stems from an imagined analogy with the distinction between 'almost complex structure' and 'complex structure'.

However, for a subgroup $G \subset GL(n,\mathbb{R})$, the use of '$G$-structure' on an n-manifold M to mean a G-subbundle of the $GL(n,\mathbb{R})$-bundle of frames (or coframes) on M is well established. It seems unwise to tamper with this usage, especially since 'almost G-structure' suggests a structure that lacks some property of actual G-structures. Making an exception for the case $G = G_2$ merely invites confusion.

This use of 'G-structure' does not conflict with the 'almost complex structure' vs. 'complex structure' usage since a complex structure on a $2n$-manifold is not simply a $\mathrm{GL}(n,\mathbb{C})$-structure, but is (by the Newlander-Nirenberg theorem, equivalent to) a $\mathrm{GL}(n,\mathbb{C})$-structure with an assumed integrability property, whereas an 'almost complex structure' actually is (equivalent to) a $\mathrm{GL}(n,\mathbb{C})$-structure, not an 'almost $\mathrm{GL}(n,\mathbb{C})$-structure'.

Some authors speak of an 'integrable G_2-structure', meaning a G_2-structure $\sigma \in \Omega^3_+(M)$ satisfying some differential equations, such as $\mathrm{d}\sigma = 0$ (the exact differential equation intended varies with the author). Again, this usage appears to stem from an imagined analogy with a symplectic structure, which is defined by a nondegenerate 2-form ω that is closed, i.e., $\mathrm{d}\omega = 0$. In the symplectic case, Darboux' Theorem says that ω is, indeed, locally equivalent to the flat model, i.e., is 'integrable' in the standard terminology of the theory of Lie pseudo-groups. (In a similar way, one speaks of 'integrable almost complex structures'.) This usage of 'integrable' for G_2-structures also seems ill-advised to me since, as will be seen below, no first order condition on a G_2-structure implies that it is locally equivalent to the flat model (which is the only interpretation of 'integrable' in this context that would be consistent with the established usage in the theory of Lie pseudo-groups). Moreover, this encourages the confusing shift of terminology in which 'G_2-structure' is used to mean 'integrable G_2-structure' and 'almost G_2-structure' is used to mean an actual G_2-structure.

For this reason, none of the modifiers 'integrable', 'almost', 'nearly', or their ilk will be used in this article when referring to G_2-structures.

However, since it seems to be harmless, the terminology 'G_2-manifold' will sometimes be used to denote a manifold endowed with a G_2-structure that is flat to first order (i.e., 'torsion-free' in the usual terminology).

Definition 3 (Associated metric, orientation, and vector cross product). For any $\sigma \in \Omega^3_+(M)$, denote by g_σ, $*_\sigma$, and $\times_\sigma$ the metric, Hodge star operator, and vector cross product on M that are canonically associated to σ. When it is needed, the oriented orthonormal frame bundle of g_σ with this orientation will be denoted $\mathsf{F}_\sigma = F_\sigma \cdot \mathrm{SO}(7)$.

Remark 3 (Existence of G_2-structures). Because G_2 is both connected and simply connected, a connected 7-manifold M can support a G_2-structure only if it is both orientable and spinnable, i.e., if the first two Stiefel-Whitney classes of M vanish.

Conversely, by an observation due to Gray [9], these two necessary conditions are also sufficient:

Since G_2 is simply connected, it is the image under the standard double covering map $\rho : \mathrm{Spin}(7) \to \mathrm{SO}(7)$ of a unique subgroup of $\mathrm{Spin}(7)$, which, by abuse of language, will also be called G_2. Now, $\mathrm{Spin}(7)$ has a faithful representation on $\mathbb{R}^8$ and hence can be regarded as a subgroup of $\mathrm{SO}(8)$. The restriction of this representation to G_2 must also be faithful and hence, for dimension reasons, it must be isomorphic to $\mathsf{V}_{0,0} \oplus \mathsf{V}_{1,0}$. In particular, G_2 fixes a vector in $\mathbb{R}^8$ and acts transitively on the unit 6-sphere orthogonal to this vector. Consequently, $\mathrm{Spin}(7)$ must act transitively on the unit 7-sphere in $\mathbb{R}^8$ with stabilizer subgroup G_2.

Now, suppose M^7 to be orientable and spinnable. Choose a Riemannian metric g, an orientation, and a spin structure $\tilde{\mathsf{F}} \to M$, i.e., a spin double cover of the SO(7)-bundle $\mathsf{F} \to M$ consisting of oriented, g-orthonormal coframes on M. The associated spinor bundle $\mathbb{S} = \tilde{\mathsf{F}} \times_{\mathrm{Spin}(7)} \mathbb{R}^8$ is a vector bundle of rank 8 over the 7-manifold M and therefore has a nonvanishing unit section $s : M \to \mathbb{S}$. This allows one to reduce the structure group of $\tilde{\mathsf{F}}$ (and hence F) from Spin(7) to G_2 (since, by the previous paragraph, this is, up to conjugacy, the Spin(7)-stablizer of any nonzero vector in $\mathbb{R}^8$). Thus, M admits a G_2-structure whose associated metric and orientation are the chosen ones.

3.2. Type decomposition

Since G_2 acts reducibly on $\Lambda^p(V^*)$ for $2 \le p \le 5$, one can associate to any G_2-structure σ on M natural splittings of the p-form bundles $\Lambda^p(T^*M)$ into direct summands. These will be labeled as $\Lambda^p_d(T^*M, \sigma)$, or more simply, $\Lambda^p_d(T^*M)$ when the structure σ is clear from context. Denote the space of sections of $\Lambda^p_d(T^*M, \sigma)$ by $\Omega^p_d(M, \sigma)$.

Thus, for example, in view of (2.14), one has

$$\Omega^2_7(M,\sigma) = \big\{\, \beta \in \Omega^2(M) \mid \beta \wedge \sigma = 2{*_\sigma}\beta \,\big\} \tag{3.2}$$

$$\Omega^2_{14}(M,\sigma) = \big\{\, \beta \in \Omega^2(M) \mid \beta \wedge \sigma = -{*_\sigma}\beta \,\big\}\,. \tag{3.3}$$

Fortunately, the irreducible modules of dimensions 14 and 27 only occur in one dual pair of dimensions each. Meanwhile, the irreducible module of dimension 7 occurs in each degree $1 \le p \le 6$. From time to time, it is useful to be able to recognize the scale factors that can be introduced by the various different isomorphisms between these different modules. For example, for $\alpha \in \Omega^1_7(M)$ one has

$$\begin{aligned} {*_\sigma}\big({*_\sigma}(\alpha\wedge\sigma)\wedge\sigma\big) &= -4\,\alpha \\ {*_\sigma}\big({*_\sigma}(\alpha\wedge{*_\sigma}\sigma)\wedge{*_\sigma}\sigma\big) &= \ \ 3\,\alpha, \end{aligned} \tag{3.4}$$

and these identities can sometimes be useful in simplifying various expressions. One should also keep in mind that, using the metric, each 1-form α has a corresponding dual vector field $\alpha^\sharp$ and there are useful identities of the form

$$\begin{aligned} {*_\sigma}(\alpha\wedge\sigma) &= -\alpha^\sharp \lrcorner {*_\sigma}\sigma \\ {*_\sigma}(\alpha\wedge{*_\sigma}\sigma) &= \ \ \alpha^\sharp \lrcorner\, \sigma\,. \end{aligned} \tag{3.5}$$

Remark 4 (G_2-structures with the same associated metric and orientation). These type of decompositions have many uses. For example, they furnish a description of all of the G_2-structures that have the same associated metric and orientation as a given $\sigma \in \Omega^3_+(M)$:

Let a and α be a function and a 1-form, respectively, on M with $a^2 + |\alpha|^2_\sigma = 1$. Then the 3-form

$$\tilde\sigma = \big(a^2 - |\alpha|^2_\sigma\big)\,\sigma + 2a\,{*_\sigma}\big(\alpha\wedge\sigma\big) + \mathsf{i}\big(\alpha\circ\alpha\big) \tag{3.6}$$

is definite and has the same associated metric and orientation as σ. (This pointwise fact is most easily proved by checking it in the case $\sigma = \phi$ and $(a,\alpha) = (c, s\,e^1)$ where $c^2+s^2 = 1$

and then using the fact that G_2 acts transitively on the unit 6-sphere in $\mathbb{R}^7$ to reduce to this case.)

Moreover, any definite 3-form on M that has g_σ and $*_\sigma$ as associated metric and orientation is of the form (3.6) for some pair (a,α) satisfying $a^2+|\alpha|^2_\sigma=1$, unique up to replacement by $(-a,-\alpha)$. (If $H^1(M,\mathbb{Z}_2)\neq 0$, the pair (a,α) might only be defined up to sign.)

Of course, some such formula was expected, since $\mathrm{SO}(7)/\mathrm{G}_2\simeq\mathbb{RP}^7$ (a consequence of the result $\mathrm{Spin}(7)/\mathrm{G}_2\simeq S^7$ discussed in Remark 3). What (3.6) displays is a concrete isomorphism between the bundle $\mathsf{F}_\sigma/\mathrm{G}_2$ and the $\mathbb{RP}^7$-bundle $\mathbb{P}\big(\mathbb{R}\oplus T^*M\big)$ over M.

3.3. Exterior derivative formulae

The decomposition of the p-forms on M allows one to express the exterior derivatives of both σ and $*_\sigma\sigma$ in fairly simple terms:

Proposition 1 (The torsion forms). *For any* G_2*-structure* $\sigma\in\Omega^3_+(M)$*, there exist unique differential forms* $\tau_0\in\Omega^0(M)$*,* $\tau_1\in\Omega^1(M)$*,* $\tau_2\in\Omega^2_{14}(M,\sigma)$*, and* $\tau_3\in\Omega^3_{27}(M,\sigma)$ *so that the following equations hold:*

$$\begin{aligned}\mathrm{d}\,\sigma &= \tau_0 *_\sigma\sigma + 3\,\tau_1\wedge\sigma + *_\sigma\tau_3\,,\\ \mathrm{d}*_\sigma\sigma &= \qquad 4\,\tau_1\wedge *_\sigma\sigma + \tau_2\wedge\sigma\,.\end{aligned}\tag{3.7}$$

Proof. In view of the decomposition (2.14), the only part of this proposition that is not simply the definition of the τ_i is the occurrence of τ_1 in two places. In fact, by (2.14), there exist unique forms $\tau_0\in\Omega^0(M)$, $\tau_1,\tilde\tau_1\in\Omega^1(M)$, $\tau_2\in\Omega^2_{14}(M,\sigma)$, and $\tau_3\in\Omega^3_{27}(M,\sigma)$ so that the above equation for $d\,\sigma$ holds while $\mathrm{d}*_\sigma\sigma=4\,\tilde\tau_1\wedge*_\sigma\sigma+\tau_2\wedge\sigma$.

However, as is shown in [2] (see Remark 5 below for a sketch of the proof), there is an identity

$$*_\sigma\sigma\wedge *_\sigma\big(\mathrm{d}(*_\sigma\sigma)\big)+(*_\sigma\mathrm{d}\sigma)\wedge\sigma=0\tag{3.8}$$

valid for all $\sigma\in\Omega^3_+(M)$, and, in view of (3.4), this is equivalent to $\tilde\tau_1=\tau_1$. □

Definition 4 (The torsion forms). For a definite 3-form $\sigma\in\Omega^3_+(M)$, the quadruple of forms $(\tau_0,\tau_1,\tau_2,\tau_3)$ defined by (3.7) will be referred to as the *intrinsic torsion forms* of σ.

Remark 5 (General intrinsic torsion). The existence of the identity (3.8) may seem surprising at first, but the existence of such an identity can be understood by general considerations.

For any subgroup $G\subset\mathrm{SO}(n)$, the first order invariants (usually called the 'intrinsic torsion') of a G-structure F on an n-manifold M take values in a bundle over M associated to the natural G-representation on $(\mathfrak{so}(n)/\,\mathfrak{g})\otimes\mathbb{R}^n$. (See §4.2 below for a further explication of this fact.) When the first order invariants of a given G-structure vanish, it is said to be '1-flat' or 'flat to first order'. For more discussion of this notion, see [2].

In the case of $\mathrm{G}_2\subset\mathrm{SO}(7)$, this torsion representation space is

$$(\mathfrak{so}(7)/\,\mathfrak{g}_2)\otimes\mathbb{R}^7\simeq\mathsf{V}_{1,0}\otimes\mathsf{V}_{1,0}\simeq\mathsf{V}_{0,0}\oplus\mathsf{V}_{1,0}\oplus\mathsf{V}_{0,1}\oplus\mathsf{V}_{2,0}.\tag{3.9}$$

and, as has already been remarked, these four summands are isomorphic, respectively, to $\Lambda^0(V^*)$, $\Lambda^1(V^*)$, $\Lambda^2_{14}(V^*)$, and $\Lambda^3_{27}(V^*)$. Since the exterior derivatives of the defining forms σ and $*_\sigma\sigma$ can be expressed linearly in terms of the first order invariants of F_σ and since there is only one $\Lambda^1(V^*)$ in the above representation list, it follows that the two 1-forms τ_1 and $\tilde{\tau}_1$ alluded to in the above proof must satisfy some universal linear relation.

Consideration of the fact that replacing σ by $\lambda^3\sigma$ for some positive function λ will replace $*_\sigma\sigma$ by $\lambda^4 *_\sigma\sigma$ shows that this relation must be the one given in Proposition 1.

Proposition 2 (1-flatness of G_2-structures)**.** *A G_2-structure $\sigma \in \Omega^3_+(M)$ is flat to first order if and only if its torsion forms all vanish, i.e., if and only if $\mathrm{d}\sigma = \mathrm{d}*_\sigma\sigma = 0$.*

Proof. A G_2-structure $\sigma \in \Omega^3_+(M)$ is flat to first order at $p \in M$ if there exists a p-centered coordinate chart $x : U \to \mathbb{R}^7$ such that the 3-form $\sigma - x^*(\phi)$ on U vanishes to order at least 2 at p.

Recall that the map $\mathsf{S} : \Lambda^3_+(W^*) \to \Lambda^4_+(W^*)$ defined in §2.8 is a smooth double covering. This implies that if $\sigma - x^*(\phi)$ vanishes to order 2 at p, then $*_\sigma\sigma - x^*(*_\phi\phi)$ vanishes to order 2 at p as well.

Since $\mathrm{d}\phi = \mathrm{d}*_\phi\phi = 0$, if σ is flat to first order at p, then $\mathrm{d}\sigma$ and $\mathrm{d}*_\sigma\sigma$ must vanish to at least first order at p. Thus, the claim in one direction is established.

To demonstrate the claim in the converse direction, it suffices to show that any definite 3-form σ defined on a neighborhood of $0 \in \mathbb{R}^7$ that satisfies $\sigma_0 = \phi$ and $\mathrm{d}\sigma = \mathrm{d}*_\sigma\sigma = 0$ is flat to first order at $0 \in \mathbb{R}^7$.

Now, if ψ is any 3-form on $\mathbb{R}^7$ that vanishes at the origin $0 \in \mathbb{R}^7$, then, because $\Lambda^3_+(V^*)$ is an open set in $\Lambda^3(V^*)$, the 3-form $\sigma = \phi + \psi$ is a definite 3-form on some open neighborhood of $0 \in \mathbb{R}^7$. Since $\mathrm{d}\sigma = \mathrm{d}\psi$, and since, for any 4-form $\Psi \in \Lambda^4(V^*)$, there exists a 3-form ψ on $\mathbb{R}^7$ that vanishes at $0 \in \mathbb{R}^7$ and that satisfies $(\mathrm{d}\psi)_0 = \Psi$, it follows that the condition $\mathrm{d}\sigma = 0$, i.e., $\tau_0 = \tau_1 = \tau_3 = 0$, imposes 35 independent linear conditions on the intrinsic torsion of σ. Since these conditions must define some G_2-invariant subspace of the torsion representation $\mathsf{V}_{0,0}\oplus\mathsf{V}_{1,0}\oplus\mathsf{V}_{0,1}\oplus\mathsf{V}_{2,0}$, it follows by dimension count that it is the subspace $\mathsf{V}_{0,0}\oplus\mathsf{V}_{1,0}\oplus\mathsf{V}_{2,0}$.

Similarly, since $\Lambda^4_+(W^*)$ is an open subset of $\Lambda^4(W^*)$ and since $\mathsf{S} : \Lambda^3_+(W^*) \to \Lambda^4_+(W^*)$ is a smooth double covering, it follows that if ψ is any smooth 4-form vanishing at the origin $0 \in \mathbb{R}^7$, then there is an open neighborhood U of $0 \in \mathbb{R}^7$ on which there exists a definite form σ such that $\sigma_0 = \phi$ and $*_\sigma\sigma = *_\phi\phi + \psi$. Moreover, if Ψ is any 5-form in $\Lambda^5(V^*)$, then there exists a smooth 4-form ψ vanishing at $0 \in \mathbb{R}^7$ such that $(\mathrm{d}\psi)_0 = \Psi$. The corresponding definite 3-form σ will then satisfy $\mathrm{d}*_\sigma\sigma = \mathrm{d}\psi$, so that $(\mathrm{d}*_\sigma\sigma)_0 = \Psi$. It follows that the condition $\mathrm{d}*_\sigma\sigma = 0$, i.e., $\tau_1 = \tau_2 = 0$, must be 21 independent linear equations on the intrinsic torsion of σ. Since these conditions must define some G_2-invariant subspace of the torsion representation $\mathsf{V}_{0,0}\oplus\mathsf{V}_{1,0}\oplus\mathsf{V}_{0,1}\oplus\mathsf{V}_{2,0}$, it follows by dimension count that it is the subspace $\mathsf{V}_{1,0}\oplus\mathsf{V}_{0,1}$.

Thus, the conditions $\mathrm{d}\sigma = 0$ and $\mathrm{d}*_\sigma\sigma = 0$ together imply that all of the intrinsic torsion of σ vanishes, i.e., that σ is flat to first order at each point. $\square$

Remark 6 (Fernández and Gray's theorem on vector cross products). Proposition 2 implies the 1982 result of Fernández and Gray [6] that a vector cross product $\times : TM{\times}TM \to TM$ that is compatible with a Riemannian metric g on M is g-parallel if and only if the corresponding 3-form is closed and coclosed (with respect to g).

The essential difference between Proposition 2 and their result is that they assume a specific metric g and vector cross product to be given, whereas Proposition 2 starts with a definite 3-form σ and constructs a specific metric associated to σ.

4. Frame Bundle Calculations

4.1. The associated Levi-Civita connection

Let $\sigma \in \Omega^3_+(M)$ be a G_2-structure with associated G_2-bundle $F_\sigma \subset \mathcal{F}$. This bundle can be canonically enlarged to an oriented orthonormal frame bundle $\mathsf{F}_\sigma = F_\sigma \cdot SO(7) \subset \mathcal{F}$ and this larger bundle will be referred to as the associated metric frame bundle of σ.

Now $\pi\colon \mathsf{F}_\sigma \to M$ has a tautological V-valued 1-form ω defined by requiring that $\omega(v) = u(\pi_*(v))$ for all $v \in T_u\mathsf{F}$. It may help the reader to think of ω as expanded in the basis e_i in the form $\omega = \omega_1\, e_1 + \cdots + \omega_7\, e_7$ and then think of ω as a column of height 7, i.e., $\omega = (\omega_i)$.

The Levi-Civita connection is then represented on F_σ as a 1-form ψ on F_σ taking values in $\mathfrak{so}(7)$, i.e., the 7-by-7 skew-symmetric matrices. As such, $\psi = (\psi_{ij})$ where $\psi_{ij} = -\psi_{ji}$.

The defining property of ψ is that it satisfies the *first structure equation* of Cartan:

$$\mathrm{d}\,\omega = -\psi \wedge \omega. \tag{4.1}$$

In indices (i.e., components) this matrix equation becomes the system of equations $\mathrm{d}\omega_i = -\psi_{ij}{\wedge}\omega_j$.

The curvature of this connection is represented by the 2-form $\Psi = \mathrm{d}\psi + \psi{\wedge}\psi$. It satisfies the *first Bianchi identity*

$$\Psi \wedge \omega = 0 \tag{4.2}$$

and has the indicial expression

$$\Psi_{ij} = \mathrm{d}\psi_{ij} + \psi_{ik} \wedge \psi_{kj} = \tfrac{1}{2}\mathsf{R}_{ijkl}\, \omega_k \wedge \omega_l\,. \tag{4.3}$$

4.1.1. The natural connection and intrinsic torsion on F_σ

To save writing, I will denote the pullbacks of ω and ψ to F_σ by the same letters, trusting the reader to keep in mind where various equations are taking place.

The pullback of ψ to F_σ will not generally have values in $\mathfrak{g}_2 \subset \mathfrak{so}(7)$. However, keeping in mind the canonical decomposition $\mathfrak{so}(7) = \mathfrak{g}_2 \oplus [V]$, there is a unique decomposition of the form

$$\psi = \theta + 2[\tau] \tag{4.4}$$

where θ takes values in $\mathfrak{g}_2$ and τ takes values in V. (The coefficient 2 simplifies subsequent formulas.)

Then θ is a connection 1-form on F_σ and defines what will be referred to as the *natural* connection associated to the G_2-structure σ. This connection will not be torsion-free (and hence is not the Levi-Civita connection) unless τ vanishes identically.

4.2. General G-structure torsion

This construction of a natural connection for a G_2-structure σ is an instance of a general construction valid for any $G \subset O(n)$.

Letting $\mathfrak{g} \subset \mathfrak{so}(n)$ denote the Lie algebra of G, there is a unique G-equivariant splitting $\mathfrak{so}(n) = \mathfrak{g} \oplus \mathfrak{g}^\perp$ obtained by using the standard $O(n)$-invariant inner product on $\mathfrak{so}(n)$.

For any G-structure $\pi : F \to M$, one has the associated orthonormal frame bundle $\mathsf{F} = F \cdot O(n)$. One can then pull back the Levi-Civita connection ψ on F to F and decompose it uniquely in the form $\psi = \theta + \tau$ where θ takes values in $\mathfrak{g}$ and τ takes values in $\mathfrak{g}^\perp \simeq \mathfrak{so}(n)/\mathfrak{g}$. The 1-form θ defines a natural connection on F (one that is the pullback to F of a metric-compatible connection, generally with torsion, on F). The 1-form τ represents a section T of the associated *torsion bundle* $F \times_\rho (\mathfrak{g}^\perp \otimes \mathbb{R}^n)$, where $\rho : G \to \mathrm{End}\big(\mathfrak{g}^\perp \otimes \mathbb{R}^n\big)$ is the tensor product of the two obvious representations.

It is a general result (essentially due to É. Cartan) that all of the pointwise first-order diffeomorphism invariants of a G-structure $F \subset \mathcal{F}$ that are polynomial in the derivatives of the corresponding defining section σ of the bundle $\mathcal{F}/G$ are expressible as polynomials in the section T.

Moreover, for $k \geq 2$, all of the pointwise k-th order diffeomorphism invariants of a G-structure $F \subset \mathcal{F}$ that are polynomial in the first k derivatives of the corresponding defining section σ of the bundle $\mathcal{F}/G$ are expressible as polynomials in the section T, its first $k-1$ covariant derivatives with respect to the connection θ, the curvature of θ, and its first $k-2$ covariant derivatives (with respect to θ).

Consequently, for each $k \geq 1$, the polynomial pointwise invariants of order k are polynomials in a canonically defined section of a vector bundle of the form

$$F \times_{\rho_1 \times \cdots \times \rho_k} \big(V_1(\mathfrak{g}) \oplus \cdots \oplus V_k(\mathfrak{g})\big)$$

where $V_k(\mathfrak{g})$ is the unique G-representation that satisfies

$$\big(\mathfrak{gl}(n,\mathbb{R})/\mathfrak{g}\big) \otimes \mathsf{S}^k(\mathbb{R}^n) = V_k(\mathfrak{g}) \oplus \big(\mathbb{R}^n \otimes \mathsf{S}^{k+1}(\mathbb{R}^n)\big). \tag{4.5}$$

In the familiar case in which $\mathfrak{g} = \mathfrak{so}(n)$, the first torsion space $V_1\big(\mathfrak{so}(n)\big)$ vanishes (this is simply the fundamental lemma of Riemannian geometry) and one has the result (due to Cartan and Weyl) that all of the pointwise invariants of a metric can be expressed in terms of the Riemann curvature tensor and its covariant derivatives with respect to the Levi-Civita connection.

Remark 7 (Canonical connections). The use of the term 'natural' with regard to the connection θ on the G-structure F should not be construed to mean that this is the only 'canonical' connection on M that is compatible with F. In many cases, this is only one of a family of possible 'canonical' connections that can be defined in terms of the first-order invariants of the G-structure F and that are preserved under equivalence of G-structures.

For example, if the G-modules $V_1(\mathfrak{g})$ and $\mathfrak{g}\otimes\mathbb{R}^n$ have common constituents, so that the space $\mathrm{Hom}^G\big(V_1(\mathfrak{g}), \mathfrak{g}\otimes\mathbb{R}^n\big)$ of G-equivariant homomorphisms between the two spaces has dimension $r > 0$, there will be an r-parameter family of ways of modifying θ, by adding a $\mathfrak{g}$-valued 1-form whose coefficients are linear in the torsion functions, in such a way that the resulting modification defines a connection on M compatible with the G-structure F. Each element in this r-parameter family of connections can be regarded as canonical in the sense that equivalence of G-structures will induce isomorphisms between the corresponding connections in the r-parameter family.

Of course, there is no *a priori* reason to consider only connection modifications that are linear in the torsion functions; for example, any G-equivariant polynomial mapping $V_1(\mathfrak{g}) \to \mathfrak{g}\otimes\mathbb{R}^n$ could be used to define such a modification of θ. However, these 'higher' modifications do not often arise in practice.

Depending on the intended use, it could well be that one of these other connections (rather than the one being called 'natural' in the present article) is better suited for expressing identities of one kind or another.

4.3. G_2-specific calculations

In the specific case of $\mathrm{G}_2 \subset \mathrm{SO}(7)$, one finds, as has already been remarked,

$$V_1(\mathfrak{g}_2) \simeq \mathsf{V}_{0,0} \oplus \mathsf{V}_{1,0} \oplus \mathsf{V}_{0,1} \oplus \mathsf{V}_{2,0}, \tag{4.6}$$

while $V_2(\mathfrak{g}_2)$, which has dimension 392, has the decomposition

$$V_2(\mathfrak{g}_2) \simeq \mathsf{V}_{0,0} \oplus 2\mathsf{V}_{1,0} \oplus \mathsf{V}_{0,1} \oplus 3\mathsf{V}_{2,0} \oplus 2\mathsf{V}_{1,1} \oplus \mathsf{V}_{0,2} \oplus \mathsf{V}_{3,0}\,. \tag{4.7}$$

Naturally, this latter space has $V_2\big(\mathfrak{so}(7)\big)$, i.e., the curvature tensors of metrics in dimension 7, as a quotient. For comparison, note that, as G_2-modules:

$$V_2\big(\mathfrak{so}(7)\big) \simeq \mathsf{V}_{0,0} \oplus 2\mathsf{V}_{2,0} \oplus \mathsf{V}_{1,1} \oplus \mathsf{V}_{0,2}\,. \tag{4.8}$$

The Ricci tensor takes values in a subspace isomorphic to $\mathsf{V}_{0,0}\oplus\mathsf{V}_{2,0}$ while the remainder represents the Weyl tensor.

Remark 8 (Canonical G_2-connections). Since $\mathfrak{g}_2\otimes\mathsf{V}_{1,0} = \mathsf{V}_{1,0} \oplus \mathsf{V}_{2,0} \oplus \mathsf{V}_{1,1}$ shares two G_2-irreducible modules with $V_1(\mathfrak{g})$, it follows from Remark 7 that there is actually a 2-parameter family of canonical connections associated to any G_2-structure σ. Each element in this family is compatible with σ (in the sense that σ is parallel under the corresponding parallel translation). Since the common constituents $\mathsf{V}_{1,0}$ and $\mathsf{V}_{2,0}$ correspond to the torsion forms τ_1 and τ_3, respectively, it follows that the entire two-parameter family of canonical connections collapses to a single connection if and only if the G_2-structure σ satisfies $\tau_1 = \tau_3 = 0$. In this case, differentiating the equations (3.7) shows that $\tau_0\tau_2 = 0$ and $\mathrm{d}\tau_0 = 0$. In particular, when M is connected, it follows that τ_0 is constant. If $\tau_0 = 0$, then the G_2-structure is closed. If $\tau_0 \neq 0$, then $\tau_2 = 0$ and one has the equation $\mathrm{d}\sigma = \tau_0 *_\sigma \sigma$, which is the defining equation for the so-called 'nearly G_2-manifolds'.

Thus, the family of canonical G_2-connections associated to a G_2-structure σ collapses to a single G_2-connection if and only if either σ is closed or it defines a nearly G_2-manifold.

4.4. The second structure equations

It is helpful to make the following observation: The identities (2.10) imply that the 2-form $2[\tau]{\wedge}[\tau] + [\,[\tau]{\wedge}\tau\,]$ takes values in $\mathfrak{g}_2$. This motivates the definitions

$$D\tau = d\tau + \theta \wedge \tau - [\tau] \wedge \tau \tag{4.9}$$

$$D\theta = d\theta + \theta \wedge \theta + 4[\tau] \wedge [\tau] + 2[\,[\tau] \wedge \tau\,], \tag{4.10}$$

for, with these definitions, $D\theta$ takes values in $\mathfrak{g}_2$. Moreover

$$\Psi = d(\theta + 2[\tau]) + (\theta + 2[\tau]) \wedge (\theta + 2[\tau]) = D\theta + 2[D\tau] \tag{4.11}$$

so that the first Bianchi identity takes the form

$$(D\theta + 2[D\tau]) \wedge \omega = 0. \tag{4.12}$$

Remark 9 (Covariant differentials). The decisive advantage of using the forms $D\tau$ and $D\theta$ to express the curvature tensor is that these forms do *not* contain all of the information about the second order invariants of the underlying G_2-structure σ although they do contain enough information to recover the Riemann curvature tensor of the underlying metric.

4.5. Indicial calculations

The indicial expression of (4.1) in terms of (4.4) is

$$d\omega_i = -\theta_{ij} \wedge \omega_j - 2\varepsilon_{ijk}\tau_k \wedge \omega_j\,. \tag{4.13}$$

Denote $\pi^*(\sigma)$ by σ and, with a slight abuse of notation, denote $\pi^*(*_\sigma \sigma)$ by $\star\sigma$. Then

$$\sigma = \tfrac{1}{6}\varepsilon_{ijk}\, \omega_i \wedge \omega_j \wedge \omega_k \tag{4.14}$$

$$\star\sigma = \tfrac{1}{24}\varepsilon_{ijkl}\, \omega_i \wedge \omega_j \wedge \omega_k \wedge \omega_l\,. \tag{4.15}$$

These give rise, via (4.13) and the ε-identities, to the formulae

$$\begin{aligned} d\,\sigma &= \varepsilon_{ijkl}\, \tau_i \wedge \omega_j \wedge \omega_k \wedge \omega_l \\ d\star\sigma &= -\,(\tau_p \wedge \omega_p) \;\wedge\; (\varepsilon_{ijk}\, \omega_i \wedge \omega_j \wedge \omega_k) \\ &= -6\,(\tau_p \wedge \omega_p) \wedge \sigma \end{aligned} \tag{4.16}$$

4.5.1. Torsion decomposition.

There are unique functions T_{ij} on F_σ so that

$$\tau_i = \mathsf{T}_{ij}\, \omega_j\,. \tag{4.17}$$

These functions can be used to express the intrinsic torsion forms in indicial form:

$$\begin{aligned} \pi^*(\tau_0) &= \tfrac{24}{7}\,\mathsf{T}_{ii}\,, \\ \pi^*(\tau_1) &= \varepsilon_{ijk}\mathsf{T}_{ij}\; \omega_k\,, \\ \pi^*(\tau_2) &= 4\,\mathsf{T}_{ij}\, \omega_i \wedge \omega_j - \varepsilon_{ijkl}\mathsf{T}_{ij}\, \omega_k \wedge \omega_l\,, \\ \pi^*(\tau_3) &= -\tfrac{3}{2}\varepsilon_{ikl}\,(\mathsf{T}_{ij} + \mathsf{T}_{ji})\, \omega_j \wedge \omega_k \wedge \omega_l + \tfrac{18}{7}\,\mathsf{T}_{ii}\,\sigma\,. \end{aligned} \tag{4.18}$$

(In these formulae, one sums over repeated indices in any term.)

4.5.2. Curvature identities

The covariant differentials can be expressed in indices as

$$D\tau = (D\tau_i) = (\tfrac{1}{2}\mathsf{T}_{ijk}\,\omega_j \wedge \omega_k) \tag{4.19}$$

$$D\theta = (D\theta_{ij}) = (\tfrac{1}{2}\mathsf{S}_{ijkl}\,\omega_j \wedge \omega_k) \tag{4.20}$$

where each of T and S are skew-symmetric in their last two indices, S is skew-symmetric in its first two indices and, since $D\theta$ takes values in $\mathfrak{g}_2$, the functions S also satisfy

$$\varepsilon_{ijm}\mathsf{S}_{ijkl} = 0$$

for all m, k, and l.

Since $\Psi = D\theta + 2\big[D\tau\big]$, the Riemann curvature functions are expressed as

$$\mathsf{R}_{ijkl} = \mathsf{S}_{ijkl} + 2\,\varepsilon_{ijp}\mathsf{T}_{pkl}\,, \tag{4.21}$$

so that the first Bianchi identity becomes

$$\mathsf{S}_{ijkl} + \mathsf{S}_{iljk} + \mathsf{S}_{iklj} + 2\,\varepsilon_{ijp}\mathsf{T}_{pkl} + 2\,\varepsilon_{ilp}\mathsf{T}_{pjk} + 2\,\varepsilon_{ikp}\mathsf{T}_{plj} = 0. \tag{4.22}$$

The identities (4.22) impose 28 linear conditions on the T_{ijk} alone. Perhaps the easiest way to derive these 28 conditions is to expand the identities

$$\mathrm{d}(\mathrm{d}(\sigma)) = \mathrm{d}(\mathrm{d}(\star\sigma)) = 0 \tag{4.23}$$

and use the structure equations (4.13) together with the definitions (4.9), (4.19), and (4.20). This will be left as an exercise for the reader. The result is that the conditions (4.23) are equivalent to the following equations (some of which are redundant):

$$\begin{aligned} 0 &= \mathsf{T}_{iij}\,, \\ 0 &= \varepsilon_{ipq}\mathsf{T}_{jpq} - \varepsilon_{jpq}\mathsf{T}_{ipq}\,, \\ 0 &= \varepsilon_{ipq}\mathsf{T}_{pqj} - \varepsilon_{jpq}\mathsf{T}_{pqi}\,. \end{aligned} \tag{4.24}$$

This implies that the function (T_{ijk}), which nominally takes values in a G_2-module of the form

$$\begin{aligned} \mathsf{V}_{1,0} \otimes \Lambda^2\big(\mathsf{V}_{1,0}\big) &= \mathsf{V}_{1,0} \otimes \big(\mathsf{V}_{1,0} \oplus \mathsf{V}_{0,1}\big) \\ &= \mathsf{V}_{0,0} \oplus 2\mathsf{V}_{1,0} \oplus \mathsf{V}_{0,1} \oplus 2\mathsf{V}_{2,0} \oplus \mathsf{V}_{1,1}\,, \end{aligned} \tag{4.25}$$

actually takes values in a submodule of the form

$$\mathsf{V}_{0,0} \oplus 2\mathsf{V}_{2,0} \oplus \mathsf{V}_{1,1}\,. \tag{4.26}$$

4.5.3. The Ricci identity

It was Bonan [1] who first observed that the Bianchi identities imply that a G_2-structure with vanishing torsion must necessarily have vanishing Ricci tensor. On general abstract grounds, it then follows that the Bianchi identities (4.22) must allow one to express the Ricci curvature in terms of the T_{ijk}. Indeed, by combining the first Bianchi identities via the ε-identities (another exercise for the reader), one derives the following expression for the Ricci curvature components $\mathsf{R}_{ij} = \mathsf{R}_{kikj}$:

$$\mathsf{R}_{ij} = 6\,\varepsilon_{pqi}\mathsf{T}_{pqj}\,. \tag{4.27}$$

This allows one to express the Ricci curvature directly in terms of the four torsion forms and their exterior derivatives. The resulting formula for the scalar curvature of the underlying metric g_σ is

$$\mathrm{Scal}(g_\sigma) = 12\,\delta\tau_1 + \tfrac{21}{8}\,{\tau_0}^2 + 30\,|\tau_1|^2 - \tfrac{1}{2}\,|\tau_2|^2 - \tfrac{1}{2}\,|\tau_3|^2. \tag{4.28}$$

The full Ricci tensor is somewhat more complicated, but can be expressed as follows:

First, define a G_2-invariant quadratic pairing $\mathsf{Q} : \Lambda^3(T^*) \times \Lambda^3(T^*) \to \Lambda^3(T^*)$ by the following recipe: Choose a local basis $e_1, \ldots, e_7$ of orthonormal vector fields such that $\sigma(e_i, e_j, e_k) = \varepsilon_{ijk}$ (such a basis is often called a G_2-frame field). Then, for $\alpha, \beta \in \Omega^3(M)$ set

$$\mathsf{Q}(\alpha, \beta) = *_\sigma\left[\varepsilon_{ijkl}\left((e_i \wedge e_j) \lrcorner *_\sigma \alpha\right) \wedge \left((e_k \wedge e_l) \lrcorner *_\sigma \beta\right)\right]. \tag{4.29}$$

The resulting mapping Q does not depend on the choice of local G_2-frame field. With this definition (and keeping in mind the definition (2.18) of j) one finds

$$\begin{aligned}\mathrm{Ric}(g_\sigma) = &-\left(\tfrac{3}{2}\,\delta\tau_1 - \tfrac{3}{8}\,{\tau_0}^2 + 15\,|\tau_1|^2 - \tfrac{1}{4}\,|\tau_2|^2 + \tfrac{1}{2}\,|\tau_3|^2\right) g_\sigma \\ &+ \mathsf{j}\Big(-\tfrac{5}{4}\,\mathrm{d}\big(*_\sigma(\tau_1 \wedge *_\sigma\sigma)\big) - \tfrac{1}{4}\,\mathrm{d}\tau_2 + \tfrac{1}{4}\,*_\sigma \mathrm{d}\tau_3 \\ &\qquad + \tfrac{5}{2}\,\tau_1 \wedge *_\sigma(\tau_1 \wedge *_\sigma\sigma) - \tfrac{1}{8}\,\tau_0\tau_3 + \tfrac{1}{4}\,\tau_1 \wedge \tau_2 \\ &\qquad + \tfrac{3}{4}\,*_\sigma(\tau_1 \wedge \tau_3) + \tfrac{1}{8}\,*_\sigma(\tau_2 \wedge \tau_2) + \tfrac{1}{64}\,\mathsf{Q}(\tau_3, \tau_3)\Big).\end{aligned} \tag{4.30}$$

While a formula in this generality is not of much practical use, when one goes to investigate special classes of G_2-structures, this formula can simplify considerably, as will be seen.

Formulae essentially equivalent to a special case of the formulae (4.28) and (4.30) were found in [7, 8], where those authors considered what they called 'integrable G_2-structures', which, in the terminology of this article, means G_2-structures σ satisfying $\tau_2 = 0$.

Remark 10 (General identities). It is perhaps worth remarking on why the identities (4.28) and (4.30) could be expected to have the form that they do.

In the first place, one knows that the scalar curvature must be expressible in a G_2-invariant manner as a sum of a linear expression in the second order invariants, i.e., a section of a vector bundle modeled on $V_2(\mathfrak{g}_2)$, and an expression in the first order invariants, i.e., the torsion forms, that is at most quadratic. A glance at (4.7) shows that there is only one trivial summand in the representation $V_2(\mathfrak{g}_2)$ and hence there is essentially only one possible second order term up to a universal constant multiple. Since $\delta\tau_1$ is a scalar second order invariant, it must represent this copy of $\mathsf{V}_{0,0}$ in $V_2(\mathfrak{g}_2)$. As for the first order terms, since $V_1(\mathfrak{g}_2)$ consists of four mutually inequivalent G_2-modules, the space of G_2-invariant quadratic forms on this space has dimension 4 and must be represented by the square norms of the four torsion forms. Thus, a formula of the form (4.28) was inevitable; it was just a matter of determining the numerical coefficients, which was done with the aid of MAPLE.

The argument for the form of (4.30) is quite similar. Since the scalar curvature has already been determined, it is a question of writing down a formula for the trace-free part of the Ricci tensor, i.e., finding linear terms in $V_2(\mathfrak{g}_2)$ and quadratic terms in $V_1(\mathfrak{g}_2)$ that take values in the G_2-module $\mathsf{V}_{2,0}$. Again, a glance at (4.7) shows that there are at most three possible second order terms and it is not difficult to see that the three second order terms that take values in $\mathsf{V}_{2,0}$ found by taking derivatives of τ_1, τ_2, and τ_3 and projecting into a suitable $\mathsf{V}_{2,0}$ representation are, in fact, independent and generate the three copies of $\mathsf{V}_{2,0}$ that appear in $V_2(\mathfrak{g}_2)$. On the other hand, using representation theory to compute the second symmetric power of $V_1(\mathfrak{g}_2)$ shows that there exist eight copies of $\mathsf{V}_{2,0}$ in this symmetric power. Of those eight copies, five are computable via wedge product and appear in the formula for Ricci. Of the remaining three, one bilinear in τ_2 and τ_3 and the other two quadratic in τ_3, only one of the terms quadratic in τ_3 actually makes an appearance. The rest is just a matter of determining constants.

4.6. Closed G_2-structures

Now, consider the case of a *closed* $\sigma \in \Omega^3_+(M)$, i.e., $\mathrm{d}\sigma = 0$. In this case, by Proposition 1, it follows that

$$\mathrm{d}{*_\sigma}\sigma = \tau_2 \wedge \sigma, \tag{4.31}$$

where τ_2 lies in $\Omega^2_{14}(M,\sigma)$. In particular,

$$\tau_2 \wedge {*_\sigma}\sigma = 0. \tag{4.32}$$

Taking the exterior derivative of (4.31) yields

$$0 = \mathrm{d}\tau_2 \wedge \sigma\,, \tag{4.33}$$

implying that $\mathrm{d}\tau_2$ has no component in $\Omega^3_7(M,\sigma)$. Differentiating (4.32) yields

$$\begin{aligned} 0 = \mathrm{d}\big(\tau_2 \wedge {*_\sigma}\sigma\big) &= \mathrm{d}\tau_2 \wedge {*_\sigma}\sigma + \tau_2 \wedge \mathrm{d}{*_\sigma}\sigma \\ &= \mathrm{d}\tau_2 \wedge {*_\sigma}\sigma + \tau_2 \wedge \tau_2 \wedge \sigma = \mathrm{d}\tau_2 \wedge {*_\sigma}\sigma - |\tau_2|^2 {*_\sigma}1\,. \end{aligned} \tag{4.34}$$

Thus, from (4.33) and (4.34) it follows that there exists a $\gamma \in \Omega^3_{27}(M,\sigma)$ so that

$$\mathrm{d}\tau_2 = \tfrac{1}{7}|\tau_2|^2\,\sigma + \gamma. \tag{4.35}$$

In summary, formulae (4.28) and (4.30) can be simplified, in this case, to

$$\mathrm{Scal}(g_\sigma) = -\tfrac{1}{2}\,|\tau_2|^2. \tag{4.36}$$

and

$$\mathrm{Ric}(g_\sigma) = \tfrac{1}{4}\,|\tau_2|^2\,g_\sigma - \tfrac{1}{4}\,\mathsf{j}\left(\mathrm{d}\tau_2 - \tfrac{1}{2}\,{*_\sigma}(\tau_2 \wedge \tau_2)\right). \tag{4.37}$$

Remark 11 (Differential invariants of closed G_2-structures). Just as one can compute the dimension of the space of k-jets of G-structures as in §4.2, one can compute the dimension of the space of k-jets of G-structures satisfying some set of differential equations. In the case of closed G_2-structures, denote the module of k-th order differential invariants by $V_k'(\mathfrak{g}_2) \subset V_k(\mathfrak{g}_2)$. One finds, for example, that

$$V_1'(\mathfrak{g}_2) \simeq \mathsf{V}_{0,1} \qquad V_2'(\mathfrak{g}_2) \simeq \mathsf{V}_{2,0} \oplus \mathsf{V}_{1,1} \oplus \mathsf{V}_{0,2}. \tag{4.38}$$

This implies, on abstract grounds, that the scalar curvature of the underlying metric of a closed G_2-structure must be expressed in terms of the first order invariants (since there is no $\mathsf{V}_{0,0}$ component in $V_2'(\mathfrak{g}_2)$) and that the full Ricci tensor can be expressed in terms of τ_2 and $d\tau_2$. Thus, the form of (4.36) and (4.37) could have been anticipated, if not the numerical coefficients.

Of course, it is easy to 'write down' the general closed G_2-structure locally: If $\beta \in \Omega^2(\mathbb{R}^7)$ is a (smooth) 2-form that vanishes to second order at $0 \in \mathbb{R}^7$, then the 3-form $\sigma = \phi + d\beta$ will equal ϕ at 0 and hence will be a closed, definite 3-form on some open neighborhood of $0 \in \mathbb{R}^7$. Conversely, if σ is a closed G_2-structure on a manifold M^7, then any point $p \in M$ has an open neighborhood U on which there exists a p-centered coordinate chart $x : U \to \mathbb{R}^7$ such that $\sigma_U = x^*(\phi + d\beta)$ where $\beta \in \Omega^2(\mathbb{R}^7)$ is a 2-form that vanishes to second order at $0 \in \mathbb{R}^7$.

In a sense that it is possible to make precise using Cartan's notion of the generality of the space of solutions of a system of PDE, one can develop this discussion further to show that the general closed G_2-structure modulo diffeomorphism depends on 8 functions of seven variables.

An immediate consequence of (4.36) is the following:

Corollary 1. *For any closed* G_2*-structure* $\sigma \in \Omega^3_+(M)$*, the scalar curvature of the underlying metric is non-positive and vanishes identically if and only if the entire Ricci tensor of the underlying metric vanishes. Equivalently, the scalar curvature vanishes identically if and only if* σ *satisfies* $d\sigma = d{*_\sigma}\sigma = 0$.

Using the formulae for i and j, the formula (4.37) can be rewritten as

$$d\tau_2 = \tfrac{3}{14}\,|\tau_2|^2\,\sigma + \tfrac{1}{2}{*_\sigma}(\tau_2 \wedge \tau_2) - \tfrac{1}{2}\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big), \tag{4.39}$$

where $\mathrm{Ric}^0(g_\sigma)$ is the traceless Ricci tensor of g_σ.

Corollary 2. *A closed* G_2*-structure* $\sigma \in \Omega^3_+(M)$ *has an Einstein underlying metric if and only if it satisfies* $d{*_\sigma}\sigma = \tau_2 \wedge \sigma$ *where* $d\tau_2 = \frac{3}{14}\,|\tau_2|^2\,\sigma + \frac{1}{2}{*_\sigma}(\tau_2 \wedge \tau_2)$.

Remark 12 (Nonexistence of compact Einstein examples). I do not know whether there exist any closed G_2-structures that are Einstein but not Ricci-flat, even local (i.e., incomplete) ones.

After Version 1.0 of the present article was posted to the arXiv, Cleyton and Ivanov [3] gave an argument (based on a comparison of the Ricci curvatures of the Levi-Civita connection and the canonical connection of the underlying G_2-structure) showing that no *compact* 7-manifold can support a closed G_2-structure σ whose underlying metric g_σ is Einstein unless σ is also coclosed, i.e., $d{*_\sigma}\sigma = 0$. Their argument is rather involved, but Corollary 2 yields a simple proof:

Suppose that $\sigma \in \Omega^3_+(M)$ is a closed G_2-structure whose underlying metric g_σ is Einstein. Then, by Corollary 2, it follows that $d{*_\sigma}\sigma = \tau_2{\wedge}\sigma$ where $d\tau_2 = \frac{3}{14}\,|\tau_2|^2\,\sigma + \frac{1}{2}{*_\sigma}(\tau_2{\wedge}\tau_2)$. Now, using this formula together with the formula (2.21), one finds

$$\begin{aligned} d\big(\tfrac{1}{3}{\tau_2}^3\big) &= {\tau_2}^2\wedge d\tau_2 = {\tau_2}^2\wedge\big(\tfrac{3}{14}\,|\tau_2|^2\,\sigma + \tfrac{1}{2}{*_\sigma}(\tau_2\wedge\tau_2)\big) \\ &= -\tfrac{3}{14}\,|\tau_2|^4{*_\sigma}1 + \tfrac{1}{2}|\tau_2\wedge\tau_2|^2\,{*_\sigma}1 = \tfrac{2}{7}|\tau_2|^4{*_\sigma}1. \end{aligned} \tag{4.40}$$

Now, suppose that M were compact. Integrating both ends of (4.40) over M and applying Stokes' theorem yields

$$0 = \int_M d\big(\tfrac{1}{3}{\tau_2}^3\big) = \int_M \tfrac{2}{7}|\tau_2|^4{*_\sigma}1\,, \tag{4.41}$$

implying that τ_2 must vanish identically, as was to be shown.

In view of (4.39), this nonexistence can be seen as a special case of a general result about pinching of Ricci curvature:

Corollary 3. *Suppose that $\sigma \in \Omega^3_+(M)$ is a closed G_2-structure on a compact 7-manifold M that satisfies the pinching condition*

$$\big|\mathrm{Ric}^0(g_\sigma)\big|^2 \le \tfrac{4}{21}\,C\,\mathrm{Scal}(g_\sigma)^2. \tag{4.42}$$

for some constant $C \le 1$. If $C < 1$, then σ is also coclosed. If $C = 1$, then equality must hold in (4.42) *everywhere on M. Moreover, in this case, the identity*

$$\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big) = \tfrac{2}{3}\big({*_\sigma}(\tau_2\wedge\tau_2) + \tfrac{1}{7}|\tau_2|^2\,\sigma\big) \tag{4.43}$$

or, equivalently,

$$d\tau_2 = \tfrac{1}{6}\,\big(|\tau_2|^2\,\sigma + {*_\sigma}(\tau_2\wedge\tau_2)\big) \tag{4.44}$$

must hold everywhere on M.

Proof. Using (4.39), one obtains, after using (2.21), the orthogonality of $\Omega^3_1(M,\sigma)$ and $\Omega^3_{27}(M,\sigma)$, the identity (2.30), and the Cauchy-Schwartz inequality,

$$\begin{aligned} d\big(\tfrac{1}{3}{\tau_2}^3\big) &= {\tau_2}^2\wedge d\tau_2 = {\tau_2}^2\wedge\big(\tfrac{3}{14}\,|\tau_2|^2\,\sigma + \tfrac{1}{2}{*_\sigma}(\tau_2\wedge\tau_2) - \tfrac{1}{2}\,\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big)\big) \\ &= \tfrac{2}{7}|\tau_2|^4{*_\sigma}1 - \tfrac{1}{2}\,{\tau_2}^2\wedge\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big) \\ &= \tfrac{2}{7}|\tau_2|^4{*_\sigma}1 - \tfrac{1}{2}\,\big({\tau_2}^2 + \tfrac{1}{7}\,|\tau_2|^2{*_\sigma}\sigma\big)\wedge\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big) \\ &\ge \left(\tfrac{2}{7}|\tau_2|^4 - \tfrac{1}{2}\sqrt{\tfrac{6}{7}}\,|\tau_2|^2\,\big|\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big|\right)\,{*_\sigma}1. \end{aligned} \tag{4.45}$$

Now, the expression at the end of (4.45) will be a nonnegative multiple of the volume form ${*_\sigma}1$ as long as

$$\big|\mathrm{Ric}^0(g_\sigma)\big| = \sqrt{\tfrac{1}{8}}\,\big|\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big| \le \sqrt{\tfrac{1}{21}}|\tau_2|^2 = -\sqrt{\tfrac{4}{21}}\,\mathrm{Scal}(g_\sigma). \tag{4.46}$$

Since $-\operatorname{Scal}(g_\sigma) \geq 0$, the inequality (4.42) with $C < 1$ will evidently imply that the expression at the end of (4.45) is a positive multiple of $|\tau_2|^2 *_\sigma 1$. By Stokes' theorem, this will imply that τ_2 vanishes identically, as desired.

