Outer Billiards on Kites

Richard Evan Schwartz

ANNALS OF MATHEMATICS STUDIES

Annals of Mathematics Studies
Number 171

This page intentionally left blank

Outer Billiards on Kites

Richard Evan Schwartz

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD
2009

Published by Princeton University Press,
41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TW

Schwartz, Richard Evan.
Outer billiards on kites / Richard Evan Schwartz.
p. cm. – (Annals of mathematics studies ; 171)
Includes bibliographical references and index.
ISBN 978-0-691-14248-7 (cl.: alk. paper)
ISBN 978-0-691-14249-4 (pb.: alk. paper)
1. Hyperbolic spaces. 2. Singularities (Mathematics)
3. Transformations (Mathematics) 4. Geometry, Plane.
I. Title. QA685.S45 2009
516.9–dc22

British Library Cataloging-in-Publication Data is available

This book has been composed in LaTeX

Printed on acid-free paper. ∞

press.princeton.edu

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

Preface

Outer billiards is a dynamical system defined relative to a convex shape in the plane. B. H. Neumann introduced outer billiards in the 1950s, and J. Moser popularized the system in the 1970s as a toy model for celestial mechanics. When the underlying shape is smooth, outer billiards has connections to area-preserving twist maps and Kolmogorov-Arnold-Moser (KAM) theory. When the underlying shape is a polygon, outer billiards is related to interval exchange transformations and piecewise isometric actions. Outer billiards is an appealing dynamical system because it is quite simple to define and yet gives rise to a rich intricate structure.

The *Moser-Neumann question* has been one of the basic questions guiding the subject of outer billiards. This question asks, *Does there exist an outer billiards system with an unbounded orbit*? Until recently, all the results on the subject have given negative answers to the question in particular cases. That is, it has been shown that all orbits are bounded for various classes of shape.

Recently, we answered the Moser-Neumann question in the affirmative by showing that outer billiards has an unbounded orbit when defined relative to the Penrose kite, the convex quadrilateral that arises in the famous Penrose kite-and-dart tilings. Even more recently, D. Dolgopyat and B. Fayad proved, using different methods, that outer billiards has unbounded orbits when defined relative to a half-disk.

Our original unboundedness proof involves special properties of the Penrose kite and naturally raises questions about generalizations. In this book, we will prove that outer billiards has unbounded orbits when defined relative to any irrational kite. A *kite* is a convex quadrilateral having a diagonal that is also a line of symmetry. The kite is *irrational* if the other diagonal divides the kite into two triangles whose areas are not rational multiples of each other.

As we prove the unboundedness result for irrational kites, we will explore the deep structure underlying outer billiards on kites. Our analysis reveals connections between outer billiards on kites and self-similar sets, higher-dimensional polytope exchange maps, Diophantine approximation, the modular group, the universal odometer, and renormalization. The structural results in this book perhaps point the way toward a broader theory of polygonal outer billiards.

I discovered most of the phenomena discussed in this book through computer experimentation with my program Billiard King and only later found conventional proofs. I encourage the reader of this book to download Billiard King and play with it. This Java program is platform-independent and heavily documented. The reader can download Billiard King from http://press.princeton.edu/titles/9105.html or from my Brown University website, http://www.math.brown.edu/∼/res/BilliardKing. My website also has an interactive guide to this book.

I thank Sergei Tabachnikov for both encouragement and mathematical input. I first heard about the Moser-Neumann problem from Sergei and subsequently learned a lot about outer billiards from reading his excellent book, *Geometry and Billiards*.

This book owes an intellectual debt to the beautiful result of Vivaldi-Shaidenko, Kolodziej, and Gutkin-Simanyi about the periodicity of outer billiards orbits for rational polygons. This result provided the theoretical underpinnings for my initial computer investigations. This work also owes an intellectual debt to the work of Yair Minsky on the punctured torus case of the Ending Lamination Conjecture. The notion of indexing 3-manifolds by nodes of the Farey graph inspired my idea of indexing outer billiards systems on rational kites in a similar way. I would also like to acknowledge Dan Genin's boundedness result about outer billiards on trapezoids. Some of my work on kites is very similar in spirit to the work Dan did on trapezoids.

I would like to thank Peter Ashwin, Jeff Brock, Yitwah Cheung, Dmitry Dolgopyat, Peter Doyle, David Dumas, Bernold Fiedler, Giovanni Forni, Dan Genin, Arek Goetz, Eugene Gutkin, Pat Hooper, Richard Kent, Howie Masur, Yair Minsky, Curt McMullen, Jill Pipher, John Smillie, Sergei Tabachnikov, Franco Vivaldi, and Ben Wieland for various helpful conversations about this work. Thanks are also due to Vickie Kearn and Anna Pierrehumbert at Princeton University Press for their encouragement while I worked on this project. I also thank Gerree Pecht of Princeton University for her expert LaTeX advice.

I am grateful to the National Science Foundation for its continued support, currently in the form of grant DMS-0604426. I also thank the Clay Mathematics Institute for its support, in the form of a Clay Research Scholarship. I am indebted to my home institution, Brown University, for providing an excellent research environment during the writing of this book. I also extend thanks to the Institut des Hautes Études Scientifiques, Harvard University, and the California Institute of Technology, for their hospitality during various periods of my sabbatical in 2008-2009.

I especially thank my wife, Brienne Brown, and my daughters, Lucy and Lily, for their support and understanding while I worked on this project.

I dedicate this book to my parents, Karen and Uri.

Outer Billiards on Kites

This page intentionally left blank

Chapter One

Introduction

1.1 DEFINITIONS AND HISTORY

B. H. Neumann [**N**] introduced *outer billiards* in the late 1950s. In the 1970s, J. Moser [**M1**] popularized outer billiards as a toy model for celestial mechanics. See [**T1**], [**T3**], and [**DT1**] for expositions of outer billiards and many references on the subject.

Outer billiards is a dynamical system defined (typically) in the Euclidean plane. Unlike the more familiar variant, which is simply called *billiards*, outer billiards involves a discrete sequence of moves outside a convex shape rather than inside it. To define an outer billiards system, one starts with a bounded convex set $K \subset \mathbf{R}^2$ and considers a point $x_0 \in \mathbf{R}^2 - K$. One defines x_1 to be the point such that the segment $\overline{x_0 x_1}$ is tangent to K at its midpoint and K lies to the right of the ray $\overrightarrow{x_0 x_1}$. The iteration $x_0 \to x_1 \to x_2 \to \cdots$ is called the *forward outer billiards orbit* of x_0. It is defined for almost every point of $\mathbf{R}^2 - K$. The backward orbit is defined similarly.