Suppose now that (4.42) holds with $C = 1$. Then the expression at the end of (4.45) is still a nonnegative multiple of $|\tau_2|^2 *_\sigma 1$ and hence, by Stokes' theorem, must vanish identically. However, by the strong form of the Cauchy-Schwartz inequality, this can only happen if the relation

$$\mathsf{i}\big(\operatorname{Ric}^0(g_\sigma)\big) = \tfrac{2}{3}\big(*_\sigma(\tau_2 \wedge \tau_2) + \tfrac{1}{7}|\tau_2|^2\,\sigma\big) \tag{4.47}$$

holds identically on the open set where $|\tau_2| > 0$. Now, if the locus $|\tau_2| = 0$ has any interior, then $\operatorname{Ric}(g_\sigma)$ vanishes on this interior since σ is both closed and coclosed there. Thus, (4.47) holds on both the open set where $|\tau_2| > 0$ and the interior of the locus where $|\tau_2| = 0$. Consequently, it must hold on all of M, as desired. □

Remark 13 (Extremally Ricci-pinched closed G_2-structures). Note that another way of phrasing Corollary 3 is to use (4.45) to show that the inequality

$$\int_M \big|\operatorname{Ric}^0(g_\sigma)\big|^2 *_\sigma 1 \geq \frac{4}{21}\int_M \operatorname{Scal}(g_\sigma)^2 *_\sigma 1 \tag{4.48}$$

holds for any closed G_2-structure σ on a compact manifold M^7 and that equality holds in (4.48) if and only if σ satisfies

$$\mathrm{d}\sigma = 0, \qquad \mathrm{d}{*_\sigma}\sigma = \tau \wedge \sigma, \qquad \mathrm{d}\tau = \tfrac{1}{6}\big(|\tau|^2\,\sigma + *_\sigma(\tau\wedge\tau)\big). \tag{4.49}$$

Indeed, Corollary 3 suggests that the G_2-structures σ that satisfy (4.49) might be of particular interest, since these are, in some sense, the most 'extremally Ricci-pinched' that a closed G_2-structure can be on a compact 7-manifold.

One can see that there are some rather subtle restrictions on such structures on compact manifolds by developing these equations a bit further: Note that (4.49) implies

$$\begin{aligned} \mathrm{d}\big(\tau^3\big) &= 3\,\tau^2 \wedge \mathrm{d}\tau = \tau^2 \wedge \big(\tfrac{1}{2}\,|\tau|^2\,\sigma + \tfrac{1}{2}*_\sigma(\tau\wedge\tau)\big) \\ &= -\tfrac{1}{2}\,|\tau|^4 *_\sigma 1 + \tfrac{1}{2}|\tau\wedge\tau|^2 *_\sigma 1 = 0. \end{aligned} \tag{4.50}$$

On the other hand, computation using the structure equations and (4.49) yields

$$0 = \mathrm{d}\big(\mathrm{d}\tau\big) = \mathrm{d}\left(\tfrac{1}{6}\left(|\tau|^2\,\sigma + *_\sigma(\tau\wedge\tau)\right)\right) = \alpha\wedge\sigma + *_\sigma\gamma \tag{4.51}$$

where γ lies in $\Omega^3_{27}(M,\sigma)$ and

$$\alpha = \tfrac{1}{8}\left(\mathrm{d}\big(|\tau|^2\big) - \tfrac{2}{9}*_\sigma(\tau^3)\right). \tag{4.52}$$

Consequently, any solution of (4.49) must satisfy[2]

$$\mathrm{d}\big(|\tau|^2\big) = \tfrac{2}{9}*_\sigma(\tau^3). \tag{4.53}$$

[2]The vanishing of γ as defined in (4.51) imposes 27 more equations on the covariant derivative of τ, but these are not as easily stated as (4.53).

Combining this with (4.50) yields

$$\Delta_\sigma\bigl(|\tau|^2\bigr) = 0. \tag{4.54}$$

Assume now that M is compact and connected. It then follows from (4.54) that $|\tau|^2$ must be a constant.

Of course, if $|\tau|^2 = 0$, then $\tau = 0$ and σ is coclosed and hence g_σ-parallel. Thus, assume from now on that $|\tau|^2 > 0$.

Then (4.53) implies that $\tau^3 = 0$. However, $|\tau{\wedge}\tau|^2 = |\tau|^4 \neq 0$, implying that τ has constant rank 4 (instead of the *a priori* maximum of 6) and hence that $\tau{\wedge}\tau$ is a nonzero simple 4-form of constant norm. Using (2.23) and the fact that $\tau^3 = 0$ then yields

$$\begin{aligned} \mathrm{d}\bigl(\tau^2\bigr) &= 2\,\tau\wedge\mathrm{d}\tau = \tfrac{1}{3}\tau\wedge\Bigl(\,|\tau|^2\,\sigma + {*_\sigma}(\tau\wedge\tau)\Bigr)\\ &= -\tfrac{1}{3}\,|\tau|^2{*_\sigma}\tau + \tfrac{1}{3}\,|\tau|^2{*_\sigma}\tau = 0, \end{aligned} \tag{4.55}$$

So that the simple 4-form $\tau{\wedge}\tau$ is closed.

Since $\tau{\wedge}\tau$ is simple with constant norm, the 3-form ${*_\sigma}(\tau{\wedge}\tau)$ is also nonzero and simple, with constant norm. Moreover, in view of the constancy of $|\tau|^2$, expanding $\mathrm{d}\bigl(\mathrm{d}\tau\bigr) = 0$ and using (4.49) shows that ${*_\sigma}(\tau{\wedge}\tau)$ is also closed.

Consequently, the tangent bundle of M splits as an orthogonal direct sum of two integrable subbundles

$$TM = P \oplus Q \tag{4.56}$$

with $P = \{\, v\in TM \mid v \lrcorner (\tau{\wedge}\tau) = 0\,\}$ of rank 3 and $Q = \{\, v\in TM \mid v \lrcorner {*_\sigma}(\tau{\wedge}\tau) = 0\,\}$ of rank 4. The P-leaves are calibrated by $-|\tau|^{-2}\,{*_\sigma}(\tau{\wedge}\tau)$ while the Q-leaves are calibrated by $-|\tau|^{-2}\,(\tau{\wedge}\tau)$. (The reason for the minus signs is that they correctly orient the P-leaves as associative submanifolds and the Q-leaves as coassociative submanifolds.)

The Ricci curvature in this case simplifies to

$$\mathrm{Ric}(g_\sigma) = \tfrac{1}{12}\,\mathsf{j}\bigl({*_\sigma}(\tau\wedge\tau)\bigr) = -\tfrac{1}{6}\,|\tau|^2\,\bigl(g_\sigma\bigr)\big|_P\,, \tag{4.57}$$

so that, in particular, the Ricci curvature is nonpositive, with one eigenvalue $-\frac{1}{6}\,|\tau|^2$ of multiplicity 3 and the other eigenvalue 0 of multiplicity 4.

Example 1 (A homogeneous example). Just how general the G_2-structures σ satisfying (4.49) with $\tau\neq 0$ are, even locally, is an interesting question. I will now show that these equations do have a nontrivial solution, by producing a (homogeneous) example.

Let G be the group of volume-preserving affine transformations of $\mathbb{C}^2$. Thus G can be regarded as the matrix group consisting of the 3-by-3 matrices with complex entries of the form

$$g = \begin{pmatrix} a & b & x\\ c & d & y\\ 0 & 0 & 1\end{pmatrix} \tag{4.58}$$

where $ad - bc = 1$. Write the canonical left-invariant form on G as

$$\alpha = g^{-1}\,\mathrm{d}g = \begin{pmatrix} -\omega^1 + \mathrm{i}\,\eta^1 & -\omega^3 - \eta^3 + \mathrm{i}\,(\eta^2 - \omega^2) & \omega^4 + \mathrm{i}\,\omega^5\\ -\omega^3 + \eta^3 + \mathrm{i}\,(\eta^2 + \omega^2) & \omega^1 - \mathrm{i}\,\eta^1 & \omega^6 - \mathrm{i}\,\omega^7\\ 0 & 0 & 0\end{pmatrix}. \tag{4.59}$$

Then $\mathrm{d}\alpha = -\alpha{\wedge}\alpha$ implies that the left-invariant 3-form $\tilde\sigma$ defined by

$$\tilde\sigma = \omega^{123} + \omega^{145} + \omega^{167} + \omega^{246} - \omega^{257} - \omega^{347} - \omega^{356} \tag{4.60}$$

(where ω^{ijk} stands for the wedge product $\omega^i{\wedge}\omega^j{\wedge}\omega^k$, etc.) satisfies $\mathrm{d}\tilde\sigma = 0$. Consequently, $\tilde\sigma$ is the pullback to G of a definite 3-form σ on the left coset space $M^7 = G/\,\mathrm{SU}(2)$. (Here, $\mathrm{SU}(2) \subset G$ is the subgroup whose left cosets are the integral leaves of the differentially closed system $\omega^i = 0$ on G.) Moreover, letting $\pi : G \to M$ denote the coset projection, one sees that

$$\pi^*\bigl({*_\sigma}\sigma\bigr) = \omega^{4567} + \omega^{2367} + \omega^{2345} + \omega^{1357} - \omega^{1346} - \omega^{1256} - \omega^{1247} \tag{4.61}$$

while

$$\pi^*\bigl(g_\sigma\bigr) = (\omega^1)^2 + \cdots + (\omega^7)^2. \tag{4.62}$$

Finally, one finds that there exists a 2-form τ on M so that

$$\pi^*(\tau) = 6\,\omega^{45} - 6\,\omega^{67}. \tag{4.63}$$

The equation $\mathrm{d}\alpha = -\alpha{\wedge}\alpha$ then implies that the pair (σ,τ) satisfy (4.49).

Note that M is diffeomorphic to $\mathbb{R}^7$ and that the P-leaves and Q-leaves are, respectively, the fibers of maps $M \to \mathbb{C}^2 = G/\,\mathrm{SL}(2,\mathbb{C})$ and $M \to \mathrm{SL}(2,\mathbb{C})/\,\mathrm{SU}(2)$. Although M is not compact, it has compact quotients on which σ is well-defined. To see this, let $\mathsf{\Gamma} \subset \mathrm{SL}(2,\mathbb{C})$ be a cocompact, discrete, torsion-free subgroup that preserves a lattice $\mathsf{L} \subset \mathbb{C}^2$. (Such $\mathsf{\Gamma}$ do exist. For example, let $q \in \mathbb{Z}[\mathrm{i}]$ be a prime in the Gaussian integers satisfying $|q|^2 = q\bar q > 4$ and let $\mathsf{\Gamma}_q \subset \mathrm{SL}\bigl(2,\mathbb{Z}[\mathrm{i}]\bigr)$ be the finite index subgroup consisting of the elements $\gamma \in \mathrm{SL}\bigl(2,\mathbb{Z}[\mathrm{i}]\bigr)$ that satisfy $\gamma \equiv \mathrm{I}_2 \mod (q)$. Then $\mathsf{\Gamma}_q$ has the required properties and preserves the lattice $\mathbb{Z}[\mathrm{i}]^2 \subset \mathbb{C}^2$.) Now consider the discrete subgroup $\Gamma \subset G$ consisting of elements of the form

$$g = \begin{pmatrix} \gamma & \ell \\ 0 & 1 \end{pmatrix} \tag{4.64}$$

where γ lies in $\mathsf{\Gamma}$ and ℓ lies in L. Then Γ acts on $M = G/\,\mathrm{SU}(2)$ on the left preserving σ and it is not difficult to see that this action is both free and properly discontinuous. The quotient $\bar M = \Gamma\backslash M$ is compact and supports a closed extremally Ricci-pinched G_2-structure $\bar\sigma$ that pulls back to M to equal σ.

Remark 14 (Natural equations for closed G_2-structures). Let λ be a constant and consider the system of equations

$$\mathrm{d}\sigma = 0, \qquad \mathrm{d}{*_\sigma}\sigma = \tau\wedge\sigma, \qquad \mathrm{d}\tau = \tfrac{1}{7}\,|\tau|^2\,\sigma + \lambda\left(\tfrac{1}{7}\,|\tau|^2\,\sigma + {*_\sigma}(\tau\wedge\tau)\right). \tag{4.65}$$

for a G_2-structure σ on a manifold M^7. This family includes both the Einstein condition ($\lambda = \tfrac{1}{2}$) and the 'extremally pinched Ricci' condition ($\lambda = \tfrac{1}{6}$). Indeed, in view of (4.35) and (2.28) and since $\mathsf{S}^2(\mathsf{V}_{0,1}) \simeq \mathsf{V}_{0,0}{\oplus}\mathsf{V}_{2,0}{\oplus}\mathsf{V}_{0,2}$ while $\Lambda^3(\mathsf{V}_{1,0}) \simeq \mathsf{V}_{0,0}{\oplus}\mathsf{V}_{0,1}{\oplus}\mathsf{V}_{2,0}$, the 1-parameter family of natural equations (4.65) for closed G_2-structures describes the most general way in which $\mathrm{d}\tau$ can be prescribed naturally and quadratically in terms of τ. In view of the fact that $\mathrm{d}\tau$ can have no component in $\Omega^3_7(M,\sigma)$ and that the component of $\mathrm{d}\tau$ in $\Omega^3_1(M,\sigma)$ is determined by (4.35), it follows that (4.65) is a system of 27 ($= \dim \mathsf{V}_{2,0}$) equations for a closed G_2-structure σ. In view of the discussion in Remark 11, one should regard (4.65) as an overdetermined system of PDE. This system is not involutive for any value of λ, as the following discussion will show.

First, the computation (4.50) can be redone for G_2-structures satisfying (4.65), yielding

$$\mathrm{d}\big(\tau^3\big) = \frac{3(6\lambda-1)}{7}\,|\tau|^4 *_\sigma 1\,. \tag{4.66}$$

In particular, on a compact 7-manifold, the only value of λ that is possible for such a structure with τ not identically zero is $\lambda = \frac{1}{6}$.

Redoing the computation (4.51) using the structure equations and (4.65) instead of (4.49) yields

$$0 = \mathrm{d}\big(\mathrm{d}\tau\big) = \alpha \wedge \sigma + *_\sigma \gamma \tag{4.67}$$

where γ lies in $\Omega^3_{27}(M,\sigma)$ and

$$\alpha = \frac{\lambda(2\lambda-1)}{4} *_\sigma (\tau^3) - \frac{(3\lambda-4)}{28}\,\mathrm{d}\big(|\tau|^2\big). \tag{4.68}$$

Consequently, any solution of (4.65) satisfies

$$(3\lambda-4)\,\mathrm{d}\big(|\tau|^2\big) = 7\lambda(2\lambda-1) *_\sigma (\tau^3). \tag{4.69}$$

When $\lambda = \frac{4}{3}$, this condition implies $\tau^3 = 0$, which, by (4.66), then implies $|\tau| = 0$, i.e., $\tau = 0$. Thus, there are no G_2-structures σ satisfying (4.65) with $\lambda = \frac{4}{3}$ except those that are closed and coclosed.

When $\lambda \neq \frac{4}{3}$, the system (4.69) represents 7 'new' second order equations on σ that are not algebraic consequences of (4.65). The existence of these 'new' equations implies that the system (4.65) is not involutive.

Even beyond this, when $\lambda \neq 0$, the vanishing of the term γ in (4.67) yields 27 more equations of second order on σ that are not algebraic consequences of (4.65) and (4.69) combined. Whether further differentiation of these combined equations would yield more second (or even first) order equations remains to be seen. It is this phenomenon that makes the analysis of systems of type (4.65) troublesome.

5. The Torsion-free Case

A G_2-structure $\sigma \in \Omega^3_+(M)$ is said to be *torsion-free* if all of its four torsion forms vanish. There is an aspect of the geometry of torsion-free G_2-structures that is analogous to the Kähler identities in complex Riemannian geometry and that is the concern of this section.

The material in this section was the result of a joint project with F. Reese Harvey and was carried out between 1991 and 1994.

5.1. Reference modules

It will be convenient to chose a 'reference' representation for each of the irreducible G_2-modules that appear in the exterior algebra on $\mathsf{V}_{1,0}$.

Given any G_2-structure $\sigma \in \Omega^3_+(M)$, these will be chosen to correspond to the spaces of differential forms $\Omega^0(M)$, $\Omega^1(M)$, $\Omega^2_{14}(M,\sigma)$, and $\Omega^3_{27}(M,\sigma)$. For simplicity, these spaces will be referred to as Ω_1, Ω_7, Ω_{14}, and Ω_{27} when M and σ are clear from context.

5.2. Exterior derivative identities

When a G_2-structure σ has vanishing intrinsic torsion, the fundamental forms σ and $*_\sigma\sigma$ are parallel with respect to the natural connection (which is torsion-free) and so are all of the various natural isomorphisms between the different constituents of the bundle of exterior differential forms. Consequently, the various differential operators that one can define by decomposing the exterior derivative into its constituent components are really manifestations of first order differential operators between the abstract bundles. Thus, there will be identities (analogous to the identities one proves in Kähler geometry) between these different manifestations. In this subsection, these will be made explicit. Essentially, the proof of the following proposition is a matter of checking constants.

Proposition 3 (Exterior derivative identities)**.** *Suppose that σ is a torsion-free G_2-structure on M. Then, for all $p, q \in \{1, 7, 14, 27\}$, there exists a first order differential operator $d^p_q \colon \Omega_p \to \Omega_q$, so that the exterior derivative formulas given in Table 1 hold for all $f \in \Omega_1$, $\alpha \in \Omega_7$, $\beta \in \Omega_{14}$, and $\gamma \in \Omega_{27}$. These operators are non-zero except for d^1_{27}, d^{27}_1, d^1_{14}, d^{14}_1, d^1_1, and d^{14}_{14}. With respect to the natural metrics on the underlying bundles, $(d^p_q)^* = d^q_p$. The identity $d^2 = 0$ is equivalent to the second order identities on the operators d^p_q listed in Table 2. Finally, the formulas for the Hodge Laplacians in terms of the operators d^p_q are as given in Table 3.*

Proof. The operators d^p_q are defined by decomposing the exterior derivative operator into types (much as ∂ and $\bar\partial$ are defined in Kähler geometry by the projection of the exterior derivative into types). For example, take the formula $d^7_7\alpha = *_\sigma(d(\alpha\wedge *_\sigma\sigma))$ as the *definition* of $d^7_7 \colon \Omega_7 \to \Omega_7$ and define $d^7_{27}\alpha$ to be the $\Omega^3_{27}(M,\sigma)$-component of $d\big(*_\sigma(\alpha\wedge *_\sigma\sigma)\big)$. Verifying the exterior derivative formulas is a routine matter that is best left to the reader. Once these have been established, the second order identities and the Laplacian formulas follow by routine computation. □

Remark 15 (Torsion perturbations). In the general case of a G_2-structure with torsion, all of the formulae in the tables listed above must be modified by lower order terms. For example, in Table 1 the second line would be modified to

$$d(f\,\sigma) = d^1_7 f\wedge\sigma + f\tau_0 *_\sigma\sigma + 3f\tau_1\wedge\sigma + f *_\sigma\tau_3. \tag{5.1}$$

The zero right hand sides in Table 2 have to be replaced by first order operators whose coefficients depend on the torsion terms and, in Table 3, one must take into account which particular part of the exterior algebra a given form occupies before writing down the appropriate formula for the Laplacian. For example, it is not true, in general, that $\Delta(f\,\sigma) = \Delta f\,\sigma$.

$\mathrm{d}\, f$	$=$		$\mathrm{d}^1_7 f$		
$\mathrm{d}\,(f\,\sigma)$	$=$		$\mathrm{d}^1_7 f \wedge \sigma$		
$\mathrm{d}\,(f *_\sigma \sigma)$	$=$		$\mathrm{d}^1_7 f \wedge *_\sigma \sigma$		
$\mathrm{d}\,\alpha$	$=$		$\frac{1}{3} *_\sigma (\mathrm{d}^7_7\alpha \wedge *_\sigma \sigma)$	$+\mathrm{d}^7_{14}\alpha$	
$\mathrm{d} *_\sigma (\alpha \wedge *_\sigma \sigma)$	$=$	$-\frac{3}{7}\,\mathrm{d}^7_1\alpha\,\sigma$	$-\frac{1}{2} *_\sigma (\mathrm{d}^7_7\alpha \wedge \sigma)$		$+\mathrm{d}^7_{27}\alpha$
$\mathrm{d} *_\sigma (\alpha \wedge \sigma)$	$=$	$\frac{4}{7}\,\mathrm{d}^7_1\alpha *_\sigma \sigma$	$+\frac{1}{2}\,\mathrm{d}^7_7\alpha \wedge \sigma$		$+ *_\sigma \mathrm{d}^7_{27}\alpha$
$\mathrm{d}\,(\alpha \wedge \sigma)$	$=$		$\frac{2}{3}\,\mathrm{d}^7_7\alpha \wedge *_\sigma \sigma$	$- *_\sigma \mathrm{d}^7_{14}\alpha$	
$\mathrm{d}\,(\alpha \wedge *_\sigma \sigma)$	$=$		$- *_\sigma \mathrm{d}^7_7\alpha$		
$\mathrm{d}\,(*_\sigma \alpha)$	$=$	$-\mathrm{d}^7_1\alpha *_\sigma 1$			
$\mathrm{d}\,\beta$	$=$		$\frac{1}{4} *_\sigma (\mathrm{d}^{14}_7\beta \wedge \sigma)$		$+\mathrm{d}^{14}_{27}\beta$
$\mathrm{d}\,(*_\sigma \beta)$	$=$		$*_\sigma \mathrm{d}^{14}_7\beta$		
$\mathrm{d}\,\gamma$	$=$		$\frac{1}{4}\,\mathrm{d}^{27}_7\gamma \wedge \sigma$		$+ *_\sigma \mathrm{d}^{27}_{27}\gamma$
$\mathrm{d}\,(*_\sigma \gamma)$	$=$		$-\frac{1}{3}\,\mathrm{d}^{27}_7\gamma \wedge *_\sigma \sigma$	$- *_\sigma \mathrm{d}^{27}_{14}\gamma$	

Table 1. Exterior derivative formulae

	$\mathrm{d}^7_7\, d^1_7 = 0$	$\mathrm{d}^7_{14}\,\mathrm{d}^1_7 = 0$	
$\mathrm{d}^7_1\,\mathrm{d}^7_7 = 0$	$\mathrm{d}^{14}_7\,\mathrm{d}^7_{14} = \frac{2}{3}(\mathrm{d}^7_7)^2$ $\mathrm{d}^{27}_7\,\mathrm{d}^7_{27} = (\mathrm{d}^7_7)^2 + \frac{12}{7}\mathrm{d}^1_7\,\mathrm{d}^7_1$	$\mathrm{d}^7_{14}\,\mathrm{d}^7_7 + 2\,\mathrm{d}^{27}_{14}\,\mathrm{d}^7_{27} = 0$	$3\,\mathrm{d}^{14}_{27}\,\mathrm{d}^7_{14} + \mathrm{d}^7_{27}\,\mathrm{d}^7_7 = 0$ $2\,\mathrm{d}^{27}_{27}\,\mathrm{d}^7_{27} - \mathrm{d}^7_{27}\,\mathrm{d}^7_7 = 0$
$\mathrm{d}^7_1\,\mathrm{d}^{14}_7 = 0$	$\mathrm{d}^7_7\,\mathrm{d}^{14}_7 + 2\,\mathrm{d}^{27}_7\,\mathrm{d}^{14}_{27} = 0$		$\mathrm{d}^7_{27}\,\mathrm{d}^{14}_7 + 4\,\mathrm{d}^{27}_{27}\,\mathrm{d}^{14}_{27} = 0$
	$3\,\mathrm{d}^{14}_7\,\mathrm{d}^{27}_{14} + \mathrm{d}^7_7\,\mathrm{d}^{27}_7 = 0$ $2\,\mathrm{d}^{27}_7\,\mathrm{d}^{27}_{27} - \mathrm{d}^7_7\,\mathrm{d}^{27}_7 = 0$	$\mathrm{d}^7_{14}\,\mathrm{d}^{27}_7 + 4\,\mathrm{d}^{27}_{14}\,\mathrm{d}^{27}_{27} = 0$	

Table 2. Second order identities

6. Deformation and Evolution of G_2-structures

The material in this section was the result of a joint project with Steve Altschuler and was carried out between 1992 and 1994. Our goal was to understand the long time behavior of the Laplacian heat flow defined below for closed G_2-structures on compact 7-manifolds, specifically, to understand conditions under which one could prove that this

$$
\begin{aligned}
\Delta f &= \mathrm{d}^7_1\,\mathrm{d}^1_7\, f\\
\Delta \alpha &= \big((\mathrm{d}^7_7)^2 + \mathrm{d}^1_7\,\mathrm{d}^7_1\big)\ \alpha\\
\Delta \beta &= \big(\tfrac{5}{4}\mathrm{d}^7_{14}\,\mathrm{d}^{14}_7 + \mathrm{d}^{27}_{14}\,\mathrm{d}^{14}_{27}\big)\ \beta\\
\Delta \gamma &= \big(\tfrac{7}{12}\mathrm{d}^7_{27}\,\mathrm{d}^{27}_7 + \mathrm{d}^{14}_{27}\,\mathrm{d}^{27}_{14} + (\mathrm{d}^{27}_{27})^2\big)\ \gamma
\end{aligned}
$$

TABLE 3. Laplacians

flow converged to a G_2-structure that is both closed and coclosed. Nowadays, this flow is called the *Hitchin flow* after Hitchin's fundamental paper [12].

We were never able to prove long-time existence under any reasonable hypotheses, so we wound up not publishing anything on the subject, although we did get some interesting results and formulae that I have not seen so far in the literature.[3]

6.1. The deformation forms

It turns out to be quite easy to describe deformations of G_2-structures. The following result is well-known and can be found most explicitly in Joyce's treatment [14, §10.3], though the notation is somewhat different. It is included here to establish notation for the discussion to follow.

Proposition 4 (Deformation forms). *Let $\sigma_t \in \Omega^3_+(M)$ be a smooth 1-parameter family of G_2-structures on M. Let g_t and $*_t$ denote the underlying metric and Hodge star operator associated to σ_t. Then there exist three differential forms $f^0_t \in \Omega^0(M)$, $f^1_t \in \Omega^1(M)$, and $f^3_t \in \Omega^3_{27}(M,\sigma_t) \subset \Omega^3(M)$ that depend differentiably on t and that are uniquely characterized by the equation (in which the t-dependence has been suppressed for notational clarity)*

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\sigma\big) = 3f^0\,\sigma + *_\sigma\big(f^1\wedge\sigma\big) + f^3\,. \tag{6.1}$$

Moreover, the associated metric and dual 4-forms satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(g\big) = 2f^0\,g + \tfrac{1}{2}\,\mathsf{j}\big(f^3\big) \tag{6.2}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(*_\sigma\sigma\big) = 4f^0\,*_\sigma\sigma + f^1\wedge\sigma - *_\sigma f^3\,. \tag{6.3}$$

Definition 5 (The deformation forms). The forms f^0_t, f^1_t, and f^3_t associated to the family σ_t will be referred to as the *deformation forms* of the family.

[3] I would be happy to learn of any places where these results have appeared so that I can properly acknowledge them in future versions of this article.

One immediate consequence of Proposition 4 is a formula for the variation of the volume form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(*_\sigma 1\big) = 7f^0 *_\sigma 1\,. \tag{6.4}$$

By the same techniques, one can derive a second order expansion:

Proposition 5 (Taylor expansion formula)**.** *Let $\phi \in \Omega^3_+(M)$ be a G_2-structure. Then for all $b_0 \in \Omega^0(M)$, $b_1 \in \Omega^1(M)$, and $b_3 \in \Omega^3_{27}(M,\phi)$ of sufficiently small C^0-norm, the 3-form*

$$\sigma = \phi + \left(3b_0\,\phi + *_\phi(b_1 \wedge \phi) + b_3\right) \tag{6.5}$$

is definite. Moreover, there is an expansion of the form

$$\begin{aligned} *_\sigma\sigma = *_\phi\phi &+ \big(4b_0 *_\phi\phi + b_1 \wedge \phi - *_\phi b_3\big) + \big(\, 2(b_0)^2 + \tfrac{2}{21}\,|b_1|^2_\phi - \tfrac{1}{42}\,|b_3|^2_\phi\,\big) *_\phi\phi \\ &+ Q_1(b_0,b_1,b_3)\wedge\phi + *_\phi Q_3(b_0,b_1,b_3) + R(b_0,b_1,b_3) \end{aligned} \tag{6.6}$$

where Q_1 (a 1-form) and Q_3 (a 3-form in $\Omega^3_{27}(M,\phi)$) are quadratic in the coefficients of the b_i and R is a 4-form that vanishes to order 3 in the coefficients of the b_i. Consequently, there is an expansion of the form

$$*_\sigma 1 = \big(1 + 7b_0 + \big(\,14(b_0)^2 + \tfrac{2}{3}|b_1|^2_\phi - \tfrac{1}{6}|b_3|^2_\phi\,\big) + r(b_0,b_1,b_3)\,\big) *_\phi 1 \tag{6.7}$$

where r vanishes to order 3 in (b_0,b_1,b_3).

6.2. The Laplacian evolution

A natural evolution equation for G_2-structures is the (nonlinear) Laplacian evolution equation for $\sigma \in \Omega^3_+(M)$ defined as follows:

$$\frac{d}{dt}(\sigma) = \Delta_\sigma\sigma\,. \tag{6.8}$$

This equation is diffeomorphism invariant and hence cannot be elliptic in the strict sense. However, it is not difficult to compute the linearization and see that it is transversely elliptic, i.e., elliptic transverse to the action of the diffeomorphism group.

Thus, the by-now standard methods of DeTurck and Hamilton can be applied to show that, if M is compact, then for any smooth $\sigma_0 \in \Omega^3_+(M)$ there exists an extended number T satisfying $0 < T \leq \infty$ and a 1-parameter family $\sigma(t) \in \Omega^3_+(M)$ defined for all t such that $0 < t < T$ so that the family satisfies (6.8) and so that $\sigma(t)$ approaches σ_0 uniformly as t approaches 0 from above. The fundamental issue then becomes to understand the behavior of the family as t approaches T.

For general σ, the formula for the Laplacian in terms of the torsion forms is not too illuminating:

$$\Delta_\sigma\sigma = \mathrm{d}\big(\tau_2 - 4\,\tau_1^\sharp \lrcorner\, \sigma\big) + *_\sigma\mathrm{d}\big(\tau_0\,\sigma + 3\,\tau_1^\sharp \lrcorner *_\sigma\sigma + \tau_3\big). \tag{6.9}$$

This can be further expanded, but the general formula becomes unwieldy rather quickly.

6.2.1. Evolution of closed forms

Suppose now that the initial form σ is closed, i.e., that τ_0, τ_1 and τ_3 are all zero initially. It is not difficult to show that the Laplacian flow preserves this condition, i.e., that the family $\sigma(t)$ consists of closed forms.

For notational simplicity, for the rest of this section, τ_2 will be denoted simply as τ. Also, in the calculations to follow, t will be treated as a parameter, i.e., I will regard dt as zero when computing exterior derivatives. Thus, the assumptions are that

$$\begin{aligned} \mathrm{d}\sigma &= 0 \\ \mathrm{d}{*_\sigma}\sigma &= \tau\wedge\sigma \end{aligned} \tag{6.10}$$

and that

$$\frac{d}{dt}(\sigma) = \mathrm{d}\tau. \tag{6.11}$$

As has already been shown in (4.35),

$$d\tau = \tfrac{1}{7}|\tau|^2\,\sigma + \gamma \tag{6.12}$$

for some $\gamma \in \Omega^3_{27}(M,\sigma)$. In particular, it follows from Proposition 4 that

$$\frac{d}{dt}({*_\sigma}\sigma) = \tfrac{4}{21}|\tau|^2\,{*_\sigma}\sigma - {*_\sigma}\gamma = \tfrac{1}{3}|\tau|^2\,{*_\sigma}\sigma - {*_\sigma}\mathrm{d}\tau\,. \tag{6.13}$$

Moreover, (6.4) now becomes

$$\frac{d}{dt}({*_\sigma}1) = \tfrac{1}{3}|\tau|^2\,{*_\sigma}1. \tag{6.14}$$

In particular, note that the associated volume form ${*_\sigma}1$ is *pointwise* increasing.[4]

Finally, combining (6.11) with the formulae (4.39) and (6.2), one gets the evolution of the metric g_σ in the form

$$\frac{d}{dt}(g_\sigma) = -2\,\mathrm{Ric}(g_\sigma) + \tfrac{8}{21}|\tau|^2\,g_\sigma + \tfrac{1}{4}\,\mathsf{j}\big({*_\sigma}(\tau\wedge\tau)\big). \tag{6.15}$$

Remark 16 (Hitchin's interpretation). Hitchin [12] has given the following interpretation of this flow. Suppose that ϕ is a closed definite 3-form and on a compact 7-manifold M. Let

$$[\phi]_+ = \{\ \phi + \mathrm{d}\beta \in \Omega^3_+(M)\ \mid\ \beta\in\Omega^2(M)\ \} \tag{6.16}$$

be the open set in the cohomology class $[\phi] = \{\phi + \mathrm{d}\beta \mid \beta \in \Omega^2(M)\}$ that consists of definite 3-forms.

Define the volume function $V : [\phi]_+ \to \mathbb{R}^+$ by $V(\sigma) = \int_M {*_\sigma}1 > 0$ for $\sigma \in [\phi]_+$. Hitchin shows that $\sigma \in [\phi]_+$ is a critical point of V if and only if σ is coclosed (as well as closed) and he shows that the flow (6.8) is the gradient flow of the functional V (in the L^2 metric on $[\phi]_+$).

[4] In view of Hitchin's interpretation of this flow as the gradient flow of the volume functional on the space $[\phi]_+$, it is to be expected that the integral of ${*_\sigma}1$ over M is increasing.

Suppose that ϕ is a critical point of V, i.e., that $*_\phi\phi$ is closed. Then by Hodge theory there is a direct sum decomposition

$$\mathrm{d}\big(\Omega^2(M)\big) = \{\, \mathcal{L}_Z\,\phi \mid Z \in \mathrm{Vect}(M)\,\} \oplus \{\, \mathrm{d}\beta \mid \beta \in \Omega^2_{14}(M,\phi),\ \mathrm{d}^{14}_7\beta = 0\,\}. \tag{6.17}$$

The first summand is the tangent space to the orbit of ϕ under $\mathrm{Diff}^\circ(M)$ (i.e., the diffeomorphisms of M that act trivially on $H^*(M)$), while the second summand represents the tangent space at $\mathrm{Diff}^\circ(M){\cdot}\phi$ to the 'moduli space' $\mathrm{Diff}^\circ(M)\backslash[\phi]_+$. If $\beta \in \Omega^2_{14}$ satisfies $d^{14}_7\beta = 0$, then, setting $\sigma = \phi + t\,\mathrm{d}\beta = \phi + t\,d^{14}_{27}\beta$, one finds, by (6.7), that

$$*_\sigma 1 = \big(1 - \tfrac{1}{6}\left|d^{14}_{27}\beta\right|^2 + t^3 R(t,\mathrm{d}\beta)\big) *_\phi 1. \tag{6.18}$$

for some smooth remainder term $R(t,\mathrm{d}\beta)$. By the formulae in Table 1, the equations $\mathrm{d}^{14}_7\beta = \mathrm{d}^{14}_{27}\beta = 0$ for $\beta \in \Omega^2_{14}(M,\phi)$ imply that $\mathrm{d}\beta = \delta\beta = 0$. It follows that the Hessian of V at ϕ is negative definite on $\{\, \mathrm{d}\beta \mid \beta \in \Omega^2_{14}(M,\phi),\ \mathrm{d}^{14}_7\beta = 0\,\}$. Thus, $\mathrm{Diff}^\circ(M){\cdot}\phi$ is a local maximum of V on the moduli space $\mathrm{Diff}^\circ(M)\backslash[\phi]_+$.[5]

In particular, it seems reasonable to expect that, for $\sigma \in [\phi]_+$ 'sufficiently near' ϕ in a appropriate norm, the V-gradient flow (6.8) with σ as initial value would converge to a point on $\mathrm{Diff}^\circ(M){\cdot}\phi$.

Remark 17 (Nonconvergence). A more likely difficulty, it seems, is posed by the possibility that there may be torsion-free G_2-structures ϕ for which the volume functional is not bounded above on $[\phi]_+$, so one would not expect the Laplacian flow to converge for most closed G_2-structures in $[\phi]_+$.

Example 2 (Fernández' closed G_2-solvmanifold). Fernández [4, 5] has constructed compact 7-dimensional manifolds M^7 that support a closed G_2-structure ϕ but that cannot, for topological reasons, support a torsion-free G_2-structure. Thus, in these cases, the above flow cannot converge, since there will be no critical points of V on $[\phi]_+$.

It is instructive to look at one of her examples: Let $G \subset \mathrm{GL}(5,\mathbb{R})$ be the subgroup that consists of matrices of the form

$$g = \begin{pmatrix} 1 & 0 & x^2 & x^4 & x^6 \\ 0 & 1 & x^3 & x^5 & x^7 \\ 0 & 0 & 1 & 0 & x^1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.19}$$

where $x^i \in \mathbb{R}$ for $1 \le i \le 7$. Write the left-invariant form on G in the form

$$g^{-1}\,\mathrm{d}g = \begin{pmatrix} 0 & 0 & \omega^2 & \omega^4 & \omega^6 \\ 0 & 0 & \omega^3 & \omega^5 & \omega^7 \\ 0 & 0 & 0 & 0 & \omega^1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{6.20}$$

[5] Hitchin says that V is a 'Morse-Bott' functional on $[\phi]_+$, i.e., that V has nondegenerate critical points on the moduli space.

where $d\omega^i = 0$ for $1 \le i \le 5$ while $d\omega^6 = \omega^1{\wedge}\omega^2$ and $d\omega^7 = \omega^1{\wedge}\omega^3$.

Let $\Gamma = G{\cap}\mathrm{GL}(5,\mathbb{Z})$ and note that Γ is a co-compact discrete subgroup of G. Let $M^7 = \Gamma\backslash G$ be the space of right cosets of Γ in G. Then the ω^i are well-defined on M and it is easy to verify that the 3-form

$$\sigma = \omega^{123} + \omega^{145} + \omega^{167} + \omega^{246} - \omega^{257} - \omega^{347} - \omega^{356} \tag{6.21}$$

is a closed G_2-structure on M. It is not coclosed, but satisfies

$$d{*_\sigma}\sigma = \bigl(\omega^{27} - \omega^{36}\bigr) \wedge \sigma. \tag{6.22}$$

Since

$$d\bigl(\omega^{27} - \omega^{36}\bigr) = 2\,\omega^{123}, \tag{6.23}$$

it follows that the flow satisfies

$$\sigma(t) = e^{2t}\omega^{123} + \omega^{145} + \omega^{167} + \omega^{246} - \omega^{257} - \omega^{347} - \omega^{356}. \tag{6.24}$$

The associated metric is

$$g(t) = e^{4t/3}\bigl((\omega^1)^2 + (\omega^2)^2 + (\omega^3)^2\bigr) + e^{-2t/3}\bigl((\omega^4)^2 + (\omega^5)^2 + (\omega^6)^2 + (\omega^7)^2\bigr). \tag{6.25}$$

In particular, note that, under this flow (which exists for all time, both past and future), the volume of the metric increases without bound.

By the way, M cannot carry a metric with holonomy a subgroup of G_2 for the following reason: As Fernández shows, the first Betti number of M is 5. If there were a metric g on M with holonomy in G_2, then it would be Ricci-flat and hence the harmonic representatives of the first cohomology group would give five linearly independent g-parallel 1-forms on M. However, this would imply that the holonomy of M is trivial, which would imply that there exist seven linearly independent parallel 1-forms on M, which would in turn imply that the first Betti number was at least 7.

6.2.2. Further calculations

Return now to the flow of a general closed G_2-structure. Taking the exterior derivative of (6.13) yields

$$\frac{d}{dt}(\tau \wedge \sigma) = \tfrac{1}{3}d\bigl(|\tau|^2\bigr) \wedge {*_\sigma}\sigma + \tfrac{1}{3}|\tau|^2\,\tau \wedge \sigma - d{*_\sigma}d\tau\,. \tag{6.26}$$

Expanding the left hand side of this equation and using (6.11) yields

$$\frac{d}{dt}(\tau) \wedge \sigma = \tfrac{1}{3}d\bigl(|\tau|^2\bigr) \wedge {*_\sigma}\sigma + \tfrac{1}{3}|\tau|^2\,\tau \wedge \sigma - d{*_\sigma}d\tau - \tau \wedge d\tau\,. \tag{6.27}$$

(This equation can be solved for the time-derivative of τ since wedging with σ is an isomorphism between Ω^2 and Ω^5.) Recalling that $\tau{\wedge}{*_\sigma}\sigma = 0$ and $\tau{\wedge}\tau{\wedge}\sigma = -|\tau^2|{*_\sigma}1$, this yields

$$\frac{d}{dt}(\tau) \wedge \tau \wedge \sigma = -\tfrac{1}{3}|\tau|^4\,{*_\sigma}1 - \tau \wedge d{*_\sigma}d\tau - \tau \wedge \tau \wedge d\tau\,. \tag{6.28}$$

Finally, this can be used in the following computation

$$\begin{aligned}\frac{d}{dt}(|\tau^2| *_\sigma 1) &= \frac{d}{dt}(-\tau\wedge\tau\wedge\sigma) = -2\frac{d}{dt}(\tau)\wedge\tau\wedge\sigma - \tau\wedge\tau\wedge \mathrm{d}\tau \\ &= \tfrac{2}{3}|\tau|^4 *_\sigma 1 + 2\tau\wedge \mathrm{d}{*_\sigma}\mathrm{d}\tau + \tau\wedge\tau\wedge\mathrm{d}\tau \\ &= \left(\tfrac{2}{3}|\tau|^4 - 2|\mathrm{d}\tau|^2\right) *_\sigma 1 + \mathrm{d}\left(2\tau\wedge *_\sigma \mathrm{d}\tau + \tfrac{1}{3}\tau^3\right).\end{aligned} \tag{6.29}$$

Integrating this equation over M yields

$$\frac{d}{dt}\int_M |\tau^2| *_\sigma 1 = \int_M \left(\tfrac{2}{3}|\tau|^4 - 2|\mathrm{d}\tau|^2\right) *_\sigma 1. \tag{6.30}$$

This equation can be rewritten by using (4.39), which yields

$$\frac{d}{dt}\int_M (|\tau^2| *_\sigma 1) = \int_M \left(\tfrac{2}{3}|\tau|^4 - 2\left|\tfrac{3}{14}|\tau|^2\sigma + \tfrac{1}{2}{*_\sigma}(\tau\wedge\tau) - \tfrac{1}{2}\mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big)\right|^2\right) *_\sigma 1. \tag{6.31}$$

Now, going back to (4.45) and integrating this over M yields

$$0 = \int_M \tfrac{2}{7}|\tau_2|^4 *_\sigma 1 - \tfrac{1}{2}\left({\tau_2}^2 + \tfrac{1}{7}|\tau_2|^2 *_\sigma \sigma\right)\wedge \mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big), \tag{6.32}$$

i.e.,

$$\int_M \left\langle *_\sigma\left({\tau_2}^2 + \tfrac{1}{7}|\tau_2|^2 *_\sigma\sigma\right), \mathsf{i}\big(\mathrm{Ric}^0(g_\sigma)\big)\right\rangle *_\sigma 1 = \frac{4}{7}\int_M |\tau_2|^4 *_\sigma 1. \tag{6.33}$$

Using this relation and the algebraic identities (2.30) and (2.28), one sees that (6.31) can be rewritten in the form

$$\frac{d}{dt}\int_M (|\tau^2| *_\sigma 1) = 4\int_M \left(\tfrac{11}{21}\,\mathrm{Scal}(g_\sigma)^2 - \left|\mathrm{Ric}^0(g_\sigma)\right|^2\right) *_\sigma 1. \tag{6.34}$$

This equation is suggestive. One of the reasons for wanting to study the Laplacian flow on closed G_2-structures is that it might provide a means of constructing metrics with holonomy G_2 by starting with a closed G_2-structure $\sigma \in \Omega^3_+(M)$ with 'sufficiently small' torsion and then running the Laplacian flow to move it closer to a G_2-structure that is both closed and coclosed.

However, if such a procedure is to work, then the volume function along the flow line must approach a constant and one would certainly expect the second derivative to become negative if the volume were to approach the 'local maximum' target volume. However, (6.34) shows that, in this case, the relative separation of the eigenvalues of the Ricci tensor cannot decrease too much during the flow. This 'forced separation' is somewhat stronger than the separation implied by Corollary 3.

References

[1] E. Bonan, *Sur les variétés Riemanniennes à groupe d'holonomie* G_2 *ou* Spin(7), C. R. Acad. Sci. Paris **262** (1966), 127–129. MR 33 4855

[2] R. Bryant, *Metrics with exceptional holonomy*, Ann. of Math. (2) **126** (1987), 525–576. MR 1989b:53084

[3] R. Cleyton, S. Ivanov, *On the geometry of closed* G_2*-structures*, arXiv article, June 25, 2003. arXiv:math.DG/0306362

[4] M. Fernández, *A family of compact solvable* G_2*-calibrated manifolds*, Tôhoku Math. J. 39 (1987), 287–289.

[5] ———, *An example of a compact calibrated manifold associated with the exceptional Lie group* G_2. J. Differential Geom. 26 (1987), 367–370.

[6] M. Fernández, A. Gray *Riemannian manifolds with structure group* G_2, Ann. Mat. Pura Appl. 4 **132** (1982), 19–45. MR1984e:53056

[7] T. Friedrich and S. Ivanov, *Parallel spinors and connections with skew-symmetric torsion in string theory*, Asian J. of Math. **6** (2002), 303–336. arXiv:math.DG/0102142

[8] ———, *Killing spinor equations in dimension* 7 *and geometry of integrable* G_2*-manifolds*, arXiv:math.DG/0112201

[9] A. Gray, *Vector cross products on manifolds*, Trans. Amer. Math. Soc. **141** (1969), 465–504. MR 39 #4790

[10] F. R. Harvey, *Spinors and calibrations*, Academic Press Inc., Boston, MA, 1990. MR 1991e:53056

[11] F. R. Harvey, H. B. Lawson, Jr., *Calibrated geometries*, Acta Math. **148** (1982), 47–157. MR 1985i:53058

[12] N. Hitchin, *The geometry of three-forms in* 6 *and* 7 *dimensions*, J. Differential Geom. **55** (2000), 547–576. MR 2002m:53070

[13] J. Humphreys, *Introduction to Lie algebras and representation theory*, second printing, revised. Graduate Texts in Mathematics, 9. Springer-Verlag, New York-Berlin, 1978. MR 81b:17007

[14] D. Joyce, *Compact manifolds with special holonomy*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2000. MR 2001k:53093

[15] S. Salamon, *Riemannian geometry and holonomy groups*, Pitman Research Notes in Math., **201**. Longman, Harlow, 1989. MR 1990g:53058

[16] J. A. Schouten, *Kalssifizierung der Alternierenden Gröszen Dritten Grades in* 7 *Dimensionen*, Rend. Circ. Maten. Palermo **55** (1931), 131–156.

Duke University Mathematics Department, P.O. Box 90320, Durham, NC 27708-0320
E-mail address: `bryant@math.duke.edu`

Proceedings of 12^{th} Gökova
Geometry-Topology Conference
pp. 110 – 139

Published online at
GokovaGT.org

The exceptional holonomy groups and calibrated geometry

Dominic Joyce

Dedicated to the memory of Raoul Bott.

Abstract. The exceptional holonomy groups are G_2 in 7 dimensions, and Spin(7) in 8 dimensions. Riemannian manifolds with these holonomy groups are Ricci-flat. This is a survey paper on exceptional holonomy, in two parts. Part I introduces the exceptional holonomy groups, and explains constructions for compact 7- and 8-manifolds with holonomy G_2 and Spin(7). The simplest such constructions work by using techniques from complex geometry and Calabi–Yau analysis to resolve the singularities of a torus orbifold T^7/Γ or T^8/Γ, for Γ a finite group preserving a flat G_2 or Spin(7)-structure on T^7 or T^8. There are also more complicated constructions which begin with a Calabi–Yau manifold or orbifold. Part II discusses the calibrated submanifolds of G_2 and Spin(7)-manifolds: associative 3-folds and coassociative 4-folds for G_2, and Cayley 4-folds for Spin(7). We explain the general theory, following Harvey and Lawson, and the known examples. Finally we describe the deformation theory of compact calibrated submanifolds, following McLean.

1. Introduction

In the theory of Riemannian holonomy groups, perhaps the most mysterious are the two exceptional cases, the holonomy group G_2 in 7 dimensions and the holonomy group Spin(7) in 8 dimensions. This is a survey paper on the exceptional holonomy groups, in two parts. Part I collects together useful facts about G_2 and Spin(7) in §2, and explains constructions of compact 7-manifolds with holonomy G_2 in §3, and of compact 8-manifolds with holonomy Spin(7) in §4.

Part II discusses the *calibrated submanifolds* of manifolds of exceptional holonomy, namely *associative* 3*-folds* and *coassociative* 4*-folds* in G_2-manifolds, and *Cayley* 4*-folds* in Spin(7)-manifolds. We introduce calibrations in §5, defining the three geometries and giving examples. Finally, §6 explains their *deformation theory*.

Sections 3 and 4 describe my own work. On this the exhaustive and nearly infallible reference is my book [18], which also makes an excellent Christmas present. Part II describes work by other people, principally the very important papers by Harvey and Lawson [12] and McLean [28], but also more recent developments.

This paper was written to accompany lectures at the 11^{th} Gökova Geometry and Topology Conference in May 2004, sponsored by TÜBİTAK. In keeping with the theme of the conference, I have focussed mostly on G_2, at the expense of Spin(7). In writing it I have plagiarized shamelessly from my previous works, notably the books [18] and [11, Part I], and the survey paper [21].

Part I. Exceptional Holonomy

2. Introduction to G_2 and Spin(7)

We introduce the notion of *Riemannian holonomy groups*, and their classification by Berger. Then we give short descriptions of the holonomy groups G_2, Spin(7) and $\mathrm{SU}(m)$, and the relations between them. All the results below can be found in my book [18].

2.1. Riemannian holonomy groups

Let M be a connected n-dimensional manifold, g a Riemannian metric on M, and ∇ the Levi-Civita connection of g. Let x, y be points in M joined by a smooth path γ. Then *parallel transport* along γ using ∇ defines an isometry between the tangent spaces T_xM and T_yM at x and y.

Definition 2.1. The *holonomy group* $\mathrm{Hol}(g)$ of g is the group of isometries of T_xM generated by parallel transport around piecewise-smooth closed loops based at x in M. We consider $\mathrm{Hol}(g)$ to be a subgroup of $\mathrm{O}(n)$, defined up to conjugation by elements of $\mathrm{O}(n)$. Then $\mathrm{Hol}(g)$ is independent of the base point x in M.

Let ∇ be the *Levi-Civita connection* of g. A tensor S on M is *constant* if $\nabla S = 0$. An important property of $\mathrm{Hol}(g)$ is that it *determines the constant tensors on* M.

Theorem 2.2. *Let (M, g) be a Riemannian manifold, and ∇ the Levi-Civita connection of g. Fix a base point $x \in M$, so that $\mathrm{Hol}(g)$ acts on T_xM, and so on the tensor powers $\bigotimes^k T_xM \otimes \bigotimes^l T_x^*M$. Suppose $S \in C^\infty\bigl(\bigotimes^k TM \otimes \bigotimes^l T^*M\bigr)$ is a constant tensor. Then $S|_x$ is fixed by the action of $\mathrm{Hol}(g)$. Conversely, if $S|_x \in \bigotimes^k T_xM \otimes \bigotimes^l T_x^*M$ is fixed by $\mathrm{Hol}(g)$, it extends to a unique constant tensor $S \in C^\infty\bigl(\bigotimes^k TM \otimes \bigotimes^l T^*M\bigr)$.*

The main idea in the proof is that if S is a constant tensor and $\gamma : [0,1] \to M$ is a path from x to y, then $P_\gamma(S|_x) = S|_y$, where P_γ is the *parallel transport map* along γ. Thus, constant tensors are invariant under parallel transport. In particular, they are invariant under parallel transport around closed loops based at x, that is, under elements of $\mathrm{Hol}(g)$.

The classification of holonomy groups was achieved by Berger [1] in 1955.

Theorem 2.3. *Let M be a simply-connected, n-dimensional manifold, and g an irreducible, nonsymmetric Riemannian metric on M. Then either*

(i) $\mathrm{Hol}(g) = \mathrm{SO}(n)$,
(ii) $n = 2m$ *and* $\mathrm{Hol}(g) = \mathrm{SU}(m)$ *or* $\mathrm{U}(m)$,
(iii) $n = 4m$ *and* $\mathrm{Hol}(g) = \mathrm{Sp}(m)$ *or* $\mathrm{Sp}(m)\,\mathrm{Sp}(1)$,
(iv) $n = 7$ *and* $\mathrm{Hol}(g) = G_2$, *or*
(v) $n = 8$ *and* $\mathrm{Hol}(g) = \mathrm{Spin}(7)$.

Here are some brief remarks about each group on Berger's list.

(i) $\mathrm{SO}(n)$ is the holonomy group of generic Riemannian metrics.

(ii) Riemannian metrics g with $\mathrm{Hol}(g) \subseteq \mathrm{U}(m)$ are called *Kähler metrics*. Kähler metrics are a natural class of metrics on complex manifolds, and generic Kähler metrics on a given complex manifold have holonomy $\mathrm{U}(m)$.

Metrics g with $\mathrm{Hol}(g) = \mathrm{SU}(m)$ are called *Calabi–Yau metrics*. Since $\mathrm{SU}(m)$ is a subgroup of $\mathrm{U}(m)$, all Calabi–Yau metrics are Kähler. If g is Kähler and M is simply-connected, then $\mathrm{Hol}(g) \subseteq \mathrm{SU}(m)$ if and only if g is Ricci-flat. Thus Calabi–Yau metrics are locally more or less the same as Ricci-flat Kähler metrics.

(iii) metrics g with $\mathrm{Hol}(g) = \mathrm{Sp}(m)$ are called *hyperkähler*. As $\mathrm{Sp}(m) \subseteq \mathrm{SU}(2m) \subset \mathrm{U}(2m)$, hyperkähler metrics are Ricci-flat and Kähler.

Metrics g with holonomy group $\mathrm{Sp}(m)\,\mathrm{Sp}(1)$ for $m \geqslant 2$ are called *quaternionic Kähler*. (Note that quaternionic Kähler metrics are not in fact Kähler.) They are Einstein, but not Ricci-flat.

(iv), (v) G_2 and $\mathrm{Spin}(7)$ are the exceptional cases, so they are called the *exceptional holonomy groups*. Metrics with these holonomy groups are Ricci-flat.

The groups can be understood in terms of the four *division algebras*: the *real numbers* $\mathbb{R}$, the *complex numbers* $\mathbb{C}$, the *quaternions* $\mathbb{H}$, and the *octonions* or *Cayley numbers* $\mathbb{O}$.

- $\mathrm{SO}(n)$ is a group of automorphisms of $\mathbb{R}^n$.
- $\mathrm{U}(m)$ and $\mathrm{SU}(m)$ are groups of automorphisms of $\mathbb{C}^m$
- $\mathrm{Sp}(m)$ and $\mathrm{Sp}(m)\,\mathrm{Sp}(1)$ are automorphism groups of $\mathbb{H}^m$.
- G_2 is the automorphism group of $\mathrm{Im}\,\mathbb{O} \cong \mathbb{R}^7$. $\mathrm{Spin}(7)$ is a group of automorphisms of $\mathbb{O} \cong \mathbb{R}^8$, preserving part of the structure on $\mathbb{O}$.

For some time after Berger's classification, the exceptional holonomy groups remained a mystery. In 1987, Bryant [6] used the theory of exterior differential systems to show that locally there exist many metrics with these holonomy groups, and gave some explicit, incomplete examples. Then in 1989, Bryant and Salamon [8] found explicit, *complete* metrics with holonomy G_2 and $\mathrm{Spin}(7)$ on noncompact manifolds.

In 1994-5 the author constructed the first examples of metrics with holonomy G_2 and $\mathrm{Spin}(7)$ on *compact* manifolds [14, 15, 16]. These, and the more complicated constructions developed later by the author [17, 18] and by Kovalev [22], are the subject of Part I.

2.2. The holonomy group G_2

Let $(x_1, \ldots, x_7)$ be coordinates on $\mathbb{R}^7$. Write $\mathrm{d}\mathbf{x}_{ij\ldots l}$ for the exterior form $\mathrm{d}x_i \wedge \mathrm{d}x_j \wedge \cdots \wedge \mathrm{d}x_l$ on $\mathbb{R}^7$. Define a metric g_0, a 3-form φ_0 and a 4-form $*\varphi_0$ on $\mathbb{R}^7$ by $g_0 = \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_7^2$,

$$\begin{aligned} \varphi_0 &= \mathrm{d}\mathbf{x}_{123} + \mathrm{d}\mathbf{x}_{145} + \mathrm{d}\mathbf{x}_{167} + \mathrm{d}\mathbf{x}_{246} - \mathrm{d}\mathbf{x}_{257} - \mathrm{d}\mathbf{x}_{347} - \mathrm{d}\mathbf{x}_{356} \quad \text{and} \\ *\varphi_0 &= \mathrm{d}\mathbf{x}_{4567} + \mathrm{d}\mathbf{x}_{2367} + \mathrm{d}\mathbf{x}_{2345} + \mathrm{d}\mathbf{x}_{1357} - \mathrm{d}\mathbf{x}_{1346} - \mathrm{d}\mathbf{x}_{1256} - \mathrm{d}\mathbf{x}_{1247}. \end{aligned} \tag{1}$$

The subgroup of $GL(7, \mathbb{R})$ preserving φ_0 is the *exceptional Lie group* G_2. It also preserves $g_0, *\varphi_0$ and the orientation on $\mathbb{R}^7$. It is a compact, semisimple, 14-dimensional Lie group, a subgroup of $\mathrm{SO}(7)$.

A G_2-*structure* on a 7-manifold M is a principal subbundle of the frame bundle of M, with structure group G_2. Each G_2-structure gives rise to a 3-form φ and a metric g on M, such that every tangent space of M admits an isomorphism with $\mathbb{R}^7$ identifying φ and g with φ_0 and g_0 respectively. By an abuse of notation, we will refer to (φ, g) as a G_2-structure.

Proposition 2.4. *Let M be a 7-manifold and (φ, g) a G_2-structure on M. Then the following are equivalent:*

(i) $\mathrm{Hol}(g) \subseteq G_2$, *and φ is the induced 3-form,*
(ii) $\nabla\varphi = 0$ *on M, where ∇ is the Levi-Civita connection of g, and*
(iii) $\mathrm{d}\varphi = \mathrm{d}^*\varphi = 0$ *on M.*

Note that $\mathrm{Hol}(g) \subseteq G_2$ if and only if $\nabla\varphi = 0$ follows from Theorem 2.2. We call $\nabla\varphi$ the *torsion* of the G_2-structure (φ, g), and when $\nabla\varphi = 0$ the G_2-structure is *torsion-free*. A triple (M, φ, g) is called a G_2-*manifold* if M is a 7-manifold and (φ, g) a torsion-free G_2-structure on M. If g has holonomy $\mathrm{Hol}(g) \subseteq G_2$, then g is Ricci-flat.

Theorem 2.5. *Let M be a compact 7-manifold, and suppose that (φ, g) is a torsion-free G_2-structure on M. Then $\mathrm{Hol}(g) = G_2$ if and only if $\pi_1(M)$ is finite. In this case the moduli space of metrics with holonomy G_2 on M, up to diffeomorphisms isotopic to the identity, is a smooth manifold of dimension $b^3(M)$.*

2.3. The holonomy group Spin(7)

Let $\mathbb{R}^8$ have coordinates $(x_1, \dots, x_8)$. Define a 4-form Ω_0 on $\mathbb{R}^8$ by

$$\begin{aligned}\Omega_0 = {} & \mathrm{d}\mathbf{x}_{1234} + \mathrm{d}\mathbf{x}_{1256} + \mathrm{d}\mathbf{x}_{1278} + \mathrm{d}\mathbf{x}_{1357} - \mathrm{d}\mathbf{x}_{1368} - \mathrm{d}\mathbf{x}_{1458} - \mathrm{d}\mathbf{x}_{1467} \\ & - \mathrm{d}\mathbf{x}_{2358} - \mathrm{d}\mathbf{x}_{2367} - \mathrm{d}\mathbf{x}_{2457} + \mathrm{d}\mathbf{x}_{2468} + \mathrm{d}\mathbf{x}_{3456} + \mathrm{d}\mathbf{x}_{3478} + \mathrm{d}\mathbf{x}_{5678}.\end{aligned} \tag{2}$$

The subgroup of $\mathrm{GL}(8, \mathbb{R})$ preserving Ω_0 is the holonomy group Spin(7). It also preserves the orientation on $\mathbb{R}^8$ and the Euclidean metric $g_0 = \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_8^2$. It is a compact, semisimple, 21-dimensional Lie group, a subgroup of SO(8).

A Spin(7)-structure on an 8-manifold M gives rise to a 4-form Ω and a metric g on M, such that each tangent space of M admits an isomorphism with $\mathbb{R}^8$ identifying Ω and g with Ω_0 and g_0 respectively. By an abuse of notation we will refer to the pair (Ω, g) as a Spin(7)-structure.

Proposition 2.6. *Let M be an 8-manifold and (Ω, g) a* Spin(7)*-structure on M. Then the following are equivalent:*

(i) $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$, *and Ω is the induced 4-form,*
(ii) $\nabla\Omega = 0$ *on M, where ∇ is the Levi-Civita connection of g, and*
(iii) $\mathrm{d}\Omega = 0$ *on M.*

We call $\nabla\Omega$ the *torsion* of the Spin(7)-structure (Ω, g), and (Ω, g) *torsion-free* if $\nabla\Omega = 0$. A triple (M, Ω, g) is called a Spin(7)-*manifold* if M is an 8-manifold and (Ω, g) a torsion-free Spin(7)-structure on M. If g has holonomy $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$, then g is Ricci-flat.

Here is a result on *compact* 8-manifolds with holonomy Spin(7).

Theorem 2.7. *Let (M, Ω, g) be a compact* Spin(7)*-manifold. Then* $\mathrm{Hol}(g) = \mathrm{Spin}(7)$ *if and only if M is simply-connected, and $b^3(M) + b^4_+(M) = b^2(M) + 2b^4_-(M) + 25$. In this case the moduli space of metrics with holonomy* Spin(7) *on M, up to diffeomorphisms isotopic to the identity, is a smooth manifold of dimension $1 + b^4_-(M)$.*

2.4. The holonomy groups SU(m)

Let $\mathbb{C}^m \cong \mathbb{R}^{2m}$ have complex coordinates $(z_1, \ldots, z_m)$, and define the metric g_0, Kähler form ω_0 and complex volume form θ_0 on $\mathbb{C}^m$ by

$$\begin{aligned} g_0 = |\mathrm{d}z_1|^2 + \cdots + |\mathrm{d}z_m|^2, \quad \omega_0 &= \frac{i}{2}(\mathrm{d}z_1 \wedge \mathrm{d}\bar{z}_1 + \cdots + \mathrm{d}z_m \wedge \mathrm{d}\bar{z}_m), \\ \text{and} \quad \theta_0 &= \mathrm{d}z_1 \wedge \cdots \wedge \mathrm{d}z_m. \end{aligned} \tag{3}$$

The subgroup of $\mathrm{GL}(2m, \mathbb{R})$ preserving g_0, ω_0 and θ_0 is the special unitary group $\mathrm{SU}(m)$. Manifolds with holonomy $\mathrm{SU}(m)$ are called *Calabi–Yau manifolds*.

Calabi–Yau manifolds are automatically Ricci-flat and Kähler, with trivial canonical bundle. Conversely, any Ricci-flat Kähler manifold (M, J, g) with trivial canonical bundle has $\mathrm{Hol}(g) \subseteq \mathrm{SU}(m)$. By Yau's proof of the Calabi Conjecture [**31**], we have:

Theorem 2.8. *Let (M, J) be a compact complex m-manifold admitting Kähler metrics, with trivial canonical bundle. Then there is a unique Ricci-flat Kähler metric g in each Kähler class on M, and* $\mathrm{Hol}(g) \subseteq \mathrm{SU}(m)$.

Using this and complex algebraic geometry one can construct many examples of compact Calabi–Yau manifolds. The theorem also applies in the orbifold category, yielding examples of *Calabi–Yau orbifolds*.

2.5. Relations between G_2, Spin(7) and SU(m)

Here are the inclusions between the holonomy groups $\mathrm{SU}(m), G_2$ and Spin(7):

$$\begin{array}{ccccc} \mathrm{SU}(2) & \longrightarrow & \mathrm{SU}(3) & \longrightarrow & G_2 \\ \downarrow & & \downarrow & & \downarrow \\ \mathrm{SU}(2) \times \mathrm{SU}(2) & \longrightarrow & \mathrm{SU}(4) & \longrightarrow & \mathrm{Spin}(7). \end{array}$$

We shall illustrate what we mean by this using the inclusion $\mathrm{SU}(3) \hookrightarrow G_2$. As SU(3) acts on $\mathbb{C}^3$, it also acts on $\mathbb{R} \oplus \mathbb{C}^3 \cong \mathbb{R}^7$, taking the SU(3)-action on $\mathbb{R}$ to be trivial. Thus we embed SU(3) as a subgroup of $\mathrm{GL}(7, \mathbb{R})$. It turns out that SU(3) is a subgroup of the subgroup G_2 of $\mathrm{GL}(7, \mathbb{R})$ defined in §2.2.

Here is a way to see this in terms of differential forms. Identify $\mathbb{R} \oplus \mathbb{C}^3$ with $\mathbb{R}^7$ in the obvious way in coordinates, so that $\big(x_1, (x_2 + ix_3, x_4 + ix_5, x_6 + ix_7)\big)$ in $\mathbb{R} \oplus \mathbb{C}^3$ is identified with $(x_1, \ldots, x_7)$ in $\mathbb{R}^7$. Then $\varphi_0 = \mathrm{d}x_1 \wedge \omega_0 + \mathrm{Re}\,\theta_0$, where φ_0 is defined in (1) and ω_0, θ_0 in (3). Since SU(3) preserves ω_0 and θ_0, the action of SU(3) on $\mathbb{R}^7$ preserves φ_0, and so $\mathrm{SU}(3) \subset G_2$.

It follows that if (M, J, h) is Calabi–Yau 3-fold, then $\mathbb{R} \times M$ and $\mathcal{S}^1 \times M$ have torsion-free G_2-structures, that is, are G_2-manifolds.

Proposition 2.9. *Let (M, J, h) be a Calabi–Yau 3-fold, with Kähler form ω and complex volume form θ. Let x be a coordinate on $\mathbb{R}$ or $\mathcal{S}^1$. Define a metric $g = \mathrm{d}x^2 + h$ and a 3-form $\varphi = \mathrm{d}x \wedge \omega + \mathrm{Re}\,\theta$ on $\mathbb{R} \times M$ or $\mathcal{S}^1 \times M$. Then (φ, g) is a torsion-free G_2-structure on $\mathbb{R} \times M$ or $\mathcal{S}^1 \times M$, and $*\varphi = \frac{1}{2}\omega \wedge \omega - \mathrm{d}x \wedge \mathrm{Im}\,\theta$.*

Similarly, the inclusions $\mathrm{SU}(2) \hookrightarrow G_2$ and $\mathrm{SU}(4) \hookrightarrow \mathrm{Spin}(7)$ give:

Proposition 2.10. *Let (M, J, h) be a Calabi–Yau 2-fold, with Kähler form ω and complex volume form θ. Let (x_1, x_2, x_3) be coordinates on $\mathbb{R}^3$ or T^3. Define a metric $g = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 + h$ and a 3-form φ on $\mathbb{R}^3 \times M$ or $T^3 \times M$ by*

$$\varphi = \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3 + \mathrm{d}x_1 \wedge \omega + \mathrm{d}x_2 \wedge \mathrm{Re}\,\theta - \mathrm{d}x_3 \wedge \mathrm{Im}\,\theta. \tag{4}$$

Then (φ, g) is a torsion-free G_2-structure on $\mathbb{R}^3 \times M$ or $T^3 \times M$, and

$$*\,\varphi = \tfrac{1}{2}\omega \wedge \omega + \mathrm{d}x_2 \wedge \mathrm{d}x_3 \wedge \omega - \mathrm{d}x_1 \wedge \mathrm{d}x_3 \wedge \mathrm{Re}\,\theta - \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{Im}\,\theta. \tag{5}$$

Proposition 2.11. *Let (M, J, g) be a Calabi–Yau 4-fold, with Kähler form ω and complex volume form θ. Define a 4-form Ω on M by $\Omega = \frac{1}{2}\omega \wedge \omega + \mathrm{Re}\,\theta$. Then (Ω, g) is a torsion-free* Spin(7)*-structure on M.*

3. Constructing G_2-manifolds from orbifolds T^7/Γ

We now explain the method used in [14, 15] and [18, §11–§12] to construct examples of compact 7-manifolds with holonomy G_2. It is based on the *Kummer construction* for Calabi–Yau metrics on the $K3$ surface, and may be divided into four steps.

Step 1. Let T^7 be the 7-torus and (φ_0, g_0) a flat G_2-structure on T^7. Choose a finite group Γ of isometries of T^7 preserving (φ_0, g_0). Then the quotient T^7/Γ is a singular, compact 7-manifold, an *orbifold*.

Step 2. For certain special groups Γ there is a method to resolve the singularities of T^7/Γ in a natural way, using complex geometry. We get a nonsingular, compact 7-manifold M, together with a map $\pi : M \to T^7/\Gamma$, the resolving map.

Step 3. On M, we explicitly write down a 1-parameter family of G_2-structures (φ_t, g_t) depending on $t \in (0, \epsilon)$. They are not torsion-free, but have small torsion when t is small. As $t \to 0$, the G_2-structure (φ_t, g_t) converges to the singular G_2-structure $\pi^*(\varphi_0, g_0)$.

Step 4. We prove using analysis that for sufficiently small t, the G_2-structure (φ_t, g_t) on M, with small torsion, can be deformed to a G_2-structure $(\tilde{\varphi}_t, \tilde{g}_t)$, with zero torsion. Finally, we show that $\tilde{g}_t$ is a metric with holonomy G_2 on the compact 7-manifold M.

We will now explain each step in greater detail.

3.1. Step 1: Choosing an orbifold

Let (φ_0, g_0) be the Euclidean G_2-structure on $\mathbb{R}^7$ defined in §2.2. Suppose Λ is a *lattice* in $\mathbb{R}^7$, that is, a discrete additive subgroup isomorphic to $\mathbb{Z}^7$. Then $\mathbb{R}^7/\Lambda$ is the torus T^7, and (φ_0, g_0) pushes down to a torsion-free G_2-structure on T^7. We must choose a finite group Γ acting on T^7 preserving (φ_0, g_0). That is, the elements of Γ are the push-forwards to T^7/Λ of affine transformations of $\mathbb{R}^7$ which fix (φ_0, g_0), and take Λ to itself under conjugation.

Here is an example of a suitable group Γ, taken from [18, §12.2].

Example 3.1. Let $(x_1, \ldots, x_7)$ be coordinates on $T^7 = \mathbb{R}^7/\mathbb{Z}^7$, where $x_i \in \mathbb{R}/\mathbb{Z}$. Let (φ_0, g_0) be the flat G_2-structure on T^7 defined by (1). Let α, β and γ be the involutions of T^7 defined by

$$\alpha : (x_1, \ldots, x_7) \mapsto (x_1, x_2, x_3, -x_4, -x_5, -x_6, -x_7), \tag{6}$$

$$\beta : (x_1, \ldots, x_7) \mapsto (x_1, -x_2, -x_3, x_4, x_5, \tfrac{1}{2} - x_6, -x_7), \tag{7}$$

$$\gamma : (x_1, \ldots, x_7) \mapsto \bigl(-x_1, x_2, -x_3, x_4, \tfrac{1}{2} - x_5, x_6, \tfrac{1}{2} - x_7\bigr). \tag{8}$$

By inspection, α, β and γ preserve (φ_0, g_0), because of the careful choice of exactly which signs to change. Also, $\alpha^2 = \beta^2 = \gamma^2 = 1$, and α, β and γ commute. Thus they generate a group $\Gamma = \langle \alpha, \beta, \gamma \rangle \cong \mathbb{Z}_2^3$ of isometries of T^7 preserving the flat G_2-structure (φ_0, g_0).

Having chosen a lattice Λ and finite group Γ, the quotient T^7/Γ is an *orbifold*, a singular manifold with only quotient singularities. The singularities of T^7/Γ come from the fixed points of non-identity elements of Γ. We now describe the singularities in our example.

Lemma 3.2. *In Example 3.1, $\beta\gamma, \gamma\alpha, \alpha\beta$ and $\alpha\beta\gamma$ have no fixed points on T^7. The fixed points of α, β, γ are each 16 copies of T^3. The singular set S of T^7/Γ is a disjoint union of 12 copies of T^3, 4 copies from each of α, β, γ. Each component of S is a singularity modelled on that of $T^3 \times \mathbb{C}^2/\{\pm 1\}$.*

The most important consideration in choosing Γ is that we should be able to resolve the singularities of T^7/Γ within holonomy G_2. We will explain how to do this next.

3.2. Step 2: Resolving the singularities

Our goal is to resolve the singular set S of T^7/Γ to get a compact 7-manifold M with holonomy G_2. How can we do this? In general we cannot, because we have no idea of how to resolve general orbifold singularities with holonomy G_2. However, suppose we can arrange that every connected component of S is locally isomorphic to either

(a) $T^3 \times \mathbb{C}^2/G$, where G is a finite subgroup of SU(2), or
(b) $\mathcal{S}^1 \times \mathbb{C}^3/G$, where G is a finite subgroup of SU(3) acting freely on $\mathbb{C}^3 \setminus \{0\}$.

One can use complex algebraic geometry to find a *crepant resolution* X of $\mathbb{C}^2/G$ or Y of $\mathbb{C}^3/G$. Then $T^3 \times X$ or $\mathcal{S}^1 \times Y$ gives a local model for how to resolve the corresponding component of S in T^7/Γ. Thus we construct a nonsingular, compact 7-manifold M by using the patches $T^3 \times X$ or $\mathcal{S}^1 \times Y$ to repair the singularities of T^7/Γ. In the case of

Example 3.1, this means gluing 12 copies of $T^3 \times X$ into T^7/Γ, where X is the blow-up of $\mathbb{C}^2/\{\pm 1\}$ at its singular point.

Now the point of using crepant resolutions is this. In both case (a) and (b), there exists a Calabi–Yau metric on X or Y which is asymptotic to the flat Euclidean metric on $\mathbb{C}^2/G$ or $\mathbb{C}^3/G$. Such metrics are called *Asymptotically Locally Euclidean (ALE)*. In case (a), the ALE Calabi–Yau metrics were classified by Kronheimer [23, 24], and exist for all finite $G \subset \mathrm{SU}(2)$. In case (b), crepant resolutions of $\mathbb{C}^3/G$ exist for all finite $G \subset \mathrm{SU}(3)$ by Roan [29], and the author [19], [18, §8] proved that they carry ALE Calabi–Yau metrics, using a noncompact version of the Calabi Conjecture.

By Propositions 2.9 and 2.10, we can use the Calabi–Yau metrics on X or Y to construct a torsion-free G_2-structure on $T^3 \times X$ or $\mathcal{S}^1 \times Y$. This gives a local model for how to resolve the singularity $T^3 \times \mathbb{C}^2/G$ or $\mathcal{S}^1 \times \mathbb{C}^3/G$ with holonomy G_2. So, this method gives not only a way to smooth out the singularities of T^7/Γ as a manifold, but also a family of torsion-free G_2-structures on the resolution which show how to smooth out the singularities of the G_2-structure.

The requirement above that S be divided into connected components of the form (a) and (b) is in fact unnecessarily restrictive. There is a more complicated and powerful method, described in [18, §11–§12], for resolving singularities of a more general kind. We require only that the singularities should *locally* be of the form $\mathbb{R}^3 \times \mathbb{C}^2/G$ or $\mathbb{R} \times \mathbb{C}^3/G$, for a finite subgroup G of SU(2) or SU(3), and when $G \subset \mathrm{SU}(3)$ we do *not* require that G act freely on $\mathbb{C}^3 \setminus \{0\}$.

If X is a crepant resolution of $\mathbb{C}^3/G$, where G does not act freely on $\mathbb{C}^3 \setminus \{0\}$, then the author shows [18, §9], [20] that X carries a family of Calabi–Yau metrics satisfying a complicated asymptotic condition at infinity, called *Quasi-ALE* metrics. These yield the local models necessary to resolve singularities locally of the form $\mathbb{R} \times \mathbb{C}^3/G$ with holonomy G_2. Using this method we can resolve many orbifolds T^7/Γ, and prove the existence of large numbers of compact 7-manifolds with holonomy G_2.

3.3. Step 3: Finding G_2-structures with small torsion

For each resolution X of $\mathbb{C}^2/G$ in case (a), and Y of $\mathbb{C}^3/G$ in case (b) above, we can find a 1-parameter family $\{h_t : t > 0\}$ of metrics with the properties

(a) h_t is a Kähler metric on X with $\mathrm{Hol}(h_t) = \mathrm{SU}(2)$. Its injectivity radius satisfies $\delta(h_t) = O(t)$, its Riemann curvature satisfies $\|R(h_t)\|_{C^0} = O(t^{-2})$, and $h_t = h + O(t^4 r^{-4})$ for large r, where h is the Euclidean metric on $\mathbb{C}^2/G$, and r the distance from the origin.

(b) h_t is Kähler on Y with $\mathrm{Hol}(h_t) = \mathrm{SU}(3)$, where $\delta(h_t) = O(t)$, $\|R(h_t)\|_{C^0} = O(t^{-2})$, and $h_t = h + O(t^6 r^{-6})$ for large r.

In fact we can choose h_t to be isometric to $t^2 h_1$, and then (a), (b) are easy to prove.

Suppose one of the components of the singular set S of T^7/Γ is locally modelled on $T^3 \times \mathbb{C}^2/G$. Then T^3 has a natural flat metric h_{T^3}. Let X be the crepant resolution of $\mathbb{C}^2/G$ and let $\{h_t : t > 0\}$ satisfy property (a). Then Proposition 2.10 gives a 1-parameter family of torsion-free G_2-structures $(\hat{\varphi}_t, \hat{g}_t)$ on $T^3 \times X$ with $\hat{g}_t = h_{T^3} + h_t$. Similarly, if a component of S is modelled on $\mathcal{S}^1 \times \mathbb{C}^3/G$, using Proposition 2.9 we get a family of torsion-free G_2-structures $(\hat{\varphi}_t, \hat{g}_t)$ on $\mathcal{S}^1 \times Y$.

The idea is to make a G_2-structure (φ_t, g_t) on M by gluing together the torsion-free G_2-structures $(\hat{\varphi}_t, \hat{g}_t)$ on the patches $T^3 \times X$ and $\mathcal{S}^1 \times Y$, and (φ_0, g_0) on T^7/Γ. The gluing is done using a partition of unity. Naturally, the first derivative of the partition of unity introduces 'errors', so that (φ_t, g_t) is not torsion-free. The size of the torsion $\nabla\varphi_t$ depends on the difference $\hat{\varphi}_t - \varphi_0$ in the region where the partition of unity changes. On the patches $T^3 \times X$, since $h_t - h = O(t^4 r^{-4})$ and the partition of unity has nonzero derivative when $r = O(1)$, we find that $\nabla\varphi_t = O(t^4)$. Similarly $\nabla\varphi_t = O(t^6)$ on the patches $\mathcal{S}^1 \times Y$, and so $\nabla\varphi_t = O(t^4)$ on M.

For small t, the dominant contributions to the injectivity radius $\delta(g_t)$ and Riemann curvature $R(g_t)$ are made by those of the metrics h_t on X and Y, so we expect $\delta(g_t) = O(t)$ and $\|R(g_t)\|_{C^0} = O(t^{-2})$ by properties (a) and (b) above. In this way we prove the following result [18, Th. 11.5.7], which gives the estimates on (φ_t, g_t) that we need.

Theorem 3.3. *On the compact 7-manifold M described above, and on many other 7-manifolds constructed in a similar fashion, one can write down the following data explicitly in coordinates:*

- *Positive constants A_1, A_2, A_3 and ϵ,*
- *A G_2-structure (φ_t, g_t) on M with $\mathrm{d}\varphi_t = 0$ for each $t \in (0, \epsilon)$, and*
- *A 3-form ψ_t on M with $\mathrm{d}^*\psi_t = \mathrm{d}^*\varphi_t$ for each $t \in (0, \epsilon)$.*

These satisfy three conditions:

(i) $\|\psi_t\|_{L^2} \leqslant A_1 t^4$, $\|\psi_t\|_{C^0} \leqslant A_1 t^3$ *and* $\|\mathrm{d}^*\psi_t\|_{L^{14}} \leqslant A_1 t^{16/7}$,
(ii) *the injectivity radius $\delta(g_t)$ satisfies $\delta(g_t) \geqslant A_2 t$,*
(iii) *the Riemann curvature $R(g_t)$ of g_t satisfies $\|R(g_t)\|_{C^0} \leqslant A_3 t^{-2}$.*

Here the operator d^* and the norms $\|\,.\,\|_{L^2}$, $\|\,.\,\|_{L^{14}}$ and $\|\,.\,\|_{C^0}$ depend on g_t.

Here one should regard ψ_t as a *first integral* of the torsion $\nabla\varphi_t$ of (φ_t, g_t). Thus the norms $\|\psi_t\|_{L^2}$, $\|\psi_t\|_{C^0}$ and $\|\mathrm{d}^*\psi_t\|_{L^{14}}$ are measures of $\nabla\varphi_t$. So parts (i)–(iii) say that $\nabla\varphi_t$ is small compared to the injectivity radius and Riemann curvature of (M, g_t).

3.4. Step 4: Deforming to a torsion-free G_2-structure

We prove the following analysis result.

Theorem 3.4. *Let A_1, A_2, A_3 be positive constants. Then there exist positive constants κ, K such that whenever $0 < t \leqslant \kappa$, the following is true.*

Let M be a compact 7-manifold, and (φ, g) a G_2-structure on M with $\mathrm{d}\varphi = 0$. Suppose ψ is a smooth 3-form on M with $\mathrm{d}^*\psi = \mathrm{d}^*\varphi$, and

(i) $\|\psi\|_{L^2} \leqslant A_1 t^4$, $\|\psi\|_{C^0} \leqslant A_1 t^{1/2}$ *and* $\|\mathrm{d}^*\psi\|_{L^{14}} \leqslant A_1$,
(ii) *the injectivity radius $\delta(g)$ satisfies $\delta(g) \geqslant A_2 t$, and*
(iii) *the Riemann curvature $R(g)$ satisfies $\|R(g)\|_{C^0} \leqslant A_3 t^{-2}$.*

Then there exists a smooth, torsion-free G_2-structure $(\tilde\varphi, \tilde g)$ on M with $\|\tilde\varphi - \varphi\|_{C^0} \leqslant Kt^{1/2}$.

Basically, this result says that if (φ, g) is a G_2-structure on M, and the torsion $\nabla\varphi$ is sufficiently small, then we can deform (φ, g) to a nearby G_2-structure $(\tilde\varphi, \tilde g)$ that is torsion-free. Here is a sketch of the proof of Theorem 3.4, ignoring several technical points. The detailed proof is given in [18, §11.6–§11.8], which is an improved version of the proof in [14].

We have a 3-form φ with $\mathrm{d}\varphi = 0$ and $\mathrm{d}^*\varphi = \mathrm{d}^*\psi$ for small ψ, and we wish to construct a nearby 3-form $\tilde\varphi$ with $\mathrm{d}\tilde\varphi = 0$ and $\tilde{\mathrm{d}}^*\tilde\varphi = 0$. Set $\tilde\varphi = \varphi + \mathrm{d}\eta$, where η is a small 2-form. Then η must satisfy a nonlinear p.d.e., which we write as

$$\mathrm{d}^*\mathrm{d}\eta = -\mathrm{d}^*\psi + \mathrm{d}^*F(\mathrm{d}\eta), \tag{9}$$

where F is nonlinear, satisfying $F(\mathrm{d}\eta) = O\big(|\mathrm{d}\eta|^2\big)$.

We solve (9) by iteration, introducing a sequence $\{\eta_j\}_{j=0}^\infty$ with $\eta_0 = 0$, satisfying the inductive equations

$$\mathrm{d}^*\mathrm{d}\eta_{j+1} = -\mathrm{d}^*\psi + \mathrm{d}^*F(\mathrm{d}\eta_j), \qquad \mathrm{d}^*\eta_{j+1} = 0. \tag{10}$$

If such a sequence exists and converges to η, then taking the limit in (10) shows that η satisfies (9), giving us the solution we want.

The key to proving this is an *inductive estimate* on the sequence $\{\eta_j\}_{j=0}^\infty$. The inductive estimate we use has three ingredients, the equations

$$\|\mathrm{d}\eta_{j+1}\|_{L^2} \leqslant \|\psi\|_{L^2} + C_1\|\mathrm{d}\eta_j\|_{L^2}\|\mathrm{d}\eta_j\|_{C^0}, \tag{11}$$

$$\|\nabla\mathrm{d}\eta_{j+1}\|_{L^{14}} \leqslant C_2\big(\|\mathrm{d}^*\psi\|_{L^{14}} + \|\nabla\mathrm{d}\eta_j\|_{L^{14}}\|\mathrm{d}\eta_j\|_{C^0} + t^{-4}\|\mathrm{d}\eta_{j+1}\|_{L^2}\big), \tag{12}$$

$$\|\mathrm{d}\eta_j\|_{C^0} \leqslant C_3\big(t^{1/2}\|\nabla\mathrm{d}\eta_j\|_{L^{14}} + t^{-7/2}\|\mathrm{d}\eta_j\|_{L^2}\big). \tag{13}$$

Here C_1, C_2, C_3 are positive constants independent of t. Equation (11) is obtained from (10) by taking the L^2-inner product with η_{j+1} and integrating by parts. Using the fact that $\mathrm{d}^*\varphi = \mathrm{d}^*\psi$ and $\|\psi\|_{L^2} = O(t^4)$, $|\psi| = O(t^{1/2})$ we get a powerful estimate of the L^2-norm of $\mathrm{d}\eta_{j+1}$.

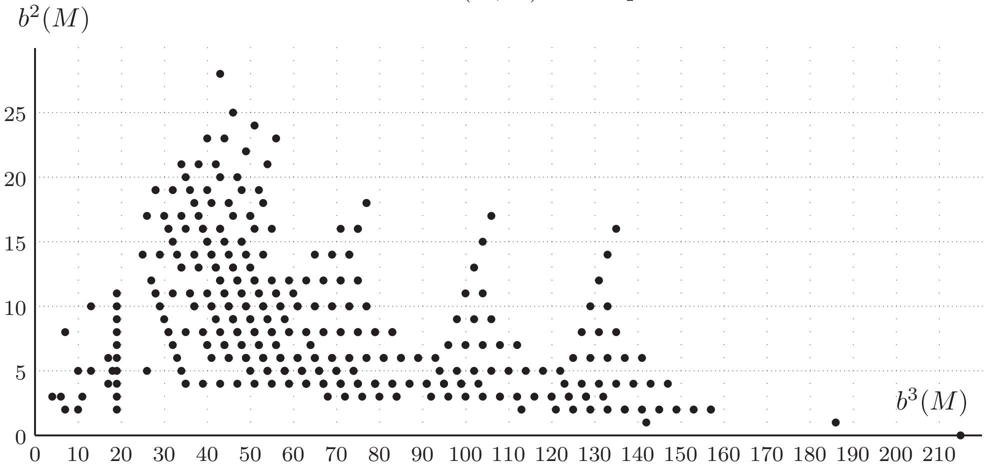
Equation (12) is derived from an *elliptic regularity estimate* for the operator $\mathrm{d} + \mathrm{d}^*$ acting on 3-forms on M. Equation (13) follows from the *Sobolev embedding theorem*, since $L^{14}_1(M) \hookrightarrow C^0(M)$. Both (12) and (13) are proved on small balls of radius $O(t)$ in M, using parts (ii) and (iii) of Theorem 3.3, and this is where the powers of t come from.

Using (11)-(13) and part (i) of Theorem 3.3 we show that if

$$\|\mathrm{d}\eta_j\|_{L^2} \leqslant C_4 t^4, \ \ \|\nabla\mathrm{d}\eta_j\|_{L^{14}} \leqslant C_5, \ \text{ and } \ \|\mathrm{d}\eta_j\|_{C^0} \leqslant Kt^{1/2}, \tag{14}$$

where C_4, C_5 and K are positive constants depending on C_1, C_2, C_3 and A_1, and if t is sufficiently small, then the same inequalities (14) apply to $\mathrm{d}\eta_{j+1}$. Since $\eta_0 = 0$, by induction (14) applies for all j and the sequence $\{\mathrm{d}\eta_j\}_{j=0}^\infty$ is bounded in the Banach space $L^{14}_1(\Lambda^3 T^*M)$. One can then use standard techniques in analysis to prove that this sequence converges to a smooth limit $\mathrm{d}\eta$. This concludes the proof of Theorem 3.4.

Figure 1. Betti numbers (b^2, b^3) of compact G_2-manifolds



¿From Theorems 3.3 and 3.4 we see that the compact 7-manifold M constructed in Step 2 admits torsion-free G_2-structures $(\tilde{\varphi}, \tilde{g})$. Theorem 2.5 then shows that $\mathrm{Hol}(\tilde{g}) = G_2$ if and only if $\pi_1(M)$ is finite. In the example above M is simply-connected, and so $\pi_1(M) = \{1\}$ and M has metrics with holonomy G_2, as we want.

By considering different groups Γ acting on T^7, and also by finding topologically distinct resolutions $M_1, \ldots, M_k$ of the same orbifold T^7/Γ, we can construct many compact Riemannian 7-manifolds with holonomy G_2. A good number of examples are given in [18, §12]. Figure 1 displays the Betti numbers of compact, simply-connected 7-manifolds with holonomy G_2 constructed there. There are 252 different sets of Betti numbers.

Examples are also known [18, §12.4] of compact 7-manifolds with holonomy G_2 with finite, nontrivial fundamental group. It seems likely to the author that the Betti numbers given in Figure 1 are only a small proportion of the Betti numbers of all compact, simply-connected 7-manifolds with holonomy G_2.

3.5. Other constructions of compact G_2-manifolds

Here are two other methods, taken from [18, §11.9], of constructing compact 7-manifolds with holonomy G_2. The first was outlined by the author in [15, §4.3].

Method 1. Let (Y, J, h) be a Calabi–Yau 3-fold, with Kähler form ω and holomorphic volume form θ. Suppose $\sigma : Y \to Y$ is an involution, satisfying $\sigma^*(h) = h$, $\sigma^*(J) = -J$ and $\sigma^*(\theta) = \bar{\theta}$. We call σ a *real structure* on Y. Let N be the fixed point set of σ in Y. Then N is a real 3-dimensional submanifold of Y, and is in fact a special Lagrangian 3-fold.

Let $\mathcal{S}^1 = \mathbb{R}/\mathbb{Z}$, and define a torsion-free G_2-structure (φ, g) on $\mathcal{S}^1 \times Y$ as in Proposition 2.9. Then $\varphi = \mathrm{d}x \wedge \omega + \mathrm{Re}\,\theta$, where $x \in \mathbb{R}/\mathbb{Z}$ is the coordinate on $\mathcal{S}^1$. Define $\hat{\sigma} : \mathcal{S}^1 \times Y \to$

$\mathcal{S}^1 \times Y$ by $\hat{\sigma}\big((x,y)\big) = \big(-x, \sigma(y)\big)$. Then $\hat{\sigma}$ preserves (φ, g) and $\hat{\sigma}^2 = 1$. The fixed points of $\hat{\sigma}$ in $\mathcal{S}^1 \times Y$ are $\{\mathbb{Z}, \frac{1}{2} + \mathbb{Z}\} \times N$. Thus $(\mathcal{S}^1 \times Y)/\langle\hat{\sigma}\rangle$ is an orbifold. Its singular set is 2 copies of N, and each singular point is modelled on $\mathbb{R}^3 \times \mathbb{R}^4/\{\pm 1\}$.

We aim to resolve $(\mathcal{S}^1 \times Y)/\langle\hat{\sigma}\rangle$ to get a compact 7-manifold M with holonomy G_2. Locally, each singular point should be resolved like $\mathbb{R}^3 \times X$, where X is an ALE Calabi–Yau 2-fold asymptotic to $\mathbb{C}^2/\{\pm 1\}$. There is a 3-dimensional family of such X, and we need to choose one member of this family for each singular point in the singular set.

Calculations by the author indicate that the data needed to do this is a closed, coclosed 1-form α on N that is nonzero at every point of N. The existence of a suitable 1-form α depends on the metric on N, which is the restriction of the metric g on Y. But g comes from the solution of the Calabi Conjecture, so we know little about it. This may make the method difficult to apply in practice.

The second method has been successfully applied by Kovalev [22], and is based on an idea due to Simon Donaldson.

Method 2. Let X be a projective complex 3-fold with canonical bundle K_X, and s a holomorphic section of K_X^{-1} which vanishes to order 1 on a smooth divisor D in X. Then D has trivial canonical bundle, so D is T^4 or $K3$. Suppose D is a $K3$ surface. Define $Y = X \setminus D$, and suppose Y is simply-connected.

Then Y is a noncompact complex 3-fold with K_Y trivial, and one infinite end modelled on $D \times \mathcal{S}^1 \times [0,\infty)$. Using a version of the proof of the Calabi Conjecture for noncompact manifolds one constructs a complete Calabi–Yau metric h on Y, which is asymptotic to the product on $D \times \mathcal{S}^1 \times [0,\infty)$ of a Calabi–Yau metric on D, and Euclidean metrics on $\mathcal{S}^1$ and $[0,\infty)$. We call such metrics *Asymptotically Cylindrical*.

Suppose we have such a metric on Y. Define a torsion-free G_2-structure (φ, g) on $\mathcal{S}^1 \times Y$ as in Proposition 2.9. Then $\mathcal{S}^1 \times Y$ is a noncompact G_2-manifold with one end modelled on $D \times T^2 \times [0,\infty)$, whose metric is asymptotic to the product on $D \times T^2 \times [0,\infty)$ of a Calabi–Yau metric on D, and Euclidean metrics on T^2 and $[0,\infty)$.

Donaldson and Kovalev's idea is to take two such products $\mathcal{S}^1 \times Y_1$ and $\mathcal{S}^1 \times Y_2$ whose infinite ends are isomorphic in a suitable way, and glue them together to get a compact 7-manifold M with holonomy G_2. The gluing process swaps round the $\mathcal{S}^1$ factors. That is, the $\mathcal{S}^1$ factor in $\mathcal{S}^1 \times Y_1$ is identified with the asymptotic $\mathcal{S}^1$ factor in $Y_2 \sim D_2 \times \mathcal{S}^1 \times [0,\infty)$, and vice versa.

4. Compact Spin(7)-manifolds from Calabi–Yau 4-orbifolds

In a very similar way to the G_2 case, one can construct examples of compact 8-manifolds with holonomy Spin(7) by resolving the singularities of torus orbifolds T^8/Γ. This is done in [16] and [18, §13–§14]. In [18, §14], examples are constructed which realize 181 different sets of Betti numbers. Two compact 8-manifolds with holonomy Spin(7) and the same Betti numbers may be distinguished by the cup products on their cohomologies (examples of this are given in [16, §3.4]), so they probably represent rather more than 181 topologically distinct 8-manifolds.

The main differences with the G_2 case are, firstly, that the technical details of the analysis are different and harder, and secondly, that the singularities that arise are typically more complicated and more tricky to resolve. One reason for this is that in the G_2 case the singular set is made up of 1 and 3-dimensional pieces in a 7-dimensional space, so one can often arrange for the pieces to avoid each other, and resolve them independently.

But in the Spin(7) case the singular set is typically made up of 4-dimensional pieces in an 8-dimensional space, so they nearly always intersect. There are also topological constraints arising from the $\hat{A}$-genus, which do not apply in the G_2 case. The moral appears to be that when you increase the dimension, things become more difficult.

Anyway, we will not discuss this further, as the principles are very similar to the G_2 case above. Instead, we will discuss an entirely different construction of compact 8-manifolds with holonomy Spin(7) developed by the author in [17] and [18, §15], a little like Method 1 of §3.5. In this we start from a *Calabi–Yau 4-orbifold* rather than from T^8. The construction can be divided into five steps.

Step 1. Find a compact, complex 4-orbifold (Y, J) satisfying the conditions:

(a) Y has only finitely many singular points $p_1, \ldots, p_k$, for $k \geqslant 1$.

(b) Y is modelled on $\mathbb{C}^4/\langle i \rangle$ near each p_j, where i acts on $\mathbb{C}^4$ by complex multiplication.

(c) There exists an antiholomorphic involution $\sigma : Y \to Y$ whose fixed point set is $\{p_1, \ldots, p_k\}$.

(d) $Y \setminus \{p_1, \ldots, p_k\}$ is simply-connected, and $h^{2,0}(Y) = 0$.

Step 2. Choose a σ-invariant Kähler class on Y. Then by Theorem 2.8 there exists a unique σ-invariant Ricci-flat Kähler metric g in this Kähler class. Let ω be the Kähler form of g. Let θ be a holomorphic volume form for (Y, J, g). By multiplying θ by $e^{i\phi}$ if necessary, we can arrange that $\sigma^*(\theta) = \bar{\theta}$.

Define $\Omega = \frac{1}{2}\omega \wedge \omega + \operatorname{Re}\theta$. Then, by Proposition 2.11, (Ω, g) is a torsion-free Spin(7)-structure on Y. Also, (Ω, g) is σ-invariant, as $\sigma^*(\omega) = -\omega$ and $\sigma^*(\theta) = \bar{\theta}$. Define $Z = Y/\langle\sigma\rangle$. Then Z is a compact real 8-orbifold with isolated singular points $p_1, \ldots, p_k$, and (Ω, g) pushes down to a torsion-free Spin(7)-structure (Ω, g) on Z.

Step 3. Z is modelled on $\mathbb{R}^8/G$ near each p_j, where G is a certain finite subgroup of Spin(7) with $|G| = 8$. We can write down two explicit, topologically distinct ALE Spin(7)-manifolds X_1, X_2 asymptotic to $\mathbb{R}^8/G$. Each carries a 1-parameter family of homothetic ALE metrics h_t for $t > 0$ with $\operatorname{Hol}(h_t) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4) \subset$ Spin(7).

For $j = 1, \ldots, k$ we choose $i_j = 1$ or 2, and resolve the singularities of Z by gluing in X_{i_j} at the singular point p_j for $j = 1, \ldots, k$, to get a compact, nonsingular 8-manifold M, with projection $\pi : M \to Z$.

Step 4. On M, we explicitly write down a 1-parameter family of Spin(7)-structures (Ω_t, g_t) depending on $t \in (0, \epsilon)$. They are not torsion-free, but have small torsion when t is small. As $t \to 0$, the Spin(7)-structure (Ω_t, g_t) converges to the singular Spin(7)-structure $\pi^*(\Omega_0, g_0)$.

Step 5. We prove using analysis that for sufficiently small t, the Spin(7)-structure (Ω_t, g_t) on M, with small torsion, can be deformed to a Spin(7)-structure $(\tilde{\Omega}_t, \tilde{g}_t)$, with zero torsion.

It turns out that if $i_j = 1$ for $j = 1, \ldots, k$ we have $\pi_1(M) \cong \mathbb{Z}_2$ and $\mathrm{Hol}(\tilde{g}_t) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4)$, and for the other $2^k - 1$ choices of $i_1, \ldots, i_k$ we have $\pi_1(M) = \{1\}$ and $\mathrm{Hol}(\tilde{g}_t) = \mathrm{Spin}(7)$. So $\tilde{g}_t$ is a metric with holonomy Spin(7) on the compact 8-manifold M for $(i_1, \ldots, i_k) \neq (1, \ldots, 1)$.

Once we have completed Step 1, Step 2 is immediate. Steps 4 and 5 are analogous to Steps 3 and 4 of §3, and can be done using the techniques and analytic results developed by the author for the first T^8/Γ construction of compact Spin(7)-manifolds, [16], [18, §13]. So the really new material is in Steps 1 and 3, and we will discuss only these.

4.1. Step 1: An example

We do Step 1 using complex algebraic geometry. The problem is that conditions (a)–(d) above are very restrictive, so it is not that easy to find *any* Y satisfying all four conditions. All the examples Y the author has found are constructed using *weighted projective spaces*, an important class of complex orbifolds.

Definition 4.1. Let $m \geqslant 1$ be an integer, and $a_0, a_1, \ldots, a_m$ positive integers with highest common factor 1. Let $\mathbb{C}^{m+1}$ have complex coordinates $(z_0, \ldots, z_m)$, and define an action of the complex Lie group $\mathbb{C}^*$ on $\mathbb{C}^{m+1}$ by

$$(z_0, \ldots, z_m) \overset{u}{\longmapsto} (u^{a_0} z_0, \ldots, u^{a_m} z_m), \qquad \text{for } u \in \mathbb{C}^*.$$

The *weighted projective space* $\mathbb{CP}^m_{a_0,\ldots,a_m}$ is $\bigl(\mathbb{C}^{m+1} \setminus \{0\}\bigr)/\mathbb{C}^*$. The $\mathbb{C}^*$-orbit of $(z_0, \ldots, z_m)$ is written $[z_0, \ldots, z_m]$.

Here is the simplest example the author knows.

Example 4.2. Let Y be the hypersurface of degree 12 in $\mathbb{CP}^5_{1,1,1,1,4,4}$ given by

$$Y = \bigl\{[z_0, \ldots, z_5] \in \mathbb{CP}^5_{1,1,1,1,4,4} : z_0^{12} + z_1^{12} + z_2^{12} + z_3^{12} + z_4^3 + z_5^3 = 0\bigr\}.$$

Calculation shows that Y has trivial canonical bundle and three singular points $p_1 = [0, 0, 0, 0, 1, -1]$, $p_2 = [0, 0, 0, 0, 1, e^{\pi i/3}]$ and $p_3 = [0, 0, 0, 0, 1, e^{-\pi i/3}]$, modelled on $\mathbb{C}^4/\langle i \rangle$.

Now define a map $\sigma : Y \to Y$ by

$$\sigma : [z_0, \ldots, z_5] \longmapsto [\bar{z}_1, -\bar{z}_0, \bar{z}_3, -\bar{z}_2, \bar{z}_5, \bar{z}_4].$$

Note that $\sigma^2 = 1$, though this is not immediately obvious, because of the geometry of $\mathbb{CP}^5_{1,1,1,1,4,4}$. It can be shown that conditions (a)–(d) of Step 1 above hold for Y and σ.

More suitable 4-folds Y may be found by taking hypersurfaces or complete intersections in other weighted projective spaces, possibly also dividing by a finite group, and then doing a crepant resolution to get rid of any singularities that we don't want. Examples are given in [17], [18, §15].

4.2. Step 3: Resolving $\mathbb{R}^8/G$

Define $\alpha, \beta : \mathbb{R}^8 \to \mathbb{R}^8$ by

$$\alpha : (x_1, \dots, x_8) \mapsto (-x_2, x_1, -x_4, x_3, -x_6, x_5, -x_8, x_7),$$
$$\beta : (x_1, \dots, x_8) \mapsto (x_3, -x_4, -x_1, x_2, x_7, -x_8, -x_5, x_6).$$

Then α, β preserve Ω_0 given in (2), so they lie in Spin(7). Also $\alpha^4 = \beta^4 = 1$, $\alpha^2 = \beta^2$ and $\alpha\beta = \beta\alpha^3$. Let $G = \langle \alpha, \beta \rangle$. Then G is a finite nonabelian subgroup of Spin(7) of order 8, which acts freely on $\mathbb{R}^8 \setminus \{0\}$. One can show that if Z is the compact Spin(7)-orbifold constructed in Step 2 above, then $T_{p_j}Z$ is isomorphic to $\mathbb{R}^8/G$ for $j = 1, \dots, k$, with an isomorphism identifying the Spin(7)-structures (Ω, g) on Z and (Ω_0, g_0) on $\mathbb{R}^8/G$, such that β corresponds to the σ-action on Y.

In the next two examples we shall construct two different ALE Spin(7)-manifolds (X_1, Ω_1, g_1) and (X_2, Ω_2, g_2) asymptotic to $\mathbb{R}^8/G$.

Example 4.3. Define complex coordinates $(z_1, \dots, z_4)$ on $\mathbb{R}^8$ by

$$(z_1, z_2, z_3, z_4) = (x_1 + ix_2, x_3 + ix_4, x_5 + ix_6, x_7 + ix_8),$$

Then $g_0 = |\mathrm{d}z_1|^2 + \cdots + |\mathrm{d}z_4|^2$, and $\Omega_0 = \frac{1}{2}\omega_0 \wedge \omega_0 + \mathrm{Re}(\theta_0)$, where ω_0 and θ_0 are the usual Kähler form and complex volume form on $\mathbb{C}^4$. In these coordinates, α and β are given by

$$\begin{aligned} \alpha &: (z_1, \dots, z_4) \mapsto (iz_1, iz_2, iz_3, iz_4), \\ \beta &: (z_1, \dots, z_4) \mapsto (\bar{z}_2, -\bar{z}_1, \bar{z}_4, -\bar{z}_3). \end{aligned} \tag{15}$$

Now $\mathbb{C}^4/\langle \alpha \rangle$ is a complex singularity, as $\alpha \in \mathrm{SU}(4)$. Let (Y_1, π_1) be the blow-up of $\mathbb{C}^4/\langle \alpha \rangle$ at 0. Then Y_1 is the unique crepant resolution of $\mathbb{C}^4/\langle \alpha \rangle$. The action of β on $\mathbb{C}^4/\langle \alpha \rangle$ lifts to a *free* antiholomorphic map $\beta : Y_1 \to Y_1$ with $\beta^2 = 1$. Define $X_1 = Y_1/\langle \beta \rangle$. Then X_1 is a nonsingular 8-manifold, and the projection $\pi_1 : Y_1 \to \mathbb{C}^4/\langle \alpha \rangle$ pushes down to $\pi_1 : X_1 \to \mathbb{R}^8/G$.

There exist ALE Calabi–Yau metrics g_1 on Y_1, which were written down explicitly by Calabi [9, p. 285], and are invariant under the action of β on Y_1. Let ω_1 be the Kähler form of g_1, and $\theta_1 = \pi_1^*(\theta_0)$ the holomorphic volume form on Y_1. Define $\Omega_1 = \frac{1}{2}\omega_1 \wedge \omega_1 + \mathrm{Re}(\theta_1)$. Then (Ω_1, g_1) is a torsion-free Spin(7)-structure on Y_1, as in Proposition 2.11.