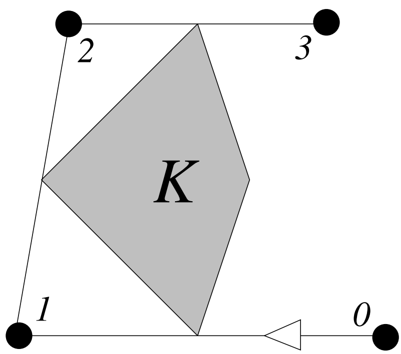

Figure 1.1: Outer billiards relative to K.

One important feature of outer billiards is that it is an affinely invariant system. Since affine transformations carry lines to lines and respect the property of bisection, an affine transformation carrying one shape to another conjugates the one outer billiards system to the other.

It is worth recalling here a few basic definitions about orbits. An orbit is called *periodic* if it eventually repeats itself, and otherwise *aperiodic*. An orbit is called *bounded* if the whole orbit lies in a bounded portion of the plane. Otherwise, the orbit is called *unbounded*. Sometimes (un)bounded orbits are called *(un)stable*.

J. Moser [**M2**, p. 11] attributes the following question[1] to Neumann ca. 1960, though it is sometimes called Moser's question. *Is there an outer billiards system with an unbounded orbit?* This is an idealized version of the question about the stability of the solar system. Here is a chronological list of much of the work related to this question.

- J. Moser [**M2**] sketches a proof, inspired by KAM theory, that outer billiards on K has all bounded orbits provided that ∂K is at least C^6 smooth and positively curved. R. Douady gives a complete proof in his thesis [**D**].
- In Vivaldi-Shaidenko [**VS**], Kolodziej [**Ko**], and Gutkin-Simanyi [**GS**], it is proved (each with different methods) that outer billiards on a *quasirational polygon* has all orbits bounded. This class of polygons includes rational polygons – i.e., polygons with rational-coordinate vertices – and also regular polygons. In the rational case, all defined orbits are periodic.
- S. Tabachnikov [**T3**] analyzes the outer billiards system for a regular pentagon and shows that there are some nonperiodic (but bounded) orbits.
- P. Boyland [**B**] gives examples of C^1 smooth convex domains for which an orbit can contain the domain boundary in its ω-limit set.
- F. Dogru and S. Tabachnikov [**DT2**] show that, for a certain class of polygons in the hyperbolic plane, called *large*, all outer billiards orbits are unbounded. (One can define outer billiards in the hyperbolic plane, though the dynamics has a somewhat different feel to it.)
- D. Genin [**G**] shows that all orbits are bounded for the outer billiards systems associated to trapezoids. See §A.4. Genin also makes a brief numerical study of a particular irrational kite based on the square root of 2, observes possibly unbounded orbits, and indeed conjectures that this is the case.
- In [**S**] we prove that outer billiards on the Penrose kite has unbounded orbits, thereby answering the Moser-Neumann question in the affirmative. The Penrose kite is the convex quadrilateral that arises in the Penrose tiling.
- Recently, D. Dolgopyat and B. Fayad [**DF**] showed that outer billiards on a half-disk has some unbounded orbits. Their proof also works for regions obtained from a disk by nearly cutting it in half with a straight line. This is a second affirmative answer to the Moser-Neumann question.

The result in [**S**] naturally raises questions about generalizations. The purpose of this book is to develop the theory of outer billiards on kites and show that the phenomenon of unbounded orbits for polygonal outer billiards is (at least for kites) quite robust.

[1] It is worth pointing out that outer billiards relative to a line segment has unbounded orbits. This trivial case is meant to be excluded from the question.

1.2 THE ERRATIC ORBITS THEOREM

A *kite* is a convex quadrilateral K having a diagonal that is a line of symmetry. We say that K is *(ir)rational* if the other diagonal divides K into two triangles whose areas are (ir)rational multiples of each other. Equivalently, K is rational iff it is affinely equivalent to a quadrilateral with rational vertices. To avoid trivialities, we require that exactly one of the two diagonals of K is a line of symmetry. This means that a rhombus does not count as a kite.

Since outer billiards is an affinely natural system, we find it useful to normalize kites in a particular way. Any kite is affinely equivalent to the quadrilateral $K(A)$ having vertices

$$(-1,0), \qquad (0,1), \qquad (0,-1), \qquad (A,0), \qquad A \in (0,1). \tag{1.1}$$

Figure 1.1 shows an example. The omitted case $A = 1$ corresponds to rhombuses. Henceforth, when we say *kite*, we mean $K(A)$ for some A. The kite $K(A)$ is (ir)rational iff A is (ir)rational.

Let $\mathbf{Z}_{\text{odd}}$ denote the set of odd integers. Reflection in each vertex of $K(A)$ preserves $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$. Hence outer billiards on $K(A)$ preserves $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$. We call an outer billiards orbit on $K(A)$ *special* if (and only if) it is contained in $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$. We discuss only special orbits in this book. The special orbits are hard enough for us already. In the appendix, we will say something about the general case. See §A.3.

We call an orbit *forward erratic* if the forward orbit is unbounded and also returns to every neighborhood of a kite vertex. We state the same definition for the backward direction. We call an orbit *erratic* if it is both forward and backward erratic. In Parts 1–4 of the book we will prove the following result.

Theorem 1.1 (Erratic Orbits) *The following hold for any irrational kite.*

1. *There are uncountably many erratic special orbits.*

2. *Every special orbit is either periodic or unbounded in both directions.*

3. *The set of periodic special orbits is open dense in* $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$.

It follows from the work on quasirational polygons cited above that all orbits are periodic relative to a rational kite. (The analysis in this book gives another proof of this fact, at least for special orbits. See the remark at the end of §3.2.) Hence the Erratic Orbits Theorem has the following corollary.

Corollary 1.2 *Outer billiards on a kite has an unbounded orbit if and only if the kite is irrational.*

The Erratic Orbits Theorem is an intermediate result included so that the reader can learn a substantial theorem without having to read the whole book. We will describe our main result in the next two sections.

1.3 COROLLARIES OF THE COMET THEOREM

In Parts 5 and 6 of the book we will go deeper into the subject and establish our main result, the Comet Theorem. The Comet Theorem and its corollaries considerably sharpen the Erratic Orbits Theorem. We defer statement of the Comet Theorem until the next section. In this section, we describe some of its corollaries.

Given a Cantor set C contained in a line L, we let $C^{\#}$ be the set obtained from C by deleting the endpoints of the components of $L - C$. We call $C^{\#}$ a *trimmed Cantor set*. Note that $C - C^{\#}$ is countable.

The interval

$$I = [0, 2] \times \{-1\} \tag{1.2}$$

turns out to be a very useful interval. Figure 1.2 shows I and its first 3 iterates under the outer billiards map.

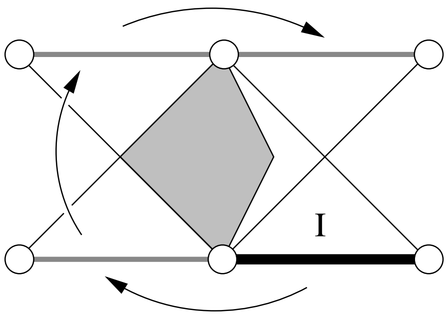

Figure 1.2: I and its first 3 iterates.