As $\beta^*(\omega_1) = -\omega_1$ and $\beta^*(\theta_1) = \bar{\theta}_1$, we see that β preserves (Ω_1, g_1). Thus (Ω_1, g_1) pushes down to a torsion-free Spin(7)-structure (Ω_1, g_1) on X_1. Then (X_1, Ω_1, g_1) is an *ALE* Spin(7)*-manifold* asymptotic to $\mathbb{R}^8/G$.

Example 4.4. Define new complex coordinates $(w_1, \dots, w_4)$ on $\mathbb{R}^8$ by

$$(w_1, w_2, w_3, w_4) = (-x_1 + ix_3, x_2 + ix_4, -x_5 + ix_7, x_6 + ix_8).$$

Again we find that $g_0 = |\mathrm{d}w_1|^2 + \cdots + |\mathrm{d}w_4|^2$ and $\Omega_0 = \frac{1}{2}\omega_0 \wedge \omega_0 + \mathrm{Re}(\theta_0)$. In these coordinates, α and β are given by

$$\begin{aligned} \alpha &: (w_1, \dots, w_4) \mapsto (\bar{w}_2, -\bar{w}_1, \bar{w}_4, -\bar{w}_3), \\ \beta &: (w_1, \dots, w_4) \mapsto (iw_1, iw_2, iw_3, iw_4). \end{aligned} \tag{16}$$

Observe that (15) and (16) are the same, except that the rôles of α, β are reversed. Therefore we can use the ideas of Example 4.3 again.

Let Y_2 be the crepant resolution of $\mathbb{C}^4/\langle\beta\rangle$. The action of α on $\mathbb{C}^4/\langle\beta\rangle$ lifts to a free antiholomorphic involution of Y_2. Let $X_2 = Y_2/\langle\alpha\rangle$. Then X_2 is nonsingular, and carries a torsion-free Spin(7)-structure (Ω_2, g_2), making (X_2, Ω_2, g_2) into an ALE Spin(7)-manifold asymptotic to $\mathbb{R}^8/G$.

We can now explain the remarks on holonomy groups at the end of Step 5. The holonomy groups $\mathrm{Hol}(g_i)$ of the metrics g_1, g_2 in Examples 4.3 and 4.4 are both isomorphic to $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$, a subgroup of Spin(7). However, they are two *different* inclusions of $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$ in Spin(7), as in the first case the complex structure is α and in the second β.

The Spin(7)-structure (Ω, g) on Z also has holonomy $\mathrm{Hol}(g) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4)$. Under the natural identifications we have $\mathrm{Hol}(g_1) = \mathrm{Hol}(g)$ but $\mathrm{Hol}(g_2) \neq \mathrm{Hol}(g)$ as subgroups of Spin(7). Therefore, if we choose $i_j = 1$ for all $j = 1, \ldots, k$, then Z and X_{i_j} all have the same holonomy group $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$, so they combine to give metrics $\tilde{g}_t$ on M with $\mathrm{Hol}(\tilde{g}_t) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4)$.

However, if $i_j = 2$ for some j then the holonomy of g on Z and g_{i_j} on X_{i_j} are *different* $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$ subgroups of Spin(7), which together generate the whole group Spin(7). Thus they combine to give metrics $\tilde{g}_t$ on M with $\mathrm{Hol}(\tilde{g}_t) = \mathrm{Spin}(7)$.

4.3. Conclusions

The author was able in [17] and [18, Ch. 15] to construct compact 8-manifolds with holonomy Spin(7) realizing 14 distinct sets of Betti numbers, which are given in Table 1. Probably there are many other examples which can be produced by similar methods.

Table 1. Betti numbers (b^2, b^3, b^4) of compact Spin(7)-manifolds

(4, 33, 200)	(3, 33, 202)	(2, 33, 204)	(1, 33, 206)	(0, 33, 208)
(1, 0, 908)	(0, 0, 910)	(1, 0, 1292)	(0, 0, 1294)	(1, 0, 2444)
(0, 0, 2446)	(0, 6, 3730)	(0, 0, 4750)	(0, 0, 11 662)	

Comparing these Betti numbers with those of the compact 8-manifolds constructed in [18, Ch. 14] by resolving torus orbifolds T^8/Γ, we see that these examples the middle Betti number b^4 is much bigger, as much as 11 662 in one case.

Given that the two constructions of compact 8-manifolds with holonomy Spin(7) that we know appear to produce sets of 8-manifolds with rather different 'geography', it is tempting to speculate that the set of all compact 8-manifolds with holonomy Spin(7) may be rather large, and that those constructed so far are a small sample with atypical behaviour.

Part II. Calibrated Geometry

5. Introduction to calibrated geometry

Calibrated geometry was introduced in the seminal paper of Harvey and Lawson [12]. We introduce the basic ideas in §5.1–§5.2, and then discuss the G_2 calibrations in more detail in §5.3–§5.5, and the Spin(7) calibration in §5.6.

5.1. Calibrations and calibrated submanifolds

We begin by defining *calibrations* and *calibrated submanifolds*, following Harvey and Lawson [12].

Definition 5.1. 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* [12, Th. II.4.2]. We prove this in the compact case, but noncompact calibrated submanifolds are locally volume-minimizing as well.

Proposition 5.2. *Let (M, g) be a Riemannian manifold, φ a calibration on M, and N a compact φ-submanifold in M. Then N is volume-minimizing in its homology class.*

Proof. Let $\dim N = k$, and let $[N] \in H_k(M, \mathbb{R})$ and $[\varphi] \in H^k(M, \mathbb{R})$ be the homology and cohomology classes of N and φ. Then

$$[\varphi] \cdot [N] = \int_{x \in N} \varphi\big|_{T_xN} = \int_{x \in N} \mathrm{vol}_{T_xN} = \mathrm{Vol}(N),$$

since $\varphi|_{T_xN} = \mathrm{vol}_{T_xN}$ for each $x \in N$, as N is a calibrated submanifold. If N' is any other compact k-submanifold of M with $[N'] = [N]$ in $H_k(M, \mathbb{R})$, then

$$[\varphi] \cdot [N] = [\varphi] \cdot [N'] = \int_{x \in N'} \varphi\big|_{T_xN'} \leqslant \int_{x \in N'} \mathrm{vol}_{T_xN'} = \mathrm{Vol}(N'),$$

since $\varphi|_{T_xN'} \leqslant \mathrm{vol}_{T_xN'}$ because φ is a calibration. The last two equations give $\mathrm{Vol}(N) \leqslant \mathrm{Vol}(N')$. Thus N is volume-minimizing in its homology class. □

Now let (M, g) be a Riemannian manifold with a calibration φ, and let $\iota : N \to M$ be an immersed submanifold. Whether N is a φ-submanifold depends upon the tangent spaces of N. That is, it depends on ι and its first derivative. So, to be calibrated with respect to φ is a *first-order* partial differential equation on ι. But if N is calibrated then N is minimal, and to be minimal is a *second-order* partial differential equation on ι.

One moral is that the calibrated equations, being first-order, are often easier to solve than the minimal submanifold equations, which are second-order. So calibrated geometry is a fertile source of examples of minimal submanifolds.

5.2. Calibrated submanifolds and special holonomy

Next we explain the connection with Riemannian holonomy. Let $G \subset \mathrm{O}(n)$ be a possible holonomy group of a Riemannian metric. In particular, we can take G to be one of the holonomy groups $\mathrm{U}(m)$, $\mathrm{SU}(m)$, $\mathrm{Sp}(m)$, G_2 or $\mathrm{Spin}(7)$ from Berger's classification. Then G acts on the k-forms $\Lambda^k(\mathbb{R}^n)^*$ on $\mathbb{R}^n$, so we can look for G-invariant k-forms on $\mathbb{R}^n$.

Suppose φ_0 is a nonzero, G-invariant k-form on $\mathbb{R}^n$. By rescaling φ_0 we can arrange that for each oriented k-plane $U \subset \mathbb{R}^n$ we have $\varphi_0|_U \leqslant \mathrm{vol}_U$, and that $\varphi_0|_U = \mathrm{vol}_U$ for at least one such U. Then $\varphi_0|_{\gamma\cdot U} = \mathrm{vol}_{\gamma\cdot U}$ by G-invariance, so $\gamma \cdot U$ is a calibrated k-plane for all $\gamma \in G$. Thus the family of φ_0-calibrated k-planes in $\mathbb{R}^n$ is reasonably large, and it is likely the calibrated submanifolds will have an interesting geometry.

Now let M be a manifold of dimension n, and g a metric on M with Levi-Civita connection ∇ and holonomy group G. Then by Theorem 2.2 there is a k-form φ on M with $\nabla\varphi = 0$, corresponding to φ_0. Hence $\mathrm{d}\varphi = 0$, and φ is closed. Also, the condition $\varphi_0|_U \leqslant \mathrm{vol}_U$ for all oriented k-planes U in $\mathbb{R}^n$ implies that $\varphi|_V \leqslant \mathrm{vol}_V$ for all oriented tangent k-planes V in M. Thus φ is a *calibration* on M.

This gives us a general method for finding interesting calibrations on manifolds with reduced holonomy. Here are the most significant examples of this.

- Let $G = \mathrm{U}(m) \subset \mathrm{O}(2m)$. Then G preserves a 2-form ω_0 on $\mathbb{R}^{2m}$. If g is a metric on M with holonomy $\mathrm{U}(m)$ then g is *Kähler* with complex structure J, and the 2-form ω on M associated to ω_0 is the *Kähler form* of g.

  One can show that ω is a calibration on (M, g), and the calibrated submanifolds are exactly the *holomorphic curves* in (M, J). More generally $\omega^k/k!$ is a calibration on M for $1 \leqslant k \leqslant m$, and the corresponding calibrated submanifolds are the complex k-dimensional submanifolds of (M, J).
- Let $G = \mathrm{SU}(m) \subset \mathrm{O}(2m)$. Then G preserves a *complex volume form* $\Omega_0 = \mathrm{d}z_1\wedge\cdots\wedge\mathrm{d}z_m$ on $\mathbb{C}^m$. Thus a *Calabi–Yau m-fold* (M, g) with $\mathrm{Hol}(g) = \mathrm{SU}(m)$ has a *holomorphic volume form* Ω. The real part $\mathrm{Re}\,\Omega$ is a calibration on M, and the corresponding calibrated submanifolds are called *special Lagrangian submanifolds*.
- The group $G_2 \subset \mathrm{O}(7)$ preserves a 3-form φ_0 and a 4-form $*\varphi_0$ on $\mathbb{R}^7$. Thus a Riemannian 7-manifold (M, g) with holonomy G_2 comes with a 3-form φ and 4-form $*\varphi$, which are both calibrations. The corresponding calibrated submanifolds are called *associative 3-folds* and *coassociative 4-folds*.
- The group $\mathrm{Spin}(7) \subset \mathrm{O}(8)$ preserves a 4-form Ω_0 on $\mathbb{R}^8$. Thus a Riemannian 8-manifold (M, g) with holonomy $\mathrm{Spin}(7)$ has a 4-form Ω, which is a calibration. We call Ω-submanifolds *Cayley 4-folds*.

It is an important general principle that to each calibration φ on an n-manifold (M, g) with special holonomy we construct in this way, there corresponds a constant calibration φ_0 on $\mathbb{R}^n$. Locally, φ-submanifolds in M will look very like φ_0-submanifolds in $\mathbb{R}^n$, and have many of the same properties. Thus, to understand the calibrated submanifolds in a manifold with special holonomy, it is often a good idea to start by studying the corresponding calibrated submanifolds of $\mathbb{R}^n$.

In particular, singularities of φ-submanifolds in M will be locally modelled on singularities of φ_0-submanifolds in $\mathbb{R}^n$. (In the sense of Geometric Measure Theory, the *tangent cone* at a singular point of a φ-submanifold in M is a conical φ_0-submanifold in $\mathbb{R}^n$.) So by studying singular φ_0-submanifolds in $\mathbb{R}^n$, we may understand the singular behaviour of φ-submanifolds in M.

5.3. Associative and coassociative submanifolds

We now discuss the calibrated submanifolds of G_2-manifolds.

Definition 5.3. Let (M, φ, g) be a G_2*-manifold*, as in §2.2. Then the 3-form φ is a *calibration* on (M, g). We define an *associative* 3*-fold* in M to be a 3-submanifold of M calibrated with respect to φ. Similarly, the Hodge star $*\varphi$ of φ is a calibration 4-form on (M, g). We define a *coassociative* 4*-fold* in M to be a 4-submanifold of M calibrated with respect to $*\varphi$.

To understand these, it helps to begin with some calculations on $\mathbb{R}^7$. Let the metric g_0, 3-form φ_0 and 4-form $*\varphi_0$ on $\mathbb{R}^7$ be as in §2.2. Define an *associative* 3*-plane* to be an oriented 3-dimensional vector subspace V of $\mathbb{R}^7$ with $\varphi_0|_V = \mathrm{vol}_V$, and a *coassociative* 4*-plane* to be an oriented 4-dimensional vector subspace V of $\mathbb{R}^7$ with $*\varphi_0|_V = \mathrm{vol}_V$. ¿From [12, Th. IV.1.8, Def. IV.1.15] we have:

Proposition 5.4. *The family $\mathcal{F}^3$ of associative* 3*-planes in $\mathbb{R}^7$ and the family $\mathcal{F}^4$ of coassociative* 4*-planes in $\mathbb{R}^7$ are both isomorphic to $G_2/SO(4)$, with dimension* 8*.*

Examples of an associative 3-plane U and a coassociative 4-plane V are

$$U = \{(x_1, x_2, x_3, 0, 0, 0, 0) : x_j \in \mathbb{R}\} \text{ and } V = \{(0, 0, 0, x_4, x_5, x_6, x_7) : x_j \in \mathbb{R}\}. \quad (17)$$

As G_2 acts *transitively* on the set of associative 3-planes by Proposition 5.4, every associative 3-plane is of the form $\gamma \cdot U$ for $\gamma \in G_2$. Similarly, every coassociative 4-plane is of the form $\gamma \cdot V$ for $\gamma \in G_2$.

Now $\varphi_0|_V \equiv 0$. As φ_0 is G_2-invariant, this gives $\varphi_0|_{\gamma \cdot V} \equiv 0$ for all $\gamma \in G_2$, so φ_0 restricts to zero on all coassociative 4-planes. In fact the converse is true: if W is a 4-plane in $\mathbb{R}^7$ with $\varphi_0|_W \equiv 0$, then W is coassociative with some orientation. From this we deduce an alternative characterization of coassociative 4-folds:

Proposition 5.5. *Let (M, φ, g) be a G_2-manifold, and L a* 4*-dimensional submanifold of M. Then L admits an orientation making it into a coassociative* 4*-fold if and only if $\varphi|_L \equiv 0$.*

Trivially, $\varphi|_L \equiv 0$ implies that $[\varphi|_L] = 0$ in $H^3(L,\mathbb{R})$. Regard L as an immersed 4-submanifold, with immersion $\iota : L \to M$. Then $[\varphi|_L] \in H^3(L,\mathbb{R})$ is unchanged under continuous variations of the immersion ι. Thus, $[\varphi|_L] = 0$ is a necessary condition not just for L to be coassociative, but also for any isotopic 4-fold N in M to be coassociative. This gives a *topological restriction* on coassociative 4-folds.

Corollary 5.6. *Let (φ, g) be a torsion-free G_2-structure on a 7-manifold M, and L a real 4-submanifold in M. Then a necessary condition for L to be isotopic to a coassociative 4-fold N in M is that $[\varphi|_L] = 0$ in $H^3(L,\mathbb{R})$.*

5.4. Examples of associative 3-submanifolds

Here are some sources of examples of associative 3-folds in $\mathbb{R}^7$:

- Write $\mathbb{R}^7 = \mathbb{R}\oplus\mathbb{C}^3$. Then $\mathbb{R}\times\Sigma$ is an associative 3-fold in $\mathbb{R}^7$ for any *holomorphic curve* Σ in $\mathbb{C}^3$. Also, if L is any *special Lagrangian 3-fold* in $\mathbb{C}^3$ and $x \in \mathbb{R}$ then $\{x\}\times L$ is associative 3-fold in $\mathbb{R}^7$. For examples of special Lagrangian 3-folds see [11, §9], and references therein.
- Bryant [5, §4] studies compact Riemann surfaces Σ in $\mathcal{S}^6$ which are (pseudo)-holomorphic with respect to the almost complex structure J on $\mathcal{S}^6$ induced by its inclusion in $\operatorname{Im}\mathbb{O} \cong \mathbb{R}^7$. Then the cone on Σ is an *associative cone* on $\mathbb{R}^7$. He shows that any Σ has a *torsion* τ, a holomorphic analogue of the Serret–Frenet torsion of real curves in $\mathbb{R}^3$.

  The torsion τ is a section of a holomorphic line bundle on Σ, and $\tau = 0$ if $\Sigma \cong \mathbb{CP}^1$. If $\tau = 0$ then Σ is the projection to $\mathcal{S}^6 = G_2/\,\mathrm{SU}(3)$ of a *holomorphic curve* $\tilde{\Sigma}$ in the *projective complex manifold* $G_2/\mathrm{U}(2)$. This reduces the problem of understanding null torsion associative cones in $\mathbb{R}^7$ to that of finding holomorphic curves $\tilde{\Sigma}$ in $G_2/\,\mathrm{U}\,(2)$ satisfing a *horizontality condition*, which is a problem in *complex algebraic geometry*. In integrable systems language, null torsion curves are called *superminimal*.

  Bryant also shows that *every* Riemann surface Σ may be embedded in $\mathcal{S}^6$ with null torsion in infinitely many ways, of arbitrarily high degree. This shows that there are *many associative cones in* $\mathbb{R}^7$, on *oriented surfaces of every genus*. These provide many local models for *singularities* of associative 3-folds.

  Perhaps the simplest nontrivial example of a pseudoholomorphic curve Σ in $\mathcal{S}^6$ with null torsion is the *Borůvka sphere* [4], which is an $\mathcal{S}^2$ orbit of an SO(3) subgroup of G_2 acting irreducibly on $\mathbb{R}^7$. Other examples are given by Ejiri [10, §5–§6], who classifies pseudoholomorphic $\mathcal{S}^2$'s in $\mathcal{S}^6$ invariant under a U(1) subgroup of G_2, and Sekigawa [30].
- Bryant's paper is one of the first steps in the study of associative cones in $\mathbb{R}^7$ using the theory of *integrable systems*. Bolton et al. [2], [3, §6] use integrable systems methods to prove important results on pseudoholomorphic curves Σ in $\mathcal{S}^6$. When Σ is a torus T^2, they show it is of *finite type* [3, Cor. 6.4], and so

can be classified in terms of algebro-geometric *spectral data*, and perhaps even in principle be written down explicitly.

- *Curvature properties* of pseudoholomorphic curves in $\mathcal{S}^6$ are studied by Hashimoto [13] and Sekigawa [30].
- Lotay [25] studies constructions for associative 3-folds N in $\mathbb{R}^7$. These generally involve writing N as the total space of a 1-parameter family of surfaces P_t in $\mathbb{R}^7$ of a prescribed form, and reducing the condition for N to be associative to an o.d.e. in t, which can be (partially) solved fairly explicitly.

  Lotay also considers *ruled associative 3-folds* [25, §6], which are associative 3-folds N in $\mathbb{R}^7$ fibred by a 2-parameter family of affine straight lines $\mathbb{R}$. He shows that any *associative cone* N_0 on a Riemann surface Σ in $\mathcal{S}^6$ is the limit of a 6-dimensional family of *Asymptotically Conical* ruled associative 3-folds if $\Sigma \cong \mathbb{CP}^1$, and of a 2-dimensional family if $\Sigma \cong T^2$.

  Combined with the results of Bryant [5, §4] above, this yields many examples of generically nonsingular Asymptotically Conical associative 3-folds in $\mathbb{R}^7$, diffeomorphic to $\mathcal{S}^2 \times \mathbb{R}$ or $T^2 \times \mathbb{R}$.

Examples of associative 3-folds in other explicit G_2-manifolds, such as those of Bryant and Salamon [8], may also be constructed using similar techniques. For finding associative 3-folds in *nonexplicit* G_2-manifolds, such as the compact examples of §3 which are known only through existence theorems, there is one method [18, §12.6], which we now explain.

Suppose $\gamma \in G_2$ with $\gamma^2 = 1$ but $\gamma \neq 1$. Then γ is conjugate in G_2 to

$$(x_1, \ldots, x_7) \longmapsto (x_1, x_2, x_3, -x_4, -x_5, -x_6, -x_7).$$

The fixed point set of this involution is the associative 3-plane U of (17). It follows that any $\gamma \in G_2$ with $\gamma^2 = 1$ but $\gamma \neq 1$ has fixed point set an associative 3-plane. Thus we deduce [18, Prop. 10.8.1]:

Proposition 5.7. *Let (M, φ, g) be a G_2-manifold, and $\sigma : M \to M$ be a nontrivial isometric involution with $\sigma^*(\varphi) = \varphi$. Then $N = \{p \in M : \sigma(p) = p\}$ is an associative 3-fold in M.*

Here a *nontrivial isometric involution* of (M, g) is a diffeomorphism $\sigma : M \to M$ such that $\sigma^*(g) = g$, and $\sigma \neq \mathrm{id}$ but $\sigma^2 = \mathrm{id}$, where id is the identity on M. Following [18, Ex. 12.6.1], we can use the proposition to construct *examples* of compact associative 3-folds in the compact 7-manifolds with holonomy G_2 constructed in §3.

Example 5.8. Let $T^7 = \mathbb{R}^7/\mathbb{Z}^7$ and Γ be as in Example 3.1. Define $\sigma : T^7 \to T^7$ by

$$\sigma : (x_1, \ldots, x_7) \mapsto (x_1, x_2, x_3, \tfrac{1}{2} - x_4, -x_5, -x_6, -x_7).$$

Then σ preserves (φ_0, g_0) and commutes with Γ, and so its action pushes down to T^7/Γ. The fixed points of σ on T^7 are 16 copies of T^3, and $\sigma\delta$ has no fixed points in T^7 for all $\delta \neq 1$ in Γ. Thus the fixed points of σ in T^7/Γ are the image of the 16 T^3 fixed by σ in T^7.

Calculation shows that these 16 T^3 do not intersect the fixed points of α, β or γ, and that Γ acts freely on the set of 16 T^3 fixed by σ. So the image of the 16 T^3 in T^7 is 2 T^3 in T^7/Γ, which do not intersect the singular set of T^7/Γ, and which are *associative 3-folds* in T^7/Γ by Proposition 5.7.

Now the resolution of T^7/Γ to get a compact G_2-manifold $(M, \tilde\varphi, \tilde g)$ with $\mathrm{Hol}(\tilde g) = G_2$ described in §3 may be done in a *σ-equivariant way*, so that σ lifts to $\sigma : M \to M$ with $\sigma^*(\tilde\varphi) = \tilde\varphi$. The fixed points of σ in M are again 2 copies of T^3, which are *associative 3-folds* by Proposition 5.7.

5.5. Examples of coassociative 4-submanifolds

Here are some sources of examples of coassociative 4-folds in $\mathbb{R}^7$:

- Write $\mathbb{R}^7 = \mathbb{R} \oplus \mathbb{C}^3$. Then $\{x\} \times S$ is a coassociative 4-fold in $\mathbb{R}^7$ for any *holomorphic surface* S in $\mathbb{C}^3$ and $x \in \mathbb{R}$. Also, $\mathbb{R} \times L$ is a coassociative 4-fold in $\mathbb{R}^7$ for any *special Lagrangian 3-fold L in $\mathbb{C}^3$ with phase i*. For examples of special Lagrangian 3-folds see [11, §9], and references therein.
- Harvey and Lawson [12, §IV.3] give examples of coassociative 4-folds in $\mathbb{R}^7$ invariant under SU(2), acting on $\mathbb{R}^7 \cong \mathbb{R}^3 \oplus \mathbb{C}^2$ as $\mathrm{SO}(3) = \mathrm{SU}(2)/\{\pm 1\}$ on the $\mathbb{R}^3$ and SU(2) on the $\mathbb{C}^2$ factor. Such 4-folds correspond to solutions of an o.d.e., which Harvey and Lawson solve.
- Mashimo [27] classifies *coassociative cones* N in $\mathbb{R}^7$ with $N \cap \mathcal{S}^6$ homogeneous under a 3-dimensional simple subgroup H of G_2.
- Lotay [26] studies *2-ruled coassociative 4-folds* in $\mathbb{R}^7$, that is, coassociative 4-folds N which are fibred by a 2-dimensional family of affine 2-planes $\mathbb{R}^2$ in $\mathbb{R}^7$, with base space a Riemann surface Σ. He shows that such 4-folds arise locally from data $\phi_1, \phi_2 : \Sigma \to \mathcal{S}^6$ and $\psi : \Sigma \to \mathbb{R}^7$ satisfying nonlinear p.d.e.s similar to the Cauchy–Riemann equations.

  For ϕ_1, ϕ_2 fixed, the remaining equations on ψ are *linear*. This means that the family of 2-ruled associative 4-folds N in $\mathbb{R}^7$ asymptotic to a fixed 2-ruled coassociative cone N_0 has the structure of a *vector space*. It can be used to generate families of examples of coassociative 4-folds in $\mathbb{R}^7$.

We can also use the fixed-point set technique of §5.4 to find examples of coassociative 4-folds in other G_2-manifolds. If $\alpha : \mathbb{R}^7 \to \mathbb{R}^7$ is linear with $\alpha^2 = 1$ and $\alpha^*(\varphi_0) = -\varphi_0$, then either $\alpha = -1$, or α is conjugate under an element of G_2 to the map

$$(x_1, \ldots, x_7) \longmapsto (-x_1, -x_2, -x_3, x_4, x_5, x_6, x_7).$$

The fixed set of this map is the coassociative 4-plane V of (17). Thus, the fixed point set of α is either $\{0\}$, or a coassociative 4-plane in $\mathbb{R}^7$. So we find [18, Prop. 10.8.5]:

Proposition 5.9. *Let (M, φ, g) be a G_2-manifold, and $\sigma : M \to M$ an isometric involution with $\sigma^*(\varphi) = -\varphi$. Then each connected component of the fixed point set $\bigl\{p \in M : \sigma(p) = p\bigr\}$ of σ is either a coassociative 4-fold or a single point.*

Bryant [7] uses this idea to construct many *local* examples of compact coassociative 4-folds in G_2-manifolds.

Theorem 5.10 (Bryant [7]). *Let (N, g) be a compact, real analytic, oriented Riemannian 4-manifold whose bundle of self-dual 2-forms is trivial. Then N may be embedded isometrically as a coassociative 4-fold in a G_2-manifold (M, φ, g), as the fixed point set of an involution σ.*

Note here that M need not be *compact*, nor (M, g) *complete*. Roughly speaking, Bryant's proof constructs (φ, g) as the sum of a power series on $\Lambda^2_+ T^*N$ converging near the zero section $N \subset \Lambda^2 T^*N$, using the theory of *exterior differential systems*. The involution σ acts as -1 on $\Lambda^2_+ T^*N$, fixing the zero section. One moral of Theorem 5.10 is that to be coassociative places no significant local restrictions on a 4-manifold, other than orientability.

Examples of *compact* coassociative 4-folds in *compact* G_2-manifolds with holonomy G_2 are constructed in [18, §12.6], using Proposition 5.9. Here [18, Ex. 12.6.4] are examples in the G_2-manifolds of §3.

Example 5.11. Let $T^7 = \mathbb{R}^7/\mathbb{Z}^7$ and Γ be as in Example 3.1. Define $\sigma : T^7 \to T^7$ by

$$\sigma : (x_1, \ldots, x_7) \mapsto (\tfrac{1}{2} - x_1, x_2, x_3, x_4, x_5, \tfrac{1}{2} - x_6, \tfrac{1}{2} - x_7).$$

Then σ commutes with Γ, preserves g_0 and takes φ_0 to $-\varphi_0$. The fixed points of σ in T^7 are 8 copies of T^4, and the fixed points of $\sigma\alpha\beta$ in T^7 are 128 points. If $\delta \in \Gamma$ then $\sigma\delta$ has no fixed points unless $\delta = 1, \alpha\beta$. Thus the fixed points of σ in T^7/Γ are the image of the fixed points of σ and $\sigma\alpha\beta$ in T^7.

Now Γ acts freely on the sets of 8 σ T^4 and 128 $\sigma\alpha\beta$ points. So the fixed point set of σ in T^7/Γ is the union of T^4 and 16 isolated points, none of which intersect the singular set of T^7/Γ. When we resolve T^7/Γ to get $(M, \tilde{\varphi}, \tilde{g})$ with $\mathrm{Hol}(\tilde{g}) = G_2$ in a σ-equivariant way, the action of σ on M has $\sigma^*(\tilde{\varphi}) = -\tilde{\varphi}$, and again fixes T^4 and 16 points. By Proposition 5.9, this T^4 is *coassociative*.

More examples of associative and coassociative submanifolds with different topologies are given in [18, §12.6].

5.6. Cayley 4-folds

The calibrated geometry of Spin(7) is similar to the G_2 case above, so we shall be brief.

Definition 5.12. Let (M, Ω, g) be a Spin(7)-*manifold*, as in §2.3. Then the 4-form Ω is a *calibration* on (M, g). We define a *Cayley 4-fold* in M to be a 4-submanifold of M calibrated with respect to Ω.

Let the metric g_0, and 4-form Ω_0 on $\mathbb{R}^8$ be as in §2.3. Define a *Cayley 4-plane* to be an oriented 4-dimensional vector subspace V of $\mathbb{R}^8$ with $\Omega_0|_V = \mathrm{vol}_V$. Then we have an analogue of Proposition 5.4:

Proposition 5.13. *The family $\mathcal{F}$ of Cayley 4-planes in $\mathbb{R}^8$ is isomorphic to* $\mathrm{Spin}(7)/K$, *where* $K \cong \big(\mathrm{SU}(2)\times\mathrm{SU}(2)\times\mathrm{SU}(2)\big)/\mathbb{Z}_2$ *is a Lie subgroup of* $\mathrm{Spin}(7)$*, and* $\dim\mathcal{F} = 12$.

Here are some sources of examples of Cayley 4-folds in $\mathbb{R}^8$:

- Write $\mathbb{R}^8 = \mathbb{C}^4$. Then any *holomorphic surface* S in $\mathbb{C}^4$ is Cayley in $\mathbb{R}^8$, and any *special Lagrangian* 4*-fold* N in $\mathbb{C}^4$ is Cayley in $\mathbb{R}^8$.
  Write $\mathbb{R}^8 = \mathbb{R}\times\mathbb{R}^7$. Then $\mathbb{R}\times L$ is Cayley for any *associative* 3*-fold* L in $\mathbb{R}^7$.
- Lotay [26] studies 2*-ruled Cayley* 4*-folds* in $\mathbb{R}^8$, that is, Cayley 4-folds N fibred by a 2-dimensional family Σ of affine 2-planes $\mathbb{R}^2$ in $\mathbb{R}^8$, as for the coassociative case in §5.5. He constructs explicit families of 2-ruled Cayley 4-folds in $\mathbb{R}^8$, including some depending on an arbitrary holomorphic function $w:\mathbb{C}\to\mathbb{C}$, [26, Th. 5.1].

By the method of Propositions 5.7 and 5.9 one can prove [18, Prop. 10.8.6]:

Proposition 5.14. *Let* (M,Ω,g) *be a* $\mathrm{Spin}(7)$*-manifold, and* $\sigma: M\to M$ *a nontrivial isometric involution with* $\sigma^*(\Omega)=\Omega$*. Then each connected component of the fixed point set* $\{p\in M:\sigma(p)=p\}$ *is either a Cayley* 4*-fold or a single point.*

Using this, [18, §14.3] constructs examples of *compact* Cayley 4-folds in compact 8-manifolds with holonomy Spin(7).

6. Deformation theory of calibrated submanifolds

Finally we discuss *deformations* of associative, coassociative and Cayley submanifolds. In §6.1 we consider the local equations for such submanifolds in $\mathbb{R}^7$ and $\mathbb{R}^8$, following Harvey and Lawson [12, §IV.2]. Then §6.2 explains the deformation theory of *compact* coassociative 4-folds, following McLean [28, §4]. This has a particularly simple structure, as coassociative 4-folds are defined by the vanishing of φ. The deformation theory of compact associative 3-folds and Cayley 4-folds is more complex, and is sketched in §6.3.

6.1. Parameter counting and the local equations

We now study the local equations for 3- or 4-folds to be (co)associative or Cayley.

Associative 3-folds. The set of all 3-planes in $\mathbb{R}^7$ has dimension 12, and the set of associative 3-planes in $\mathbb{R}^7$ has dimension 8 by Proposition 5.4. Thus the associative 3-planes are of *codimension* 4 in the set of all 3-planes. Therefore the condition for a 3-fold L in $\mathbb{R}^7$ to be associative is 4 equations on each tangent space. The freedom to vary L is the sections of its normal bundle in $\mathbb{R}^7$, which is 4 real functions. Thus, the deformation problem for associative 3-folds involves 4 *equations on* 4 *functions*, so it is a *determined* problem.

To illustrate this, let $f:\mathbb{R}^3\to\mathbb{H}$ be a smooth function, written

$$f(x_1,x_2,x_3) = f_0(x_1,x_2,x_3) + f_1(x_1,x_2,x_3)i + f_2(x_1,x_2,x_3)j + f_3(x_1,x_2,x_3)k.$$

Define a 3-submanifold L in $\mathbb{R}^7$ by

$$L = \big\{\big(x_1,x_2,x_3,f_0(x_1,x_2,x_3),\ldots,f_3(x_1,x_2,x_3)\big): x_j\in\mathbb{R}\big\}.$$

Then Harvey and Lawson [12, §IV.2.A] calculate the conditions on f for L to be associative. With the conventions of §2.1, the equation is

$$i\frac{\partial f}{\partial x_1}+j\frac{\partial f}{\partial x_2}-k\frac{\partial f}{\partial x_3}=C\Bigl(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\frac{\partial f}{\partial x_3}\Bigr), \tag{18}$$

where $C:\mathbb{H}\times\mathbb{H}\times\mathbb{H}\rightarrow\mathbb{H}$ is a trilinear cross product.

Here (18) is 4 equations on 4 functions, as we claimed, and is a *first order nonlinear elliptic p.d.e.* When $f,\partial f$ are small, so that L approximates the associative 3-plane U of (17), equation (18) reduces approximately to the linear equation $i\frac{\partial f}{\partial x_1}+j\frac{\partial f}{\partial x_2}-k\frac{\partial f}{\partial x_3}=0$, which is equivalent to the *Dirac equation* on $\mathbb{R}^3$. More generally, first order deformations of an associative 3-fold L in a G_2-manifold (M,φ,g) correspond to solutions of a *twisted Dirac equation* on L.

Coassociative 4-folds. The set of all 4-planes in $\mathbb{R}^7$ has dimension 12, and the set of coassociative 4-planes in $\mathbb{R}^7$ has dimension 8 by Proposition 5.4. Thus the coassociative 4-planes are of *codimension* 4 in the set of all 4-planes. Therefore the condition for a 4-fold N in $\mathbb{R}^7$ to be coassociative is 4 equations on each tangent space. The freedom to vary N is the sections of its normal bundle in $\mathbb{R}^7$, which is 3 real functions. Thus, the deformation problem for coassociative 4-folds involves 4 *equations on* 3 *functions*, so it is an *overdetermined* problem.

To illustrate this, let $f:\mathbb{H}\rightarrow\mathbb{R}^3$ be a smooth function, written

$$f(x_0+x_1i+x_2j+x_3k)=(f_1,f_2,f_3)(x_0+x_1i+x_2j+x_3k).$$

Define a 4-submanifold N in $\mathbb{R}^7$ by

$$N=\bigl\{\bigl(f_1(x_0,\ldots,x_3),f_2(x_0,\ldots,x_3),f_3(x_0,\ldots,x_3),x_0,\ldots,x_3\bigr):x_j\in\mathbb{R}\bigr\}.$$

Then Harvey and Lawson [12, §IV.2.B] calculate the conditions on f for N to be coassociative. With the conventions of §2.1, the equation is

$$i\partial f_1+j\partial f_2-k\partial f_3=C(\partial f_1,\partial f_2,\partial f_3), \tag{19}$$

where the derivatives $\partial f_j=\partial f_j(x_0+x_1i+x_2j+x_3k)$ are interpreted as functions $\mathbb{H}\rightarrow\mathbb{H}$, and C is as in (18). Here (19) is 4 equations on 3 functions, as we claimed, and is a *first order nonlinear overdetermined elliptic p.d.e.*

Cayley 4-folds. The set of all 4-planes in $\mathbb{R}^8$ has dimension 16, and the set of Cayley 4-planes in $\mathbb{R}^8$ has dimension 12 by Proposition 5.13, so the Cayley 4-planes are of *codimension* 4 in the set of all 4-planes. Therefore the condition for a 4-fold K in $\mathbb{R}^8$ to be Cayley is 4 equations on each tangent space. The freedom to vary K is the sections of its normal bundle in $\mathbb{R}^8$, which is 4 real functions. Thus, the deformation problem for Cayley 4-folds involves 4 *equations on* 4 *functions*, so it is a *determined* problem.

Let $f = f_0 + f_1 i + f_2 j + f_3 k = f(x_0 + x_1 i + x_2 j + x_3 k) : \mathbb{H} \to \mathbb{H}$ be smooth. Choosing signs for compatibility with (2), define a 4-submanifold K in $\mathbb{R}^8$ by

$$K = \big\{\big(-x_0, x_1, x_2, x_3, f_0(x_0 + x_1 i + x_2 j + x_3 k), -f_1(x_0 + x_1 i + x_2 j + x_3 k), \\ -f_2(x_0 + x_1 i + x_2 j + x_3 k), f_3(x_0 + x_1 i + x_2 j + x_3 k)\big) : x_j \in \mathbb{R}\big\}.$$

Following [12, §IV.2.C], the equation for K to be Cayley is

$$\frac{\partial f}{\partial x_0} + i\frac{\partial f}{\partial x_1} + j\frac{\partial f}{\partial x_2} + k\frac{\partial f}{\partial x_3} = C(\partial f), \tag{20}$$

for $C : \mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \to \mathbb{H}$ a homogeneous cubic polynomial. This is 4 equations on 4 functions, as we claimed, and is a first-order nonlinear elliptic p.d.e. on f. The linearization at $f = 0$ is equivalent to the *positive Dirac equation* on $\mathbb{R}^4$. More generally, first order deformations of a Cayley 4-fold K in a Spin(7)-manifold (M, Ω, g) correspond to solutions of a *twisted positive Dirac equation* on K.

6.2. Deformation theory of coassociative 4-folds

Here is the main result in the deformation theory of coassociative 4-folds, proved by McLean [28, Th. 4.5]. As our sign conventions for $\varphi_0, *\varphi_0$ in (1) are different to McLean's, we use self-dual 2-forms in place of McLean's anti-self-dual 2-forms.

Theorem 6.1. *Let (M, φ, g) be a G_2-manifold, and N a compact coassociative 4-fold in M. Then the moduli space $\mathscr{M}_N$ of coassociative 4-folds isotopic to N in M is a smooth manifold of dimension $b^2_+(N)$.*

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. There is a natural isomorphism $\nu \cong \Lambda^2_+ T^*N$, constructed as follows. Let $x \in N$ and $V \in \nu_x$. Then $V \in T_x M$, so $V \cdot \varphi|_x \in \Lambda^2 T^*_x M$, and $(V \cdot \varphi|_x)|_{T_x N} \in \Lambda^2 T^*_x N$. It turns out that $(V \cdot \varphi|_x)|_{T_x N}$ actually lies in $\Lambda^2_+ T^*_x N$, the bundle of *self-dual 2-forms* on N, and that the map $V \mapsto (V \cdot \varphi|_x)|_{T_x N}$ defines an *isomorphism* $\nu \xrightarrow{\cong} \Lambda^2_+ T^*N$.

Let T be a small *tubular neighbourhood* of N in M. Then we can identify T with a neighbourhood of the zero section in ν, using the exponential map. The isomorphism $\nu \cong \Lambda^2_+ T^*N$ then identifies T with a neighbourhood U of the zero section in $\Lambda^2_+ T^*N$. 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* $\Gamma(\alpha)$ of small smooth sections α of $\Lambda^2_+ T^*N$ lying in U. Write $C^\infty(U)$ for the subset of the vector space of smooth self-dual 2-forms $C^\infty(\Lambda^2_+ T^*N)$ on N lying in $U \subset \Lambda^2_+ T^*N$. Then for each $\alpha \in C^\infty(U)$ the graph $\Gamma(\alpha)$ is a 4-submanifold of U, and so is identified with a 4-submanifold of T. We need to know: which 2-forms α correspond to *coassociative* 4-folds $\Gamma(\alpha)$ in T?

Well, N' is coassociative if $\varphi|_{N'} \equiv 0$. Now $\pi|_{N'} : N' \to N$ is a diffeomorphism, so we can push $\varphi|_{N'}$ down to N, and regard it as a function of α. That is, we define

$$P : C^\infty(U) \longrightarrow C^\infty(\Lambda^3 T^*N) \quad \text{by} \quad P(\alpha) = \pi_*(\varphi|_{\Gamma(\alpha)}). \tag{21}$$

Then the moduli space $\mathscr{M}_N$ is locally isomorphic near N to the set of small self-dual 2-forms α on N with $\varphi|_{\Gamma(\alpha)} \equiv 0$, that is, to a neighbourhood of 0 in $P^{-1}(0)$.

To understand the equation $P(\alpha) = 0$, note that at $x \in N$, $P(\alpha)|_x$ depends on the tangent space to $\Gamma(\alpha)$ at $\alpha|_x$, and so on $\alpha|_x$ and $\nabla\alpha|_x$. Thus the functional form of P is

$$P(\alpha)|_x = F\big(x, \alpha|_x, \nabla\alpha|_x\big) \quad \text{for } x \in N,$$

where F is a smooth function of its arguments. Hence $P(\alpha) = 0$ is a *nonlinear first order p.d.e.* in α. The *linearization* $\mathrm{d}P(0)$ of P at $\alpha = 0$ turns out to be

$$\mathrm{d}P(0)(\beta) = \lim_{\epsilon \to 0}\big(\epsilon^{-1}P(\epsilon\beta)\big) = \mathrm{d}\beta.$$

Therefore $\mathrm{Ker}(\mathrm{d}P(0))$ is the vector space $\mathscr{H}^2_+$ of *closed self-dual* 2*-forms* β on N, which by Hodge theory is a finite-dimensional vector space isomorphic to $H^2_+(N, \mathbb{R})$, with dimension $b^2_+(N)$. This is the *Zariski tangent space* of $\mathscr{M}_N$ at N, the *infinitesimal deformation space* of N as a coassociative 4-fold.

To complete the proof we must show that $\mathscr{M}_N$ is locally isomorphic to its Zariski tangent space $\mathscr{H}^2_+$, and so is a smooth manifold of dimension $b^2_+(N)$. To do this rigorously requires some technical analytic machinery, which is passed over in a few lines in [28, p. 731]. Here is one way to do it.

Because $C^\infty(\Lambda^2_+T^*N), C^\infty(\Lambda^3T^*N)$ are not *Banach spaces*, we extend P in (21) to act on *Hölder spaces* $C^{k+1,\gamma}(\Lambda^2_+T^*N), C^{k,\gamma}(\Lambda^3T^*N)$ for $k \geqslant 1$ and $\gamma \in (0,1)$, giving

$$P_{k,\gamma} : C^{k+1,\gamma}(U) \longrightarrow C^{k,\gamma}(\Lambda^3T^*N) \quad \text{defined by} \quad P_{k,\gamma}(\alpha) = \pi_*(\varphi|_{\Gamma(\alpha)}).$$

Then $P_{k,\gamma}$ is a smooth map of Banach manifolds. Let $V_{k,\gamma} \subset C^{k,\gamma}(\Lambda^3T^*N)$ be the Banach subspace of *exact* $C^{k,\gamma}$ 3-forms on N.

As φ is closed, $\varphi|_N \equiv 0$, and $\Gamma(\alpha)$ is isotopic to N, we see that $\varphi|_{\Gamma(\alpha)}$ is an *exact* 3-form on $\Gamma(\alpha)$, so that $P_{k,\gamma}$ maps into $V_{k,\gamma}$. The linearization

$$\mathrm{d}P_{k,\gamma}(0) : C^{k+1,\gamma}(\Lambda^2_+T^*N) \longrightarrow V_{k,\gamma}, \qquad \mathrm{d}P_{k,\gamma}(0) : \beta \longmapsto \mathrm{d}\beta$$

is then *surjective* as a map of Banach spaces. (To prove this requires a discursion, using elliptic regularity results for $\mathrm{d} + \mathrm{d}^*$.)

Thus, $P_{k,\gamma} : C^{k+1,\gamma}(U) \to V_{k,\gamma}$ is a smooth map of Banach manifolds, with $\mathrm{d}P_{k,\gamma}(0)$ surjective. The *Implicit Function Theorem for Banach spaces* now implies that $P^{-1}_{k,\gamma}(0)$ is near 0 a smooth submanifold of $C^{k+1,\gamma}(U)$, locally isomorphic to $\mathrm{Ker}(\mathrm{d}P_{k,\gamma}(0))$. But $P_{k,\gamma}(\alpha) = 0$ is an *overdetermined elliptic equation* for small α, and so elliptic regularity implies that solutions α are smooth. Therefore $P^{-1}_{k,\gamma}(0) = P^{-1}(0)$ near 0, and similarly $\mathrm{Ker}(\mathrm{d}P_{k,\gamma}(0)) = \mathrm{Ker}(\mathrm{d}P(0)) = \mathscr{H}^2_+$. This completes the proof. □

Here are some remarks on Theorem 6.1.

- This proof relies heavily on Proposition 5.5, that a 4-fold N in M is coassociative if and only if $\varphi|_N \equiv 0$, for φ a closed 3-form on M. The consequence of this is that the deformation theory of compact coassociative 4-folds is *unobstructed*, and the moduli space is *always* a smooth manifold with dimension given by a topological formula.

Special Lagrangian m-folds of Calabi-Yau m-folds can also be defined in terms of the vanishing of closed forms, and their deformation theory is also unobstructed, as in [28, §3] and [11, §10.2]. However, associative 3-folds and Cayley 4-folds cannot be defined by the vanishing of closed forms, and we will see in §6.3 that this gives their deformation theory a different flavour.

- We showed in §6.1 that the condition for a 4-fold N in M to be coassociative is locally 4 equations on 3 functions, and so is *overdetermined*. However, Theorem 6.1 shows that coassociative 4-folds have *unobstructed* deformation theory, and often form *positive-dimensional* moduli spaces. This seems very surprising for an overdetermined equation.

  The explanation is that the condition $\mathrm{d}\varphi = 0$ acts as an *integrability condition* for the existence of coassociative 4-folds. That is, since closed 3-forms on N essentially depend locally only on 3 real parameters, not 4, as φ is closed the equation $\varphi|_N \equiv 0$ is in effect only 3 equations on N rather than 4, so we can think of the deformation theory as really controlled by a determined elliptic equation.

  Therefore $\mathrm{d}\varphi = 0$ is essential for Theorem 6.1 to work. In 'almost G_2-manifolds' (M, φ, g) with $\mathrm{d}\varphi \neq 0$, the deformation problem for coassociative 4-folds is overdetermined and obstructed, and generically there would be no coassociative 4-folds.
- In Example 5.11 we constructed an example of a compact coassociative 4-fold N diffeomorphic to T^4 in a compact G_2-manifold (M, φ, g). By Theorem 6.1, N lies in a *smooth* 3*-dimensional family* of coassociative T^4's in M. Locally, these may form a *coassociative fibration* of M.

Now suppose $\{(M, \varphi_t, g_t) : t \in (-\epsilon, \epsilon)\}$ is a smooth 1-parameter family of G_2-manifolds, and N_0 a compact coassociative 4-fold in (M, φ_0, g_0). When can we extend N_0 to a smooth family of coassociative 4-folds N_t in (M, φ_t, g_t) for small t? By Corollary 5.6, a necessary condition is that $[\varphi_t|_{N_0}] = 0$ for all t. Our next result shows that locally, this is also a *sufficient* condition. It can be proved using similar techniques to Theorem 6.1, though McLean did not prove it.

Theorem 6.2. *Let* $\{(M, \varphi_t, g_t) : t \in (-\epsilon, \epsilon)\}$ *be a smooth* 1*-parameter family of* G_2*-manifolds, and* N_0 *a compact coassociative* 4*-fold in* (M, φ_0, g_0)*. Suppose that* $[\varphi_t|_{N_0}] = 0$ *in* $H^3(N_0, \mathbb{R})$ *for all* $t \in (-\epsilon, \epsilon)$*. Then* N_0 *extends to a smooth* 1*-parameter family* $\{N_t : t \in (-\delta, \delta)\}$*, where* $0 < \delta \leqslant \epsilon$ *and* N_t *is a compact coassociative* 4*-fold in* (M, φ_t, g_t)*.*

6.3. Deformation theory of associative 3-folds and Cayley 4-folds

Associative 3-folds and Cayley 4-folds cannot be defined in terms of the vanishing of closed forms, and this gives their deformation theory a different character to the coassociative case. Here is how the theories work, drawn mostly from McLean [28, §5–§6].

Let N be a compact associative 3-fold or Cayley 4-fold in a 7- or 8-manifold M. Then there are vector bundles $E, F \to N$ with $E \cong \nu$, the normal bundle of N in M, and a first-order elliptic operator $D_N : C^\infty(E) \to C^\infty(F)$ on N. The *kernel* $\operatorname{Ker} D_N$ is the set of *infinitesimal deformations* of N as an associative 3-fold or Cayley 4-fold. The *cokernel* $\operatorname{Coker} D_N$ is the *obstruction space* for these deformations.

Both are finite-dimensional vector spaces, and

$$\dim \operatorname{Ker} D_N - \dim \operatorname{Coker} D_N = \operatorname{ind}(D_N),$$

the *index* of D_N. It is a topological invariant, given in terms of characteristic classes by the *Atiyah–Singer Index Theorem*. In the associative case we have $E \cong F$, and D_N is anti-self-adjoint, so that $\operatorname{Ker}(D_N) \cong \operatorname{Coker}(D_N)$ and $\operatorname{ind}(D_N) = 0$ automatically. In the Cayley case we have

$$\operatorname{ind}(D_N) = \tau(N) - \tfrac{1}{2}\chi(N) - \tfrac{1}{2}[N] \cdot [N],$$

where τ is the signature, χ the Euler characteristic and $[N] \cdot [N]$ the self-intersection of N.

In a *generic* situation we expect $\operatorname{Coker} D_N = 0$, and then deformations of N will be unobstructed, so that the moduli space $\mathscr{M}_N$ of associative or Cayley deformations of N will locally be a smooth manifold of dimension $\operatorname{ind}(D_N)$. However, in nongeneric situations the obstruction space may be nonzero, and then the moduli space may not be smooth, or may have a larger than expected dimension.

This general structure is found in the deformation theory of other important mathematical objects — for instance, pseudo-holomorphic curves in almost complex manifolds, and instantons and Seiberg–Witten solutions on 4-manifolds. In each case, the moduli space is only smooth with a topologically determined dimension under a *genericity assumption* which forces the obstructions to vanish.

Acknowledgements: I would like to thank the conference organizers Turgut Önder and Selman Akbulut for their hospitality. Many people have helped me develop my ideas on exceptional holonomy and calibrated geometry; I would particularly like to thank Simon Salamon and Robert Bryant.

References

[1] M. Berger, *Sur les groupes d'holonomie homogène des variétés à connexion affines et des variétés Riemanniennes*, Bull. Soc. Math. France 83 (1955), 279–330.

[2] J. Bolton, L. Vrancken and L.M. Woodward, *On almost complex curves in the nearly Kähler* 6-*sphere*, Quart. J. Math. Oxford 45 (1994), 407–427.

[3] J. Bolton, F. Pedit and L. Woodward, *Minimal surfaces and the affine Toda field model*, J. reine angew. Math. 459 (1995), 119–150.

[4] O. Borůvka, *Sur les surfaces represéntées par les fonctions sphériques de prèmiere espèce*, J. Math. Pures Appl. 12 (1933), 337–383.

[5] R.L. Bryant, *Submanifolds and special structures on the octonians*, J. Diff. Geom. 17 (1982), 185–232.

[6] R.L. Bryant, *Metrics with exceptional holonomy*, Ann. Math. 126 (1987), 525–576.

[7] R.L. Bryant, *Calibrated embeddings: the special Lagrangian and coassociative cases*, Ann. Global Anal. Geom. 18 (2000), 405–435. math.DG/9912246.

[8] R.L. Bryant and S.M. Salamon, *On the construction of some complete metrics with exceptional holonomy*, Duke Math. J. 58 (1989), 829–850.