Let U_A denote the set of unbounded special orbits relative to A.

Theorem 1.3 *Relative to any irrational $A \in (0, 1)$, the following are true.*

1. *U_A is minimal: Every orbit in U_A is dense in U_A and all but at most 2 orbits in U_A are both forward dense and backward dense in U_A.*
2. *U_A is locally homogeneous: Every two points in U_A have arbitrarily small neighborhoods that are isometric to each other.*
3. *$U_A \cap I = C_A^{\#}$ for some Cantor set C_A.*

Remarks:
(i) One endpoint of C_A is the kite vertex $(0, -1)$. Hence Statement 1 implies that all but at most 2 unbounded special orbits are erratic. The remaining special orbits, if any, are each erratic in one direction.
(ii) Statements 2 and 3 combine to say that every point in U_A lies in an interval that intersects U_A in a trimmed Cantor set. This gives us a good local picture of U_A. One thing we are missing is a good global picture of U_A.
(iii) The Comet Theorem describes C_A explicitly.

Given Theorem 1.3, it makes good sense to speak of the first return map to any interval in $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$. From the minimality result, the local nature of the return map is essentially the same around any point of U_A. To give a crisp picture of this first return map, we consider the interval I discussed above.

For $j = 1, 2$, let $f_j \colon X_j \to X_j$ be a map such that f_j and f_j^{-1} are defined on all but perhaps a finite subset of X_j. We call f_1 and f_2 *essentially conjugate* if there are countable sets $C_j \subset X_j$, each one contained in a finite union of orbits, and a homeomorphism

$$h \colon X_1 - C_1 \to X_2 - C_2$$

that conjugates f_1 to f_2.

An *odometer* is the map $x \to x + 1$ on the inverse limit of the system

$$\cdots \to \mathbf{Z}/D_3 \to \mathbf{Z}/D_2 \to \mathbf{Z}/D_1, \qquad D_k | D_{k+1} \qquad \forall k. \tag{1.3}$$

The *universal odometer* is the map $x \to x+1$ on the *profinite completion* of $\mathbf{Z}$. This is the inverse limit taken over the system of all finite cyclic groups. For concreteness, Equation 1.3 defines the universal odometer when $D_k = k$ factorial. See [**H**] for a detailed discussion of the universal odometer.

Theorem 1.4 *Let ρ_A be the first return map to $U_A \cap I$.*

1. *For any irrational $A \in (0, 1)$, the map ρ_A is defined on all but at most one point and is essentially conjugate to an odometer $\mathcal{Z}_A$.*

2. *Any given odometer is essentially conjugate to ρ_A for uncountably many difference choices of A.*

3. *ρ_A is essentially conjugate to the universal odometer for almost all A.*

Remarks:

(i) The Comet Theorem explicitly describes $\mathcal{Z}_A$ in terms of a sequence we call the *remormalization sequence*. This sequence is related to the continued fraction expansion of A. We will give a description of this sequence in the next section.

(ii) Theorem 1.4 is part of a larger result. There is a certain suspension flow over the odometer, which we call *geodesic flow on the cusped solenoid*. It turns out that the time-one map for this flow serves as a good model, in a certain sense, for the dynamics on U_A. §24.3.

Our next result highlights an unexpected connection between outer billiards on kites and the modular group $SL_2(\mathbf{Z})$. The group $SL_2(\mathbf{Z})$ acts naturally on the upper half-plane model of the hyperbolic plane, $\mathbf{H}^2$, by linear fractional transformations. Closely related to $SL_2(\mathbf{Z})$ is the $(2, \infty, \infty)$-triangle group Γ generated by reflections in the sides of the geodesic triangle with vertices $(0, 1, i)$. The points 0 and 1 are the *cusps*, and the point i is the internal vertex corresponding to the right angle of the triangle. See §25.2 for more details. Γ and $SL_2(\mathbf{Z})$ are commensurable: Their intersection has finite index in both groups. In our next result, we interpret our kite parameter interval $(0, 1)$ as the subset of the ideal boundary of $\mathbf{H}^2$.

Theorem 1.5 *Let $S = [0, 1] - \mathbf{Q}$. Let $u(A)$ be the Hausdorff dimension of U_A.*

1. *For all $A \in S$, the set U_A has length 0. Hence almost all points in $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ have periodic orbits relative to outer billiards on $K(A)$.*

2. *If $A, A' \in S$ are in the same Γ-orbit, then U_A and $U_{A'}$ are locally similar. In particular, $u(A) = u(A')$.*

3. *If $A \in S$ is quadratic irrational, then every point of U_A lies in an interval that intersects U_A in a self-similar trimmed Cantor set.*

4. *The function u is almost everywhere equal to some constant u_0 and yet maps every open subset of S onto $[0, 1]$.*

Remarks:
(i) We do not know the value of u_0. We guess that $0 < u_0 < 1$. Theorem 25.9 gives a formula for $u(A)$ in many cases.
(ii) The word *similar* in statement 2 means that the two sets have neighborhoods that are related by a similarity. In statement 3, a *self-similar* set is a disjoint finite union of similar copies of itself.
(iii) We will see that statement 2 essentially implies both statements 3 and 4. Statement 2 is the first hint that outer billiards on kites is connected to the modular group. The Comet Theorem says more about this.
(iv) Statement 3 of Theorem 1.4 combines with statement 4 of Theorem 1.5 to say that there is a "typical behavior" for outer billiards on kites, in a certain sense. For almost every parameter A, the dimension of U_A is the (unknown) constant u_0 and the return map ρ_A is essentially conjugate to the universal odometer.

We end this section by comparing our results here with the main theorems in [**S**] concerning the Penrose kite. The Penrose kite parameter is

$$A = \sqrt{5} - 2 = \phi^{-3},$$

where ϕ is the golden ratio. In [**S**], we prove[2] that $C_A^{\#} \subset U_A$ and that the first return map to $C_A^{\#}$ is essentially conjugate to the 2-adic odometer. Theorems 1.3 and 1.4 subsume these results about the Penrose kite.

As in §25.5.2, we might have computed in [**S**] that $\dim(C_A) = \log(2)/\log(\phi^3)$. However, at the time we did not know how this number was related to $\dim(U_A)$, the real quantity of interest to us. From Theorem 1.3, we know additionally that $C_A^{\#} = U_A \cap I$ and $\dim(U_A) = \dim(C_A)$.

While we recover and improve all the main *theorems* in [**S**], there is one way that the work we do in [**S**] for the Penrose kite goes deeper than what we do here (for every irrational kite). The work in [**S**] establishes a deeper kind of self-similarity for the Penrose kite orbits than we have established in statement 3 of Theorem 1.5. See §A.2 for a discussion.