[9] E. Calabi, *Métriques kählériennes et fibrés holomorphes*, Ann. scient. éc. norm. sup. 12 (1979), 269–294.
[10] N. Ejiri, *Equivariant Minimal Immersions of* $\mathcal{S}^2$ *into* $\mathcal{S}^{2m}(1)$, Trans. A.M.S. 297 (1986), 105–124.
[11] M. Gross, D. Huybrechts and D. Joyce, *Calabi–Yau Manifolds and Related Geometries*, Universitext, Springer, Berlin, 2003.
[12] R. Harvey and H. B. Lawson, *Calibrated geometries*, Acta Mathematica 148 (1982), 47–157.
[13] H. Hashimoto, *J-Holomorphic Curves of a* 6*-Dimensional Sphere*, Tokyo Math J. 23 (2000), 137–159.
[14] D.D. Joyce, *Compact Riemannian* 7*-manifolds with holonomy* G_2*. I*, J. Diff. Geom. 43 (1996), 291–328.
[15] D.D. Joyce, *Compact Riemannian* 7*-manifolds with holonomy* G_2*. II*, J. Diff. Geom. 43 (1996), 329–375.
[16] D.D. Joyce, *Compact Riemannian* 8*-manifolds with holonomy* Spin(7), Invent. math. 123 (1996), 507–552.
[17] D.D. Joyce, *A new construction of compact* 8*-manifolds with holonomy* Spin(7), J. Diff. Geom. 53 (1999), 89–130. math.DG/9910002.
[18] D.D. Joyce, *Compact manifolds with special holonomy*, Oxford Mathematical Monographs Series, Oxford University Press, 2000.
[19] D.D. Joyce, *Asymptotically Locally Euclidean metrics with holonomy* $\mathrm{SU}(m)$, Ann. Global Anal. Geom. 19 (2001), 55–73. math.AG/9905041.
[20] D.D. Joyce, *Quasi-ALE metrics with holonomy* $\mathrm{SU}(m)$ *and* $\mathrm{Sp}(m)$, Ann. Global Anal. Geom. 19 (2001), 103–132. math.AG/9905043.
[21] D.D. Joyce, *Constructing compact manifolds with exceptional holonomy*, math.DG/0203158, 2002.
[22] A.G. Kovalev, *Twisted connected sums and special Riemannian holonomy*, J. Reine Angew. Math. 565 (2003), 125–160. math.DG/0012189.
[23] P.B. Kronheimer, *The construction of ALE spaces as hyperkähler quotients*, J. Diff. Geom. 29 (1989), 665–683.
[24] P.B. Kronheimer, *A Torelli-type theorem for gravitational instantons*, J. Diff. Geom. 29 (1989), 685–697.
[25] J. Lotay, *Constructing Associative* 3*-folds by Evolution Equations*, math.DG/0401123, 2004.
[26] J. Lotay, 2*-Ruled Calibrated* 4*-folds in* $\mathbb{R}^7$ *and* $\mathbb{R}^8$, math.DG/0401125, 2004.
[27] K. Mashimo, *On some stable minimal cones in* $\mathbb{R}^7$, pages 107–115 in *Differential geometry of submanifolds (Kyoto, 1984)*, Lecture Notes in Math. 1090, Springer, Berlin, 1984.
[28] R.C. McLean, *Deformations of calibrated submanifolds*, Comm. Anal. Geom. 6 (1998), 705–747.
[29] S.-S. Roan, *Minimal resolution of Gorenstein orbifolds*, Topology 35 (1996), 489–508.
[30] K. Sekigawa, *Almost complex submanifolds of a* 6*-dimensional sphere*, Kodai Math. J. 6 (1983), 174–185.
[31] 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
E-mail address: dominic.joyce@lincoln.ox.ac.uk

Proceedings of 12^{th} Gökova
Geometry-Topology Conference
pp. 140 – 156

Published online at
GokovaGT.org

Ricci-flat deformations of asymptotically cylindrical Calabi–Yau manifolds

Alexei Kovalev

Dedicated to the memory of Raoul Bott.

Abstract. We study a class of asymptotically cylindrical Ricci-flat Kähler metrics arising on quasiprojective manifolds. Using the Calabi–Yau geometry and analysis and the Kodaira–Kuranishi–Spencer theory and building up on results of N.Koiso, we show that under rather general hypotheses any local asymptotically cylindrical Ricci-flat deformations of such metrics are again Kähler, possibly with respect to a perturbed complex structure. We also find the dimension of the moduli space for these local deformations. In the class of asymptotically cylindrical Ricci-flat metrics on $2n$-manifolds, the holonomy reduction to $SU(n)$ is an open condition.

Let M be a compact smooth manifold with integrable complex structure J and g a Ricci-flat Kähler metric with respect to J. A theorem due to N.Koiso [10] asserts that if the deformations of the complex structure of M are unobstructed then the Ricci-flat Kähler metrics corresponding to the nearby complex structures and Kähler classes fill in an open neighbourhood in the moduli space of Ricci-flat metrics on M. The proof of this result relies on Hodge theory and Kodaira–Spencer–Kuranishi theory and Koiso also found the dimension of the moduli space.

The purpose of this paper is to extend the above result to a class of complete Ricci-flat Kähler manifolds with asymptotically cylindrical ends (see §1 for precise definitions). A suitable version of Hodge theory was developed as part of elliptic theory for asymptotically cylindrical manifolds in [13, 14, 15, 16]. A complex manifold underlying an asymptotically cylindrical Ricci-flat Kähler manifold admits a compactification by adding a 'divisor at infinity'. There is an extension of Kodaira–Spencer–Kuranishi theory for this class of non-compact complex manifolds using the cohomology of logarithmic sheaves [8]. On the other hand, manifolds with asymptotically cylindrical ends appear as an essential step in the gluing constructions of compact manifolds endowed with special Riemannian structures. In particular, the Ricci-flat Kähler asymptotically cylindrical manifolds were prominent in [11] in the construction of compact 7-dimensional Ricci-flat manifolds with special holonomy G_2.

We introduce the class of Ricci-flat Kähler asymptotically cylindrical manifolds in §1, where we also state our first main Theorem 1.3 and give interpretation in terms of special

This article was presented at the 12th Gökova Geometry-Topology Conference.

holonomy. We review basic facts about the Ricci-flat deformations in §2. §§3–5 contain the proof of Theorem 1.3 and our second main result Theorem 5.1 on the dimension of the moduli space for the Ricci-flat asymptotically cylindrical deformations of a Ricci-flat Kähler asymptotically cylindrical manifold. Some examples (motivated by [11]) are considered in §6.

1. Asymptotically cylindrical manifolds

A non-compact Riemannian manifold (M, g) is called *asymptotically cylindrical* with cross-section Y if

(1) M can be decomposed as a union $M = M_{\mathrm{cpt}} \cup_Y M_e$ of a compact manifold M_{cpt} with boundary Y and an end M_e diffeomorphic to half-cylinder $[1,\infty) \times Y$, the two pieces attached via $\partial M_{\mathrm{cpt}} \cong \{1\} \times Y$, and

(2) The metric g on M is asymptotic, along the end, to a product cylindrical metric $g_0 = dt^2 + g_Y$ on $[1,\infty) \times Y$,

$$\lim_{t\to\infty} (g - g_0) = 0, \qquad \lim_{t\to\infty} \nabla_0^k g = 0, \quad k = 1, 2, \dots,$$

where t is the coordinate on $[1,\infty)$ and ∇_0 denotes the Levi–Civita connection of g_0.

Note that the cross-section Y is always a compact manifold. We shall sometimes assume that t is extended to a smooth function defined on all of M, so that $t \geq 1$ on the end and $0 \leq t \leq 1$ on the compact piece of M_{cpt}.

Remark 1.1. Setting $x = e^{-t}$, one can attach to M a copy of Y corresponding to $x = 0$ and obtain a compactification $\mathbf{M} = M \cap Y$ 'with boundary at infinity'. Then x defines a normal coordinate near the boundary of $\mathbf{M}$. The metric g is defined on the interior of $\mathbf{M}$ and blows up in a particular way at the boundary,

$$g = \Big(\frac{dx}{x}\Big)^2 + \tilde{g}, \tag{1}$$

for some semi-positive definite symmetric bilinear $\tilde{g}$ smooth on M and continuous on $\mathbf{M}$, such that $\tilde{g}|_{x=0} = g_Y$. Metrics of this latter type are called 'exact b-metrics' and are studied in [16].

Our main result concerns a Kähler version of the asymptotically cylindrical Riemannian manifolds which we now define. Suppose that M has an integrable complex structure J and write Z for the resulting complex manifold. The basic idea is to replace a real parameter t along the cylindrical end by a complex parameter, $t + i\theta$ say, where $\theta \in S^1$. Thus in the complex setting the asymptotic model for a cylindrical end of Z takes a slightly special form $\mathbb{R}_{>0} \times S^1 \times D$, for some compact complex manifold D. Respectively, the normal coordinate $x = e^{-t}$ becomes the real part of a holomorphic local coordinate $z = e^{-t-i\theta}$ taking values in the punctured unit disc $\Delta^* = \{0 < |z| < 1\} \subset \mathbb{C}$. It follows that the complex structure on the cylindrical end is asymptotic to the product $\Delta^* \times D$ and the complex manifold Z is *compactifiable*, $Z = \overline{Z} \setminus D$, where $\overline{Z}$ is a compact complex manifold of the same dimension as Z and D is a complex submanifold of codimension 1 in $\overline{Z}$ with holomorphically trivial normal bundle $N_{D/\overline{Z}}$.

The local complex coordinate z on $\overline{Z}$ vanishes to order one precisely on D and a tubular neighbourhood $\overline{Z}_e = \{|z| < 1\}$ is a local deformation family for D,

$$\pi : \overline{Z}_e \to \Delta, \qquad D = \pi^{-1}(0), \tag{2}$$

where π denotes the holomorphic map defining the coordinate z. Note that the cylindrical end $Z_e = \overline{Z}_e \setminus D$ is diffeomorphic (as a real manifold) but not in general biholomorphic to $\mathbb{R}_{>0} \times S^1 \times D$ as the complex structure on the fibre $\pi^{-1}\{z\}$ depends on z.

Remark 1.2. If $H^{0,1}(\overline{Z}) = 0$ then the local map (2) extends to a holomorphic fibration $\overline{Z} \to \mathbb{C}P^1$ (cf. [6, pp.34–35]).

A product Kähler metric, with respect to a product complex structure on $\mathbb{R} \times S^1 \times D$, has Kähler form $a^2 dt \wedge d\theta + \omega_D$, where ω_D is a Kähler form on D and a is a positive function of t, θ. We shall be interested in the situation when the product Kähler metric is *Ricci-flat*; then a is a constant and can be absorbed by rescaling the variable t.

We say that a Kähler metric on Z is *asymptotically cylindrical* if its Kähler form ω can be expressed on the end $Z_e = \overline{Z}_e \setminus D \subset Z$ as

$$\omega|_{Z_e} = \omega_D + dt \wedge d\theta + \varepsilon,$$

for some closed form $\varepsilon \in \Omega^2(Z_e)$ decaying, with all derivatives, to zero uniformly on $S^1 \times D$ as $t \to \infty$. An asymptotically cylindrical Kähler metric defines an asymptotically cylindrical Riemannian metric on the underlying real manifold.

We shall sometimes refer to Kähler metrics by their Kähler forms.

Proposition 1.1. *Let Z be a compactifiable complex manifold as defined above. If ω is an asymptotically cylindrical Kähler metric on M then the decaying term on Z_e is* exact*,*

$$\omega|_{Z_e} = \omega_D + dt \wedge d\theta + d\psi. \tag{3}$$

Proof. We can write $\varepsilon = \varepsilon_0(t) + dt \wedge \varepsilon_1(t)$, where $\varepsilon_0(t), \varepsilon_1(t)$ are 1-parameter families of, respectively, 2-forms and 1-forms on the cross-section $S^1 \times D$. As ε is closed, $\varepsilon_0(t)$ must be closed for each t and $\frac{\partial}{\partial t}\varepsilon_0(t) = d_{S^1 \times D}\varepsilon_1(t)$. As ε_1 decays exponentially fast, we have $\varepsilon_0(t) = \int_\infty^t d_{S^1 \times D}\varepsilon_1(s)ds$ and the integral converges absolutely. So we can write

$$\varepsilon = d_{S^1 \times D} \int_\infty^t \varepsilon_1(s)ds + dt \wedge \varepsilon_1(t)$$

which is an exact differential of a 1-form $\psi = -\int_t^\infty \varepsilon_1(s)ds$ on Z_e. □

Recall that by Yau's solution of the Calabi conjecture a compact Kähler manifold admits Ricci-flat Kähler metrics if and only if its first Chern class vanishes [21]. Moreover, the Ricci-flat Kähler metric is uniquely determined by the cohomology class of its Kähler form. Ricci-flat Kähler manifolds are sometimes called Calabi–Yau manifolds.

Remark 1.3. There is an alternative way to define the Calabi–Yau manifolds using the holonomy reduction. The holonomy group of a Riemannian $2n$-manifold is the group of isometries of a tangent space generated by parallel transport using the Levi–Civita connection over closed paths based at a point. The holonomy group can be identified with a subgroup of $SO(2n)$ if the manifold is orientable. If the holonomy of a Riemannian $2n$-manifold is contained in $SU(n) \subset SO(2n)$ then the manifold has an integrable complex structure J, so that with respect to J the metric is Ricci-flat Kähler. The converse is in general not true unless the manifold is simply-connected.

A version of the Calabi conjecture for asymptotically cylindrical Kähler manifolds is proved in [20, Thm. 5.1] and [11, §§2–3]. It can be stated as the following.

Theorem 1.2. (cf. [11, Thms. 2.4 and 2.7]) *Suppose that $Z = \overline{Z} \smallsetminus D$ is a compactifiable complex n-fold as defined above, such that D is an anticanonical divisor on $\overline{Z}$ and the normal bundle of D is holomorphically trivial and $b^1(\overline{Z}) = 0$. Let $\overline{g}$ be a Kähler metric on $\overline{Z}$ and denote by g_D the Ricci-flat Kähler metric on D in the Kähler class defined by the embedding in $\overline{Z}$.*

Then $Z = \overline{Z} \smallsetminus D$ admits a complete Ricci-flat Kähler metric g_Z. The Kähler form of g_Z can be written, on the cylindrical end of Z, as in (3) *with ω_D the Kähler form of g_D.*

If, in addition, $\overline{Z}$ and D are simply-connected and there is a closed real 2-dimensional submanifold of $\overline{Z}$ meeting D transversely with non-zero intersection number then the holonomy of g is $SU(n)$.

Note that an anticanonical divisor D admits Ricci-flat Kähler metrics as $c_1(D) = 0$ by the adjunction formula. The result in [11] is stated for threefolds, but the proof generalizes to an arbitrary dimension by a change of notation. We consider examples arising by application of the above theorem in §6. A consequence of the arguments in [11] is that if an asymptotically cylindrical Kähler metric ω is Ricci-flat then the 1-form $\psi \in \Omega^1(M_e)$ in (3) can be taken to be decaying, with all derivatives, at an *exponential* rate $O(e^{-\lambda t})$ as $t \to \infty$, for some $0 < \lambda < 1$ depending on g_D. Furthermore, if ω and $\tilde{\omega}$ are asymptotically cylindrical Ricci-flat metrics on Z such that $\tilde{\omega} = \omega + i\partial\bar{\partial}u$ for some $u \in C^\infty(Z)$ decaying to zero as $t \to \infty$ then $\omega = \tilde{\omega}$ [11, Prop. 3.11].

Let (M, g) be an asymptotically cylindrical Riemannian manifold. A local deformation $g + h$ of g is given by a field of symmetric bilinear forms satisfying $|h|_g < 1$ at each point, so that $g + h$ is a well-defined metric. Suppose that $g + h$ is asymptotically cylindrical. Then there is a well-defined symmetric bilinear form h_Y on Y obtained as the limit of h as $t \to \infty$ and h_Y is a deformation of the limit g_Y of g, in particular $|h_Y|_{g_Y} < 1$. The h_Y defines via the obvious projection $\mathbb{R} \times Y \to Y$ a t-independent symmetric bilinear form on the cylinder, still denoted by h_Y. Let $\rho : \mathbb{R} \to [0, 1]$ denote a smooth function, such that $\rho(t) = 1$, for $t \geq 2$, and $\rho(t) = 0$, for $t \leq 1$. In view of the remarks in the previous paragraph we shall be interested in the class of metrics which are asymptotically cylindrical at an exponential rate and deformations h satisfying $h - \alpha h_Y = e^{-\mu t}\tilde{h}$ for some $\mu > 0$ and a bounded $\tilde{h}$. Given an exponentially asymptotically cylindrical metric g, a deformation $g + h$ 'sufficiently close' to g is understood in the sense of sufficiently small Sobolev norms of $\tilde{h}$ and h_Y, where the Sobolev norms are chosen to dominate the uniform norms on M and Y, respectively.

We now state our first main result in this paper.

Theorem 1.3. *Let $W = \overline{W} \setminus D$, where $\overline{W}$ is a compact complex manifold and D is a smooth anticanonical divisor on $\overline{W}$ with holomorphically trivial normal bundle. Let g be an asymptotically cylindrical Ricci-flat Kähler metric on W. Suppose that all the compactifiable infinitesimal deformations of the complex manifold W are integrable (arise as tangent vectors to paths of deformations).*

Then any Ricci-flat asymptotically cylindrical metric on W sufficiently close to g is Kähler with respect to some compactifiable deformation of the complex structure on W.

The additional conditions for the holonomy reduction given in Theorem 1.2 are topological and we deduce from Theorem 1.3.

Corollary 1.4. *Assume that $W = \overline{W} \setminus D$ satisfies the hypotheses of Theorem 1.3. Suppose further that W, $\overline{W}$, and D are simply-connected and so the metric g has holonomy $SU(n)$, $n = \dim_{\mathbb{C}} W$. Then any Ricci-flat asymptotically cylindrical metric on W close to g also has holonomy $SU(n)$.*

Our second main result determines the dimension of the moduli space of the asymptotically cylindrical Ricci-flat Kähler metrics and is given by Theorem 5.1 below.

2. Infinitesimal Ricci-flat deformations

Before dealing with the moduli of asymptotically cylindrical Ricci-flat metrics we recall, in summary, some results on the moduli problem for the Ricci-flat metrics on a compact manifold. The case of a compact manifold is standard and further details can be found in [3, Ch. 12] and references therein.

A natural symmetry group of the equation $\mathrm{Ric}(g) = 0$ for a metric g on a compact manifold X is the group $\mathsf{Diff}X$ of diffeomorphisms of X. It is also customary to identify a metric g with $a^2 g$, for any positive constant a. This is equivalent to considering only the metrics such that X has volume 1. The moduli space of Ricci-flat metrics on X is defined as the space of orbits of all the solutions g of $\mathrm{Ric}(g) = 0$ in the action of $\mathsf{Diff}(X) \times \mathbb{R}_{>0}$,

$$g \mapsto a^2 \phi^* g, \qquad \phi \in \mathsf{Diff}X,\ a > 0,$$

or, equivalently, the space of all ($\mathsf{Diff}X$)-orbits of the solutions of $\mathrm{Ric}(g) = 0$ such that $\mathrm{vol}_g(X) = 1$. The tangent space at g to an orbit of g under the action of $\mathsf{Diff}X$ is the image of the first order linear differential operator

$$\delta_g^* : V^\flat \in \Omega^1(X) \to \tfrac{1}{2}\mathcal{L}_V g \in \mathrm{Sym}^2 T^*X, \tag{4}$$

where $\mathcal{L}$ denotes the Lie derivative. The operator δ_g^* may be equivalently expressed as the symmetric component of the Levi–Civita covariant derivative $\nabla_g : \Omega^1(X) \to \Omega^1 \otimes \Omega^1(X)$, for the metric g,

$$\nabla_g \eta = \delta_g^* \eta + \frac{1}{2} d\eta, \quad \eta \in \Omega^1(X). \tag{5}$$

The L^2 formal adjoint of δ_g^* is therefore given by

$$\delta_g : h \in \mathrm{Sym}^2 T^*X \to \nabla_g^* h \in \Omega^1(X).$$

The operator δ_g^* is overdetermined-elliptic with finite-dimensional kernel and closed image and there is an L^2-orthogonal decomposition

$$\mathrm{Sym}^2 T^*X = \mathrm{Ker}\,\delta_g \oplus \mathrm{Im}\,\delta_g^*.$$

The equation $\delta_g h = 0$ defines a local transverse slice for the action of $\mathsf{Diff}(X)$.

The infinitesimal Ricci-flat deformations h of a Ricci-flat g preserving the volume are obtained by linearizing the equation $\mathrm{Ric}(g+h) = 0$ at $h = 0$, imposing an additional condition $\int_X \mathrm{tr}_g\, h\, \nu_g = 0$, where ν_g is the volume form of g. By a theorem of Berger and Ebin, the space of infinitesimal Ricci-flat deformations of g is given by a system of linear PDEs

$$(\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h = 0, \qquad \delta_g h = 0, \qquad \mathrm{tr}_g\, h = 0. \tag{6}$$

Here $\overset{\circ}{R}_g$ is a linear map induced by the Riemann curvature and acting on symmetric bilinear forms, $\overset{\circ}{R}_g h(X,Y) = \sum_i h(R_g(X,e_i)Y, e_i)$ (e_i is an orthonormal basis). The first equation in (6) is elliptic and so the solutions of (6) form a finite-dimensional space.

Suppose that every infinitesimal deformation h satisfying (6) arises as the tangent vector to a path of Ricci-flat metrics. Then it turns out that a neighbourhood of g in the moduli space of Ricci-flat metrics on X is diffeomorphic to the quotient of the solutions space of (6) by a finite group. This finite group depends on the isometry group of g and the moduli space is an orbifold of dimension equal to the dimension of the solution space of (6).

Now suppose that the manifold X has an integrable complex structure, J say, and the Ricci-flat metric g on X is Kähler, with respect to J. Then any deformation h of g may be written as a sum $h = h_+ + h_-$ of Hermitian form h_+ and skew-Hermitian form h_- defined by the conditions $h_\pm(Jx, Jy) = \pm h(x,y)$. Furthermore, the operator $\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g$ preserves the subspaces of Hermitian and skew-Hermitian forms.

The skew-Hermitian forms h_- may be identified, via

$$g(x, Iy) = h_-(x, Jy), \tag{7}$$

with the symmetric real endomorphisms I satisfying $IJ + JI = 0$. Thus $J + I$ is an almost complex structure and the endomorphism I may be regarded as a $(0,1)$-form with values in the holomorphic tangent bundle $T^{1,0}X$. Then one has

$$\delta_g h_- = -J \circ (\bar{\partial}^* I). \tag{8}$$

In particular, $\delta h_- = 0$ if and only if I defines an class in $H^1(X, T^{1,0}X)$, that is I defines an infinitesimal deformation of the complex manifold (X, J) (see [9]). With the help of

Weitzenböck formula one can replace $\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g$ by the complex Laplacian for $(0,q)$-forms with values in $T^{1,0}X$

$$((\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h_-)(\cdot, J\cdot) = g(\cdot, (\Delta_{\bar\partial} I)\cdot).$$

Thus $(\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h_- = 0$ precisely when $I \in \Omega^{0,1}(T^{1,0}X)$ is harmonic.

Hermitian forms h_+ are equivalent, with the help of the complex structure, to the real differential (1,1)-forms

$$\psi(\cdot,\cdot) = h_+(\cdot, J\cdot). \tag{9}$$

The Weitzenböck formula yields

$$((\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h_+)(\cdot, J\cdot) = \Delta\psi,$$

for a Ricci-flat metric g, thus h_+ satisfies the first equation in (6) if and only if $\psi \in \Omega^{1,1}$ is harmonic. The other two equations in (6) become

$$\operatorname{tr}_g h_+ = \langle \psi, \omega\rangle_g, \quad \text{and} \quad \delta_g h_+ = -d^*\psi, \tag{10}$$

where ω denotes is the Kähler form of g.

3. The moduli problem and a transverse slice

We want to extend the set-up of the moduli space for Ricci-flat metrics outlined in §2 to the case when (M,g) is an asymptotically cylindrical Ricci-flat manifold. For this, we require a Banach space completion for sections of vector bundles associated to the tangent bundle of M and we use Sobolev spaces with exponential weights. A weighted Sobolev space $e^{-\mu t}L^p_k(M)$ is, by definition, the space of all functions $e^{-\mu t}f$ such that $f \in L^p_k(M)$. The norm of $e^{-\mu t}f$ in $e^{-\mu t}L^p_k(M)$ is defined to be the L^p_k-norm of f. The definition generalizes in the usual way to vector fields, differential forms, and, more generally, tensor fields on M with the help of the Levi–Civita connection. Note that if $k - \dim M/p > \ell$, for some integer $\ell \geq 0$, then there is a bounded inclusion map between Banach spaces $L^p_k(M) \to C^\ell(M)$ because (M,g) is complete and has bounded curvature [2, §2.7].

The weighted Sobolev spaces $e^{-\mu t}L^p_k(M)$ are not quite convenient for working with bounded sections that are asymptotically t-independent but not necessarily decaying to zero on the end of M. We shall use slightly larger spaces which we call, following a prototype in [1], the *extended weighted Sobolev spaces*, denoted $W^p_{k,\mu}(M)$.

As before, use Y to denote the cross-section of the end of M. Fix once and for all a smooth cut-off function $\rho(t)$ such that $0 \leq \rho(t) \leq 1$, $\rho(t) = 0$ for $t \leq 1$, and $\rho(t) = 1$ for $t \geq 2$. Define

$$W^p_{k,\mu}(M) = e^{\mu t}L^p_k(M) + \rho(t)L^p_k(Y)$$

where, by abuse of notation, $L^p_k(Y)$ in the above formula is understood as a space of t-independent functions on the cylinder $\mathbb{R}\times Y$ pulled back from Y. Elements in $\rho(t)L^p_k(Y)$ are well-defined as functions supported on the end of M. The norm of $f + \rho(t)f_Y$ in $W^p_{k,\mu}(M)$ is defined as the sum of the $e^{\mu t}L^p_k(M)$-norm of f and the L^p_k-norm of f_Y (where f_Y is interchangeably considered as a function on Y). More generally, the extended weighted Sobolev space of $W^p_{k,\mu}$ sections of a bundle associated to TM is defined in a similar manner using parallel transport in the t direction defined by the Levi–Civita connection.

We shall need some results of the elliptic theory and Hodge theory for an asymptotically cylindrical manifold (M, g). The Hodge Laplacian Δ on M is an instance of an *asymptotically translation invariant* elliptic operator. That it, Δ can be written locally on the end of M in the form $a(t, y, \partial_t, \partial_y)$, where a is smooth in $(t, y) \in \mathbb{R} \times Y$ and polynomial in ∂_t, ∂_y. The coefficients $a(t, y, \partial_t, \partial_y)$ have a t-independent asymptotic model $a_0(y, \partial_t, \partial_y)$ on the cylinder $\mathbb{R} \times Y$, so that $a(t, y, \partial_t, \partial_y) - a_0(y, \partial_t, \partial_y)$ decays to zero, together with all derivatives, as $t \to \infty$.

Proposition 3.1. *Let (M, g) be an oriented asymptotically cylindrical manifold with Y a cross-section of M and let Δ denote the Hodge Laplacian on M. Then there exists $\mu_1 > 0$ such that for $0 < \mu < \mu_1$ the following holds.*

(i) The Hodge Laplacian defines bounded Fredholm linear operators

$$\Delta_{\pm\mu} : e^{\pm\mu t} L^p_{k+2} \Omega^r(M) \to e^{\pm\mu t} L^p_k \Omega^r(M)$$

with index, respectively, $\pm(b^r(Y) + b^{r-1}(Y))$. The image of $\Delta_{\pm\mu}$ is, respectively, the subspace of the forms in $e^{\pm\mu t} L^p_k \Omega^r(M)$ which are L^2-orthogonal to the kernel of $\Delta_{\mp\mu}$.

(ii) Any r-form $\eta \in \operatorname{Ker} \Delta_g \cap e^{\mu t} L^p_{k+2} \Omega^r(M)$ is smooth and can be written on the end $\mathbb{R}_+ \times Y$ of M as

$$\eta|_{\mathbb{R}\times Y} = \eta_{00} + t\eta_{10} + dt \wedge (\eta_{01} + t\eta_{11}) + \eta', \tag{11}$$

where η_{ij} are harmonic forms on Y of degree $r - j$ and the r-form η' is $O(e^{-\mu_1 t})$ with all derivatives. In particular, any L^2 harmonic form on M is $O(e^{-\mu_1 t})$. The harmonic form η is closed and co-closed precisely when $\eta_{10} = 0$ and $\eta_{11} = 0$, i.e. when η is bounded.

Proof. For (i) see [13] or [16]. In particular, the last claim is just a Fredholm alternative for elliptic operators on weighted Sobolev spaces.

The clause (ii) is an application of [15, Theorem 6.2]. Cf. also [16, Prop. 5.61 and 6.14] proved with an assumption that the b-metric corresponding to g is smooth up to and on the boundary of M at infinity. The last claim is verified by the standard integration by parts argument. □

Corollary 3.2. *Assume the hypotheses and notation of Proposition 3.1. Suppose also that the metric g on M is asymptotic to a product cylindrical metric on $\mathbb{R}_+ \times Y$ at an* exponential *rate $O(e^{-\mu_1 t})$. Then for $\xi \in e^{-\mu t} L^p_k \Omega^r(M)$, the equation $\Delta\eta = \xi$ has a solution $\eta \in e^{-\mu t} L^p_{k+2} \Omega^r(M) + \rho(t)(\eta_{00} + dt \wedge \eta_{01})$ if and only if ξ is L^2-orthogonal to $\mathcal{H}^r_{\mathrm{bd}}(M)$, where $0 < \mu < \mu_1$ and $\mathcal{H}^r_{\mathrm{bd}}(M)$ denotes the space of bounded harmonic r-forms on M.*

Proof. The hypotheses on g and μ implies that the Laplacian defines a Fredholm operator

$$e^{-\mu t} L^p_{k+2} \Omega^r(M) + \rho(t)(\eta_{00} + dt \wedge \eta_{01}) \to e^{-\mu t} L^p_{k+2} \Omega^r(M). \tag{12}$$

It follows from Proposition 3.1 that the index of (12) is zero and the kernel is $\mathcal{H}^r_{\mathrm{bd}}(M)$. Further, if $\xi \in e^{-\mu t} L^p_{k+2} \Omega^r(M) + \rho(t)(\eta_{00} + dt \wedge \eta_{01})$ then we find from (11) that $d\xi$ and

$d^*\xi$ decay to zero as $t \to \infty$. Recall from Proposition 3.1 that any bounded harmonic form is closed and co-closed and then the standard Hodge theory argument using integration by parts is valid and shows that the image of (12) is L^2-orthogonal to $\xi \in \mathcal{H}^r_{\mathrm{bd}}(M)$. But as the codimension of the image of (12) is equal to $\dim \mathcal{H}^r_{\mathrm{bd}}(M)$ the image must be precisely the L^2-orthogonal complement of $\mathcal{H}^r_{\mathrm{bd}}(M)$ in $e^{-\mu t}L^p_{k+2}\Omega^r(M)$. □

For an asymptotically cylindrical n-dimensional manifold (M, g), let $\mathsf{Diff}_{p,k,\mu}M$ (where $k - n/p > 1$, $0 < \mu < \mu_1$) denote the group of locally L^p_k diffeomorphisms of M generated by $\exp_V$, for all vector fields V on M that can be written as $V = V_0 + \rho(t)V_Y$, where $V_0 \in e^{-\mu t}L^p_k$ and a t-independent V_Y is defined by a Killing field for g_Y. Respectively, $V_Y^\flat$ is defined by a harmonic 1-form on Y (cf. (14) below). Also require that V has a sufficiently small C^1 norm on M, so that that $\exp_V$ is a well-defined diffeomorphism. Denote by $\mathsf{Metr}_{p,k,\mu}(g)$ (where $k - n/p > 0$, $0 < \mu < \mu_1$) the space of deformations $h + \rho(t)h_Y$ of g where $h \in e^{-\mu t}L^p_k$, $|h|_g < 1$ at every point of M, and $\delta_{g_Y}h_Y = 0$, $\mathrm{tr}_{g_Y} h_Y = 0$. ($g_Y$ is the limit of g as defined in §1.) Then $\mathsf{Diff}_{p,k+1,\mu}M$ acts on $\mathsf{Metr}_{p,k,\mu}(g)$ by pull-backs and the linearization of the action is given by the operator δ^*_g on weighted Sobolev spaces,

$$\delta^*_g : e^{-\mu t}L^p_{k+1}\Omega^1(M) + \rho(t)\mathcal{H}^1(Y) \to e^{-\mu t}L^p_k \operatorname{Sym}^2 T^*M. \tag{13}$$

It will be convenient to replace the last two equations in (6) and instead use another local slice equation for the action of $\mathsf{Diff}_{p,k,\mu}M$

$$\delta_g h + \frac{1}{2} d\,\mathrm{tr}_g\, h = 0.$$

A transverse slice defined by the operator $\delta_g + \frac{1}{2}d\,\mathrm{tr}_g$ was previously used for different classes of complete non-compact manifolds in [4, I.1.C and I.4.B]. The operator $\delta_g + \frac{1}{2}d\,\mathrm{tr}_g$ satisfies a useful relation:

$$(2\delta_g + d\,\mathrm{tr}_g)\delta^*_g = 2\nabla^*_g\nabla_g - \nabla^*_g d + d\,\mathrm{tr}_g\,\delta^*_g = 2\nabla^*_g\nabla_g - d^*_g d - dd^*_g = \Delta_g, \tag{14}$$

where Δ_g is the Hodge Laplacian and we used the Weitzenböck formula for 1-forms on a Ricci-flat manifold in the last equality.

Proposition 3.3. *Assume that Y is connected and that $k - \dim M/p > 1$, $0 < \mu < \mu_1$, where μ_1 is defined in Prop. 3.1 for the Laplacian on differential forms on M. Then there is a direct sum decomposition into closed subspaces*

$$\mathsf{Metr}_{p,k,\mu}(g) = \delta^*_g(e^{-\mu t}L^p_{k+1}\Omega^1(M) + \rho(t)\mathcal{H}^1(Y)) \oplus \big(\mathrm{Ker}(\delta_g + \frac{1}{2}d\,\mathrm{tr}_g) \cap \mathsf{Metr}_{p,k,\mu}(g)\big). \tag{15}$$

Proof. Any bounded harmonic 1-form on M is in $e^{-\mu t}L^p_{k+1}\Omega^1(M) + \rho(t)\mathcal{H}^1(Y)$ by [16, Prop. 6.16 and 6.18] (see also Prop. 4.3 below) and because Y is connected. For any $\eta \in e^{-\mu t}L^p_{k+1}\Omega^1(M) + \rho(t)\mathcal{H}^1(Y)$, $\nabla_g\eta$ decays on the end of M, so the standard integration by parts applies to show that the bounded harmonic 1-forms on M are parallel with respect to g. As the bounded harmonic 1-forms on M are closed we obtain using (5) and (14) that these are in the kernel of δ^*. It follows that the two subspaces in (15) have trivial intersection.

By the definition of $\mathsf{Metr}_{p,k,\mu}(g)$ the 'constant term' h_Y of h satisfies $\delta_Y h_Y = 0$ and $\operatorname{tr}_g h_Y = 0$. If $\eta \in \mathcal{H}^1_{\mathrm{bd}}(M)$ and $h \in \mathsf{Metr}_{p,k,\mu}(g)$ then the 1-form $\langle \eta, h\rangle_g$ decays along the end of M and we can integrate by parts

$$\langle \eta, \delta_g h + \frac{1}{2} d \operatorname{tr}_g h\rangle_{L^2} = \langle \delta_g^* \eta, h\rangle_{L^2} + \frac{1}{2}\langle d^*\eta, \operatorname{tr}_g h\rangle_{L^2} = 0.$$

Thus the image $(\delta_g + \frac{1}{2} d \operatorname{tr}_g)\mathsf{Metr}_{p,k,\mu}(g)$ is L^2-orthogonal to $\mathcal{H}^1_{\mathrm{bd}}$. By Corollary 3.2 the equation $\Delta\eta = (\delta + \frac{1}{2} d \operatorname{tr}_g)h$ has a solution η in $e^{-\mu t} L^p_{k+1,\mu}\Omega^1(M) + \rho(t)\mathcal{H}^1_Y$ and so

$$\delta_g^*\eta - h \in \operatorname{Ker}(\delta_g + \frac{1}{2} d \operatorname{tr}_g) \cap \mathsf{Metr}_{p,k,\mu}(g)$$

which gives the required decomposition $h = \delta_g^*\eta + (\delta_g^*\eta - h)$. □

Proposition 3.4. *Assume that p, k, μ are as in Proposition 3.3. Let $\tilde{g}$ an asymptotically cylindrical deformation of g. If $\tilde{h} = \tilde{g} - g \in \mathsf{Metr}_{pk,\mu}(g)$ is sufficiently small in $W^p_{k,\mu} \operatorname{Sym}^2 T^*M$ then there exists $\phi \in \mathsf{Diff}_{p,k+1,\mu} M$ such that $\phi^*\tilde{g} = g + h$, for some $h \in \mathsf{Metr}_{p,k,\mu}(g)$ with $(\delta_g + \frac{1}{2} d \operatorname{tr}_g)h = 0$.*

Proof. If the desired ϕ is close to the identity then $\phi = \exp_V$ for a vector field V on M with small $e^{-\mu t} L^p_k$ norm. We want to show that the map

$$\mathsf{Diff}_{p,k+1,\mu} \times \{h \in \mathsf{Metr}_{p,k,\mu}(g) : (\delta_g + \frac{1}{2} d \operatorname{tr}_g)h = 0\} \to \mathsf{Metr}_{p,k,\mu}(g)$$

defined by

$$(V, h) \mapsto \exp_V^*(g + h) - g$$

is a onto a neighbourhood of $(0, 0)$. The linearization of $(DF)_{(0,0)}$ is given by $(V, h) \mapsto \delta_g^*(V^\flat) + h$ and is surjective by (15). By the implicit function theorem for Banach spaces, a solution (V, h) of $F(V, h) = \tilde{g}$ exists, whenever $\tilde{g} - g$ is sufficiently small. □

Finally, we obtain the system of linear PDEs describing the infinitesimal Ricci-flat deformations of an asymptotically cylindrical metric transverse to the action of the diffeomorphism group on the asymptotically cylindrical metrics.

Theorem 3.5. *Suppose that (M, g) is a Ricci-flat asymptotically cylindrical Riemannian manifold, but not a cylinder $\mathbb{R} \times Y$, and $g(s)$, $|s| < \varepsilon$ ($\varepsilon > 0$) is a smooth path of asymptotically cylindrical Ricci-flat metrics on M with $g(0) = g$. Suppose also that $g(s) - g \in \mathsf{Metr}_{p,k,\mu}(g)$, with p, k, μ as in Proposition 3.3. Then there is a smooth path $\psi(s) \in \mathsf{Diff}_{p,k,\mu} M$, so that $h = \frac{d}{ds}|_{s=0}\big[\psi(s)^* g(s)\big]$ satisfies the equations*

$$(\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h = 0, \tag{16a}$$

$$(\delta_g + \frac{1}{2} d \operatorname{tr}_g)h = 0, \tag{16b}$$

Furthermore, if every bounded solution h of (16) is the tangent vector at g to a path of Ricci-flat asymptotically cylindrical metrics on M then the moduli space is an orbifold.

The dimension of this orbifold is equal to the dimension of the space of solutions of (16) *that are bounded on* M*.*

Proof. Applying Proposition 3.4 for each $g(s)$, we can find a path of diffeomorphisms in $\psi(s) \in \mathsf{Diff}_{p,k,\mu}M$ so that the slice equation (16b) holds for h.

The linearization of $\mathrm{Ric}(g+h) = 0$ in h is $\nabla_g^*\nabla_g h - 2\delta_g^*\delta_g h - \nabla_g d\,\mathrm{tr}_g\, h - 2\overset{\circ}{R}_g h = 0$. which becomes equivalent to $(\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h = 0$ in view of of (16b) and (5).

The last claim follows similarly to the case of a compact base manifold, cf. [3, 12.C]. It can be shown using Proposition 3.3 that the infinitesimal action of the identity component of the group $I(M,g)$ of isometries of g in $\mathsf{Diff}^p_{k,\mu}M$ is trivial on the slice $(\delta_g + \frac{1}{2}d\,\mathrm{tr}_g)h = 0$. As M is not a cylinder, it has only one end [18] and we show in Lemma 3.6 below that $I(M,g)$ is compact. It follows that a neighbourhood of the orbit of g in the orbit space $\mathsf{Metr}_{p,k,\mu}(g)/\mathsf{Diff}_{p,k,\mu}M$ is homeomorphic to a finite quotient of the kernel of $\delta_g + \frac{1}{2}d\,\mathrm{tr}_g$. □

Lemma 3.6. *Let* (M,g) *be an asymptotically cylindrical manifold with a connected cross-section* Y *(that is,* M *has only one end). Then the group* $I(M,g)$ *of isometries of* M *is compact.*

Proof. It is a well-known result the isometry group $I(M,g)$ of any Riemannian manifold (M,g) is a finite-dimensional Lie group and if a sequence $T_k \in I(M,g)$ is such that, for some $P \in M$, $T_k(P)$ is convergent then T_k has a convergent subsequence [17].

For an asymptotically cylindrical (M,g), it is not difficult to check that there is a choice of point P_0 on the end of M and $r > 0$, so that $M_{0,r} = \{P \in M : \mathrm{dist}(P_0,P) > r\}$ is connected but for any P_1 such that $\mathrm{dist}(P_0,P_1) > 3r$ the set $M_{1,r} = \{P \in M : \mathrm{dist}(P_1,P) > r\}$ is not connected. It follows that for any sequence $\tilde{T}_k \in I(M,g)$ we must have $\mathrm{dist}(P_0,\tilde{T}_k(P_0)) \le 3r$ and hence $\tilde{T}_k$ has a convergent subsequence. □

4. Infinitesimal Ricci-flat deformations of asymptotically cylindrical Kähler manifolds

We now specialize to the *Kähler* Ricci-flat metrics. It is known [10] that if an infinitesimal deformation h of a Ricci-flat Kähler metric on a *compact* manifold satisfies the Berger–Ebin equations (6) then the Hermitian and skew-Hermitian components h_+ and h_- of h also satisfy (6). In this section we prove a version of this result for the asymptotically cylindrical manifolds.

Proposition 4.1. *Let* W *be a compactifiable complex manifold with* g *an asymptotically cylindrical Ricci-flat Kähler metric on* W*, as defined in* §*1. Suppose that an asymptotically cylindrical deformation* $h \in \mathsf{Metr}$ *of* g *satisfies* (16)*. Then the skew-Hermitian component* h_- *of* h *also satisfies* (16)*.*

Proof. The proof is uses the same ideas as in the case of for a compact manifold ([10, §7] or [3, Lemma 12.94]). The operator $\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g$, for a Kähler metric g, preserves the subspaces of Hermitian and skew-Hermitian forms, so $(\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h_- = 0$. Recall from §2 that the latter equation implies that the form $I \in \Omega^{0,1}(T^{1,0})$ corresponding to h_- via (7) is harmonic, $\Delta_{\bar\partial} I = 0$. An argument similar to that of Proposition 3.1 shows that a bounded harmonic section I satisfies $\bar\partial I = 0$ which implies $\delta_g h_- = 0$ by (8) and, further, $\delta_g - \frac{1}{2} d\,\mathrm{tr}_g\, h_- = 0$ as a skew-Hermitian deformation h_- is automatically trace-free. □

Proposition 4.2. *Any infinitesimal Ricci-flat asymptotically cylindrical deformation $h \in \mathsf{Metr}_{p,k,\mu}(g)$ of a Ricci-flat Kähler asymptotically cylindrical metric g on W is the sum of a Hermitian and a skew-Hermitian infinitesimal deformation.*

The space of skew-Hermitian infinitesimal Ricci-flat asymptotically cylindrical deformations of g is isomorphic to the space of bounded harmonic $(0,1)$-forms on W with values in $T^{1,0}(W)$.

The space of Hermitian infinitesimal Ricci-flat asymptotically cylindrical deformations of g is isomorphic to the orthogonal complement of the Kähler form of g in the space of bounded harmonic real $(1,1)$-forms on W.

Proof. Only the last statement requires justification. Let ω denote the Kähler form of g.

Recall from §2 that the equation $(\nabla_g^*\nabla_g - 2\overset{\circ}{R}_g)h_+ = 0$ satisfied by a Hermitian infinitesimal Ricci-flat asymptotically cylindrical deformations h_+ is equivalent to the condition that $\psi \in \Omega^{1,1}(W)$ defined in (9) is harmonic $\Delta\psi = 0$. Hence $\delta_g h_+ = 0$ by (10) and Proposition 3.1 and so the second equation in (16) tells us that $\langle\psi, \omega\rangle_g = \mathrm{const}$, in view of (10). Considering the limit as $t \to \infty$ and the definition of $\mathsf{Metr}_{p,k,\mu}(g)$ we find that the latter constant must be zero. □

Thus in order to find the dimension of the space of infinitesimal Ricci-flat deformations of an asymptotically cylindrical Kähler metric, we may consider the Hermitian and a skew-Hermitian cases separately. This is done in the next subsection.

4.1. Bounded harmonic forms and logarithmic sheaves

It is well-known that harmonic forms on a compact manifold are identified with the de Rham cohomology classes via Hodge theorem. On a non-compact manifold W one can consider the usual de Rham cohomology $H^*(W)$ and also the de Rham cohomology $H^*_c(W)$ *with compact support*. The latter is the cohomology of the de Rham complex of compactly supported differential forms. We shall write $b^r(W) = \dim H^r(W)$ and $b^r_c(W) = \dim H^r_c(W)$, for the respective Betti numbers. There is a natural inclusion homomorphism $H^r_c(W) \to H^r(W)$ whose image is the subspace of the de Rham cohomology classes representable by closed forms with compact support; the dimension of this subspace will be denoted by $b^r_0(W)$.

Proposition 4.3. *Let (W, g) be an oriented asymptotically cylindrical manifold. Then the space $\mathcal{H}_{L^2}(W)$ of L^2 harmonic r-forms on W has dimension $b^r_0(W)$. The space $\mathcal{H}_{\mathrm{bd}}(W)$ of bounded harmonic r-forms on W has dimension $b^r(W) + b^r_c(W) - b^r_0(W)$.*

Proof. For the claim on L^2 harmonic forms see [1, Prop. 4.9] or [14, §7]. In the case when an asymptotically cylindrical metric g corresponds to an exact b-metric smooth up to the boundary at infinity (see Remark 1.1), the dimension of bounded harmonic forms is a direct consequence of [16, Prop. 6.18] identifying a Hodge-theoretic version of the long exact sequence

$$\ldots \to H^{r-1}(Y) \to H^r_c(W) \to H^r(W) \to h^r(Y) \to \ldots . \tag{17}$$

The argument of [16, Prop. 6.18] can be adapted for arbitrary asymptotically cylindrical metrics; the details will appear in [12]. □

If W is an asymptotically cylindrical Kähler manifold then there is a well-defined subspace $\mathcal{H}^{1,1}_{\mathrm{bd},\mathbb{R}}(W) \subset \mathcal{H}^2_{\mathrm{bd}}(W)$ of bounded harmonic real forms of type $(1,1)$. The bounded harmonic 2-forms in the orthogonal complement of $\mathcal{H}^{1,1}_{\mathrm{bd},\mathbb{R}}(W)$ are the real and imaginary parts of bounded harmonic $(0,2)$-forms. We shall denote the complex vector space of bounded harmonic $(0,2)$-forms on W by $\mathcal{H}^{0,2}_{\mathrm{bd}}(W)$.

The space of bounded harmonic real (1,1)-forms on W orthogonal to the Kähler form ω therefore has dimension $b^r(W) + b^r_c(W) - b^r_0(W) - 1 - 2\dim_{\mathbb{C}} \mathcal{H}^{0,2}_{\mathrm{bd}}(W)$.

Now for the skew-Hermitian infinitesimal deformations. Recall from §1 that the definition of an asymptotically cylindrical Ricci-flat Kähler manifold (M, J, ω) includes the condition that a complex manifold $W = (M, J)$ is compactifiable. That is, there exist a compact complex n-fold $\overline{W}$ and a compact complex $(n-1)$-dimensional submanifold D in $\overline{W}$, so that W is isomorphic to $\overline{W} \setminus D$. We saw in Proposition 4.1 that any skew-Hermitian Ricci-flat asymptotically translation-invariant deformation of ω can be expressed as a $\bar\partial$- and $\bar\partial^*$-closed symmetric $(0,1)$-form I with values in the holomorphic tangent bundle of W. A $\bar\partial$- and $\bar\partial^*$-closed such I, not necessarily symmetric, defines an infinitesimal deformation $J + I$ of the integrable complex structure J on W. The deformations given by skew-symmetric such forms I correspond to the bounded harmonic $(2,0)$-forms on W.

Let z denote a complex coordinate on $\overline{W}$ so that D is defined by the equation $z = 0$, as in §1. Let $T_{\overline{W}}$ denote the sheaf of holomorphic local vector fields on $\overline{W}$. The subsheaf of the holomorphic local vector fields whose restrictions to D are tangent to D is denoted by $T_{\overline{W}}(\log D)$ and called the *logarithmic tangent sheaf.* The form I in general has a simple pole precisely along D and defines a class in the Čech cohomology $H^1(T_{\overline{W}}(\log D))$. The classical Kodaira–Spencer–Kuranishi theory of deformations of the holomorphic structures on compact manifolds [9] has an extension for the compactifiable complex manifolds; the details can be found in [8]. In this latter theory, the cohomology groups $H^i(T_{\overline{W}}(\log D))$ have the same role as the cohomology of tangent sheaves for the compact manifolds. In particular, the isomorphisms classes of infinitesimal deformations of W are canonically parameterized by classes in $H^1(T_{\overline{W}}(\log D))$. These classes arise from the actual deformations of W is the obstruction space $H^2(T_{\overline{W}}(\log D))$ vanishes.

Thus the space of the skew-Hermitian Ricci-flat asymptotically cylindrical deformations I of the Ricci-flat Kähler asymptotically cylindrical metric g on W is identified as a subspace of the infinitesimal compactifiable deformations of W. The real dimension of this subspace is $2(\dim_{\mathbb{C}} H^1(T_{\overline{W}}(\log D)) - \dim_{\mathbb{C}} \mathcal{H}^{0,2}_{\mathrm{bd}}(W))$.

5. The asymptotically cylindrical Ricci-flat deformations

In this section, we show that every infinitesimal Ricci-flat deformation of an asymptotically cylindrical Ricci-flat Kähler manifold is tangent to a genuine deformation.

Theorem 5.1. *Let* (W, g) *be as in Theorem 1.3. Then every bounded solution* h *of* (16) *arises as* $h = \frac{d}{ds}|_{s=0} g(s)$ *for some path of asymptotically cylindrical Ricci-flat metrics on* W *with* $g(0) = g$. *The moduli space of asymptotically cylindrical Ricci-flat deformations of* g *is an orbifold of real dimension*

$$2 \dim_{\mathbb{C}} H^1(T_W(\log D)) + b^2(W) + b^2_c(W) - b^2_0(W) - 1 - 4 \dim_{\mathbb{C}} \mathcal{H}^{2,0}_{\mathrm{bd}}(W).$$

Proof. By the hypotheses of Theorem 1.3, there is a manifold $\mathcal{M}$ of small compactifiable deformations of W, so that $H^1(T_{\overline{W}}(\log D))$ is the tangent space to $\mathcal{M}$ at W. The data of the compactifiable deformations of W includes the deformations of $\overline{W}$ [8]. Let ω' be a Kähler metric on $\overline{W}$. By the results of Kodaira and Spencer [9], for a family of sufficiently small deformations $\overline{I}$ of a compact complex manifold $\overline{W}$, there is a family of forms $\omega'(\overline{J}+\overline{I})$ on $\overline{W}$ depending smoothly on $\overline{I}$ and such that $\omega'(\overline{J}) = \omega'$ and $\omega'(\overline{J} + \overline{I})$ defines a Kähler metric with respect to a perturbed complex structure $\overline{J}+\overline{I}$. Using the methods of [11, §3], we can construct from $\omega'(\overline{J} + \overline{I})$ a smooth family $\omega(J + I)$ of asymptotically cylindrical Kähler metrics (not necessarily Ricci-flat) on the respective deformations of $W = \overline{W} \setminus D$.

Consider a vector bundle $\mathcal{V}$ over $\mathcal{M}$ whose fibre over $\overline{I} \in \mathcal{M}_{\overline{W}}$ is the space of bounded harmonic (1,1)-forms with respect to the Kähler metric $\omega(\overline{J} + \overline{I})$. The task of integrating an infinitesimal Ricci-flat deformation of the given asymptotically cylindrical Kähler metric ω on W is expressed by the complex Monge–Ampére equation (with parameters) for a function u on W

$$(\omega(J + I) + \beta + i\partial\bar{\partial}u)^n - e^{f_{I,\beta}}(\omega(J + I) + \beta)^n = 0, \tag{18}$$

where $n = \dim_{\mathbb{C}} W$ and $\beta \in \mathcal{V}$ is a bounded harmonic real (1,1)-form with respect to the Kähler metric $\omega(J + I)$ and orthogonal to $\omega(J + I)$. The operators ∂,$\bar{\partial}$ in (18) are those defined by $J + I$.

If $I = 0$ and $\beta = 0$ then $u = 0$ is a solution of (18) as the metric ω is Ricci-flat. Consider the right-hand side of (18) as a function $f(I, \beta, u)$ where the domain of u is a version of extended weighted Sobolev space $E^p_{k,\mu}(W) = e^{-\mu t} L^p_{k+2}(W) + \{\rho(t)(at + b) \mid a, b \in \mathbb{R}\}$ for a sufficiently small $\mu > 0$ ($Y = S^1 \times D$ is the cross-section of W in the present case and D is connected). The linearization of f in u at $u = 0$ is the Laplacian for functions on the asymptotically cylindrical Kähler manifold (W, ω). A dimension counting argument similar to that in Corollary 3.2 shows that this latter Laplacian defines a *surjective* linear map $E^p_{k,\mu}(W) \to e^{-\mu t} L^p_k(W)$. The Laplacian has a one-dimensional kernel given by the constant functions on W, so we reduce the domain for u by taking the L^2 orthogonal complement of the constants. Then the implicit function theorem applies to $f(I, \beta, u)$

and defines a smooth family $u = u(I, \beta)$ so that $f(I, \beta, u(I, \beta)) = 0$ for every small I, β in the respective spaces of bounded harmonic forms. This defines a smooth family of Ricci-flat metrics $\omega(J + I) + \beta + i\partial\bar{\partial}u(I, \beta)$ tangent to the infinitesimal deformations identified in the previous section. □

6. Examples

In this section, we consider some examples of asymptotically cylindrical Ricci-flat Kähler manifolds arising by application of Theorem 1.2 and compute the dimension of the moduli space for their asymptotically cylindrical Ricci-flat deformations. This is done by considering appropriate long exact sequences and applying vanishing theorems to determine the dimensions of cohomology groups appearing in Theorem 5.1.

6.1. Rational elliptic surfaces

An elliptic curve $C = \mathbb{C}/\Lambda$ embeds in the complex projective plane as a cubic curve in the anticanonical class. Choosing another non-singular elliptic curve C' in $\mathbb{C}P^2$ we obtain a pencil $aC + bC'$, $a : b \in \mathbb{C}P^1$. Assuming that C' is chosen generically and blowing up the 9 intersection points $C \cap C'$ we obtain an algebraic surface $\tilde{S}$ so that the proper transform $\tilde{C}$ of C is in the anticanonical class, $\tilde{C} \in |-K_{\tilde{S}}|$, and $\tilde{C}$ has a holomorphically trivial normal bundle, in particular $\tilde{C} \cdot \tilde{C} = 0$. Then, by Theorem 1.2, the quasiprojective surface $S = \tilde{S} \setminus \tilde{C}$ has a complete Ricci-flat Kähler metric asymptotic to the flat metric on the half-cylinder $\mathbb{R}_{>0} \times S^1 \times \mathbb{C}/\Lambda$ with cross-section a 3-dimensional torus. Although in this example the divisor at infinity is not simply-connected it can be easily checked that S is simply-connected and the asymptotically cylindrical Ricci-flat Kähler metric on S has holonomy $SU(2)$ (cf. [11, Theorem 2.7]). It is well-known that a Ricci-flat Kähler metric on a complex surface is hyper-Kähler.

Furthermore, S is topologically a 'half of the K3 surface' in the sense that there is an embedding of a 3-torus T^3 in the K3 surface so that the complement of this T^3 consists of two components, each homeomorphic to S. From the arising Mayer–Vietoris exact sequence, we find that $b^2(S) = b^2_c(S) = 11$ using also the Poincaré duality. The long exact sequence (17) with $W = S$ and $Y = T^3$ yields $b^2_0 = 8$.

As S is simply-connected with holonomy $SU(2)$ there is a nowhere-vanishing parallel (hence holomorphic) $(2, 0)$-form Ω on S. Any other $(2, 0)$-form on S can be written as $f\Omega$ for some complex function f and $f\Omega$ will be a bounded harmonic form if and only if the real and imaginary parts of f are bounded harmonic functions, hence constants by the maximum principle. Thus $\dim_{\mathbb{C}} \mathcal{H}^{2,0}_{\mathrm{bd}}(S) = 1$.

The dimensions of $H^1(T_S(\log \tilde{C}))$ and $H^2(T_S(\log \tilde{C}))$ are obtained by taking the cohomology of the exact sequences

$$0 \to T_{\tilde{S}}(-C) \to T_{\tilde{S}}(\log \tilde{C}) \to T_{\tilde{C}} \to 0$$

and

$$0 \to T_{\tilde{S}}(\log \tilde{C}) \to T_{\tilde{S}} \to N_{\tilde{C}/\tilde{S}} \to 0$$

(see [8]). Using Serre duality [6] we find that $H^2(T_{\tilde S}(-C)) = H^0(\Omega^1_{\tilde S}) = H^{0,1}(S) = 0$, hence $H^2(T_S(\log \tilde C))$ vanishes and the compactifiable deformations of S are unobstructed. Note that any small deformation of $\tilde S$ is the blow-up of a small deformation of the cubic $\tilde C$ in $\mathbb{C}P^2$ ([5] or [7, Theorem 9.1]). Therefore, $\dim_{\mathbb{C}} H^1_{\tilde S} = 10$ and we deduce that $\dim_{\mathbb{C}} H^1(T_S(\log \tilde C)) = 10$.

Now by Theorem 5.1 the moduli space of asymptotically cylindrical Ricci-flat deformations of S has dimension 29. All these deformations are hyper-Kähler with holonomy $SU(2)$.

6.2. Blow-ups of Fano threefolds

A family of examples of asymptotically cylindrical Ricci-flat Kähler threefolds is constructed in [11, §6] using Fano threefolds. A Fano threefold is a non-singular complex threefold V with $c_1(V) > 0$. Any Fano threefold is necessarily projective and simply-connected. A generically chosen anticanonical divisor D_0 in V is a K3 surface [19]. Let $D_1 \in |-K_V|$ be another K3 surface such that $D_0 \cap D_1 = C$ is a smooth curve.

The blow-up of V along C is a Kähler complex threefold $(\overline{W}, \omega')$ and the proper transform $D \subset \overline{W}$ of D_0 is an anticanonical divisor on $\overline{W}$ with the normal bundle of D holomorphically trivial. The complement $W = \overline{W} \setminus D$ is simply-connected.

Thus W is topologically a manifold with a cylindrical end $\mathbb{R}_{>0} \times S^1 \times D$. By Theorem 1.2 W admits a complete Ricci-flat Kähler metric ω, with holonomy $SU(3)$. The metric ω is asymptotic on the end of W to the product of the standard flat metric on $\mathbb{R}_{>0} \times S^1$ and a Yau's hyper-Kähler metric on D.

By the Weitzenböck formula, the Hodge Laplacian Δ for the $(2,0)$-forms on a Ricci-flat Kähler manifold can be expressed as $\Delta = \nabla_g^* \nabla_g$. The quantity $\langle \nabla\eta, \eta\rangle_g$ for a bounded harmonic form η decays on the end of W, so we can integrate by parts to show that a bounded harmonic $(2,0)$-form is parallel. But the holonomy of the metric ω is $SU(3)$ which has no invariant elements in $\Lambda^{2,0}\mathbb{C}^3$. Therefore, W admits no parallel $(2,0)$-forms and thus no bounded harmonic $(2,0)$-forms.

The dimension of the moduli space for asymptotically cylindrical Ricci-flat deformations of ω then becomes $2\dim_{\mathbb{C}} H^1(T_W(\log D)) + b^2(W) + b^2_c(W) - b^2_0(W) - 1$.

The dimensions of $H^i(T_W(\log D))$, $i = 1, 2$, are obtained from the two long exact sequences similar to §6.1. To verify that the compactifiable deformations of W are unobstructed note that $H^2(T_{\overline{W}}) = H^1(\Omega^1_{\overline{W}}(-D)) = 0$ by the Kodaira vanishing theorem and $H^1(N_{D/\overline{W}}) = H^{1,0}(D) = 0$. It is shown in [11, §8] that $b^2(W) = \rho(V)$ and $h^{2,1}(\overline{W}) = h^{2,1}(V) + g(V)$, where $g(V) = -K_V^3/2 + 1$ is the genus of V and $\rho(V)$ is the Picard number. Taking the cohomology of $0 \to T_{\overline{W}}(-D) \to T_{\overline{W}}(\log D) \to T_D \to 0$ we obtain $\dim_{\mathbb{C}} H^1(T_W(\log D)) = 20 + h^{2,1}(V) + g(V) - \rho(V)$. From the long exact sequence (17) we find that $b^2(W) + b^2_c(W) - b^2_0(W) = b^2(W) + 1$.

Thus the dimension of the moduli space for W in this example is given by

$$b^3(V) + 2g(V) - \rho(V) + 40$$

in terms of standard invariants of the Fano threefold.

Acknowledgements: The work on this paper began while the author was visiting Université Paris XII. I am grateful to Frank Pacard for the invitation and many helpful discussions. I also thank Akira Fujiki and Nick Shepherd-Barron for discussions on the logarithmic sheaves. I would like to thank the organizers for the invitation to speak at the Gökova Geometry-Topology Conference, supported by TÜBİTAK and NSF.

References

[1] M.F. Atiyah, V.K. Patodi, and I.M. Singer. *Spectral asymmetry and Riemannian geometry, I.* Math. Proc. Camb. Phil. Soc. **77** (1975), 97–118.

[2] T. Aubin. *Some nonlinear problems in Riemannian geometry.* Springer, 1998.

[3] A. Besse. *Einstein manifolds.* Springer, 1987.

[4] O. Biquard. *Métriques d'Einstein asymptotiquement symétriques.* Astérisque **265** (2000).

[5] A. Fujiki and S. Nakano. Supplement to "On the inverse of monoidal transformation". *Publ. Res. Inst. Math. Sci.* **7** (1971/72), 637–644.

[6] P. Griffiths and J. Harris. *Principles of algebraic geometry.* John Wiley & Sons, 1978.

[7] E. Horikawa. On deformations of holomorphic maps. III. *Math. Ann.* **222** (1976), 275–282.

[8] Y. Kawamata. *On deformations of compactifiable complex manifolds.* Math. Ann. **235** (1978), 247–265.

[9] K. Kodaira and D.C. Spencer. *On deformations of complex analytic structures. I, II, III.* Ann. of Math. (2) **67** (1958), 328–466 and **71** (1960), 43–76.

[10] N. Koiso. *Einstein metrics and complex structures.* Inventiones Math. **73** (1983), 71–106.

[11] A.G. Kovalev. *Twisted connected sums and special Riemannian holonomy.* J. Reine Angew. Math. **565** (2003), 125–160.

[12] A.G. Kovalev. *Metrics of special holonomy and gluing theorems*, in preparation.

[13] R.B. Lockhart and R.C. McOwen. *Elliptic differential operators on noncompact manifolds.* Ann. Scuola Norm. Sup. Pisa Cl. Sci. **12** (1985), 409–447.

[14] R.B. Lockhart. *Fredholm, Hodge and Liouville theorems on noncompact manifolds.* Trans. Amer. Math. Soc. **301** (1987), 1–35.

[15] V.G. Maz'ja and B.A. Plamenevski. *Estimates on L^p and Hölder classes on the Miranda–Agmon maximum principle for solutions of elliptic boundary value problems with singular points on the boundary.* Math. Nachr. **81** (1978), 25–82. English translation: Amer. Math. Soc. Transl. Ser. 2 **123** (1984) 1–56.

[16] R.B. Melrose. *The Atiyah-Patodi-Singer index theorem.* A K Peters Ltd., Wellesley, MA, 1993.

[17] S. B. Myers and N. E. Steenrod. The group of isometries of a Riemannian manifold. *Ann. of Math. (2)* **40** (1939), 400–416.

[18] S. Salur. *Asymptotically cylindrical Ricci-flat manifolds.* math.DG/0410063 version 2, January 2006, to appear in Proc. Amer. Math. Soc.

[19] V.V. Shokurov. *Smoothness of a general anticanonical divisor on a Fano variety.* Izv. Akad. Nauk SSSR Ser. Mat. **43** (1979), 430–441. English translation: Math. USSR-Izvestia **14** (1980), no. 2, 395–405.

[20] G. Tian and S.-T. Yau. *Complete Kähler manifolds with zero Ricci curvature. I.* J. Amer. Math. Soc., **3** (1990), 579–609.

[21] S.-T. Yau. *On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I.* Comm. Pure Appl. Math., **31** (1978), 339–411.

DPMMS, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, England

E-mail address: a.kovalev@dpmms.cam.ac.uk

Proceedings of 12^{th} Gökova
Geometry-Topology Conference
pp. 157 – 164

Published online at
GokovaGT.org

Deformations of special Lagrangian submanifolds; An approach via Fredholm alternative

Sema Salur

Dedicated to the memory of Raoul Bott.

Abstract. In an earlier paper, [9], we showed that the moduli space of deformations of a smooth, compact, orientable special Lagrangian submanifold L in a symplectic manifold X with a non-integrable almost complex structure is a smooth manifold of dimension $H^1(L)$, the space of harmonic 1-forms on L. We proved this first by showing that the linearized operator for the deformation map is surjective and then applying the Banach space implicit function theorem. In this paper, we obtain the same surjectivity result by using a different method, the Fredholm Alternative, which is a powerful tool for compact operators in linear functional analysis.

1. Introduction

In [8], McLean showed that the moduli space of nearby submanifolds of a smooth, compact special Lagrangian submanifold L in a Calabi-Yau manifold X is a smooth manifold and its dimension is equal to the dimension of $H^1(L)$, the space of harmonic 1-forms on L. Special Lagrangian submanifolds have attracted much attention after Strominger, Yau and Zaslow proposed a mirror Calabi-Yau construction using special Lagrangian fibration [11]. For more information about special Lagrangian submanifolds and examples, see [3], [4], [6].

One can also define special Lagrangian submanifolds of symplectic manifolds equipped with a nowhere vanishing complex valued $(n, 0)$-form, [9]. Such symplectic manifolds were studied recently by Smith, Thomas and Yau in [10].

In [9], we showed that the moduli space of special Lagrangian deformations of L in a symplectic manifold with non-integrable almost complex structure is also a smooth manifold of dimension $b_1(L)$, the first Betti number of L. In order to prove this result we first modified the definition of a special Lagrangian submanifold for symplectic manifolds, extended the parameter space of deformations and showed that the linearization of the deformation map is onto and finally applied the infinite dimensional Banach space implicit function theorem.

In this paper, we obtain the same result by a different approach. In particular, we show that the linearized operator for the deformation map is invertible by using Fredholm Alternative, a technique from linear functional analysis.

2. Deformations of special Lagrangian submanifolds

Let $(M^{2n}, \omega, J, g, \Omega)$ be a Calabi-Yau manifold with a Kähler 2-form ω, a complex structure J, a compatible Riemannian metric g and a nowhere vanishing holomorphic $(n, 0)$-form Ω which is normalized with respect to ω. Then one can define a special Lagrangian submanifold of M.

Definition 2.1. An n-dimensional submanifold $L \subseteq M$ is *special Lagrangian* if L is Lagrangian (i.e. $\omega|_L \equiv 0$) and $Im(\Omega)$ restricts to zero on L. Equivalently, $\mathrm{Re}(\Omega)$ restricts to be the volume form on L with respect to the induced metric.

McLean studied the deformations of compact special Lagrangian submanifolds in Calabi-Yau manifolds and proved the following theorem, [8].

Theorem 2.2. *The moduli space of all deformations of a smooth, compact, orientable special Lagrangian submanifold L in a Calabi-Yau manifold M within the class of special Lagrangian submanifolds is a smooth manifold of dimension equal to dim*$(H^1(L))$.

One natural generalization of McLean's result is for symplectic manifolds. Now let (X, ω, J, g, ξ) denote a $2n$-dimensional symplectic manifold X with symplectic 2-form ω, an almost complex structure J which is tamed by ω, the compatible Riemannian metric g and a nowhere vanishing complex valued $(n, 0)$-form $\xi = \mu + i\beta$, where μ and β are real valued n-forms. Here we also take ξ to be normalized with respect to ω.

Note that the holomorphic form Ω is a closed form on Calabi-Yau manifolds. However, on a symplectic manifold which is equipped with a nowhere vanishing complex-valued $(n, 0)$ form ξ, it is not necessarily closed.

For more general special Lagrangian calibrations, one can introduce an additional term $e^{i\theta}$, where for each fixed angle θ we have a corresponding form $e^{i\theta}\xi$ and its associated geometry. Here θ is the phase factor of the calibration and using this as the new parameter one can define special Lagrangian submanifolds in a symplectic manifold and study their deformations, [9].

Definition 2.3. An n-dimensional submanifold $L \subseteq X$ is *special Lagrangian* if L is Lagrangian (i.e. $\omega|_L \equiv 0$) and $Im(e^{i\theta}\xi)$ restricts to zero on L, for some $\theta \in \mathbb{R}$. Equivalently, $Re(e^{i\theta}\xi)$ restricts to be the volume form on L with respect to the induced metric.

Now we recall the basics of a technique for compact operators in linear functional analysis, known as Fredholm Alternative. One can find more information about the subject in [5] and [7].

Let $\mathcal{X}$ and $\mathcal{Y}$ be two real Banach spaces.

Definition 2.4. A bounded linear operator $\mathcal{K} : \mathcal{X} \to \mathcal{Y}$ is called *compact* provided for each bounded sequence $\{u_k\}_{k=1}^{\infty}$ is precompact in $\mathcal{Y}$.

Now let H denote a real Hilbert space, with inner product $\langle \cdot, \cdot \rangle$.

Theorem 2.5. *Let $\mathcal{K} : H \to H$ be a compact operator. Then*

(i) $\ker(I - \mathcal{K})$ *is finite dimensional,*
(ii) Range$(I - \mathcal{K})$ *is closed,*
(iii) Range$(I - \mathcal{K}) = \ker(I - \mathcal{K}^*)^\perp$,
(iv) $\ker(I - \mathcal{K}) = \{0\}$ *if and only if Range*$(I - \mathcal{K}) = H$
(v) dim $\ker(I - \mathcal{K}) =$ *dim* $\ker(I - \mathcal{K}^*)$.

Remark 2.6. Theorem 2.5 asserts in particular either

(a) for each $f \in H$, the equation $u - \mathcal{K}u = f$ has a unique solution

or else

(b) the homogeneous equation $u - \mathcal{K}u = 0$ has solutions $u \neq 0$.

In addition, should (a) obtain the space of solutions of the homogeneous problem is finite dimensional and the nonhomogeneous equation $u - \mathcal{K}u = f$ has a solution if and only if $f \in \ker(I - \mathcal{K}^*)^\perp$.

Now we prove the following theorem, [9], using the Fredholm Alternative:

Theorem 2.7. *Let L be a smooth, compact, orientable special Lagrangian submanifold of a symplectic manifold X. Then the moduli space of all deformations of L in X within the class of special Lagrangian submanifolds is a smooth manifold of dimension $H^1(L)$.*

Proof. Given a domain Ω, let $C^{k,\alpha}(\Omega)$ denote the Hölder norms defined as

$$C^{k,\alpha}(\Omega) = \{f \in C^k(\Omega) | \ [D^\gamma f]_{\alpha,\Omega} < \infty, |\gamma| \leqq k\}$$

where

$$[f]_{\alpha,\Omega} = \underset{x,y\in\Omega,\, x\neq y}{Sup} \frac{dist(f(x), f(y))}{(dist(x,y))^\alpha} \quad \text{in } \Omega.$$

Then for a small vector field V and a scalar $\theta \in \mathbb{R}$, we define the deformation map as follows,

$$F : C^{1,\alpha}(\Gamma(N(L))) \times \mathbb{R} \to C^{0,\alpha}(\Omega^2(L)) \oplus C^{0,\alpha}(\Omega^n(L))$$

$$F(V,\theta) = ((\exp_V)^*(-\omega), (\exp_V)^*(Im(e^{i\theta}\xi)).$$

Here $N(L)$ denotes the normal bundle of L, $\Gamma(N(L))$ the space of sections of the normal bundle, and $\Omega^2(L)$, $\Omega^n(L)$ denote the differential 2-forms and n-forms, respectively.

Since the symplectic form ω is closed on X and the restriction of $Im(e^{i\theta}\xi)$ is a top dimensional form on L the image of the deformation map F lies in the closed 2-forms and closed n-forms. So by Hodge decomposition we get

$$F : C^{1,\alpha}(\Gamma(N(L))) \times \mathbb{R} \to C^{0,\alpha}(d\Omega^1(L)) \oplus C^{0,\alpha}(d\Omega^{n-1}(L) \oplus \mathcal{H}^n(L))$$

where $d\Omega^{n-1}(L)$ denotes the space of exact n-forms and $\mathcal{H}^n(L)$ denotes the space of harmonic n-forms on L.

In [9], we computed the linearization of F at (0,0),

$$dF(0,0): C^{1,\alpha}(\Gamma(N(L))) \times \mathbb{R} \to C^{0,\alpha}(d\Omega^1(L)) \oplus C^{0,\alpha}(d\Omega^{n-1}(L) \oplus \mathcal{H}^n(L))$$

where

$$dF(0,0)(V,\theta) = \frac{\partial}{\partial t} F(tV, s\theta)|_{t=0,s=0} + \frac{\partial}{\partial s} F(tV, s\theta)|_{t=0,s=0}$$

$$= (-d(i_V\omega)|_L, (i_V d\beta + d(i_V\beta))|_L + \theta)$$

$$= (dv, \zeta + d*v + \theta), \quad \text{where } \zeta = i_V(d\beta)|_L.$$

Here i_V is the interior derivative and v is the dual 1-form to the vector field V with respect to the induced metric. For the details of local calculations see [8], [9].

Let $x_1, x_2, ..., x_n$ and $x_1, x_2, ..., x_{2n}$ be the local coordinates on L and X, respectively. Then for any given normal vector field $V = (V_1\frac{\partial}{\partial x_{n+1}}, ..., V_n\frac{\partial}{\partial x_{2n}})$ to L we can show that

$$\zeta = i_V(d\beta)|_L = -n(V_1 \cdot g_1 + ... + V_n \cdot g_n)\text{dvol}$$

where g_i $(0 < i \leq n)$ are combinations of coefficient functions in the connection-one forms.

One can decompose the n-form $\zeta = da + d^*b + h_2$ by using Hodge Theory and because ζ is a top dimensional form on L, ζ is closed and the equation becomes

$$dF(0,0)(V,\theta) = (dv, da + d*v + h_2 + \theta)$$

for some $(n-1)$-form a and harmonic n-form h_2. Also the harmonic projection for $\zeta = -n(V_1.g_1 + ... + V_n.g_n)\text{dvol}$ is given by $(\int_L -n(V_1.g_1 + ... + V_n.g_n)\text{dvol})\text{dvol}$ and therefore one can show that

$$da = -n(V_1.g_1 + ... + V_n.g_n)\text{dvol} + (n\int_L (V_1.g_1 + ... + V_n.g_n)\text{dvol})\text{dvol}$$

and

$$h_2 = (-n\int_L (V_1.g_1 + ... + V_n.g_n)\text{dvol})\text{dvol}.$$

One should note that the differential forms a and h_2 both depend on V.

The Implicit Function Theorem says that $F^{-1}(0,0)$ is a manifold and its tangent space at $(0,0)$ can be identified with the kernel of dF.

$$(dv) \oplus (\zeta + d*v + \theta) = (0,0)$$

which implies

$$dv = 0 \quad \text{and} \quad \zeta + d*v + \theta = da + d*v + h_2 + \theta = 0.$$

The space of harmonic n-forms $\mathcal{H}^n(L)$, and the space of exact n-forms $d\Omega^{n-1}(L)$, on L are orthogonal vector spaces by Hodge Theory. Therefore, $dv = 0$ and $da+d*v+h_2+\theta = 0$ is equivalent to $dv = 0$ and $d*v + da = 0$ and $h_2 + \theta = 0$.

One can see that the special Lagrangian deformations (the kernel of dF) can be identified with the 1-forms on L which satisfy the following equations:

$$\begin{aligned} (i) &\;\; dv = 0 \\ (ii) &\;\; d*(v+\kappa(v)) = 0 \\ (iii) &\;\; h_2 + \theta = 0. \end{aligned}$$

Here, $\kappa(v)$ is a linear functional that depends on v and h_2 is the harmonic part of ζ which also depends on v. These equations can be formulated in a slightly different way in terms of decompositions of v and $*a$.

If $v = dp + d^*q + h_1$ and $*a = dm + d^*n + h_3$ then we have

$$\begin{aligned} (i) &\;\; dd^*q = 0 \\ (ii) &\;\; \Delta(p \pm m) = 0 \\ (iii) &\;\; h_2 + \theta = 0. \end{aligned}$$

This formulation of the solutions will provide the proof of the surjectivity of the linearized operator without using $\kappa(v)$.

Now we show that the linearized operator is surjective at $(0,0)$. Recall that the deformation map is given as

$$F : C^{1,\alpha}(\Gamma(N(L))) \times \mathbb{R} \to C^{0,\alpha}(d\Omega^1(L)) \oplus C^{0,\alpha}(d\Omega^{n-1}(L) \oplus \mathcal{H}^n(L)).$$

Therefore, for any given exact 2-form x and closed n-form $y = u + z$ in the image of the deformation map (here u is the exact part and z is the harmonic part of y), we need to show that there exists a 1-form v and a constant θ that satisfy the equations,

$$\begin{aligned} (i) &\;\; dv = x \\ (ii) &\;\; d*(v+\kappa(v)) = u \\ (iii) &\;\; h_2 + \theta = z. \end{aligned}$$

Alternatively, we can solve the following equations for p, q and θ.

$$\begin{aligned} (i) &\;\; dd^*q = x \\ (ii) &\;\; \Delta(p \pm m) = *u \\ (iii) &\;\; h_2 + \theta = z, \end{aligned}$$

where the star operator $*$ in (ii) is defined on L.

For (i), since x is an exact 2-form we can write $x = d(dr + d^*s + \text{harmonic form})$ by Hodge Theory. Then one can solve (i) for q by setting $q = s$.

For (ii), since $\Delta m = d^* dm = d^* * a = *d * *a = \pm * da$,

$$\Delta(p \pm m) = \Delta p \pm *da,$$

where a depends on p and we obtain

$$\Delta p \pm (-n(V_1.g_1 + ... + V_n.g_n) + (n \int_L (V_1.g_1 + ... + V_n.g_n) \text{dvol})) = *u \tag{1}$$

We can show the solvability as follows: Since $V = (V_1, ..., V_n)$ is the dual vector field of the one form $v = dp + d^* q + h_1$ we can write the equation (1) as

$$\Delta p \pm (-n(v \cdot g) + (n \int_L (v \cdot g) \text{dvol})) = *u \tag{2}$$

$$\Delta p \pm (-n(dp + d^* q + h_1) \cdot g + (n \int_L (dp + d^* q + h_1) \cdot g \text{ dvol})) = *u, \tag{3}$$

where $v \cdot g$ represents the action of the one form v on the vector field $g = (g_1, .., g_n)$ and $n \int_L (dp + d^* q + h_1 \cdot g)\text{dvol}$ is the harmonic projection of $-n(dp + d^* q + h_1) \cdot g$.

Then we get

$$\Delta p \pm n(-(dp \cdot g) + \int_L dp \cdot g \text{ dvol}) = *u \mp n[-(d^* q + h_1) \cdot g + \int_L (d^* q + h_1) \cdot g \text{ dvol}].$$

For simplicity we put

$$*u \mp n[-(d^* q + h_1) \cdot g + \int_L (d^* q + h_1) \cdot g \text{ dvol}] = h.$$

Since $\int_L * u = 0$ and $\int_L (d^* q + h_1) \cdot g$ dvol is equal to the harmonic projection of $(d^* q + h_1) \cdot g$, we get $\int_L h = 0$.

Since L is a compact manifold without boundary, by Stoke's Theorem,

$$\int_L dp \cdot g \text{ dvol} = -\int_L p \cdot \text{div} g \text{ dvol}$$

and the equation becomes

$$\Delta p \pm n(-(dp \cdot g) - \int_L p \cdot \text{div} g \text{ dvol}) = h.$$

Then by adding and subtracting p from the equation

$$(\Delta - Id)p = [\pm n(-(dp \cdot g) - \int_L p \cdot \text{div} g \text{ dvol}) - p + h]$$

and

$$p = (\Delta - Id)^{-1}[.....]p + \overline{h} = \mathcal{K}(p) + \overline{h}$$

where $\overline{h} = (\Delta - Id)^{-1}h$.
Since

$$||(\Delta - Id)^{-1}\int_L p \cdot \text{div}g||_{L_1^2} \leq C|\int_L p \cdot \text{div}g| \leq C||p||_{L^2}$$

$\mathcal{K}(p)$ is a compact operator which takes bounded sets in L^2 to bounded sets in L_1^2. Also note that we assumed here $1 \notin \text{spec}(\Delta)$, and if this is not the case then we can modify the above argument by adding and subtracting λp, $\lambda \notin \text{spec}(\Delta)$ from the equation.

Next we show that the set of solutions of the equation

$$\Delta p \pm n(-(dp \cdot g) - \int_L p \cdot \text{div}g \text{ dvol}) = 0 \tag{4}$$

is constant functions and therefore of dimension 1. Note that this set of solutions also satisfy the equation $(Id - \mathcal{K})(p) = 0$.

Also note that $\int_L p \cdot \text{div}g$ dvol is a constant which depends on p. We denote this as $C(p)$. At maximum values of p, Δp will be negative which imply that $C(p) \leq 0$ and at minimum values of p, Δp will be positive which imply that $C(p) \geq 0$ so $C(p)$ should be zero. Then the maximum principle holds for the equation $\Delta p \pm n(-(dp \cdot g)) = 0$ and since L is a compact manifold without boundary the solutions of this equation are constant functions. Hence the dimension of the kernel of $(Id - \mathcal{K})$ is one.

Next we find the kernel of $(Id - \mathcal{K}^*)$.

$$\int_L (\Delta p \pm n(-(dp \cdot g) - \int_L p \cdot \text{div}g))q(y)dy \tag{5}$$

$$= \int_L p\Delta q(y) \pm n\int_L -(dp \cdot g)q(y)dy - n\int_L(\int_L p \cdot \text{div}g)q(y)dy.$$

$$= \int_L p(y)\Delta q \pm n\int_L +(p\text{div}(g \cdot q)(y)dy - n\int_L p(x) \cdot \text{div}g(x)\int_L q(y)dydx$$

$$= \int_L p(y)\Delta q \pm n\int_L +(p\text{div}(g \cdot q)(y)dy - n\int_L p(y) \cdot \text{div}g(y)\int_L q(x)dxdy$$

$$= \int_L p(y)(\Delta q \pm n(+\text{div}(g \cdot q) - \text{div}g\int_L q(x)dx)dy.$$

Since we assumed that $1 \notin \text{spec}(\Delta)$, dim ker(Id$-\mathcal{K}^*$)($\Delta-$Id) =dim ker(Id$-\mathcal{K}^*$) and the kernel of (Id$-\mathcal{K}^*$) is equivalent to the solution space of the equation

$$\Delta q \pm n(+\text{div}(g \cdot q) - \text{div}g\int_L q(x)dx) = 0. \tag{6}$$

By Fredholm Alternative, the dimension of this kernel is 1 and one can check that a constant function $q = 1$ satisfies this equation, therefore the kernel consists of constant functions. Moreover these functions satisfy the compatibility condition $\int h.q = 0$.

Then by Fredholm Alternative, Theorem 2.5, we can conclude the existence of solutions of the equation

$$\Delta p \pm (-n(V_1.g_1 + ... + V_n.g_n) + (n\int_L (V_1.g_1 + ... + V_n.g_n)\ \mathrm{dvol})) = *u.$$

(iii) is straightforward.

It follows from [9] that the image of the deformation map F_1 lies in $d\Omega^1(L)$ and the image of F_2 lies in $d\Omega^{n-1}(L) \oplus \mathcal{H}^n(L)$. So we conclude that dF is surjective at $(0,0)$. Also since both the index of $d + *d^*(v)$ and $d + *d^*(v + \kappa(v))$ are equal to $b_1(L)$ the dimension of tangent space of special Lagrangian deformations in a symplectic manifold is also $b_1(L)$, the first Betti number of L. By infinite dimensional version of the implicit function theorem and elliptic regularity, the moduli space of all deformations of L within the class of special Lagrangian submanifolds is a smooth manifold and has dimension $b_1(L)$.

□

Remark 2.8. One can study the deformations of special Lagrangian submanifolds in much more general settings. In a forthcoming paper we plan to study these deformations using the techniques which we developed recently for associative submanifolds of G_2 manifolds, [1], [2].

Acknowledgements. The author is grateful to Peng Lu for many useful conversations and valuable comments.

References

[1] Akbulut, S. and Salur, S., *Calibrated Manifolds and Gauge Theory*, math.GT/0402368, 2004.

[2] Akbulut, S. and Salur, S., *Associative Submanifolds of G_2 Manifolds*, math.GT/0412032, 2004.

[3] Bryant, R.L. *Some examples of special Lagrangian Tori*, Adv. Theor. Math. Phys. **3, no. 1**, (1999) 83–90.

[4] Bryant, R.L. *Calibrated embeddings in the special Lagrangian and coassociative cases*, Special issue in memory of Alfred Gray (1939–1998). Ann. Global Anal. Geom. **18, no. 3-4**, (2000) 405–435.

[5] Evans, L.C. *Partial Differential Equations*, AMS, Graduate Studies in Mathematics, **Vol.19**, (1998).

[6] Harvey, F.R. and Lawson, H.B. *Calibrated Geometries*, Acta. Math. **148** (1982), 47-157.

[7] Lax, P.D. *Functional Analysis*, Pure and Applied Mathematics, Wiley-Interscience Series of Texts, Monographs, and Tracts, 2001.

[8] McLean, R.C. *Deformations of calibrated submanifolds*, Comm. Anal. Geom. **6,** (1998), 705-747.

[9] Salur, S. *Deformations of Special Lagrangian Submanifolds*, Communications in Contemporary Mathematics **Vol.2, No.3** (2000), 365-372.

[10] Smith, I., Thomas, R.P., and Yau, S.T. *Symplectic conifold transitions* J. Differential Geom. **62**, **no.2**, (2002), 209–242.

[11] Strominger, A., Yau, S.T. and Zaslow, E., *Mirror Symmetry is T-Duality*, Nucl. Phys. **B479** (1996), 243-259.

Department of Mathematics, Northwestern University, IL, 60208
E-mail address: salur@math.northwestern.edu

Proceedings of 12th Gökova
Geometry-Topology Conference
pp. 165 – 174

Published online at
GokovaGT.org

Plane curves and contact geometry

Lenhard Ng

Dedicated to the memory of Raoul Bott, who taught me algebraic topology.

Abstract. We apply contact homology to obtain new results in the problem of distinguishing immersed plane curves without dangerous self-tangencies.

1. Introduction

The purpose of this manuscript is to show that contact geometry, and in particular Legendrian knot theory and contact homology, can be used to give new information about plane curves without dangerous self-tangencies. Throughout, the term "plane curve" will refer to an immersion $S^1 \to \mathbb{R}^2$ up to orientation-preserving reparametrization, i.e., an oriented immersed plane curve in $\mathbb{R}^2$.

Definition 1. A self-tangency of a plane curve is *dangerous* if the orientations on the tangent directions to the curve agree at the tangency. Two plane curves without dangerous self-tangencies are *safely homotopic* if they are homotopic through plane curves without dangerous self-tangencies.

A generic homotopy of plane curves may contain three types of singularities, of which one is the dangerous self-tangency; see Figure 1. Arnold [1, 2] initiated the study of plane curves up to safe homotopy, in particular introducing a function J^+ on plane curves without dangerous self-tangencies. In the literature, any function of plane curves without dangerous self-tangencies which does not change under safe homotopy is called a *J^+-type invariant.*

The key point of interest of plane curves without dangerous self-tangencies is their close link to contact geometry, first noted by Arnold. There is a natural way to associate to any such plane curve a Legendrian knot in $\mathcal{J}^1(S^1)$, the 1-jet space of S^1, which is a contact manifold. We call this the *conormal knot* of the plane curve. For details, see Section 2.1.

The conormal knot is a special case of a construction which associates a Legendrian submanifold to any embedded submanifold of any manifold, or to any immersed submanifold without dangerous self-tangencies. This construction has recently been applied to construct new invariants of knots in S^3, and potentially yields interesting isotopy invariants of arbitrary submanifolds; see [5] or [12] for an introduction.

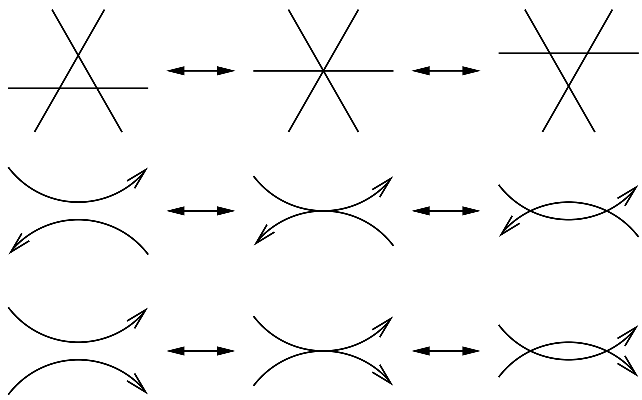


FIGURE 1. Singularities ("perestroikas") encountered in homotopies of plane curves: (a) triple point; (b) safe self-tangency; (c) dangerous self-tangency.

There are several well-known J^+-type invariants of plane curves, all arising from the conormal knot construction. The simplest is the Whitney index, or the degree of the Gauss map of the plane curve. This is invariant under safe homotopy since it is invariant more generally under regular homotopy; it also counts the number of times the conormal knot winds around the base of the solid torus $\mathcal{J}^1(S^1)$.

A more nontrivial J^+-type invariant, as observed by Arnold, is simply the knot type of the conormal knot in the solid torus. More interesting still, since the conormal is Legendrian, the contact planes along the conormal knot give it a framing, and so the framed knot type of the conormal knot is invariant under safe homotopy. The framing is measured by a number which is Arnold's original J^+ invariant.

To the author's knowledge, all previous work on J^+-type invariants is based on studying the framed knot type of the conormal knot. For instance, Goryunov [9] examined the space of finite type invariants of plane curves without dangerous self-tangencies, and Chmutov, Goryunov, and Murakami [4] introduced a J^+-type invariant in the form of a HOMFLY polynomial for the framed conormal knot.

On the other hand, two safely homotopic plane curves have conormal knots which are isotopic not just as framed knots, but as Legendrian knots. We will see that the Legendrian type of the conormal knot gives a finer classification of plane curves than the framed knot type. The fact (essentially) that Legendrian isotopy is a subtler notion than framed isotopy was famously demonstrated by Chekanov [3] for knots in $\mathbb{R}^3$, using a combinatorial form of Legendrian contact homology [6]. In this paper, we show that contact homology gives a similar result in our case.

Theorem 1 (see Propositions 3 and 4)**.** *There are (arbitrarily many) plane curves with the same framed conormal knot type which are not safely homotopic.*

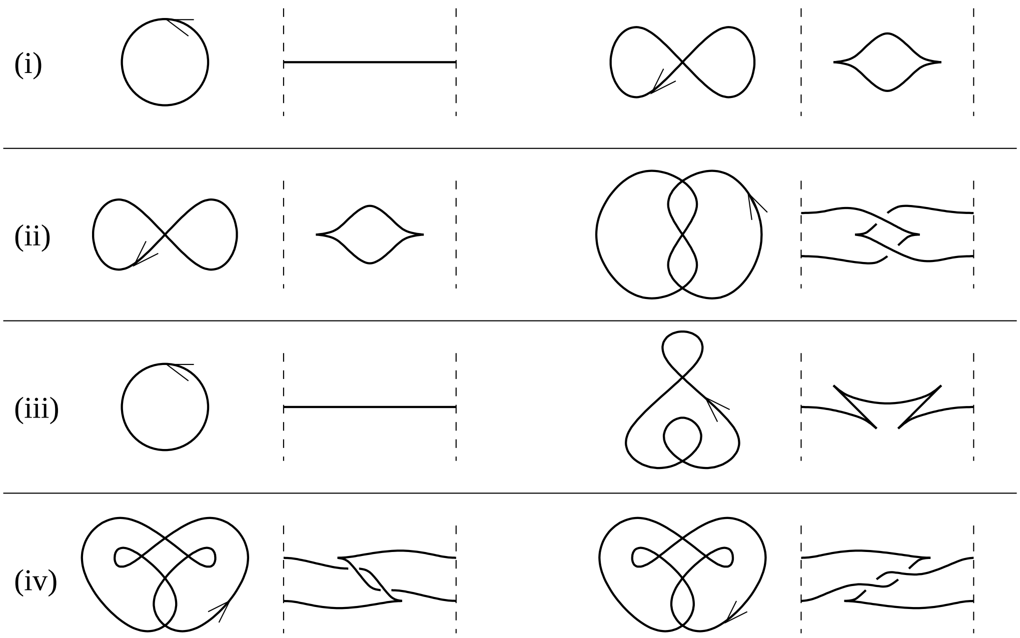


Figure 2. Pairs of plane curves, along with their conormal knots, that are distinguished by increasingly subtle invariants: (i) Whitney index; (ii) conormal knot type; (iii) framed conormal knot type (Arnold's J^+ invariant); (iv) Legendrian conormal knot type.

In the language of Legendrian knot theory, we can rephrase this result: there are arbitrarily many plane curves whose conormal knots all have the same classical invariants but are not Legendrian isotopic.

An example of a pair of plane curves satisfying the conditions in Theorem 1 is given by the bottom line of Figure 2. Note that this "pair" is actually the same plane curve but with different orientations. Proposition 3 uses contact homology to distinguish between these curves.

We review definitions in Section 2.1, and present an algorithm for drawing conormal knots in Section 2.2. Section 2.3 gives the proof of our main result, Theorem 1. In Section 2.4, we show that contact homology gives new information about loops of plane curves as well.

Acknowledgments

I am grateful to Tobias Ekholm and John Etnyre for useful discussions, and to Vladimir Chernov for correcting a substantial error in Section 2.4. I would also like to thank Selman Akbulut and Turgut Önder for organizing the very productive conference in Gökova. This work was supported by a Five-Year Fellowship from the American Institute of Mathematics.

2. Results and Proofs

2.1. The conormal knot

Let C be a plane curve. At each point $x \in C$, the orientation on C determines two unit vectors, v_x in the direction of C and w_x given by rotating v_x 90° counterclockwise.

Definition 2. The *conormal knot* of C is the subset of the unit cotangent bundle $ST^*\mathbb{R}^2$ given by

$$\{\xi \in ST^*\mathbb{R}^2 \,|\, \xi \text{ lies over some } x \in C \text{ and } \langle \xi, v_x \rangle = 0, \langle \xi, w_x \rangle = 1\}.$$

The conormal knot inherits an orientation from the orientation on C, since each point on C yields one point in the conormal knot.

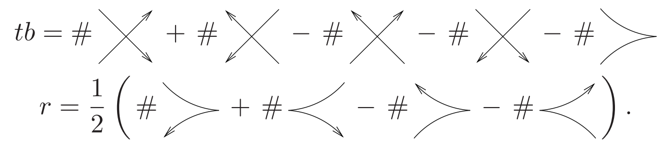
Here the metric on the fibers of $T^*\mathbb{R}^2$ used to define $ST^*\mathbb{R}^2$ is dual to the standard metric on $\mathbb{R}^2$. If C has no dangerous self-tangencies, then its conormal knot is embedded in $ST^*\mathbb{R}^2$, and so it makes sense to use the term "knot." We remark that the conormal knot is actually one half of the usual unit conormal bundle over the plane curve; the orientation of the plane curve, along with the orientation of $\mathbb{R}^2$, induces a coorientation on the curve, which picks out half of the conormal bundle.

The space $ST^*\mathbb{R}^2$ has a natural contact structure given by the kernel of the 1-form $\alpha = p_1\, dq_1 + p_2\, dq_2$, where q_1, q_2 are coordinates on $\mathbb{R}^2$ and p_1, p_2 are dual coordinates in the cotangent fibers. It is easy to check that the conormal knot K of any plane curve is Legendrian with respect to this contact structure, i.e., that $\alpha|_K = 0$.

Topologically, $ST^*\mathbb{R}^2 \cong S^1 \times \mathbb{R}^2$ is a solid torus, and it will be more useful for us to view it as the 1-jet space $\mathcal{J}^1(S^1) \cong T^*S^1 \times \mathbb{R}$. If we set coordinates θ, y, z on $\mathcal{J}^1(S^1) \cong (\mathbb{R}/2\pi\mathbb{Z}) \times \mathbb{R} \times \mathbb{R}$, then $\mathcal{J}^1(S^1)$ has a natural contact form $\alpha = dz - y\, d\theta$. We can identify $ST^*\mathbb{R}^2$ and $\mathcal{J}^1(S^1)$ by setting $\theta = \arg(p_1 + ip_2)$ (the argument of the vector (p_1, p_2)), $z = q_1 \cos\theta + q_2 \sin\theta$, $y = -q_1 \sin\theta + q_2 \cos\theta$; this map identifies the contact structures as well.

It is convenient to picture a Legendrian knot in $\mathcal{J}^1(S^1)$ in terms of its *front*, or projection to $(\mathbb{R}/2\pi\mathbb{Z}) \times \mathbb{R}$ given by the θz coordinates. In the subject, "front" is sometimes used in a different sense, namely as a cooriented plane curve with cusps; for clarity, we will avoid this connotation. A generic Legendrian knot has a front whose only singularities are double points and cusps. We can recover a Legendrian knot from its front by setting $y = dz/d\theta$; in particular, there is no need to specify over- and undercrossing information for a front. We depict $(\mathbb{R}/2\pi\mathbb{Z}) \times \mathbb{R}$ by letting θ be the horizontal axis and z the vertical axis, and drawing dashed vertical lines to represent the identified lines $\theta = 0$ and $\theta = 2\pi$. See Figure 2 for examples of fronts in $\mathcal{J}^1(S^1)$.

Any front in $\mathcal{J}^1(S^1)$ has three "classical" invariants under Legendrian isotopy. The first is the knot type of the front in $\mathcal{J}^1(S^1)$, obtained by smoothing cusps and resolving crossings in the usual way. The other two are the Thurston–Bennequin number tb and rotation number r:



We note that the Thurston–Bennequin number in $\mathcal{J}^1(S^1)$ was first introduced by Tabachnikov [14].

We now examine the front K of the conormal knot of a plane curve C. There is a simple description for K: any point (q_1, q_2) on C, with unit tangent vector $(\cos\varphi, \sin\varphi)$, gives the point $(\theta, z) = (\varphi + \pi/2, -q_1 \sin\varphi + q_2 \cos\varphi)$ in K, and K is obtained by allowing (q_1, q_2) to range over C. Points of inflection of C correspond to cusps of K, and it is easy to check that any right cusp of K is traversed upwards and any left cusp downwards; just draw a neighborhood of an inflection point of C.

As for the classical Legendrian invariants of K, since K has equal numbers of left and right cusps, it follows that $r(K) = 0$. The Thurston–Bennequin number of K measures framing and is essentially Arnold's J^+ invariant: $tb(K) = J^+(K) + n(K)^2 - 1$, where $n(K)$ is the winding number of K around S^1. Hence the framed knot type of K determines all classical information about K.

2.2. Drawing the conormal knot front

We have already discussed how to define the conormal knot front of a plane curve, but the definition is not very useful computationally. Here we present an algorithm for easily obtaining a front isotopic to the conormal knot front.

Call a plane curve *rectilinear* if it is completely composed of line segments parallel to either coordinate axis, along with arbitrarily small smoothing 90° corners, and no two line segments lie on the same (horizontal or vertical) line. Clearly any plane curve is isotopic to a rectilinear curve, and so it suffices to describe the conormal front for any rectilinear curve.

For ease of notation, label the coordinate axes x and y rather than q_1 and q_2. To each line segment L in a rectilinear plane curve, we associate the following point in $(\mathbb{R}/2\pi\mathbb{Z}) \times \mathbb{R}$:

- $(\pi/2, y)$ if L is in the $+x$ direction and y is the y coordinate of L;
- $(\pi, -x)$ if L is in the $+y$ direction and x is the x coordinate of L;
- $(3\pi/2, -y)$ if L is in the $-x$ direction and y is the y coordinate of L;
- $(0, x)$ if L is in the $-y$ direction and x is the x coordinate of L.

Next, "connect the dots" by joining the points corresponding to line segments which share an endpoint. Finally, smooth the result, rounding corners and placing cusps where necessary. See Figure 3.

Proposition 2. *The resulting front in $\mathcal{J}^1(S^1)$ is Legendrian isotopic to the front of the rectilinear plane curve.*

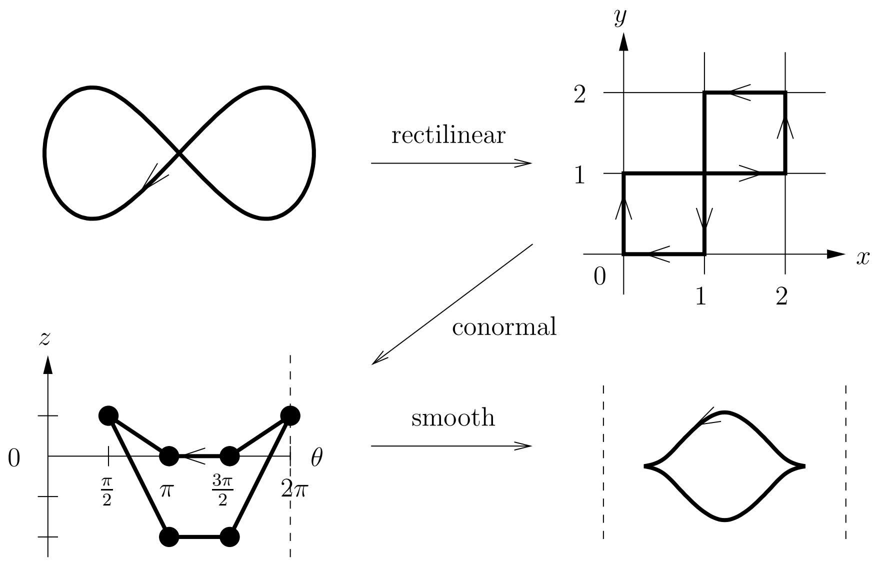


Figure 3. Algorithm for obtaining the conormal knot from a plane curve.

Proof. It is clear that the conormal front for the rectilinear curve passes through the points given in the algorithm above, since they comprise the conormal for the line segments of the rectilinear curve. The conormals of the smoothing corners interpolate between these points. The conormal front for a smoothing corner at the point (x, y) is given by $\{(\theta, x\cos\theta + y\sin\theta)\}$ for some range of θ in an interval of length $\pi/2$. Hence the conormal fronts for any two smoothing corners intersect either once or not at all. It follows that, up to Legendrian isotopy, the conormals for the smoothing corners can be approximated by the line segments joining points in the algorithm above. □

2.3. Nonhomotopic plane curves

We can use the algorithm from the previous section to show that there are plane curves whose conormal knots have the same framed knot type but which are not Legendrian isotopic.

Proposition 3. *The plane curves in the bottom line of Figure 2 have the same framed conormal knot type but are not safely homotopic.*

Proof. The two plane curves give conormal knots which are topologically Whitehead links; see Figure 4. Both conormal knots have $tb = -3$ (equivalently, $J^+ = -2$).

We claim that the two conormal knots are not Legendrian isotopic. Indeed, applying the Legendrian satellite construction (see the appendix of [13]) to the conormal fronts and the stabilized unknot in $\mathbb{R}^3$ yields two familiar Legendrian knots in $\mathbb{R}^3$: these are

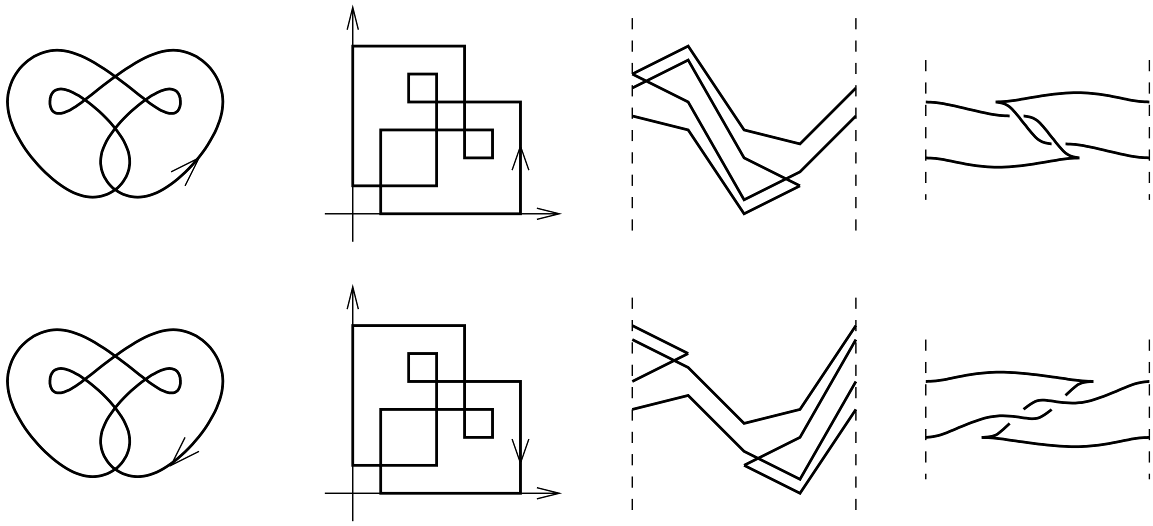


FIGURE 4. Nonhomotopic plane curves, their rectilinear approximations, conormals, and smoothed conormal fronts.