[2]Technically, we prove these results for a smaller Cantor set which is the left half of C_A. However, the arguments using C_A in place of its left half would be just about the same.

1.4 THE COMET THEOREM

Now we describe our main result. Say that p/q is *odd* or *even* according to whether pq is odd or even. There is a unique sequence $\{p_n/q_n\}$ of distinct odd rationals, converging to A, such that

$$\frac{p_0}{q_0} = \frac{1}{1}, \qquad |p_n q_{n+1} - q_n p_{n+1}| = 2, \qquad \forall n. \tag{1.4}$$

We call this sequence the *inferior sequence*. See §4.1. This sequence is closely related to continued fractions.

We define

$$d_n = \text{floor}\left(\frac{q_{n+1}}{2q_n}\right), \qquad n = 0, 1, 2, ... \tag{1.5}$$

Say that a *superior term* is a term p_n/q_n such that $d_n \geq 1$. We will show that there are infinitely many superior terms. Say that the *superior sequence* is the subsequence of superior terms. Say that the *renormalization sequence* is the corresponding subsequence of $\{d_n\}$. We reindex so that the superior and renormalization sequences are indexed by 0, 1, 2, ...

Example: To fix ideas, we demonstrate how this works for the Penrose kite parameter. $A = \phi^{-3}$. The inferior sequence for A is

$$\mathbf{\frac{1}{1}} \quad \frac{1}{3} \quad \mathbf{\frac{1}{5}} \quad \frac{3}{13} \quad \mathbf{\frac{5}{21}} \quad \frac{13}{55} \quad \mathbf{\frac{21}{89}} \quad \frac{55}{233} \quad \mathbf{\frac{89}{377}} \quad \cdots.$$

The bold terms are the terms of the superior sequence. The superior sequence obeys the recurrence relation $r_{n+2} = 4r_{n+1} + r_n$, where r stands for either p or q. The initial sequence $\{d_n\}$ is 1, 0, 1, 0, ... The renormalization sequence is 1, 1, 1, ...

The definitions that follow work entirely with the superior sequence. We define $\mathcal{Z}_A$ to be the inverse limit of the system

$$\ldots \to \mathbf{Z}/D_3 \to \mathbf{Z}/D_2 \to \mathbf{Z}/D_1, \qquad D_n = \prod_{i=0}^{n-1}(d_i + 1). \tag{1.6}$$

We equip $\mathcal{Z}_A$ with a metric, defining $d_A(x, y) = q_{n-1}^{-1}$, where n is the smallest index such that $[x]$ and $[y]$ disagree in $\mathbf{Z}/D_n$. In the Penrose kite example above, $\mathcal{Z}_A$ is naturally the 2-adic integers and d_A gives the same topology as the classical 2-adic metric.

We can identify the points of $\mathcal{Z}_A$ with the sequence space

$$\Pi_A = \prod_{i=0}^{\infty}\{0, ..., d_i\}. \tag{1.7}$$

The identification works like this.

$$\phi_1\colon \sum_{j=0}^{\infty} \tilde{k}_j D_j \in \mathcal{Z}_A \qquad \longrightarrow \qquad \{k_j\} \in \Pi_A. \tag{1.8}$$

The elements on the left hand side are formal series, and

$$\tilde{k}_j = \begin{cases} k_j & \text{if } p_j/q_j < A. \\ \\ d_j - k_j & \text{if } p_j/q_j > A. \end{cases} \tag{1.9}$$

Our identification is nonstandard in that it uses $\tilde{k}_j$ in place of the more obvious choice of k_j. Needless to say, we make this less-than-obvious choice because it reflects the structure of outer billiards.

There is a map $\phi_2\colon \Pi_A \to \mathbf{R} \times \{-1\}$, defined as follows.

$$\phi_2\colon\ \{k_j\} \quad \longrightarrow \quad \left(\sum_{j=0}^{\infty} 2k_j\lambda_j, -1\right), \qquad \lambda_j = |Aq_j - p_j|. \tag{1.10}$$

We define $C_A = \phi_2(\Pi_A)$. Equivalently,

$$C_A = \phi(\mathcal{Z}_A), \qquad \phi = \phi_2 \circ \phi_1. \tag{1.11}$$

(The map ϕ depends on A, but we suppress this from our notation.) It turns out that $\phi\colon \mathcal{Z}_A \to C_A$ is a homeomorphism and C_A is a Cantor set whose convex hull is exactly I, the interval discussed in the previous section. Let $C_A^{\#}$ denote the trimmed Cantor set based on C_A.

Define

$$\mathbf{Z}[A] = \{mA + n|\ m, n \in \mathbf{Z}\}. \tag{1.12}$$

Say that the *excursion distance* of a portion of an outer billiards orbit is the maximum distance from a point on this orbit portion to the origin.

Theorem 1.6 (Comet) *Let U_A denote the set of unbounded special orbits relative to an irrational $A \in (0, 1)$.*

1. *For any N, there is an N' with the following property. If $\zeta \in U_A$ satisfies $\|\zeta\| < N$, then the kth outer billiards iterate of ζ lies in I for some $|k| < N'$. Here N' depends only on N and A.*

2. *$U_A \cap I = C_A^{\#}$. The first return map $\rho_A\colon C_A^{\#} \to C_A^{\#}$ is defined precisely on $C_A^{\#} - \phi(-1)$. The map $\phi^{-1} \circ \rho_A \circ \phi$, wherever defined on $\mathcal{Z}_A$, equals the odometer.*

3. *For any $\zeta \in C_A^{\#} - \phi(-1)$, the orbit portion between ζ and $\rho_A(\zeta)$ has excursion distance in $\left[c_1^{-1}d^{-1}, c_1 d^{-1}\right]$ and length in $\left[c_2^{-1}d^{-2}, c_2 d^{-3}\right]$. Here c_1, c_2 are universal positive constants and $d = d_A\big(-1, \phi^{-1}(\zeta)\big)$.*

4. *$C_A^{\#} = C_A - (2\mathbf{Z}[A] \times \{-1\})$. Two points in U_A lie on the same orbit if and only if the difference between their first coordinates lies in $2\mathbf{Z}[A]$.*

Remarks:

(i) To use a celestial analogy, the unbounded special orbits are comets and I is the visible sky. Item 1 says roughly that any comet is always either approaching I or leaving I. Item 2 describes the geometry and combinatorics of the visits to I. Item 3 gives a model of the behavior between visits. Item 4 gives an algebraic view.

(ii) Lemma 23.7 replaces the bounds in item 3 with explicit estimates. The orders on all the bounds in item 3 are sharp except perhaps for the length upper bound. See the remarks following Lemma 23.7 for a discussion, and also §A.2.
(iii) The Comet Theorem has an analog for the backward orbits. The statement is the same except that the point $\phi(0)$ replaces the point $\phi(-1)$ and the map $x \to x-1$ replaces the odometer. We have the general identity $\phi(0) + \phi(-1) = (2, -2)$.
(iv) Our analysis will show that $\phi(0)$ and $\phi(-1)$ have well defined orbits iff they lie in $C_A^\#$. It turns out that this happens iff the superior sequence for A is not eventually monotone. The Comet Theorem implies that the forward orbit of $\phi(-1)$ and the backward orbit of $\phi(0)$, when defined, accumulate only at ∞. We think of $\phi(-1)$ as the "cosmic ejector." When a comet comes close to this point, it is ejected way out into space. Similarly, we think of $\phi(0)$ as the "cosmic attractor".
(v) Statement 3 of Theorem 1.5 is a hint that the sets C_A have a beautiful structure. Here is a structural result outside the scope of this book. Letting C'_A denote the scaled-in-half version of C_A that lives in the unit interval, it seems that

$$C = \bigcup_{A \in [0,1]} \Big(C'_A \times \{A\} \Big) \subset [0,1]^2 \subset \mathbf{RP}^2 \tag{1.13}$$

is the limit set of a semigroup $S \subset SL_3(\mathbf{Z})$ that acts by projective transformations. (C_A can be defined even for rational A.) The group closure of S has finite index in a maximal cusp of $SL_3(\mathbf{Z})$. Figure 1.3 shows a plot of C.

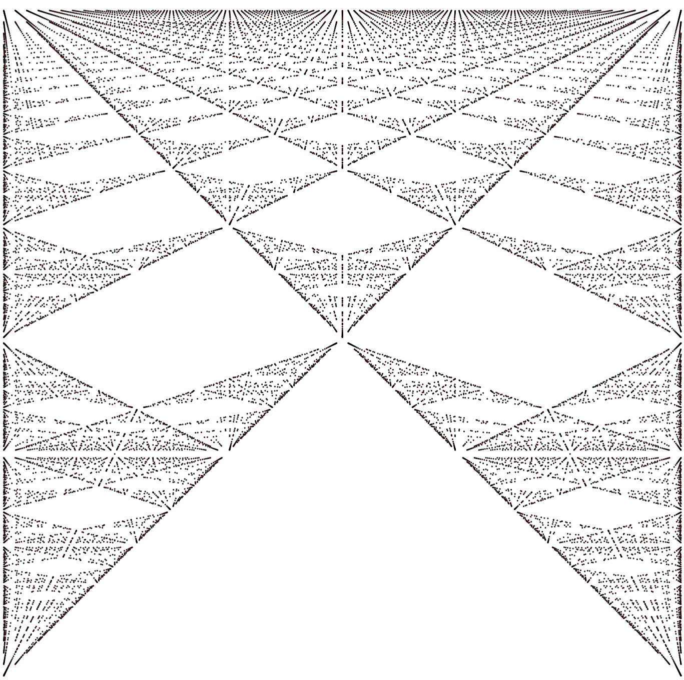

Figure 1.3: The set C. The bottom is $A = 0$ and the top is $A = 1$.

1.5 RATIONAL KITES

Like most authors who have considered outer billiards, we find it convenient to work with the square of the outer billiards map. Let $O_2(x)$ denote the square outer billiards orbit of x. Let $I = [0,2] \times \{-1\}$, as above, and let

$$\Xi = \mathbf{R}_+ \times \{-1, 1\}. \tag{1.14}$$

When $\epsilon \in (0, 2/q)$, the orbit $O_2(\epsilon, -1)$ has a combinatorial structure independent of ϵ. See Lemma 2.2. Thus $O_2(1/q, -1)$ is a natural representative of this orbit. We often call this orbit the *fundamental orbit*. The fundamental orbit plays a crucial role in our proofs. The following result is a basic mechanism for producing unbounded orbits.

Theorem 1.7 *Relative to p/q, the set $O_2(1/q, -1) \cap \Xi$ has diameter between $\lambda(p+q)/2$ and $\lambda(p+q)+2$. Here $\lambda = 1$ if p/q is odd and $\lambda = 2$ if p/q is even.*

Any odd rational p/q appears as (say) the nth term in a superior sequence $\{p_i/q_i\}$. The terms before p/q are uniquely determined by p/q. This is similar to what happens for continued fractions. Define Π_n to be the product of the first n factors of Π_A, the space from Equation 1.7.

Theorem 1.8 *Let $\mu_i = |p_n q_i - q_n p_i|$.*

$$O_2\Big(\frac{1}{q_n}, -1\Big) \cap I = \bigcup_{\kappa \in \Pi_n} \Big(X_n(\kappa), -1\Big), \qquad X_n(\kappa) = \frac{1}{q_n}\Big(1 + \sum_{i=0}^{n-1} 2k_i \mu_i\Big).$$

Example: Here we show Theorem 1.8 in action. The odd rational $19/49$ determines the inferior sequence

$$\frac{p_0}{q_0} = \frac{1}{1}, \frac{1}{3}, \frac{5}{13}, \frac{19}{49} = \frac{p_3}{q_3}.$$

All terms are superior, so this is also the superior sequence. In our example,

- $n = 3$.
- The superior sequence is $1, 2, 1$.
- The μ sequence is $30, 8, 2$.

Therefore the first coordinates of the 12 points of $O_2(1/49) \cap I$ are given by

$$\bigcup_{k_0=0}^{1} \bigcup_{k_1=0}^{2} \bigcup_{k_2=0}^{1} \frac{2(30k_0 + 8k_1 + 2k_2) + 1}{49}.$$

Writing these numbers in a suggestive way, we see that the union above works out to

$$\frac{1}{49} \times \big(1 \quad 5 \qquad 17 \quad 21 \qquad 33 \quad 37 \qquad\qquad\qquad 61 \quad 65 \qquad 77 \quad 81 \qquad 93 \quad 97\big).$$

Remarks:
(i) Theorem 1.8 is a good example of a result that is easy to check on a computer. One can check the result for the example we give, or for any other smallish parameter, using Billiard King.
(ii) A version of Theorem 1.8 holds in the even case as well. We will discuss the even case of Theorem 1.8 in §22.7.
(iii) We view statements 2 and 3 as the heart of the Comet Theorem. We will prove these two statements by combining Theorems 1.7 and 1.8 and then taking a geometric limit. The proofs for statements 1 and 4 of the Comet Theorem require some other ideas that we cannot describe without a buildup of machinery.
(iv) Theorem 1.8 has a nice conjectural extension, which describes the entire return map to I. See §A.1. A suitable geometric limit of the conjecture in §A.1 describes the structure of the orbits in $I - C_A^{\#}$ in the case when A is irrational. See Conjecture A.1.

We mention two more results about outer billiards on rational kites. These results do not play such an important role in our proof of the Comet Theorem, but they are appealing and fairly easy by-products of our analysis.

Here is an amplification of the upper bound in Theorem 1.7.