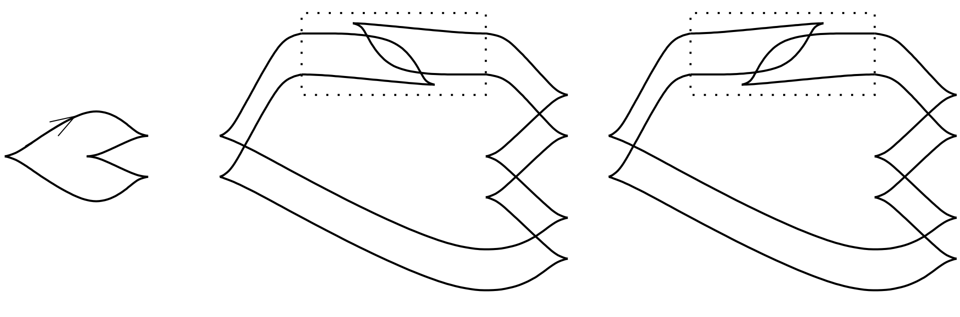


FIGURE 5. The Legendrian satellites of the conormal knots from Figure 4 to the stabilized unknot produce nonisotopic Legendrian knots.

called “Eliashberg knots” in [7] and labeled $E(2,3)$ and $E(1,4)$. See Figure 5. The two knots can be distinguished by their contact homology differential graded algebras [3]; in particular, $E(2,3)$ has Poincaré polynomial $2t+t^{-1}$ and $E(1,4)$ has Poincaré polynomial t^3+t+t^{-3}. It follows that the two conormal knots in $\mathcal{J}^1(S^1)$ are not Legendrian isotopic, as desired. □

We can use the plane curves from Proposition 3 to produce an arbitrarily large family of plane curves whose conormal knots have the same classical invariants but are not Legendrian isotopic. For $r, s \geq 0$, let $C_{r,s}$ be the plane curve shown in Figure 6, which

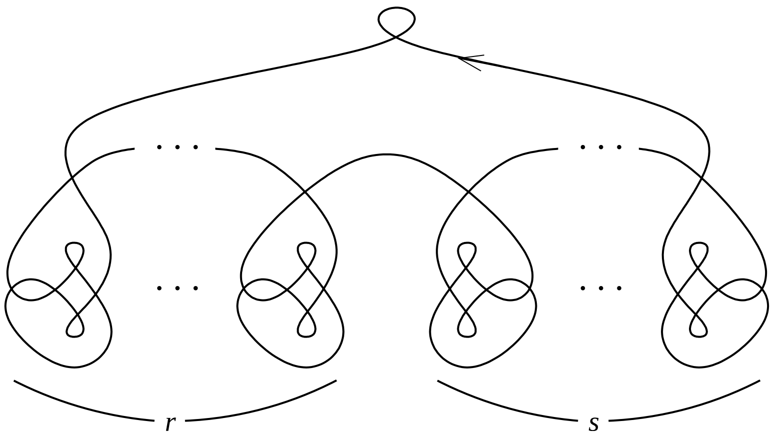


FIGURE 6. The "connected sum" plane curve $C_{r,s}$.

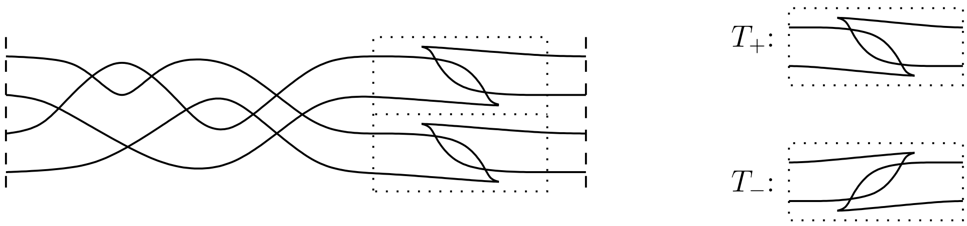


FIGURE 7. The conormal knot for $C_{2,0}$. To obtain the conormal knots for $C_{1,1}$ and $C_{0,2}$, replace one or both of the boxed tangles T_+ by T_-.

can be viewed as a connected sum of the plane curves from Proposition 3. (Note however that the connected sum operation on plane curves is not well-defined.)

Proposition 4. *For fixed $n \geq 1$, the n plane curves $C_{r,s}$, $r + s = n$, have the same framed conormal knot type but are not safely homotopic.*

Sketch of proof. The details of the proof require some working familiarity with computations in contact homology for Legendrian knots in standard contact $\mathbb{R}^3$, along the lines of [8, 11]; we provide an outline here and leave the particulars to the reader.

We can use the algorithm from Section 2.1 to find the conormal fronts for $C_{r,s}$. When $r+s = n$ is fixed, the conormal fronts for $C_{r,s}$ are identical except for n tangles. Of these tangles, r are given by the tangle T_+ defined in Figure 7, and s by T_-. The situation for $n = 2$ is shown in Figure 7; the picture for $n > 2$ is very similar. Note that the conormal fronts for $C_{r,s}$ are all isotopic as framed knots.

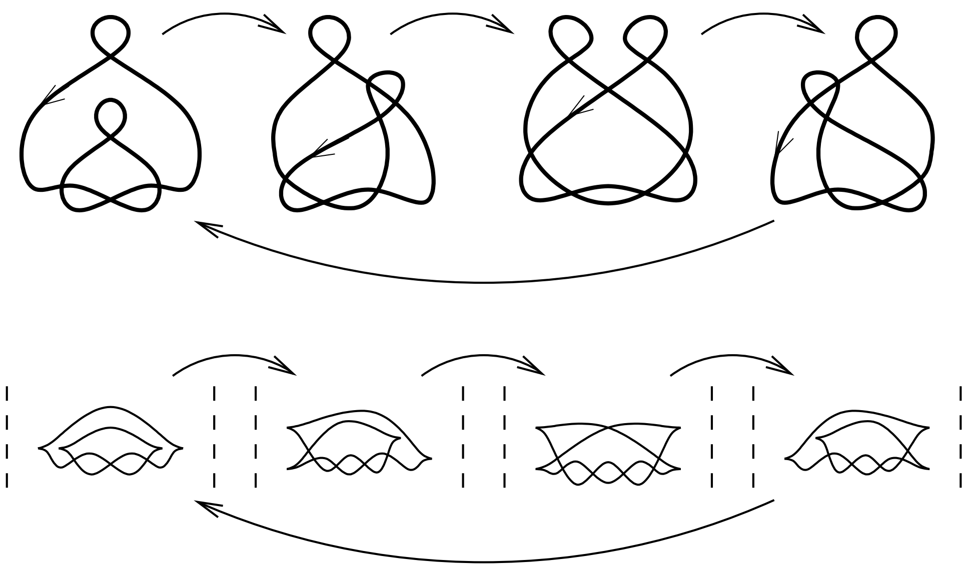


FIGURE 8. A nontrivial loop γ of plane curves, and the corresponding loop $\tilde{\gamma}$ of conormal knots.

Now consider the Legendrian satellite $K_{r,s}$ of the conormal front for $C_{r,s}$ to the stabilized unknot, as in the proof of Proposition 3. We distinguish between the knots $K_{r,s}$ using Poincaré polynomials for contact homology.

It is easy to show that the Chekanov–Eliashberg differential graded algebra for $K_{r,s}$ has a graded augmentation, for instance because it has a ruling [8]. An examination of Maslov indices shows that all crossings in $K_{r,s}$ have degrees $0, \pm 1, \pm 2$, except for the crossings in the tangles $T_{\pm}$; the two crossings in any T_+ have degree 1 and -1, while the two crossings in any T_- have degree 3 and -3. Since the (linearized) differential of the degree 3 crossing in any T_- is 0, we conclude that any Poincaré polynomial for $K_{r,s}$ has t^3 coefficient equal to s. It follows that the Legendrian knots $K_{r,s}$, $r + s = n$, are not Legendrian isotopic, and thus that the plane curves $C_{r,s}$ are not safely homotopic. □

2.4. Loops of plane curves

Here we consider loops in the space of plane curves. Let $\mathcal{C}$ denote the space of plane curves, and let $\mathcal{D} \subset \mathcal{C}$ be the discriminant of plane curves with dangerous self-tangencies. We will present a loop which is contractible in $\mathcal{C}$ but noncontractible in $\mathcal{C} \setminus \mathcal{D}$.

Consider the loop γ in $\mathcal{C} \setminus \mathcal{D}$ pictured in Figure 8. This induces a loop $\tilde{\gamma}$ of Legendrian knots in $\mathcal{J}^1(S^1)$, also shown in Figure 8. As a loop of (framed) knots in $\mathcal{J}^1(S^1)$, $\tilde{\gamma}$ is contractible; this follows from the contractibility of the corresponding loop of trefoils in S^3, which itself follows from work of Hatcher (see [10]). On the other hand, by using the contact condition and work of Kálmán [10], we can show that $\tilde{\gamma}$ is nontrivial when considered as a loop of Legendrian knots.

Proposition 5. *The loop γ is contractible in $\mathcal{C}$, but has order at least 5 in $\pi_1(\mathcal{C} \setminus \mathcal{D})$.*

Proof. It is straightforward to check that γ is contractible in $\mathcal{C}$. Note that by the h-principle, $\mathcal{C}$ is weakly homotopy equivalent to the space of free loops in $S^1 \times \mathbb{R}^2$, which is not simply connected. However, γ can be represented by a loop of *based* loops in $S^1 \times \mathbb{R}^2$, and the space of based loops is simply connected since $\pi_2(S^1 \times \mathbb{R}^2) = 0$.

Now consider γ as a loop in $\mathcal{C} \setminus \mathcal{D}$. The loop $\tilde{\gamma}$ of Legendrian knots in $\mathcal{J}^1(S^1)$ lifts to an identical looking loop $\tilde{\gamma}'$ of Legendrian knots in the universal cover $\mathbb{R}^3$ with the standard contact structure. (Just ignore the dashed lines in Figure 8.) A special case of Theorem 1.2 in [10] states that $\tilde{\gamma}'$ has order at least 5 in the Legendrian category; hence $\tilde{\gamma}$ does as well. □

References

[1] V. I. Arnold, *Plane curves, their invariants, perestroikas and classifications*, in Singularities and bifurcations, Adv. Soviet Math. **21** (Amer. Math. Soc., Providence, 1994), 33–91.

[2] V. I. Arnold, *Topological Invariants of Plane Curves and Caustics*, University Lecture Series **5** (Amer. Math. Soc., Providence, 1994).

[3] Yu. V. Chekanov, *Differential algebra of Legendrian links*, Invent. Math. **150** (2002), no. 3, 441–483.

[4] S. Chmutov, V. Goryunov, and H. Murakami, *Regular Legendrian knots and the HOMFLY polynomial of immersed plane curves*, Math. Ann. **317** (2000), no. 3, 389–413.

[5] T. Ekholm and J. Etnyre, *Invariants of knots, embeddings and immersions via contact geometry*, in Geometry and topology of manifolds, Fields Inst. Commun. **47** (Amer. Math. Soc., Providence, 1994). 77–96.

[6] Ya. Eliashberg, *Invariants in contact topology*, Doc. Math. **1998**, Extra Vol. II, 327–338 (electronic).

[7] J. Epstein, D. Fuchs, and M. Meyer, *Chekanov–Eliashberg invariants and transverse approximations of Legendrian knots*, Pacific J. Math. **201** (2001), no. 1, 89–106.

[8] D. Fuchs, *Chekanov–Eliashberg invariant of Legendrian knots: existence of augmentations*, J. Geom. Phys. **47** (2003), no. 1, 43–65.

[9] V. Goryunov, *Vassiliev type invariants in Arnold's J^+-theory of plane curves without direct self-tangencies*, Topology **37** (1998), no. 3, 603–620.

[10] T. Kálmán, *Contact homology and one parameter families of Legendrian knots*, Geom. Topol. **9** (2005), 2013–2078.

[11] L. Ng, *Computable Legendrian invariants*, Topology **42** (2003), no. 1, 55–82.

[12] L. Ng, *Conormal bundles, contact homology, and knot invariants*, to appear in the proceedings of the BIRS workshop "The interaction of finite type and Gromov–Witten invariants", Geom. Topol. Monogr. **8** (2006).

[13] L. Ng and L. Traynor, *Legendrian solid-torus links*, J. Symplectic Geom. **2** (2005), no. 3, 411–443.

[14] S. Tabachnikov, *Calculation of the Bennequin invariant of a Legendre curve from the geometry of its wave front*, Funktsional. Anal. i Prilozhen. **22** (1988), no. 3, 89–90; English translation in Funct. Anal. Appl. **22** (1988), no. 3, 246–248.

[15] H. Whitney, *On regular closed curves in the plane*, Compositio Math. **4** (1936), 276–284.

Department of Mathematics, Stanford University, Stanford, CA 94305

Proceedings of 12th Gökova
Geometry-Topology Conference
pp. 175 – 186

Published online at
GokovaGT.org

An open book decomposition compatible with rational contact surgery

Burak Ozbagci

Dedicated to the memory of Raoul Bott.

Abstract. We construct an open book decomposition compatible with a contact structure given by a rational contact surgery on a Legendrian link in the standard contact S^3. As an application we show that some rational contact surgeries on certain Legendrian knots induce overtwisted contact structures.

0. Introduction

Recently Giroux proved a central result regarding the topology of contact 3-manifolds. Namely he established a one to one correspondence between contact structures up to isotopy and open book decompositions up to positive stabilizations. This correspondence, however, does not explicitly describe an open book decomposition corresponding to a given contact structure.

In [DG], Ding and Geiges proved that every (closed) contact 3-manifold (Y, ξ) can be given by a contact (± 1)-surgery on a Legendrian link in the standard contact S^3. Here we use the parenthesis to emphasize that the surgery coefficients are measured with respect to the contact framing. If the coefficients of all the curves in a contact surgery diagram are (-1), then an open book decomposition compatible with this contact structure is given by the algorithm in [AO] coupled with the work of Plamenevskaya ([P]). Moreover, Stipsicz ([S]) showed that the same algorithm works in the general case of contact (± 1)-surgery. In this article we will review these results (giving slightly different proofs) and extend the algorithm to the case of rational contact surgery.

In fact any rational contact surgery can be turned into a sequence of contact (± 1)-surgeries ([DG, DGS]) and the algorithm above would provide an open book decomposition compatible with the resulting contact structure. However, this would give an open book decomposition with high genus and we will show that there is a short-cut in obtaining an open book decomposition (with lower genus) compatible with a rational contact surgery.

As an application we show that certain rational contact surgeries induce overtwisted contact structures by making use of the right-veering property of tight contact structures recently introduced by Honda, Kazez and Matic ([HKM]).

Here we outline the main idea in this article. Given a Legendrian link in the standard contact (S^3, ξ_{st}) with its front projection onto the yz-plane. We will show that there is an open book decomposition on S^3 compatible with a contact structure ξ_0 isotopic to ξ_{st} such that the Legendrian link (after a Legendrian isotopy) is contained in a page of this open book decomposition as described in [P]. It follows that the page framing and the contact framing coincide on each component of this link since the Reeb vector field induces both framings. Consequently when we perform contact (± 1)-surgery on this link we get an open book decomposition compatible with the resulting contact structure. The monodromy of this open book decomposition is given by a product of Dehn twists along curves explicitly drawn on a page. Now when a Legendrian link with rational surgery coefficients is given, we embed this link into the page of the open book decomposition as described above. Then we attach appropriate 1-handles to a page of this open book decomposition and extend the monodromy of our open book decomposition by Dehn twists along some push-offs of the original monodromy curves going through the attached 1-handles.

Acknowledgement: We would like to thank Olga Plamenevskaya, András Stipsicz and John Etnyre for helpful correspondence. The author was partially supported by the Turkish Academy of Sciences.

1. Open book decompositions and contact structures

We will assume that all our contact structures are positive and cooriented. In the following we describe the compatibility of an open book decomposition with a given contact structure on a 3-manifold.

Definition 1. *Suppose that for a link L in a 3-manifold Y the complement $Y - L$ fibers as $\pi \colon Y - L \to S^1$ such that the fibers are interiors of Seifert surfaces of L. Then (L, π) is an* open book decomposition *of Y. The Seifert surface $F = \overline{\pi^{-1}(t)}$ is called a* page, *while L the* binding *of the open book decomposition. The monodromy of the fibration π is called the* monodromy *of the open book decomposition.*

Any locally trivial bundle with fiber F over an oriented circle is canonically isomorphic to the fibration $I \times F/(1, x) \sim \big(0, h(x)\big) \to I/\partial I \approx S^1$ for some self-diffeomorphism h of F. In fact, the map h is determined by the fibration up to isotopy and conjugation by an orientation preserving self-diffeomorphism of F. The isotopy class represented by the map h is called the monodromy of the fibration. Conversely given a compact oriented surface F with nonempty boundary and $h \in \Gamma_F$ (the mapping class group of F) we can form the mapping torus $F(h) = I \times F/(1, x) \sim \big(0, h(x)\big)$. Since h is the identity on ∂F, the boundary $\partial F(h)$ of the mapping torus $F(h)$ can be canonically identified with r copies of $T^2 = S^1 \times S^1$, where the first S^1 factor is identified with $I/\partial I$ and the second one comes from a component of ∂F. Hence by gluing in r copies of $D^2 \times S^1$ to $F(h)$ so that ∂D^2 is identified with $S^1 = I/\partial I$ and the S^1 factor in $D^2 \times S^1$ is identified with a boundary component of ∂F, $F(h)$ can be completed to a closed 3-manifold Y equipped with an open book decomposition. In conclusion, an element $h \in \Gamma_F$ determines a 3-manifold together with an "abstract" open book decomposition on it.

Theorem 2 (Alexander)**.** *Every closed and oriented 3-manifold admits an open book decomposition.*

Suppose that an open book decomposition with page F is specified by $h \in \Gamma_F$. Attach a 1-handle to the surface F connecting two points on ∂F to obtain a new surface F'. Let α be a closed curve in F' going over the new 1-handle exactly once. Define a new open book decomposition with $h \circ t_\alpha \in \Gamma_{F'}$, where t_α denotes the right-handed Dehn twist along α. The resulting open book decomposition is called a *positive stabilization* of the one defined by h. If we use a left-handed Dehn twist instead then we call the result a *negative stabilization*. The inverse of the above process is called *positive* (*negative*) *destabilization*. Note that the topology of the underlying 3-manifold does not change when we stabilize/destabilize an open book. Also note that the resulting monodromy depends on the chosen curve α.

Definition 3. *An open book decomposition of a* 3*-manifold* Y *and a contact structure* ξ *on* Y *are called* compatible *if* ξ *can be represented by a contact form* α *such that the binding is a transverse link,* $d\alpha$ *is a volume form on every page and the orientation of the transverse binding induced by* α *agrees with the boundary orientation of the pages.*

In other words, the conditions that $\alpha > 0$ on the binding and $d\alpha > 0$ on the pages is a strengthening of the contact condition $\alpha \wedge d\alpha > 0$ in the presence of an open book decomposition on Y. The condition that $d\alpha$ is a volume form on every page is equivalent to the condition that the Reeb vector field of α is transverse to the pages. Moreover an open book decomposition and a contact structure are compatible if and only if the Reeb vector field of α is transverse to the pages (in their interiors) and tangent to the binding.

Theorem 4 (Giroux)**.** *Every contact 3-manifold admits a compatible open book decomposition.*

2. An open book decomposition compatible with a contact (± 1)-surgery

In this section we describe an explicit construction of an open book decomposition compatible with a given contact structure. The algorithm is contained in [AO] and it is proven to be compatible with the given contact structure in [P] and [S].

We will show that for a given Legendrian link $\mathbb{L}$ in $(\mathbb{R}^3, \xi_{st}) \subset (S^3, \xi_{st})$ there exists a surface $F \subset S^3$ containing $\mathbb{L}$ such that $d\alpha$ is an area form on F (where $\alpha = dz + x\,dy$), $\partial F = K$ is a torus knot which is transverse to $\xi_{st} = \ker \alpha$ and the components of $\mathbb{L}$ do not separate F. We first isotope $\mathbb{L}$ by a Legendrian isotopy so that in the front projection (onto the yz-plane) all the segments have slope (± 1) away from the points where $\mathbb{L}$ intersects the yz-plane. Then we consider narrow rectangular strips around each of these segments and connect them by small twisted bands corresponding to each point where $\mathbb{L}$ intersects the yz-plane. The small bands can be constructed in such a way that the Legendrian link lies on these bands while the bands twist along the contact planes. The narrow strips around the straight segments connected with these small twisted bands give us the Seifert surface F of a torus knot $K = \partial F$. Notice that we ensured that $\mathbb{L}$ lies in F. Moreover $d\alpha$ is an area form on F by construction since the Reeb vector field $\frac{\partial}{\partial z}$ of α is transverse to F. Furthermore we can slightly isotope $\partial F = K$ to make it transverse to ξ_{st}.

Now since K is a fibered knot with fibered surface F there is a fibration of the complement of K in S^3 where F is one of the pages of the induced open book decomposition on S^3. Note that $d\alpha$ induces an area form on the nearby pages as well since we can always keep the nearby pages transverse to $\frac{\partial}{\partial z}$. The union of these nearby pages including the binding K is a handlebody U_1 (which is a thickening of this one page F that carries the Legendrian link $\mathbb{L}$) such that $d\alpha$ is an area form on every page. But we can not guarantee that $d\alpha$ induces an area form on the rest of the pages of this open book decomposition. We would like to extend the contact structure ξ_{st} to the complementary handlebody U_2 (as some contact structure ξ_0) so that it is compatible with the pages in U_2. This can be achieved (see [P]) by an explicit construction of a contact form on U_2 similar to the one described in [TW]. Hence we get a contact structure ξ_0 on S^3 which is compatible with our open book. Moreover, by construction ξ_0 and ξ_{st} coincide on U_1 and we claim that the contact structures ξ_0 and ξ_{st} are isotopic on U_2 relative to ∂U_2. Notice that ∂U_2 can be made convex and one can check that the binding K is the dividing set on ∂U_2. Uniqueness (up to isotopy) of a tight contact structure with such boundary conditions is given by Theorem 5.

Suppose that (K, π) is a given open book decomposition on a closed 3-manifold Y. Then by presenting the circle S^1 as the union of two closed (connected) arcs $S^1 = I_1 \cup I_2$ intersecting each other in two points, the open book decomposition (K, π) naturally induces a *Heegaard decomposition* $Y = U_1 \cup_\Sigma U_2$ of the 3-manifold Y. The surface Σ along which these handlebodies are glued is simply the union of two pages $\pi^{-1}(I_1 \cap I_2)$ together with the binding.

Theorem 5 (Torisu, [To]). *Suppose that ξ_1, ξ_2 are contact structures on Y satisfying:*
(i) $\xi_i|_{U_j}$ $(i = 1, 2;\ j = 1, 2)$ *are tight, and*
(ii) Σ *is convex in* (Y, ξ_i) *and K is the dividing set for both contact structures.*
Then ξ_1 and ξ_2 are isotopic. In addition, the set of such contact structures is nonempty.

Summarizing the above discussion we get

Proposition 6 (Plamenevskaya, [P]). *For a given Legendrian link $\mathbb{L}$ in (S^3, ξ_{st}) there exists an open book decomposition of S^3 satisfying:*
(1) *the contact structure ξ_0 compatible with this open book decomposition is isotopic to ξ_{st},*
(2) $\mathbb{L}$ *is contained in one of the pages and none of the components of $\mathbb{L}$ separate F,*
(3) $\mathbb{L}$ *is Legendrian with respect to ξ_0,*
(4) *there is an isotopy which fixes $\mathbb{L}$ and takes ξ_0 to ξ_{st},*

(5) *the page framing of* $\mathbb{L}$ *(induced by* F*) is the same as its contact framing induced by* ξ_0 *(or* ξ_{st}*).*

In fact item (5) in the theorem above follows from (1)-(4) by

Lemma 7. *Let* C *be a Legendrian curve on a page of a compatible open book decomposition* $\mathfrak{ob}_\xi$ *in a contact 3-manifold* (Y, ξ)*. Then the page framing of* C *is the same as its contact framing.*

Proof. Let α be the contact 1-form for ξ such that $\alpha > 0$ on the binding and $d\alpha > 0$ on the pages of $\mathfrak{ob}_\xi$. Then the Reeb vector field R_α is transverse to the pages (in their interiors) as well as to the contact planes. Hence R_α defines both the page framing and the contact framing on C. □

Given a Legendrian link $\mathbb{L}$ in $(\mathbb{R}^3, \xi_{st}) \subset (S^3, \xi_{st})$ we described an open book decomposition on S^3 whose page is the Seifert surface of an appropriate torus knot K and $\mathbb{L}$ is included in one of the pages. When we perform contact (± 1)-surgery along $\mathbb{L}$ we get a new open book decomposition on the resulting contact 3-manifold obtained by the surgery. The monodromy of this open book decomposition is given by the composition of the monodromy of the torus knot and Dehn twists along the components of the surgery link. Here all the Dehn twists of the monodromy of the torus knot is right-handed while a $(+1)$-surgery curve (resp. (-1)) induces a left-handed (resp. right-handed) Dehn twist. Notice that the surgery curves are pairwise disjoint and they are homologically non-trivial on the Seifert surface by this construction. It turns out that the resulting contact 3-manifold and the open book decomposition are compatible by the following theorem a proof of which is can be found in [G] in case of (-1)-surgery and [E1] for the general case.

Proposition 8. *Let* C *be a Legendrian curve on a page of a compatible open book decomposition* $\mathfrak{ob}_\xi$ *with monodromy* $h \in \Gamma_F$ *on a contact 3-manifold* (Y, ξ)*. Then the contact 3-manifold obtained by contact* (± 1)*-surgery along* C *is compatible with the open book decomposition with monodromy* $h \circ (t_C)^{\mp 1} \in \Gamma_F$*, where* t_C *denotes a right-handed Dehn twist along* C*.*

3. An open book decomposition compatible with a rational contact surgery

In this section we will first outline how to turn a rational contact surgery into a sequence of contact (± 1)-surgeries. The reader is advised to consult [DG, DGS] for background on contact surgery.

Assume that we want to perform contact (r)-surgery on a Legendrian knot L in (S^3, ξ_{st}) for some rational number $r < 0$. In this case the surgery can be replaced by a sequence of contact (-1)-surgeries along Legendrian knots associated to L as follows: suppose that $r = -\frac{p}{q}$ and the continued fraction coefficients of $-\frac{p}{q}$ are equal to $[r_0 + 1, r_1, \ldots, r_k]$, with $r_i \leq -2$ $(i = 0, \ldots, k)$. Consider a Legendrian push-off of L, add $|r_0 + 2|$ zig-zags to it and get L_0. Push this knot off along the contact framing and add $|r_1 + 2|$ zig-zags to it to get L_1. Perform contact (-1)-surgery on L_0 and repeat the process with L_1. After $(k+1)$ steps we end up with a diagram involving only contact (-1)-surgeries. The result of the sequence of contact (-1)-surgeries is the same as the result of the original contact (r)-surgery according to [DG, DGS].

Proposition 9 (Ding–Geiges, [DG])**.** *Fix* $r = \frac{p}{q} > 0$ *and an integer* $k > 0$*. Then contact* (r)*-surgery on the Legendrian knot* L *is the same as contact* $(\frac{1}{k})$*-surgery on* L *followed by contact* $(\frac{p}{q-kp})$*-surgery on the Legendrian push-off* L' *of* L*.*

By choosing $k > 0$ large enough, the above proposition provides a way to reduce a contact (r)-surgery (with $r > 0$) to a contact $(\frac{1}{k})$-surgery and a negative contact (r')-surgery. This latter one can be turned into a sequence of contact (-1)-surgeries, hence the algorithm is complete once we know how to turn contact $(\frac{1}{k})$-surgery into contact (± 1)-surgeries.

Lemma 10 (Ding–Geiges, [DG])**.** *Let* $L_1, \cdots, L_k$ *denote* k *Legendrian push-offs of the Legendrian knot* L*. Then contact* $(\frac{1}{k})$*-surgery on* L *is isotopic to performing contact* $(+1)$*-surgeries on the* k *Legendrian knots* $L_1, \cdots, L_k$*.*

Given a Legendrian link $\mathbb{L}$ in $(\mathbb{R}^3, \xi_{st}) \subset (S^3, \xi_{st})$ with rational surgery coefficients. We follow the algorithm described in Section 2 to find an open book decomposition on S^3 such that $\mathbb{L}$ is embedded in one of the pages. Now use the above algorithm to turn the rational surgery into contact (± 1)-surgeries. Since the contact framing of each component of $\mathbb{L}$ agrees with the page framing by Lemma 7, a contact push-off of any component will still lie on the same page.

Let L be a Legendrian knot in $(\mathbb{R}^3, \xi_{st})$. We define the positive and negative stabilization of L as follows: First we orient the knot L and then if we replace a strand of the knot by an up (down, resp.) cusp by adding a zigzag as in Figure 1 we call the resulting Legendrian knot the negative (positive, resp.) stabilization of L. Notice that stabilization is a well defined operation, i.e., it does not depend at which point the stabilization is done.

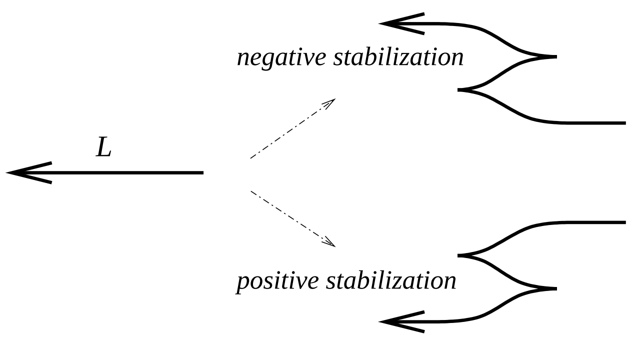


Figure 1. Positive and negative stabilization of a Legendrian knot L

Now let L be a Legendrian knot in a page of a compatible open book decomposition $\mathfrak{ob}_\xi$ in a contact 3-manifold (Y,ξ). Then by Lemma 3.3 in [E2], the stabilized knot lies in a page of an open book decomposition obtained by stabilizing $\mathfrak{ob}_\xi$ by attaching a 1-handle to the page of $\mathfrak{ob}_\xi$ and letting the stabilized knot go through the 1-handle once. Notice that there is a positive and a negative stabilization of the oriented Legendrian knot L defined by adding a down or an up cusp, and this choice corresponds to adding a *left* (i.e., to the left-hand side of the oriented curve L) or a *right* (i.e., to the right-hand side of the oriented curve L) 1-handle to the surface respectively as shown in Figure 2.

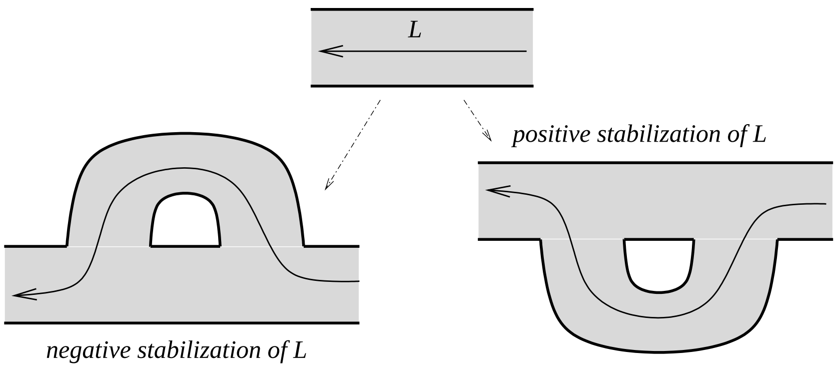


Figure 2. Stabilization of a page to include the stabilization of the Legendrian knot L

Repeating this process (by attaching appropriate right or left 1-handles to the Seifert surface of the torus knot) we will get a page of an open book decomposition where we have all the push-offs with their additional zig-zags embedded in this page. Notice that contact (r)-surgery is not uniquely defined because of the choice of adding up or down zig-zags to the push-offs and this choice can be followed in the way that we attach our 1-handles as in Figure 2. The monodromy of the resulting open book decomposition will be the composition of the monodromy of the torus knot (a product of right-handed Dehn twists), right-handed Dehn twists corresponding to the stabilizations and the Dehn twists along the push-offs. This open book decomposition is compatible with the contact (r)-surgery by Proposition 8 since we can recover the effect of this rational surgery on (S^3,ξ_{st}) by contact (±1)-surgeries along embedded curves on a page of an open book decomposition of S^3 compatible with its standard contact structure. Here notice that when we positively stabilize a compatible open book decomposition of (S^3,ξ_{st}), the resulting open book decomposition (of S^3) will be compatible with ξ_{st}. As a result we get

Theorem 11. *Given a contact 3-manifold obtained by a rational contact surgery on a Legendrian link in the standard contact S^3. Then there is an algorithm to find an open book decomposition on this 3-manifold compatible with the contact structure.*

4. An example

We illustrate the algorithm on a simple example. Consider a contact $(-\frac{5}{3})$-surgery on the right-handed Legendrian trefoil knot L in Figure 3.

FIGURE 3. Rational contact surgery on a right-handed Legendrian trefoil knot in (S^3, ξ_{st})

Observe first that the continued fraction coefficients of $-\frac{5}{3}$ are $r_0 = r_1 = -3$. Notice that this surgery is not uniquely defined and there are four different possibilities of performing this surgery. We will find an open book decomposition compatible with one of these surgeries. First orient the Legendrian knot as indicated in Figure 3. Consider a Legendrian push-off of L, add a down zig-zag to it and get L_0. Push L_0 off along the contact framing and add a down zig-zag to it to get L_1. Performing contact (-1)-surgery on both L_0 and L_1 is equivalent to performing a contact $(-\frac{5}{3})$-surgery on the right-handed Legendrian trefoil knot in Figure 3. The push-offs L_0 and L_1 are illustrated in Figure 4.

Finally in Figure 5, we depict the page of the open book decomposition which is compatible with the contact $(-\frac{5}{3})$-surgery on the right-handed Legendrian trefoil knot L. The page is the Seifert surface of the (5,6)-torus knot with two 1-handles attached. Notice that the 1-handle attachments in Figure 2 are shown abstractly but in Figure 5 this corresponds to plumbing positive Hopf-bands to the Seifert surface which is embedded in $\mathbb{R}^3$. The push-offs L_0 and L_1 are embedded on this page and the monodromy of the open book decomposition is the product of the monodromy of the (5,6)-torus knot (see [O] for details), right-handed Dehn twists along the embedded curves L_0 and L_1 and right-handed Dehn twists along the core circles of the 1-handles.

5. An application

In [O], we proved that for any positive integer k contact $(\frac{1}{k})$-surgery on a stabilized Legendrian knot in the standard contact S^3 induces an overtwisted contact structure, using sobering arcs introduced by Goodman [Go]. Note that contact $(\frac{1}{k})$-surgery is uniquely defined for any integer $k \neq 0$. In this section we will prove that for any positive rational number r, at least one of the contact (r)-surgeries on a stabilized Legendrian knot induces an overtwisted contact structure, using the *right-veering* arcs introduced by Honda,

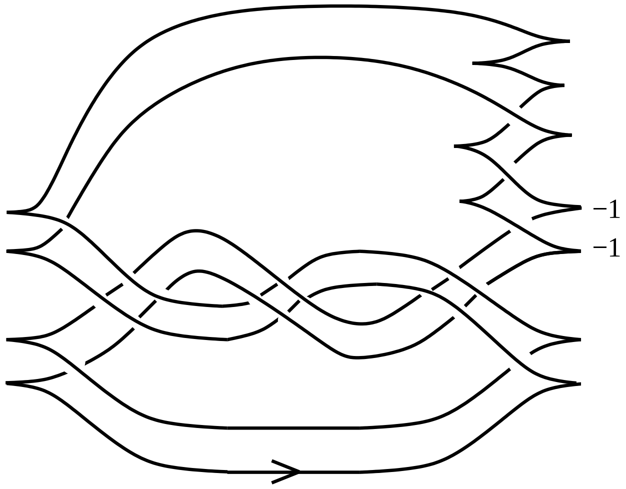


FIGURE 4. Turning rational contact surgery into contact (±1)-surgeries

Kazez and Matic. Let α and β be two properly embedded oriented arcs in a surface F with boundary. Let $\alpha(0) = \beta(0) = p \in \partial F$. Choose a hyperbolic metric on F and realize α and β as geodesics. If either $(\beta'(0), \alpha'(0))$ forms an oriented basis for T_pF or $\alpha'(0)$ is parallel to $\beta'(0)$ then we call β is to the right of α, denoted by $\beta \geq \alpha$.

Definition 12. *An element $h \in \Gamma_F$ (the mapping class group of F) is right-veering if for every $p \in \partial F$ and for every properly embedded arc α with $\alpha(0) = p$ we have $h(\alpha) \geq \alpha$. An open book decomposition is called right-veering if the monodromy of the open book decomposition is right-veering.*

Theorem 13 (Honda–Kazez–Matic, [HKM])**.** *If a contact 3-manifold (Y, ξ) is tight then every open book decomposition of Y compatible with ξ is right-veering.*

Suppose that K is a positively stabilized Legendrian knot in (S^3, ξ_{st}). Let r be a positive rational number and apply the algorithm in Section 3 to turn a contact (r)-surgery on K into a sequence of contact (± 1)-surgeries along some push-offs of K in such a way that all the push-offs are only negatively stabilized. Now consider the open book we described in Section 3 compatible with this surgery diagram. Since K is already positively stabilized, the page of the compatible open book is obtained by attaching a left 1-handle H_0 and some right 1-handles $H_1, H_2, \cdots, H_n$ (corresponding to the push-offs of K) to the Seifert surface of an appropriate torus knot. Note that all the surgery curves are embedded disjointly on the resulting page of the open book. We claim that the arc α across H_0 (depicted in Figure 6) is not right-veering at its top-end and hence showing that the induced contact structure is overtwisted by the criterion given in Theorem 13.

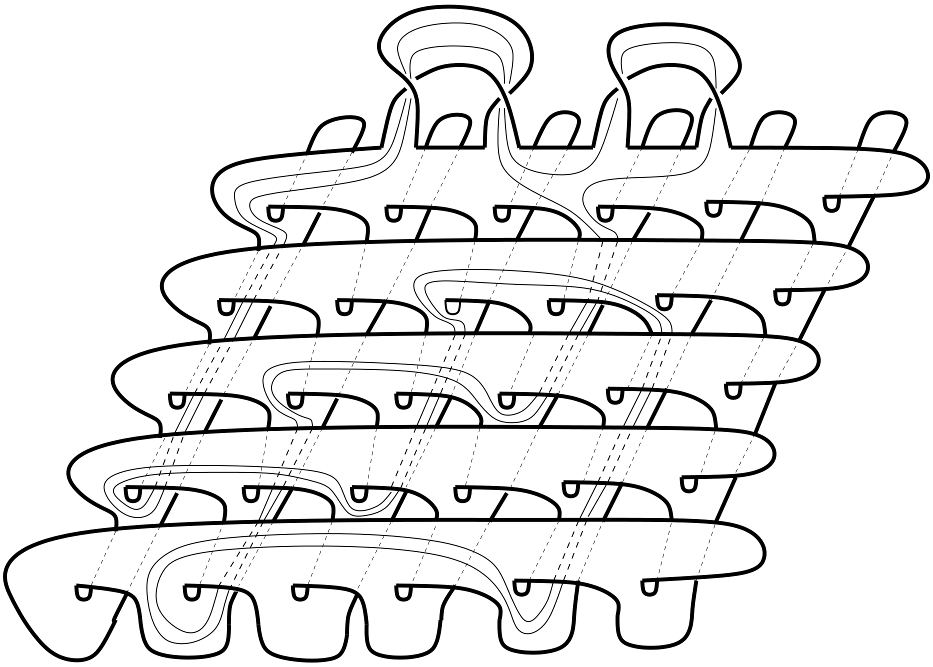


Figure 5.

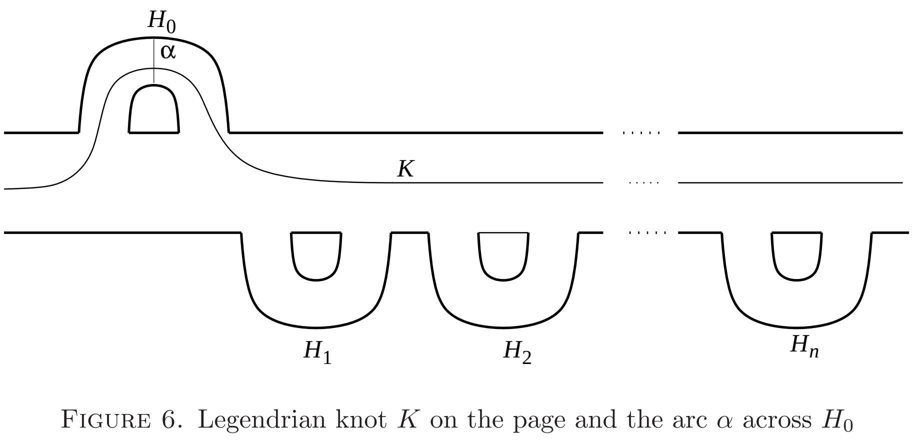


Figure 6. Legendrian knot K on the page and the arc α across H_0

To verify that α is not right-veering at its top-end we apply the monodromy h of the open book to α and observe that $h(\alpha)$ lies to the left of α on the page. Here note that the relevant part of the monodromy h consists of a product of k (described by Proposition 9) left-handed Dehn twists along K, right-handed Dehn twists along the push-offs of K (going through the 1-handles in Figure 6 in various ways) and a right-handed Dehn twist along the core circle of the handle H_0. The key point is that all the curves along which we apply right-handed Dehn twists stay only on one side of K on the surface and a cancellation of a left-handed Dehn twist by a right-handed Dehn twist is not allowed by construction. Here we would like to point out that α is not a sobering arc, so we could not use Goodman's criterion (cf. [Go]) to prove overtwistedness of the induced contact structure as we did in [O]. Thus we proved

Proposition 14. *For any positive rational number r, at least one of the contact (r)-surgeries on a stabilized Legendrian knot in the standard contact 3-sphere induces an overtwisted contact structure.*

We depict in Figure 7 an example of a contact structure which is overtwisted by Proposition 14. The contact structure in Figure 7 is obtained by a $(\frac{5}{2})$-contact surgery on the Legendrian unknot which is shown in Figure 8. The next result easily follows from Proposition 14:

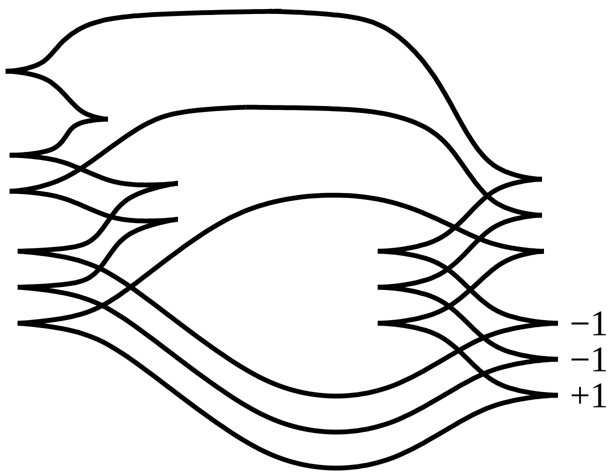


FIGURE 7. An overtwisted contact structure

Corollary 15. *If a rational contact surgery diagram contains a Legendrian knot with an isolated stabilized arc whose surgery coefficient is positive then at least one of the surgeries it represents is overtwisted.*

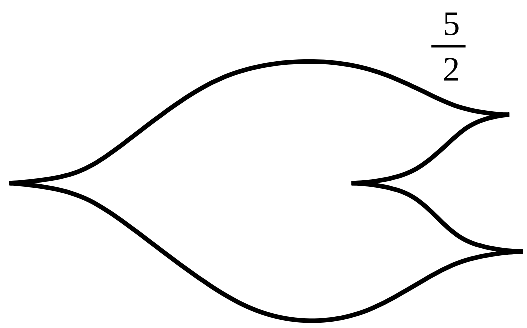


Figure 8. ($\frac{5}{2}$)-contact surgery on a Legendrian unknot

References

[AO] S. Akbulut and B. Ozbagci, *Lefschetz fibrations on compact Stein surfaces*, Geom. Topol. **5** (2001), 319–334.

[DG] F. Ding and H. Geiges, *A Legendrian surgery presentation of contact 3-manifolds*, Proc. Cambridge. Philos. Soc. **136** (2004), 583–598.

[DGS] F. Ding, H. Geiges and A. Stipsicz, *Surgery diagrams for contact 3-manifolds*, Turkish J. Math. **28** (2004), no. 1, 41–74.

[E1] J.B. Etnyre, *Lectures on open book decompositions and contact structures*, to appear in the proceedings of the "Floer Homology, Gauge Theory, and Low Dimensional Topology Workshop", arXiv: math.SG/0409402

[E2] J.B. Etnyre, *Planar open book decompositions and contact structures*, Int. Math. Res. Not. **79** (2004), 4255–4267

[G] D. Gay, *Explicit concave fillings of* 3*-manifolds,* Math. Proc. Cambridge Philos. Soc. **133** (2002), 431–441.

[Go] N. Goodman, *Contact structures and open books,* Ph.D thesis, UT Austin, 2003.

[HKM] K. Honda, W.H. Kazez, G. Matic, *Right-veering diffeomorphisms of compact surfaces with boundary I*, arXiv:math.GT/0510639

[O] B. Ozbagci, *A note on contact surgery diagrams*, Int. Jour. of Math. **16**-1 (2005), 87–99.

[P] O. Plamenevskaya, *Contact structures with distinct Heegaard-Floer invariants*, Math. Res. Letters **11**-4 (2004), 547–561.

[S] A. Stipsicz, *Surgery diagrams and open book decompositions of contact 3-manifolds*, Acta Math. Hungar. **108** (2005), no. 1-2, 71–86.

[TW] W. Thurston and H. Winkelnkemper, *On the existence of contact forms,* Proc. Amer. Math. Soc. **52** (1975), 345–347.

[To] I. Torisu, *Convex contact structures and fibered links in* 3*-manifolds,* Int. Math. Res. Not. **9** (2000), 441–454.

Department of Mathematics, Koç University, Istanbul, Turkey
E-mail address: bozbagci@ku.edu.tr

Proceedings of 12th Gökova
Geometry-Topology Conference
pp. 187 – 212

Published online at
GokovaGT.org

Virtual links, orientations of chord diagrams and Khovanov homology

Oleg Viro

Dedicated to the memory of Raoul Bott.

Abstract. By adding or removing appropriate structures to Gauss diagrams, one can create useful objects related to virtual links. In this paper few objects of this kind are studied: twisted virtual links generalizing virtual links; signed chord diagrams staying halfway between twisted virtual links and Kauffman bracket/Khovanov homology; alternatable virtual links intermediate between virtual and classical links. The most profound role here belongs to a structure that we dare to call orientation of chord diagrams. Khovanov homology is generalized to oriented signed chord diagrams and links in an oriented thickened surface such that the link projection realizes the first Stiefel-Whitney class of the surface.

1. Introduction

In the middle of nineties Lou Kauffman suggested a natural extension of classical knot theory by replacing classical links with *virtual links.* From the point of view of traditional 3-dimensional topology, as it was shown by Greg Kuperberg [13], virtual links are nothing but links in oriented thickennings of orientable closed surfaces, considered up to homeomorphisms of thickennings. Virtual links provide a way to extend the range of applications of combinatorial diagrammatic technique. It became especially interesting due to recent development of link homology theories built on link diagrams.

This paper was written in an attempt to analyze the difficulties which emerge when the construction of Khovanov homology is extended to virtual links. I analyzed few geometric objects closely related to a virtual link and found that a key role is played by an additional structure: orientation of a chord diagram, that is a *signed chord diagram*. Existence of this structure is solely responsible for the possibility to construct a Khovanov complex literally as in the classical case.

Virtual links with orientable chord diagrams appeared in various occasions: virtual links that admit checkerboard colorable diagrams, virtual links which can be made alternating by crossing changes, virtual links with zero homologous modulo 2 irreducible model, virtual links with orientable atom, etc. I will call virtual links of this kind *alternatable*.

Construction of Khovanov homology with coefficients in $\mathbb{Z}/2\mathbb{Z}$ for any virtual link is quite straightforward. It was presented by V.O.Manturov in [14], [15] and [16]. In the same papers Manturov presented also two constructions of Khovanov homology with integer coefficients, but they do not generalize the original Khovanov homology of classical links and rely on preliminary geometric constructions.

Manturov's constructions of Khovanov complexes with integer coefficients split into geometric constructions cooking from an arbitrary virtual link alternatable ones, and the straightforward construction of Khovanov complex with integer coefficients which works only for alternatable links. The first construction is defined for a framed virtual link, and proceeds by doubling it. The second one relies on a 2-fold covering obtained as restriction to the link of an orientation covering of a surface, on which the link is naturally placed as a zero homologous modulo 2 curve. This surface with the link diagram and checkerboard coloring is called the *atom* of the link. For details see Manturov [14], or [15], or [16].

The construction of Khovanov complex for alternatable virtual link does not require all the information contained in the virtual link diagram. One can pass (without any loss) to a Gauss diagram and then forget orientations of its chords. A signed chord diagram obtained in this way contains everything needed for building the Khovanov complex. Alternability of a virtual link is orientability of its chord diagram.

Virtual links admit a generalization to *twisted virtual links*, which emerge in relation to links in oriented thickennings of non-orientable surfaces. Many link invariants, in particular Kauffman bracket and Jones polynomial, are extended to twisted virtual links and links in oriented thickennings of non-orientable surfaces.

A link in an oriented thickenning of a non-orientable surface which gives rise to an oriented signed chord diagram realizes homology class dual to the first Stiefel-Whitney class of the surface. This is the widest class of links in oriented thickennings of surfaces for which the classical construction of Khovanov complex works without any modification over the integers.

This creates a peculiar situation: the construction of Khovanov complex with integer coefficients works, say, for non-zero homologous links in the real projective space, but does not work for zero homologous links in the same space. It is difficult to believe that this is not just due to a lack of technique, especially since in the theory of Heegaard-Floer homology developed by Ozsváth and Szabó, which so far was quite similar to the Khovanov homology, these two classes of links in the projective space both have homology with integer coefficients. Khovanov, Rasmussen and Manturov conjectured that there should be a twisted version of Khovanov complex which works for all virtual links.

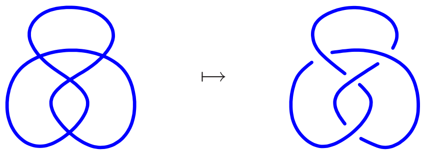
The main part of this paper starts with yet another introduction to virtual knot theory incorporating twisted virtual links. Then we review Kauffman bracket and Jones polynomial constructions, both for classical and virtual links, and show that they are defined for any signed chord diagram. The original part of the paper is devoted to the notion of orientation of chord diagrams and construction of Khovanov homology for orientable signed chord diagrams.

I am grateful to V.O.Manturov and A.N.Shumakovich for valuable information and interesting discussions.

2. Three faces of virtual knot theory

2.1. Link diagrams and Gauss diagrams

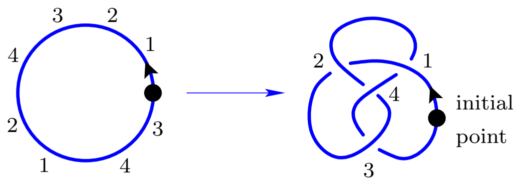
To describe graphically a classical link (that is, a closed smooth 1-dimensional submanifold of $\mathbb{R}^3$), one takes its generic projection to a plane and decorates the image at double points to show over and underpasses. This gives rise to a *link diagram:*



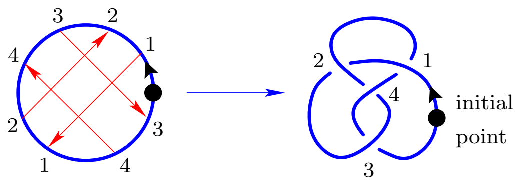
Genericity of projection means here that this is an immersion with multiple points of multiplicity at most 2 and transversality at double points.

A link diagram is a 2-dimensional picture of a link. In many cases 1-dimensional picture serves better. In particular, it is easier to convert to a combinatorial description, used as input data in computer programs.

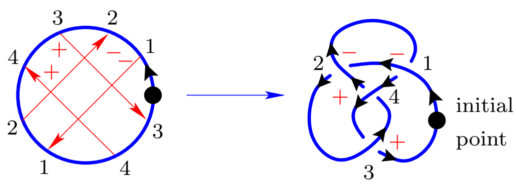
A 1-dimensional picture comes from a parametrization of the link:



The source of the parametrization is decorated. First, it is oriented. Second, each overpass is connected to the corresponding underpass with an arrow:



Third, each arrow is equipped with a sign:



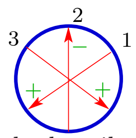
The sign is the *local writhe* of the crossing. It is + at a crossing which looks like: ,
and − at a crossing which looks like: . The signs depend on the orientation.