Theorem 1.9 *If p/q is odd, let $\lambda = 1$. If p/q is even, let $\lambda = 2$. Each special orbit intersects Ξ in exactly one set of the form $I_k \times \{-1, 1\}$, where*

$$I_k = (\lambda k(p+q), \lambda(k+1)(p+q)), \qquad k = 0, 1, 2, 3, \ldots.$$

Hence any special orbit intersects Ξ in a set of diameter at most $\lambda \cdot (p+q) + 2$.

Theorem 1.9 is similar in spirit to a result in [**K**]. See §3.4 for a discussion.

We call an outer billiards orbit on $K(A)$ *persistent*[3] if there are nearby and combinatorially identical orbits on $K(A')$ for all A' sufficiently close to A. Otherwise, we call the orbit *fleeting*. In the odd case, $O_2(1/q, \pm 1)$ is fleeting.

Theorem 1.10 *In the even rational case, all special orbits are persistent. In the odd case, the set $I_k \times \{-1, 1\}$ contains exactly two fleeting orbits, U_k^+ and U_k^-, and these are conjugate by reflection in the x-axis. In particular, we have $U_0^{\pm} = O_2(1/q, \pm 1)$.*

Remark: None of our structure theorems holds, as stated, for general quadrilaterals or even for nonspecial orbits on kites. We do not really have a good understanding of the structure of outer billiards on a general rational quadrilateral, though we can see that it promises to be quite interesting. We take up this discussion in §A.4.

[3] It would be more usual to call such orbits *stable*, but in the subject of outer billiards, the word *stable* has historically meant the same as the word *bounded*.

1.6 THE ARITHMETIC GRAPH

Here we describe the *arithmetic graph*, a central construction in the book. One should think of the first return map to $\Xi = \mathbf{R}_+ \times \{-1, 1\}$, for rational parameters, as an essentially combinatorial object. The arithmetic graph gives a 2-dimensional representation of this combinatorial object. The principle guiding our construction is that sometimes it is better to understand the Abelian group $\mathbf{Z}[A]$ as a module over $\mathbf{Z}$ rather than as a subset of $\mathbf{R}$. Our arithmetic graph is similar in spirit to the lattice vector fields studied by Vivaldi et al. in connection with interval exchange transformations. See, e.g., [**VL**].

Here we explain the idea roughly. See §2.4 for precise details. The arithmetic graph is most easily explained in the rational case. Let ψ be the square of the outer billiards map. It turns out that every orbit starting on Ξ eventually returns to Ξ. See Lemma 2.3. Thus we can define the first return map

$$\Psi\colon \Xi \to \Xi. \tag{1.15}$$

We define the map $T\colon \mathbf{Z}^2 \to 2\mathbf{Z}[A] \times \{-1, 1\}$ by the formula

$$T(m,n) = \Big(2Am + 2n + 1/q, (-1)^{p+q+1}\Big). \tag{1.16}$$

Here $A = p/q$.

Up to the reversal of the direction of the dynamics, every point of Ξ has the same orbit as a point of the form $T(m,n)$, where $(m,n) \in \mathbf{Z}^2$. For instance, the orbit of $T(0,0) = (1/q, -1)$ is what we called the fundamental orbit above. We form the graph $\widehat{\Gamma}(p/q)$ by joining the points (m_1, m_2) to (m_2, n_2) when these points are sufficiently close together and also $T(m_1, n_1) = \Psi^{\pm 1}(m_2, n_2)$. (The map T is not injective, so we have choices to make. That is the purpose of the *sufficiently close* condition.)

We let $\Gamma(p/q)$ denote the component of $\widehat{\Gamma}(p/q)$ that contains $(0,0)$. This component tracks the orbit $O_2(1/q, -1)$, the main orbit of interest to us. When p/q is odd, $\Gamma(p/q)$ is an infinite periodic polygonal arc, invariant under translation by the vector $(q, -p)$. Note that $T(q, -p) = T(0,0)$. When p/q is even, $\Gamma(p/q)$ is an embedded polygon. We prove many structural theorems about the arithmetic graph. Here we informally mention three central ones.

- **The Embedding Theorem** (Chapter 2): $\widehat{\Gamma}(p/q)$ is a disjoint union of embedded polygons and infinite embedded polygonal arcs. Every edge of $\widehat{\Gamma}(p/q)$ has length at most $\sqrt{2}$. The persistent orbits correspond to closed polygons, and the fleeting orbits correspond to infinite (but periodic) polygonal arcs.

- **The Hexagrid Theorem** (Chapter 3): The structure of $\widehat{\Gamma}(p/q)$ is controlled by 6 infinite families of parallel lines. See Figure 3.3. The *quasiperiodic* structure is similar to what one sees in DeBruijn's famous pentagrid construction of the Penrose tilings. See [**DeB**].

- **The Copy Theorem** (Chapter 18; also Lemmas 4.2 and 4.3): If A_1 and A_2 are two rationals that are close in the sense of Diophantine approximation, then the corresponding arithmetic graphs Γ_1 and Γ_2 have substantial agreement.

The Hexagrid Theorem causes $\Gamma(p/q)$ to have an oscillation (relative to the line of slope $-p/q$ through the origin) on the order of $p+q$. The Hexagrid Theorem is responsible for Theorems 1.7, 1.9, and 1.10. Referring to the superior sequence, the Copy Theorem guarantees that one period of $\Gamma(p_n/q_n)$ is copied by $\Gamma(p_{n+1}/q_{n+1})$. If we combine the Copy Theorem and the Hexgrid Theorem, we get Theorem 1.8. The Hexagrid Theorem and the Copy Theorem work as a team, with one result forcing large oscillations and the other result organizing these oscillations in a coherent way for the family of arithmetic graphs corresponding to the superior sequence.

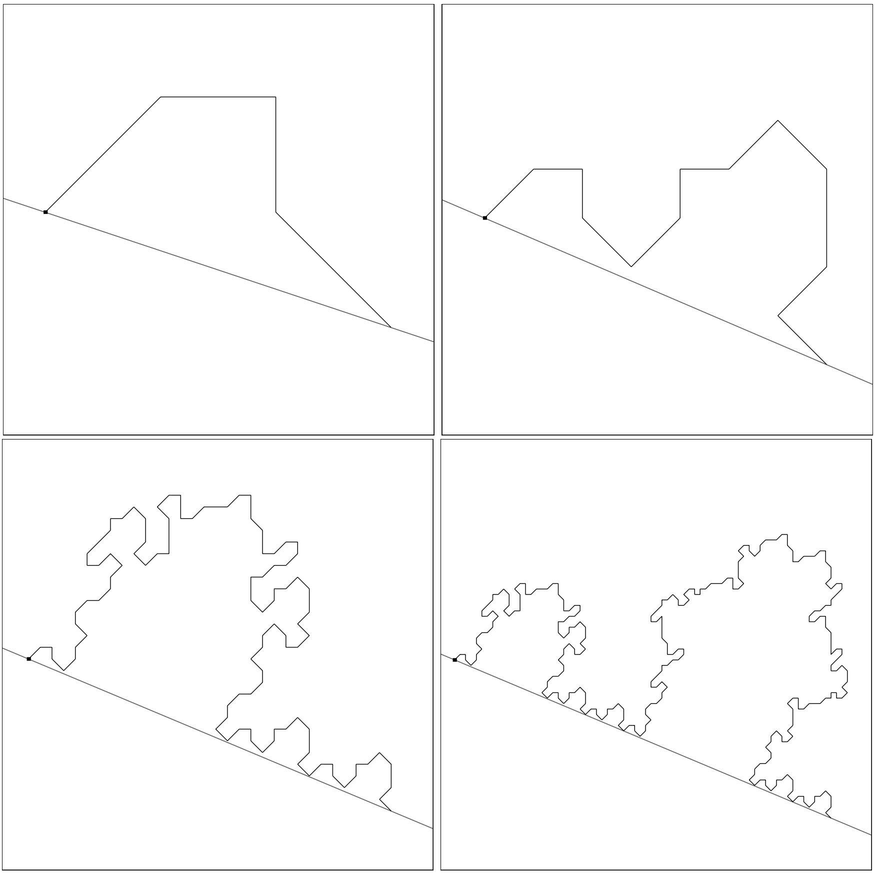

Figure 1.4: The graphs $\Gamma(1/3)$, $\Gamma(3/7)$, $\Gamma(13/31)$, $\Gamma(29/69)$.

We illustrate these ideas in Figure 1.4, where each frame shows one period of $\Gamma(p/q)$ in reference to the line of slope $-p/q$ through the origin. Here p/q depends on the box. We choose 4 consecutive terms in a superior sequence. Each graph copies at least one period of the previous one, creating the beginnings of a large-scale fractal structure.

When p/q is an even rational, $\Gamma(p/q)$ is a closed embedded polygon. A related kind of period-copying phenomenon happens in the case of even rationals. We consider arithmetic graphs associated to chains of rationals ..., p'/q', p/q, ... such that $|pq' - qp'| = 1$ for consecutive pairs. Figure 1.5 shows the 4 solid polygons bounded by the corresponding arithmetic graphs corresponding to 4 consecutive terms in such a chain of even rationals.

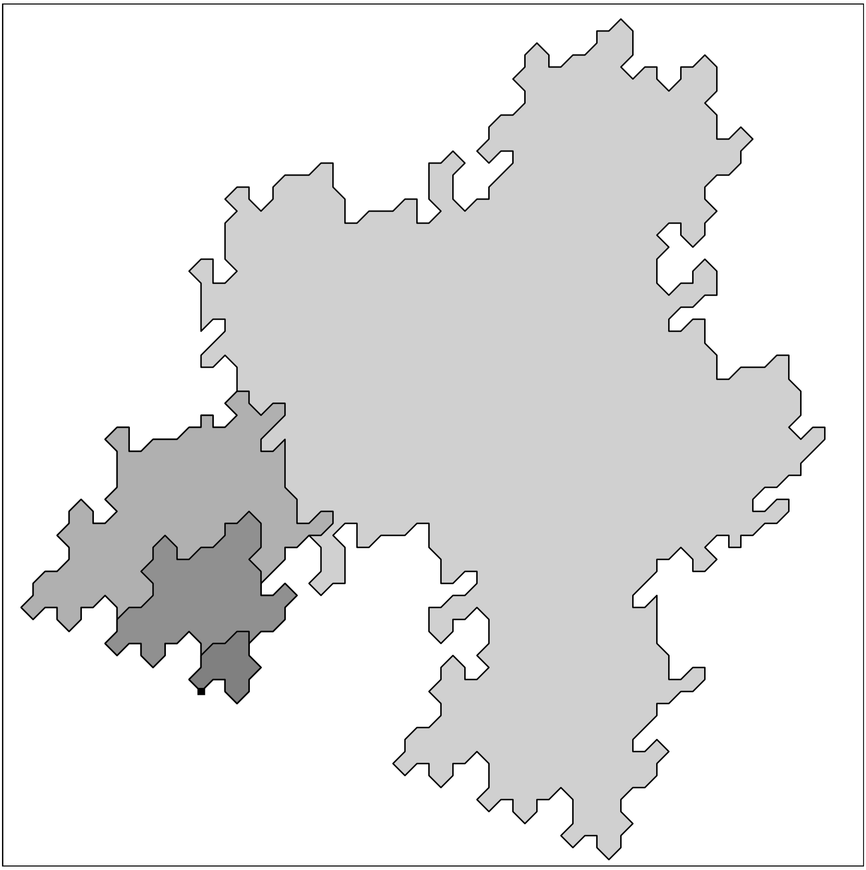

Figure 1.5: The filled-in graphs $\Gamma(2/5)$, $\Gamma(5/12)$, $\Gamma(8/19)$, $\Gamma(21/50)$.

The polygons are nested. This always seems to occur for such chains of rationals, though we do not actually know a proof. Fortunately, our actual proofs do not rely on this nesting phenomenon. Billiard King has a feature that draws figures like this automatically once the final term in the chain of rationals is supplied.

One final remark: The reader should compare the undersides of the polygons in Figure 1.5 with the graphs in Figure 1.4. The fact that the two figures so closely resemble each other is not an accident. It has to do with our careful choice of rationals. Part 6 of the book explores relationships like this.

1.7 THE MASTER PICTURE THEOREM

The logic of the book works like this. After we define the arithmetic graph, we prove a number of structural results about it. We then deduce the Comet Theorem and its corollaries from these structural results. The way we understand the arithmetic graph is to obtain a kind of closed-form expression for it. The Master Picture Theorem gives this expression. Here we will give a rough description of this result. We formulate and prove the Master Picture Theorem in Part 2 of the book.

Let us first discuss the Master Picture Theorem in vague terms. It sometimes happens that one has a dynamical system on a high-dimensional manifold M together with an embedding of a lower-dimensional manifold X into M that is, in some sense, compatible with the dynamics on M. The dynamics on M then induces a dynamical system on X. Sometimes the higher-dimensional system on M is much simpler than the system on X, and most of the complexity of the system on X comes from its complicated embedding into M. The Master Picture Theorem says that this situation happens for outer billiards on kites.

Now we will say something more precise. Recall that $\Xi = \mathbf{R}_+ \times \{-1, 1\}$. The arithmetic graph encodes the dynamics of the first return map $\Psi \colon \Xi \to \Xi$. It turns out that Ψ is an infinite interval exchange map. The Master Picture Theorem reveals the following structure for each parameter A.