The result is called a *Gauss diagram* of the link. Gauss diagrams were introduced in our joint paper with Polyak [17]. The corresponding combinatorial objects called *Gauss codes* can be traced back to Gauss' notebooks. Transition from Gauss codes to their geometric counterparts, Gauss diagrams, is encouraged by geometric structures and operations, such as orientations and surgery, traditional for geometric objects, but difficult to recognize in a combinatorial context.

A Gauss diagram consists of an oriented 1-manifold (not necessarily connected) and chords connecting disjoint pairs of points on the 1-manifold. Each of the chords are oriented and equipped with a sign. The 1-manifold is called the *base* of the Gauss diagram.

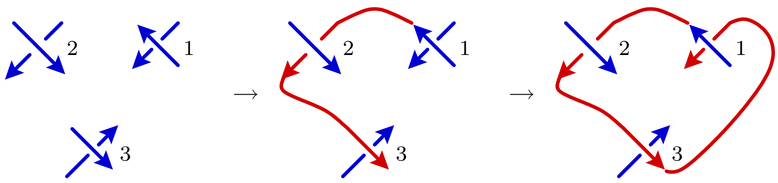
2.2. From Gauss diagrams to virtual links

Not any Gauss diagram can be obtained from a link diagram, but for any Gauss diagram one can try.

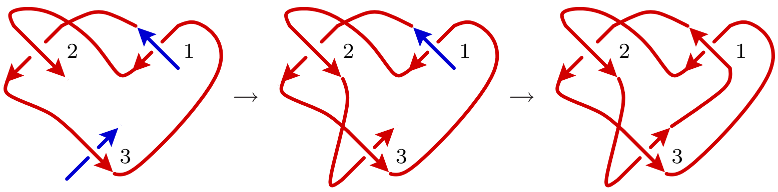
Take, for example, the Gauss diagram 

, and try to reconstruct the knot.

Let us start with crossings, as they are clearly described up to plane isotopy, then connect them step by step according to the Gauss diagram:

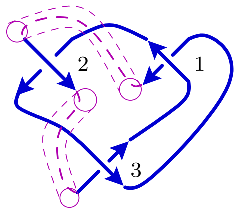


The next step meets an obstruction, as we need to penetrate through arcs that have been drawn. But let us continue neglecting the obstructions!



What is obtained looks like a knot diagram, but, besides usual crossings, it has double points which are not decorated. Such diagrams are called *virtual link diagrams*. They were introduced by Kauffman in the middle of nineties. Undecorated double points are called *virtual crossings*.

In the construction above, virtual crossings emerged inevitably. The only feasible way to avoid them is to attach handles to the plane and use them as bridges.



2.3. Link diagrams on orientable surfaces

A link diagram drawn on an orientable surface S, instead of the plane, defines a link in a thickened surface $S \times I$. It defines a Gauss diagram as well.

Any Gauss diagram appears in this way.

This is proven by the construction above, with handles added when needed. □

For each Gauss diagram, there is the smallest orientable closed surface with a link diagram defining this Gauss diagram.

Here by *smallest* I mean a surface with the greatest Euler characteristic, but without components disjoint from the link diagram. To eliminate a cheap possibility of making the Euler characteristic arbitrarily large by adding disjoint empty spheres, let us require that each connected component of the surface contains a piece of the link diagram.

To construct an orientable closed surface with greatest Euler characteristic accommodating a link diagram with a given Gauss diagram, one can first construct a germ of a link diagram on an oriented compact surface, which would contain the diagram as a deformation retract, and then cap each boundary component of the surface with a disk. For details, see [10]. □

A virtual link diagram may emerge as a projection to a plane of a link diagram on an orientable surface embedded in $\mathbb{R}^3$.

2.4. Moves

What happens to a link diagram when the link moves? It moves, too. A generic isotopy of a link can be decomposed into a sequence of isotopies each of which changes the diagram either as an isotopy of the plane or as one of the three Reidemeister moves (see Figure 1). It does not matter, if the link lies in $\mathbb{R}^3$ and its diagram lies on $\mathbb{R}^2$, or the link lies in a thickened oriented surface and its diagram lies on the surface.

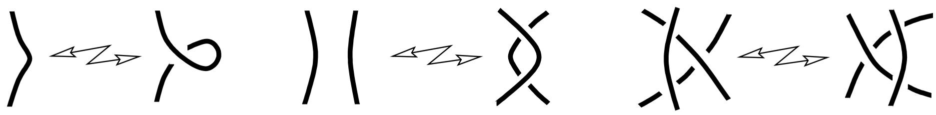


FIGURE 1. Reidemeister moves.

A virtual link diagram, which appears as a plane projection of a link diagram on a surface, moves also as shown in Figure 2 when the link moves generically in the thickened surface.

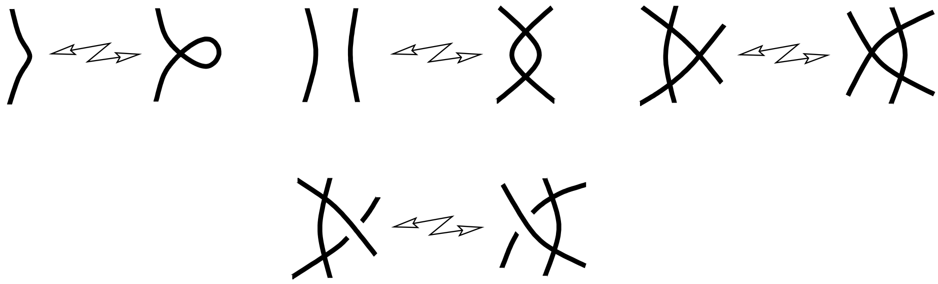


Figure 2. Virtual moves.

All virtual moves can be replaced by detour moves:

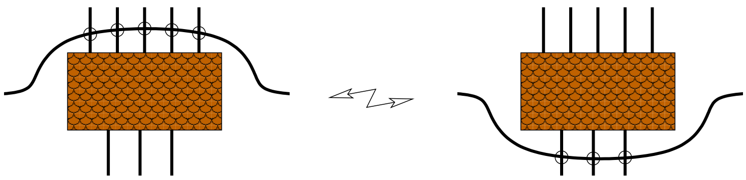


Figure 3. Detour move.

Gauss diagram does not change under virtual moves. Reidemeister moves act on Gauss diagram:

2.5. Three incarnations of virtual link theory

On the first stages of its development, the classical link theory received a completely combinatorial setup. Links were represented by link diagrams and isotopies were represented by Reidemeister moves.

We see that virtual link theory has even **two** similar **combinatorial incarnations.** Virtual links (whatever they are) are represented, on one hand, by *virtual link diagrams,* on the other hand, by *Gauss diagrams.* Furthermore, virtual isotopies (whatever they are) are represented, on one hand, by *Reidemeister and detour moves,* on the other hand, by the moves of Gauss diagrams corresponding to Reidemeister moves.

Third incarnation, truly topological one, is provided by Greg Kuperberg [13]. He proved that

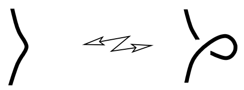
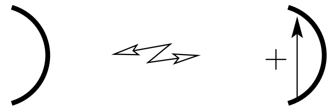
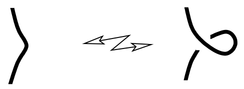
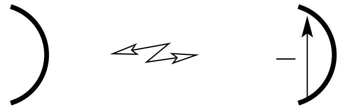
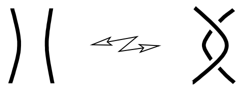
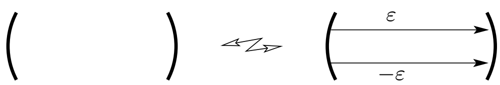
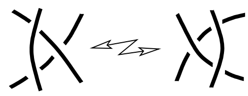
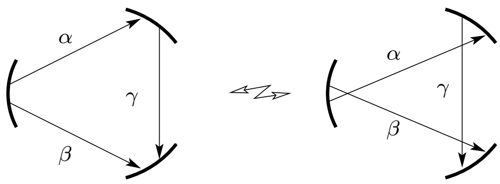
Move's name	Reidemeister move	Its action on Gauss diagram
Positive first move		
Negative first move		
Second move		
Third move		

TABLE 1. Action of Reidemeister moves on Gauss diagram.

Virtual links up to virtual isotopy are the same as irreducible links in thickened orientable surfaces up to orientation preserving homeomorphisms.

Here *irreducible* means that each connected component of the thickened surface contains some part of the link and it is impossible to find a simple curve C on the surface, which is

- disjoint from the projection of the link and
- either non-zero homologous on the surface, or separating two parts of the link projection from each other,

and it will be still impossible after any isotopy of the link.

In fact, Kuperberg [13] described more specifically how a virtual link diagram can be turned into an irreducible link in a thickened surface. He proved that a link diagram on an oriented surface can be *destabilized* to a link diagram of irreducible link on an oriented surface. A destabilization consists of embedded Reidemeister moves and Morse modifications of index 2 of the surface along a circle disjoint from the diagram.

Kuperberg's results bridge combinatorics (= 1-dimensional topology) with the 3-dimensional topology. The bridge can be used in both directions: both for extending combinatorial techniques like quantum link polynomials and link homology to links in 3-manifolds different from S^3, and for using traditional topological techniques, like signatures, in the combinatorial environment.

For instance, an oriented link in a thickened surface realizes a homology class. A homeomorphism maps a link homologous to zero to a link homologous to zero. Therefore the property of being homologous to zero is a property of the virtual link. The same holds for many other properties such as being homologous to zero modulo any number.

Furthermore, for a link homologous to zero modulo 2 in a thickened oriented surface one can define a *link signature*. Hence one can expect that there is a purely combinatorial construction of signature for virtual links of this kind.

2.6. Twisted virtual links

A non-orientable surface can also be thickened to an oriented 3-manifold. For example, take a Möbius band M embedded in $\mathbb{R}^3$ and thicken it, that is, take its regular neighborhood. See Figure 4. A neighborhood of M in $\mathbb{R}^3$ is orientable and fibers over M.

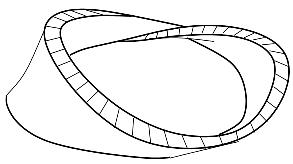
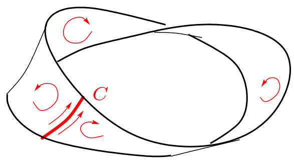


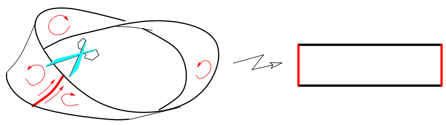
FIGURE 4.

To construct such a thickening, there is no need to embed a non-orientable surface to $\mathbb{R}^3$. Moreover, many of the non-orientable surfaces cannot be embedded into $\mathbb{R}^3$, but each of them has an orientable thickening. To thicken a non-orientable surface S:

(1) Find an *orientation change curve*[1] C (like *International date line*) on S.



(2) Cut S along C: $S \mapsto S$ ✂ C



(3) Thicken: (S ✂ C) $\times \mathbb{R}$.

(4) Paste over the sides of the cut $(x_+, t) \sim (x_-, -t)$.

[1]A curve realizing homology class Poincaré dual to the first Stiefel-Whitney class of S.

Although this construction seems to depend on the choice of the orientation change curve, the thickening does not: it is unique up to a homeomorphism and can be described as the total space of line bundles over the surface with the first Stiefel-Whitney class equal to the Stiefel-Whitney class of the surface.

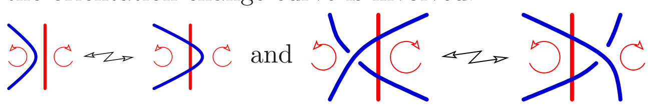
A link in an orientable thickening of a non-orientable surface has a diagram on the surface. Since the fibration of the thickened surface is not orientable one should take special care on the overpasses and underpasses. To distinguish them, one should orient the fiber over the crossing. Since it is impossible to orient all fibers coherently, the most natural solution is to use the orientation of the fibration over the complement of an orientation change curve and keep the curve as shown in the diagram.

To encode the orientation of the thickening in a diagram, one has to orient the complement of an orientation change curve on the base surface. The local orientation of the base and the global orientation of the total space determine the orientation of a fiber. This orientation is used to distinguish overpasses and underpasses.

The orientation of the complement of an orientation change curve on the base surface is not determined by the link in an oriented thickening of a non-oriented surface. It can be reversed on any component of the surface. The reversing switches over and underpassings on this component, but preserves local writhes which are defined as follows.

local writhe $= +1$ local writhe $= -1$ local writhe $= -1$ local writhe $= +1$

A generic isotopy of a link in an orientable thickening of a non-orientable surface decomposes into a sequence of isotopies each of which acts on the diagram either as an isotopy of the surface, or a Reidemeister move, or one of the following two additional moves, in which the orientation change curve is involved:



The first of these moves happens when a piece of link penetrates through the preimage of the orientation change curve. The second happens when a crossing moves through the orientation change curve. Since the orientation of the fiber changes at the moment, the overpass and underpass exchange.

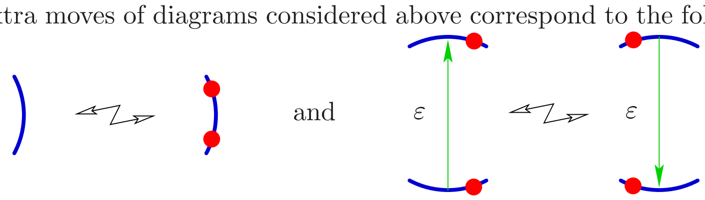
In practice, it is convenient to cut the diagram along the orientation change curve, but keep in mind the identification which would allow to recover the cut. Say, in the case of thickened projective plane, for an orientation change curve one can take the projective line. The cut along it gives a disk, which is much more comfortable to draw on than the projective plane. In this case, the theory sketched above turns into the theory of diagrams for links in the projective space developed by Julia Drobotukhina [6].

In general, links in an oriented thickening of a non-orientable surface were considered by Mario O.Bourgoin in his Research Statement [3]. He also suggested that the corresponding generalization of virtual links, twisted virtual links, and announced a generalization of Kuperberg's theorem to the twisted setup.

We consider the corresponding generalization of Gauss diagrams (rather than generalization of virtual link diagrams outlined by Bourgoin [3]).

Twisted Gauss diagram is a Gauss diagram with a finite set of points marked on the base curve. No marked point coincide with an end-point of an arrow.

The two extra moves of diagrams considered above correspond to the following moves:



2.7. Stripping of the third dimension

Recall that a classical link diagram is a decorated generic projection of the link. The generic projection is a generic immersion. A great piece of what was said above about link diagrams can be repeated, with appropriate simplifications, about generic immersions of a 1-manifold. For short, we call the image of a generic immersion of a closed 1-manifold to a surface S a *generic curve* on S. The immersion gives rise to a parametrization of the curve.

A closed 1-manifold with a finite number of chords connecting pairwise distinct points of the 1-manifold is called a *chord diagram*. The 1-manifold is called the *base* of the chord diagram. A chord diagram, each chord of which is equipped with orientation, is called an *arrow diagram*. A chord diagram, each chord of which is equipped with a sign, is called a *signed chord diagram*. Gauss diagrams considered above are signed arrow diagrams.

A generic immersion of a 1-manifold to a surface defines a chord diagram, in which the base is the source 1-manifold and each chord connects points having the same image. If the source 1-manifold and target surface of the immersion are oriented, the chords get natural orientations: direct a chord from branch A to branch B such that the orientation at the target is defined by the basis formed of vectors which are the images of tangent vectors to A and B defining the orientation of the source and taken in this order.

In the case of a link diagram, these orientations of chords can be obtained from the orientations and local writhes involved in the Gauss diagram by multiplying them: if the sign of the arrow in Gauss diagram is positive, take its orientation intact, if the sign is negative, reverse the orientation.

Not any arrow diagram or even a chord diagram can be generated by a generic immersion to a plane. This was the problem which led Gauss to think how one can recognize which diagrams appear in this way. This problem received solutions in a number of ways, but we refrain from going into this vast matter.

A step from generic curves on an oriented surface parallel to the step from link diagrams on an oriented surface to virtual link diagrams gives rise to *flat virtual knots*, see

D. Hrencecin, L. Kauffman [8]. The counterpart of Reidemeister moves are *flat Reidemeister moves*, see [8].

I am not aware about any counter-part of the Kuperberg Theorem, which would relate flat virtual knots considered up to flat Reidemeister moves to irreducible generic immersions of a 1-manifold to an orientable surface considered up to homeomorphism and homotopy.

3. Kauffman bracket of virtual links

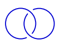
3.1. Digression on Kauffman bracket of classical links

Kauffman bracket of a classical link diagram is a Laurent polynomial in A with integer coefficients

$$\langle \text{Link diagram} \rangle \in \mathbb{Z}[A, A^{-1}]$$

For example,

$\langle \text{unknot} \rangle = \langle \bigcirc \rangle = -A^2 - A^{-2}$

$\langle \text{Hopf link} \rangle = \langle$  $\rangle = A^6 + A^2 + A^{-2} + A^{-6}$

$\langle \text{empty link} \rangle = \langle \quad \rangle = 1$

$\langle \text{trefoil} \rangle = \langle$  $\rangle = A^7 + A^3 + A^{-1} - A^{-9}$

$\langle \text{figure-eight knot} \rangle = \langle$  $\rangle = -A^{10} - A^{-10}$

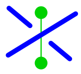
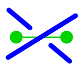
Kauffman bracket is defined by the following properties:

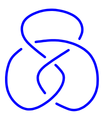
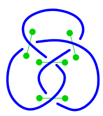
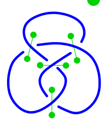
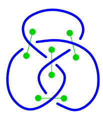
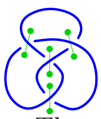
(1) $\langle \bigcirc \rangle = -A^2 - A^{-2}$,

(2) $\langle D \amalg \bigcirc \rangle = (-A^2 - A^{-2})\langle D \rangle$, (here $\amalg$ means disjoint union)

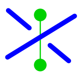
(3) $\langle$  $\rangle = A\langle$  $\rangle + A^{-1}\langle$  $\rangle$ (*Kauffman Skein Relation*) .

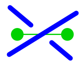
Indeed, applying the last property to each crossing of a link diagram, one reduces the diagram to a collection of embedded circles. Then the first two properties complete the job. This calculation can be summarized in the following *state sum model*.

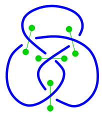
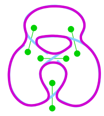
A *state* of diagram is a distribution of *markers* over all crossings. At each crossing of the diagram there should be a marker specifying a pair of vertical angles:  ,  .

For example, the knot diagram:  has states:  ,  ,  ,  , ... All together it has 2^c states, where c is the number of crossings.

Three numbers are associated to a state s of a diagram D:

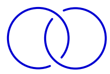
- the number $a(s)$ of *positive* markers  ,

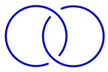
- the number $b(s)$ of *negative* markers ,
- the number $|s|$ of components of the curve D_s obtained by smoothing of D along the markers of s.

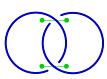
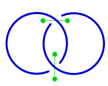
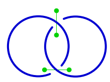
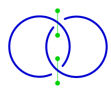
For example, state $s =$  has $a(s) = 1$, $b(s) = 3$, smoothing$(s) =$ , and $|s| = 2$.

The contribution of a state s to Kauffman bracket along with the calculation sketched above is $A^{a(s)-b(s)}(-A^2 - A^{-2})^{|s|}$. Finally, the whole Kauffman bracket is equal to the following *state sum:*

$$\langle D \rangle = \sum_{s \text{ state of } D} A^{a(s)-b(s)}(-A^2 - A^{-2})^{|s|}$$

Example: Consider the Hopf link, 

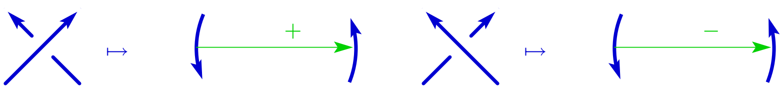
$\Big\langle$  $\Big\rangle =$

$\Big\langle$  $\Big\rangle + \Big\langle$  $\Big\rangle + \Big\langle$  $\Big\rangle + \Big\langle$  $\Big\rangle =$

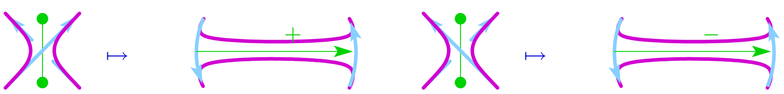
$A^2(-A^2 - A^{-2})^2 + 2(-A^2 - A^{-2}) + A^{-2}(-A^2 - A^{-2})^2 =$
$(A^6 + 2A^2 + A^{-2}) - 2A^2 - 2A^{-2} + (A^2 + 2A^{-2} + A^{-6}) =$
$A^6 + A^2 + A^{-2} + A^{-6}$.

3.2. Kauffman state sum in terms of Gauss Diagrams

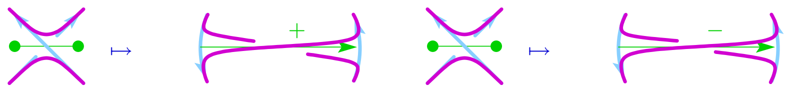
Let us rewrite the state sum above in terms of Gauss diagrams. Recall that to a crossing of a classical link diagram there corresponds an arrow equipped with a sign in the corresponding Gauss diagram.



Clearly, to a smoothing of a crossing there corresponds a surgery along the corresponding arrow:



positive marker, positive crossing   negative marker, negative crossing



negative marker, positive crossing positive marker, negative crossing

We see that the type of smoothing depends only on two signs, the sign of the marker and the sign of the crossing. Namely, the surgery preserves orientation, if these signs coincide; if the signs are opposite, the surgery reverses orientation.

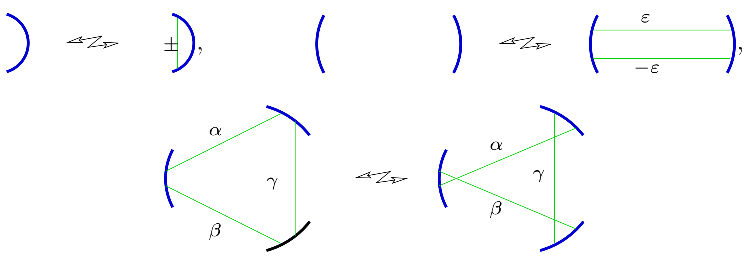
Thus, a state and all the numerical characteristic of a state involved in the Kauffman state sum can be read from the Gauss diagram. A state of Gauss diagram is a distribution of signs (signs of markers) over the set of all of its arrows. The number $|s|$ of components of the curve obtained by the smoothing along markers can be figured out from the Gauss diagram and signs of markers.

Moreover, to recover the Kauffman state sum we do not even need an important ingredient of the Gauss diagram: directions of its arrows. Therefore we can forget about them. Let us check if this is a useful possibility.

3.3. Blunted Gauss diagrams

A Gauss diagram with directions of arrows forgotten is a signed chord diagram. The forgetting of directions is called *blunting*, the result is called a *blunted Gauss diagram*.

Moves of Gauss diagrams defined by Reidemeister moves under blunting turns to the following moves of signed chord diagrams:



We shall call these moves *Reidemeister moves of signed chord diagrams*.

As shown above, Kauffman state sum is defined in terms of the corresponding blunted Gauss diagram. Such a state sum is defined for any signed chord diagram. The classical proof of invariance of the Kauffman bracket under the second and third Reidemeister moves works perfectly in the setup of signed chord diagrams. Moreover, under the first Reidemeister move the Kauffman bracket of a signed diagram behaves exactly as in the classical setup: the positive first move causes multiplication of the Kauffman bracket by $-A^3$, the negative one by $-A^{-3}$.

Thus, the Laurent polynomial

$$f_D(A) = (-A)^{-3w(D)}\langle D\rangle,$$

where $w(D)$ is the sum of signs of all chords of a signed chord diagram D, is invariant under all Reidemeister moves. As well-known, it is closely related to the Jones polynomial in the classical case. However, as everything here works just fine for blunted Gauss diagrams, it can be (and it was) taken as a definition of the Jones polynomial for virtual links.

3.4. Twisted versus blunted

Observe that under each of the additional two moves of twisted Gauss diagrams the signs of arrows do not change.

Therefore if we forget both directions of arrows and points in a twisted Gauss diagram, we get a signed chord diagram and moves of twisted Gauss diagrams turn into Reidemeister moves of signed chord diagrams.

This gives Kauffman bracket and Jones polynomial for twisted virtual links. This has been done in literature, as well. For links in oriented thickened projective plane (which is equivalent to the 3-dimensional projective space $\mathbb{R}P^3$, as the oriented thickened projective plane is the complement of a point in $\mathbb{R}P^3$), the Kauffman bracket and Jones polynomial were defined and studied by Y. V. Drobotukhina [5] in 1990. For general twisted virtual links these polynomials and their refinement was outlined by Bourgoin [3].

At first glance, everything is preserved when we pass from the classical links to virtual and even twisted virtual ones. However, this impression is a little bit misleading. Some properties of the Kauffman bracket and Jones polynomial change drastically.

3.5. Exponents in the Kauffman bracket

As we have seen in the examples

⟨unknot⟩ = ⟨◯⟩ = $-A^2 - A^{-2}$

⟨Hopf link⟩ = ⟨  ⟩ = $A^6 + A^2 + A^{-2} + A^{-6}$

⟨empty link⟩ = ⟨ ⟩ = 1

⟨trefoil⟩ = ⟨  ⟩ = $A^7 + A^3 + A^{-1} - A^{-9}$

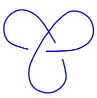
⟨figure-eight knot⟩ = ⟨  ⟩ = $-A^{10} - A^{-10}$,

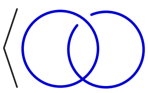
exponents in Kauffman bracket of a classical link are congruent to each other modulo 4.

This happens because:

- the contribution $A^{a(s)-b(s)}(-A^2 - A^{-2})^{|s|}$ of each state has this property,
- from each state we can get to any other state changing one marker a time,
- and changing a single marker in a state changes $a(s) - b(s)$ by 2 and $|s|$ by 1.

However, this is not the case for virtual links. For example,

⟨  ⟩ $= A^{-4} + A^{-6} - A^{-10}$

⟨  ⟩ $= A + A^{-1}$

What is wrong in the proof above if the link is virtual?

- Still, in the contribution $A^{a(s)-b(s)}(-A^2 - A^{-2})^{|s|}$ of each state all the exponents are congruent to each other modulo 4.
- Still, from each state we can get to any other state changing one marker a time.
- Still, changing a single marker changes $a(s) - b(s)$ by 2. However, $|s|$ may be preserved.

Change of a single marker causes a Morse modification of the result of smoothing. In the classical case this *Morse modification is embedded in plane, and therefore preserves orientation.*

As well-known, *an orientation preserving Morse modification of a 1-manifold changes the number of components by one. A Morse modification, which does not preserve orientation of a connected 1-manifold preserves the number of connected components.* A Morse modification which does not preserve orientation would cause a shift of exponents by 2.

4. Orientations of chord diagrams

4.1. Orientation of a chord diagram

Which structure on a blunted Gauss diagram would guarantee that, for any of its states, the corresponding smoothing has an orientation such that the change of any marker causes a Morse modification of the result of smoothing *preserving* the orientation?

Such a structure is a collection of orientations of arcs of the base 1-manifold between end points of chords such that these orientations define an orientation of each smoothing. For this, the orientations should have the following two properties:

- The orientations cannot be extended over an end point of a chord,
- For each chord, one of its end points is attractive, while the other one is repulsive.

See Figure 5.

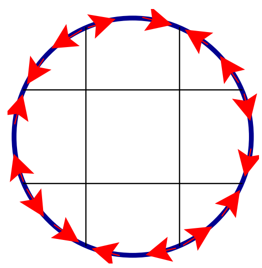


Figure 5. Orientation of a chord diagram.

The first property means that the orientations of arcs induce anonymously an orientation of each end point of each chord. The second property means that for each chord the induced orientations of its end points are opposite to each other.

Orientations of arcs of a chord diagram satisfying these two properties is called an *orientation* of a chord diagram.

Orient chords so that the induced orientations of their end points are opposite to the orientations induced by orientations of adjacent arcs. See Figure 6.

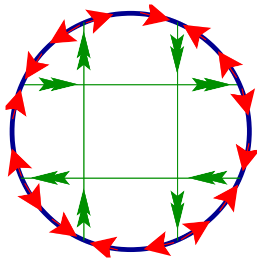


FIGURE 6. Orientations of chords in an oriented chord diagram.

Chords become arrows. Orientations of the arrows alternate in the sense that along the base of the diagram each arrowhead is followed by an arrowtail and each arrowtail is followed by an arrowhead.

Observe that an orientation of a chord diagram defines an orientation on the result of any of its smoothings. Indeed, the result of a smoothing is composed from arcs of the base and arcs which go along chords. The orientations of these pieces agree with each other.

4.2. Alternatable virtual link diagrams

Observe that directions of arrows in a Gauss diagram of a classical alternating link diagram alternate in the same way as the orientations of chords considered above.

Recall that a classical link diagram is called *alternating*, if along a branch of the link overcrossings would always follow undercrossings and undercrossings would always follow overcrossings.

Similarly, a virtual link is called *alternating*, if along a branch of the link over-crossing always follow after under-crossing and under-crossing always follow after over-crossing. By switching some overpasses and underpasses, a virtual link diagram such that its blunted Gauss diagram is orientable, can be made alternating. Vice versa, if, by switching some overpasses and underpasses, a virtual link diagram can be made alternating, its blunted Gauss diagram is orientable.

Therefore a virtual link diagram with orientable blunted Gauss diagram is called *alternatable*.

4.3. Moves of chord diagrams

Transformations of a chord diagram shown in Figure 7 are called *Reidemeister moves*.

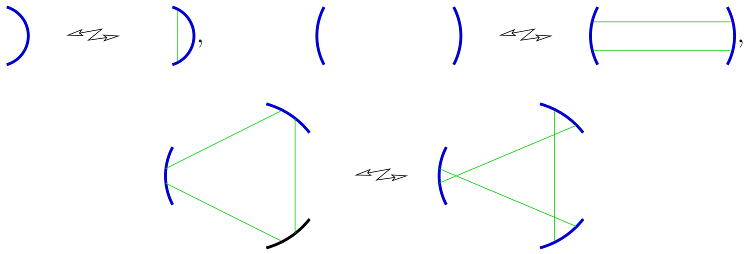


FIGURE 7. Reidemeister moves of chord diagrams

Theorem 4.A. *The result of any first or third Reidemeister move applied to an orientable chord diagram is orientable.*

Proof. An orientation of a chord diagram prior to the move in the fragment which is about to change is unique up to reversing. Its replacement admits an orientation coinciding with the original one near the boundary. See Figure 8. □

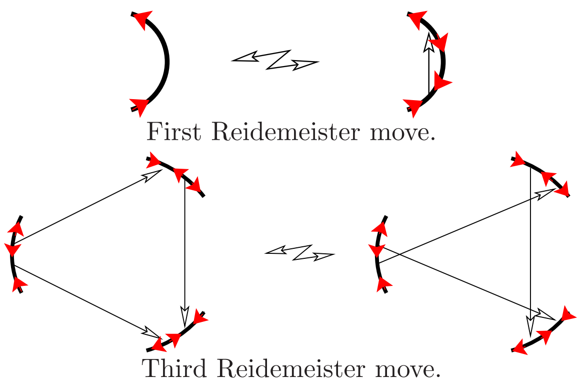
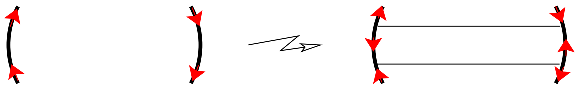
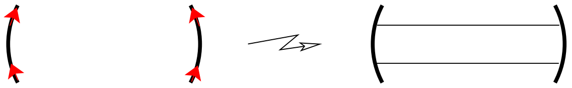


FIGURE 8. Behavior of orientation of a signed chord diagram under first and third Reidemeister moves.

The second Reidemeister move can destroy orientability. Of course, a second Reidemeister move can transform an orientable chord diagram to an orientable one. All second Reidemeister moves decreasing the number of chords and about half of all second Reidemeister moves increasing it which can be applied to a chord diagram preserve orientability. One can easily recognize if a second Reidemeister move preserves orientability, see Figure 9.



An orientation surviving under a second Reidemeister move



Second Reidemeister move distroys an orientation

Figure 9. Behavior of orientations of a chord diagram under second Reidemeister moves.

4.4. Checkerboard coloring of a classical link diagram

Let us analyze where the orientation of a chord diagram underlying a Gauss diagram of a classical link comes from.

A generic curve on a plane can be considered as a 1-cycle with coefficients in $\mathbb{Z}_2$ on the plane. As a plane has trivial homology, this 1-cycle bounds a 2-chain. The latter is described as the union of all black domains in the checkerboard coloring of the diagram.

Each connected component of the complement of a generic curve inherits an orientation from the whole plane. These orientations of the black domains induce orientations on their boundaries. The boundary of a black domain consists of pieces of the curve. Thus an arc of the curve between any two consecutive double points gets a natural orientation. See Figure 10. It is called a *checkerboard orientation.*

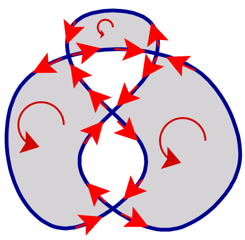


Figure 10. Checkerboard coloring and orientations.

Obviously, the checkerboard orientation gives orientation of the corresponding chord diagram. The orientation induced by the orientation of a chord diagram on the result of a smoothing of the curve can also be obtained as a checkerboard orientation.

Indeed, the result of a smoothing also admits a checkerboard coloring which coincides with the checkerboard coloring of the diagram outside small neighborhoods of its double points. The orientation induced on the smoothing by the orientations of the black domains agree with the checkerboard orientation of the diagram. See Figure 11.

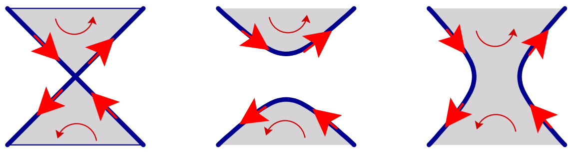


FIGURE 11. Checkerboard orientation of smoothings.

4.5. On orientable surfaces

The same arguments work for a link diagram zero homologous modulo 2 on an oriented surface. An orientable surface with a zero homologous generically immersed curve admits a checkerboard coloring. We see that:

Theorem 4.B. *The chord diagram of a generic curve on an orientable surface admitting a checkerboard coloring is orientable.*

Corollary (N. Kamada [9]). *A link diagram on an orientable surface admitting a checkerboard coloring is alternatable.* □

Corollary (N. Kamada [9]). *Exponents of monomials in the Kauffman bracket of a checkerboard colorable link diagram on an oriented surface are congruent to each other modulo 4.* □

It may happen that an alternatable link diagram on an orientable surface does not admit a checkerboard coloring. Indeed, any diagram admitting a checkerboard coloring can be spoiled by a Morse modification of index 1, that is, removing two disjoint open disks and attaching a tube connecting the boundary circles. For this, take the disks in the complementary domains colored with different colors. Of course, this stabilization of a virtual link diagram does not destroy alternatability, which is a property of Gauss diagrams.

Nonetheless, according to the following theorem, which is also basically due to Naoko Kamada [9], this cannot happen to irreducible diagram.

Theorem 4.C. *A generic curve on an orientable surface such that each connected component of its complement is a disk and its chord diagram is orientable admits a checkerboard coloring.*

Corollary (N. Kamada [9]). *An alternatable link diagram on an orientable surface such that each connected component of its complement is a disk admits a checkerboard coloring.*

Proof of Theorem 4.C. An orientation of the chord diagram gives rise to an orientation of the boundary of each connected component of the complement of the curve. Since each such component is a disk, the orientation of its boundary orients the component itself. On the other hand, the orientation of the whole surface induces an orientation on each of these components. For some of the components these two orientations coincide, for the others they are opposite to each other. The components for which the orientations coincide form a chain modulo 2 bounded by the curve. By coloring these components in black and the others in white, we get the desired checkerboard coloring of the surface. □

The condition about components of the complement can be weakened: instead of requiring that they are homeomorphic to a disk, it suffices to require that the intersection of the curve with the closure of each of the components is connected.

4.6. Alternatable virtual links

Recall that a virtual link diagram which gives rise to an orientable blunted Gauss diagram is called *alternatable*.

Theorem 4.D (Corollary of Theorem *4.A*)**.** *The result of any first or third Reidemeister move applied to an alternatable virtual link diagram is alternatable.* □

A second Reidemeister move can turn an alternatable virtual link diagram to a non-alternatable diagram, see Figure 12.

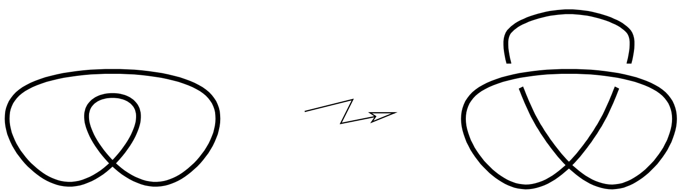


Figure 12. Creating a non-alternatable virtual knot diagram from a virtual diagram of unknot with single virtual crossing by a second Reidemeister move.

Moreover, each alternatable virtual diagram can be made non-alternatable by a single second Reidemeister move. Therefore alternatability is not a property shared by all virtual diagrams of a given virtual link.

Alternatable isotopy is a sequence of Reidemeister moves preserving alternatability and virtual moves of an alternatable virtual link diagram.

Theorem 4.E. *Alternatable virtual link diagrams are virtualy isotopic, if and only if they can be related by an alternatable isotopy.*

Proof. Let D_1 and D_2 be virtually isotopic alternatable link diagrams. Realize each of them as a diagram admitting a checkerboard coloring on an oriented closed surface.

According to the Kuperberg Theorem [13], virtually isotopic link diagrams can be destabilized to link diagrams on an oriented surface S, where they can be related by embedded moves. A destabilization consists of embedded Reidemeister moves and Morse modifications of index 2 of the ambient surface along a circle disjoint with the diagram.

Neither an embedded Reidemeister move, nor a Morse modification of index 2 can destroy a checkerboard coloring. A Morse modification of index 2 does not change the Gauss diagram. Therefore the sequence of moves connecting the Gauss diagrams corresponding to D_1 and D_2 which correspond to the Reidemeister moves existing by the Kuperberg Theorem constitute an alternatable isotopy. □

Thus alternatable virtual links do not form a new category of objects similar to virtual links, but are virtual links of special kind. They have special properties. For example, the exponents of Kauffman bracket are congruent to each other modulo 4, see the second Corollary of Theorem *4.B* above. From the point of view of 3-dimensional topology, they can be characterized as irreducible links in thickened oriented surfaces which realize trivial homology class modulo 2.

4.7. On non-orientable surfaces

Theorem *4.B* and its corollary cannot be extended literally to twisted link diagrams on non-orientable surfaces . For example, link 2_1 on the projective plane has checkerboard colorable diagram, but its Kauffman bracket is $A^{-4} + A^{-6} - A^{-10}$, see [5].

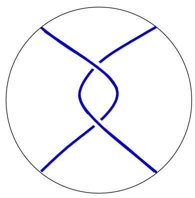
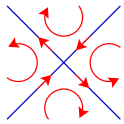


Figure 13. Projective link 2_1

However, a generalization recovers, if the property of being zero homologous modulo 2 is generalized in an other way.

Theorem 4.F. *The chord diagram of a generic curve realizing the homology class dual to the first Stiefel-Whitney class of the surface is orientable.*

Proof. Let C be a generic curve on a surface S realizing a homology class dual to $w_1(S) \in H^1(S;\mathbb{Z}_2)$. The complement $S \smallsetminus C$ admits an orientation which cannot be extended over C at any point. Each arc of C is involved in C with multiplicity 2 in the boundary of the corresponding fundamental class $[S \smallsetminus C]$. At a double point the orientations of $S \smallsetminus C$ and arcs of C look as follows:



Thus the orientations of arcs of C gives rise to an orientation of the corresponding chord diagram. □

Corollary. *Exponents of monomials in the Kauffman bracket of a link which is non-zero homologous in the real projective space are congruent to each other modulo 4.* □

Theorem 4.G (Generalization of Theorem *4.C*)***.*** *If the chord diagram of a generic curve on a surface is orientable and each connected component of its complement on the surface is a disk then the curve realizes the homology class dual to the first Stiefel-Whitney class of the surface.*

Proof. An orientation of the chord diagram gives rise to an orientation of the boundary of each connected component of the complement of the curve. Since each such component is a disk, the orientation of its boundary orients the component itself.

The orientations of the components of the complement form an orientation of the whole complement, which does not extend over the curve, because from both sides of the curve it induces the same orientation on each piece of the curve. Therefore the curve realizes the obstruction to orientability of the surface. □

4.8. Obstruction to orientability of a chord diagram

Near each chord of a chord diagram an orientation of the diagram looks standard, up to simultaneous reversing. For any choice of the orientation, on each arc of the base either the orientations at the end points extend to the whole arc, or not. Thus for any choice of orientations of chords there is a well-defined $\mathbb{Z}/2\mathbb{Z}$-valued function of the set of arcs of the base. It can be considered as a 1-cochain with values in $\mathbb{Z}/2\mathbb{Z}$ on the chord diagram. Reversing of a single chord's orientation causes change of values on each of the four adjacent arcs.

This means that there is a well-defined 1-dimensional cohomology class with values in $\mathbb{Z}/2\mathbb{Z}$ of the underlying space of the chord diagram (the union of the base and chords). This class is zero, if and only if the diagram is orientable. Therefore we call it the *obstruction to orientability.*

The class is defined by topology of the pair consisting of the underlying space and the union of all chords. It is not defined by the homotopy type of the underlying space, although it belongs to a cohomology group depending only on the homotopy type. It is not defined by the topology of the 4-valent graph obtained by contracting each of the chords, either. Indeed, it depends on division of arcs adjacent to a vertex to pairs of opposite arcs, that is, arcs adjacent to one end point of the chord in the underlying space of the chord diagram.

The simplest example is a circle with two chords intersecting each other, on one hand, and two circles connected with two chords, on the other hand. The quotient 4-valent graphs are homeomorphic.

5. Khovanov homology of oriented signed chord diagrams

5.1. Khovanov homology of classical links

In 1998 Khovanov [12] categorified Jones polynomial. His construction is a refinement of the Kauffman bracket construction. To any classical link diagram D it associates a bi-graded chain complex of abelian groups $C^{i,j}(D)$ with differential $d : C^{i,j}(D) \to C^{i+1,j}(D)$ and homology groups $\mathcal{H}^{i,j}(D)$ such that the Reidemeister moves of the diagram induce homotopy equivalences of the complex. The Khovanov homology $\mathcal{H}^{i,j}(D)$ categorifies the Jones polynomial $f_D(A)$ in the sense that

$$f_D(A) = \sum_{i,j}(-1)^{i+j} A^{-2j} \operatorname{rk} \mathcal{H}^{i,j}(D).$$

This slightly strange formula is due to the fact that Khovanov used different normalization of the Jones polynomial, namely $K(D)(q) = f_D(\frac{1}{\sqrt{-q}})$. Ranks of the Khovanov homology groups are related to K as follows:

$$K(D)(q) = \sum_{i,j}(-1)^i q^j \operatorname{rk} \mathcal{H}^{i,j}(D).$$

For any state s of D, the construction associates to each connected component C of D_s a copy $\mathcal{A}_C$ of a graded free abelian group $\mathcal{A}$ with two generators, 1 of grading 1 and x of grading -1. To the whole s the construction associates a graded group $C(s)$ which is the tensor product of all copies $\mathcal{A}_c$ of $\mathcal{A}$ associated to the connected components of D_s with grading shifted by $\frac{3w(D)-a(s)+b(s)}{2}$, where $w(D)$ is the writhe of D.

For a state s of diagram D denote $\frac{w(D)-a(s)+b(s)}{2}$ by $i(s)$ and consider $C(s)$ as a bigraded group with the first grading to be identically equal to $i(s)$ and the second grading as defined above (that is, the grading of the tensor product shifted by $\frac{3w(D)-a(s)+b(s)}{2}$). Denote by $\mathcal{C}(D)$ the bigraded group $\sum_s C(s)$. This is the total group of the Khovanov chain complex.

To define a differential, we need to fix an order of all crossings of the diagram. We also need to fix the multiplication $m : \mathcal{A} \otimes \mathcal{A} \to \mathcal{A}$ defined by the formulas $1 \otimes 1 \mapsto 1$, $1 \otimes x \mapsto x$, $x \otimes 1 \mapsto x$ and $x \otimes x \mapsto 0$, and co-multiplication $\Delta : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ defined by formulas $1 \mapsto 1 \otimes x + x \otimes 1$ and $x \mapsto x \otimes x$.

The differential $C^{i,j} \to C^{i+1,j}$ is defined as the sum of partial differentials $C^j(s) \to C^j(t)$, where s and t are states with $i(s) = i$, $i(t) = i+1$ such that t differs from s only by a marker at one crossing. Denote this crossing by x and its number by k. Denote by r the number of negative markers of s at crossings with numbers greater than k. At x the marker of s is positive, while the marker of t is negative.

Let two connected components of D_s pass near x. Denote them by C_1 and C_2. Then there is only one connected component of D_t passing near x. Denote this component by C. Then $C(s) = \mathcal{A}_{C_1} \otimes \mathcal{A}_{C_2} \otimes B$ and $C(t) = \mathcal{A}_C \otimes B$ and the partial differential $C(s) \to C(t)$ is defined by the formula $(-1)^r m \otimes id_B$.

Let only one component of D_s pass near x. Denote it by C. Then two components of D_t pass near x. Denote them by D_{C_1} and D_{C_2}. Then $C(s) = \mathcal{A}_C \otimes B$ and $C(t) = \mathcal{A}_{C_1} \otimes \mathcal{A}_{C_2} \otimes B$ and the partial differential $C(s) \to C(t)$ is defined by the formula $(-1)^r \Delta \otimes id_B$.

5.2. Khovanov complex of an oriented signed chord diagram

It is clear that the construction of Khovanov complex sketched above can be expressed completely in terms of signed chord diagram, as it was done with the Kauffman bracket in Section 3.2.

If a signed chord diagram is orientable, all properties of its Khovanov complexes proven in [12] or [2] can be repeated without any change. In particular, this is a complex. The only property of classical link diagrams that is used in the proof of this is that change of a single marker in a state s gives rise to a change of the number of connected components in D_s, and this is a corollary of orientability of the chord diagram.

Furthermore, the Reidemeister moves of signed chord diagrams induce homotopy equivalences of the Khovanov complex, provided they preserve orientation. The proof is also borrowed from the classical case without any change.

According to Theorem *4.E*, if alternatable virtual diagrams are virtualy isotopic then they can be related by an alternatable isotopy. Therefore, the homotopy type of the Khovanov complex and, in particular, Khovanov homology groups are invariants of alternatable virtual links.

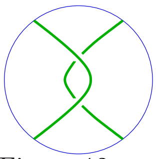
5.3. Failure in the non-orientable case

If the signed chord diagram is not orientable, there is no problem to extend the definition of the differential, because the partial differentials along a change of marker destroys orientation of D_s and hence preserving the number of components vanish for grading reasons.

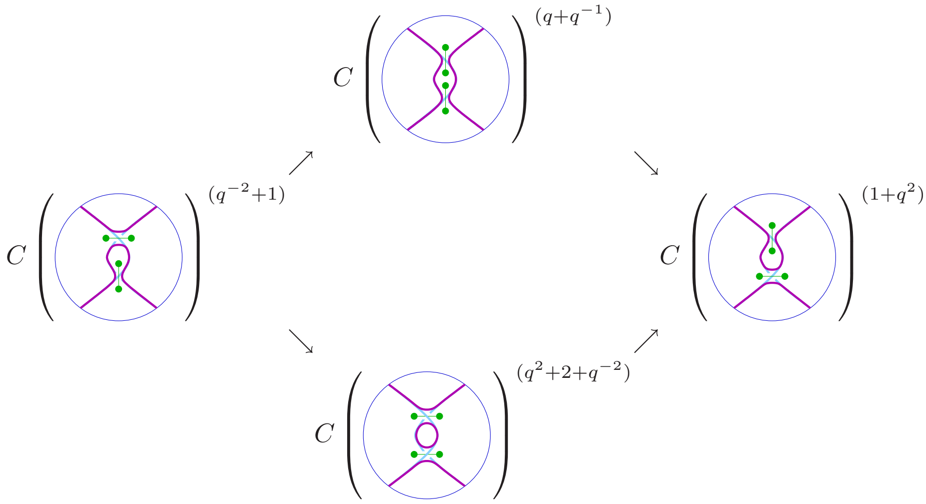
Indeed, $C(s)$ is obtained from $\mathcal{A}^{|s|}$ by a shift of the second grading by $\frac{3w(D)-a(s)+b(s)}{2}$. The graded rank [2] of $C(s)$ is $q^{\frac{3w(D)-a(s)+b(s)}{2}}(q+q^{-1})^{|s|}$. All exponents of this graded rank are congruent to each other modulo 2. When $a(s)$ and $b(s)$ change by one and $|s|$ does not change, the graded rank is multiplied by q, and parity of all exponents changes by 1. Thus any homomorphism preserving the grading is trivial.

However, the homomorphism $\mathcal{C}(D) \to \mathcal{C}(D)$ obtained in this way is not a differential: its square is not zero. This can be easily seen in the very simple example.

[2] Recall that the graded rank of a finitely generated graded group $W = \oplus_j W_j$ is the Laurent polynomial $\Sigma_j q^j \operatorname{rk} W_j$.

The unknot in $\mathbb{R}P^3$ is isotopic to . The corresponding virtual link diagram is shown on the right hand side of Figure 12.

Consider the Khovanov chain groups. There are 4 states giving rise to Khovanov chain groups which looks as follows:



Here the graded ranks are shown as exponents. For the grading reasons (the partial differentials are of degree 0), the upper arrows should be zero. Therefore the composition of the bottom two arrows should be zero. Otherwise the square of the differential would not be zero.

But it is non-trivial in the component with grading 0. Indeed, the first of them maps 1 to $\Delta(1) = 1 \otimes x + x \otimes 1$, then the second one maps $1 \otimes x$ to $m(1 \otimes x) = x$ and $x \otimes 1$ to x. Hence the composition sends 1 to $2x$.

References

[1] C.C. Adams, *The Knot Book – An Elementary Introduction to the Mathematical Theory of Knots,* W. H. Freeman and Company, New York, 1994.

[2] D. Bar-Natan, *Khovanov's categorification of the Jones polynomial,* Algebr. Geom. Topol., 2 (2002) 337–370, arXiv: math.QA/0201043.

[3] M.O. Bourgoin, *Research Statement,* http://people.brandeis.edu/~mob/ research.pdf

[4] J.S. Carter, S. Kamada, and M. Saito, *Stable equivalence of knots on surfaces and virtual knot cobordisms,* arXiv:math.GT/0008118.

[5] Y.V. Drobotukhina, *An analogue of the Jones polynomial for links in* $\mathbb{R}P^3$ *and a generalization of the Kauffman-Murasugi theorem,* Algebra i Analiz 2 (1990), no. 3, 171–191.

[6] J. Drobotukhina, *Classification of links in* $\mathbb{R}P^3$ *with at most six crossings,* Topology of manifolds and varieties, Adv. Soviet Math., vol. 18, Amer. Math. Soc., Providence, RI, 1994, pp. 87–121.

[7] M. Goussarov, M. Polyak and O. Viro, *Finite type invariants of classical and virtual knots,* Topology 39, (2000) pp. 1045–1068, arXiv:math.GT/98100073.
[8] D. Hrencecin, L.H. Kauffman, *On Filamentations and Virtual Knots,* arXiv:math.GT/0112206.
[9] N. Kamada, *On the Jones polynomials of checkerboard colorable virtual knots,* arXiv:math.GT/0008074.
[10] N. Kamada and S .Kamada, *Abstract link diagrams and virtual knots,* J. Knot Theory Ramifications 9 (2000), 93–106.
[11] L.H. Kauffman, *Virtual knot theory,* European J. Combin. 20 (1999), no. 7, 663–690, arXiv:math.GT/9811028.
[12] M. Khovanov, *A categorification of the Jones polynomial,* Duke Math. J. 101(3) (2000) 359-426, see also arXiv:math.QA/9908171.
[13] G. Kuperberg, *What is a virtual link?,* Algebraic and Geometric Topology, 2003, 3, 587–591, arXiv:math.GT/0208039.
[14] V.O. Manturov, *The Khovanov Complex for Virtual Links,* arXiv:math.GT/0501317.
[15] V.O. Manturov, *Complex Khovanova dlya virtual'nykh zacepleniy,* Fundamental'naya i prikladnaya matematika, t.11, vol.4, pp. 127-152. (Russian), English Translation in the journal Fundamental and Applied Mathematics.
[16] V.O. Manturov, *Teoriya Uzlov* (Russian), RCD, Moscow-Izhevsk, English edition: *Knot Theory,* Chapman & Hall/CRC Press, 2004.
[17] M. Polyak and O. Viro, *Gauss diagram formulas for Vassiliev invariants,* International Math. Research Notices, (1994), No. 11, 445–453.

Department of Mathematics, Uppsala University, Box 480, S-751 06 Uppsala, Sweden
E-mail address: oleg@math.uu.se