1. There is a locally affine map μ from Ξ into a union $\widehat{\Xi}$ of two 3-dimensional tori.

2. There is a polyhedron exchange map $\widehat{\Psi} \colon \widehat{\Xi} \to \widehat{\Xi}$ defined relative to a partition of $\widehat{\Xi}$ into 28 polyhedra.

3. The map μ is a semiconjugacy between Ψ and $\widehat{\Psi}$.

In other words, the return dynamics of $\widehat{\Psi}$ has a kind of compactification into a 3 dimensional polyhedron exchange map. All the objects above depend on the parameter A, but we have suppressed them from our notation.

There is one master picture, a union of two 4-dimensional convex lattice polytopes partitioned into 28 smaller convex lattice polytopes, that controls everything. For each parameter, one obtains the 3-dimensional picture by taking a suitable slice.

The fact that nearby slices give almost the same picture is the source of the Copy Theorem. The interaction between the map μ and the walls of our convex polytope partitions is the source of the Hexagrid Theorem. The Embedding Theorem follows from basic geometric properties of the polytope exchange map in an elementary way that is hard to summarize here.

My investigation of the Master Picture Theorem is really just starting, and this book has only the beginnings of a theory. First, I believe that a version of the Master Picture Theorem should hold much more generally. (This is something that John Smillie and I hope to work out together.) Second, some recent experiments convince me that there is a renormalization theory for this object grounded in real projective geometry. All of this will perhaps be the subject of a future work.

1.8 REMARKS ON COMPUTATION

As I mentioned in the preface, I discovered most of the phenomena discussed in this book using my program Billiard King. Billiard King and this book developed side by side in a kind of feedback loop. Since I am ultimately trying to verify phenomena that I discovered with the aid of a computer, one might expect some computational aspects to the formal proofs. The overall proof here uses considerably less computation than the proof in [**S**], but I still use a computer-aided proof in several places.

Mainly, I use a computer to check that various 4-dimensional convex integral polytopes have disjoint interiors. This involves a small amount of linear algebra, using exact integers, that one could in principle do by hand. One could do these calculations by hand in the same way that one could count all the coins filling up a bathtub. One could do it, but it is better left to a machine. Most of these computations come from Part 3 of the book.

The experimental method I used has the advantage that I checked essentially all the results with extensive and visually surveyable computation. The interested reader can make many of the same visual checks by downloading the program and playing with it. I suppose I cannot guarantee Billiard King does not have a subtle bug, but the output from the program makes sense in a way that would be unlikely in the presence of a serious problem. Also, the output of Billiard King matches the results I have proved in a traditional way in this book.

1.9 ORGANIZATION OF THE BOOK

The book has 6 parts. Parts 1–4 comprise the core of the book. In Part 1, we prove the Erratic Orbits Theorem modulo some auxilliary results such as the Hexagrid Theorem. In Part 2, we prove the Master Picture Theorem, our main structural result. in Parts 3 and 4, we use the Master Picture Theorem to prove the various auxilliary results assumed in Part 1.

In Part 5, we prove the Comet Theorem and its corollaries modulo various auxilliary results. In Part 6, we prove these auxilliary results.

In the Appendix, we discuss some additional phenomena, both for kites and for quadrilaterals, that we have observed but not proved.

Before each part of the book, we include an overview of that part.

Part 1. The Erratic Orbits Theorem

In this part of the book, we will prove the Erratic Orbits Theorem modulo a number of auxilliary results that we prove in Parts 2–4.

- In Chapter 2, we establish some basic results that allow for definition of the arithmetic graph. The arithmetic graph is our main object of study. We also state the Embedding Theorem, a basic structural result about the arithmetic graph that we prove in Part 3.
- In Chapter 3, we state the Hexagrid Theorem, another structural result about the arithmetic graph. We then deduce Theorems 1.7, 1.9, and 1.10 from the Hexagrid Theorem. We prove the Hexagrid Theorem in Part 3.
- In Chapter 4, we discuss the period-copying results needed to prove the Erratic Orbits Theorem. Along the way, we introduce the inferior and superior sequences, two basic ingredients in our overall theory. We prove the period-copying results in Part 4.
- In Chapter 5, we assemble the ingredients from previous chapters and prove the Erratic Orbits Theorem. We note that the arguments we use in Parts 5 and 6 to prove the Comet Theorem are independent of Chapter 5. Thus, for the reader who plans to work through the proof of the Comet Theorem, the material in Chapter 5 is redundant.

We mention several conventions that we use repeatedly throughout the book. Recall that p/q is an odd rational if pq is odd. When we say *odd rational*, we mean that the odd rational lies in (0, 1). On very rare occasions, we also consider the odd rational $1/1$. However, we never consider negative odd rationals, or odd rationals greater than 1. Also, A always stands for a kite parameter, and we write $A = p/q$. Similarly, A_n stands for p_n/q_n, and A_+ stand for p_+/q_+, etc. Sometimes we will fail to mention these conventions explicitly.

We imagine that certain readers will be interested mainly in statement 1 of the Erratic Orbits Theorem – i.e., the existence of unbounded orbits. For such readers, we sometimes add remarks indicating sections that are not necessary for this part of the proof.

This page intentionally left blank

Chapter Two

The Arithmetic Graph

2.1 POLYGONAL OUTER BILLIARDS

Let P be a convex polygon. We denote the outer billiards map relative to P by ψ', and the square of the outer billiards map by $\psi = (\psi')^2$. Our convention is that a person walking from p to $\psi'(p)$ sees the P on the right side. These maps are defined away from a countable set of line segments in $\mathbf{R}^2 - P$. This countable set of line segments is sometimes called the *limit set*.

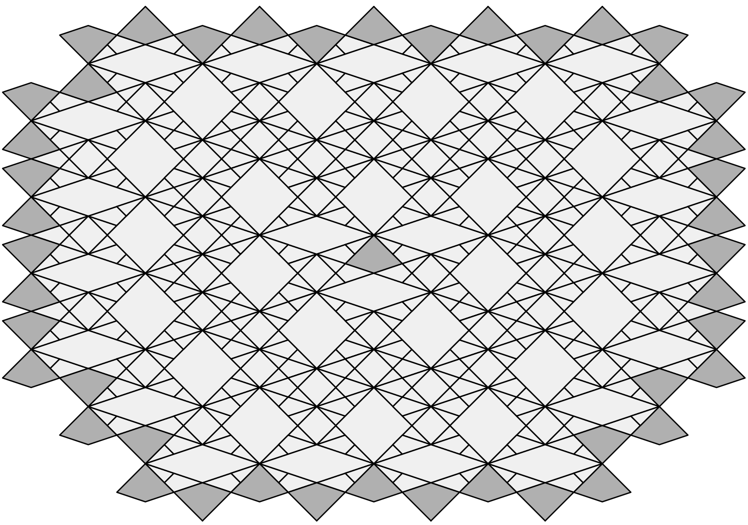

Figure 2.1: Part of the tiling for $K(1/3)$.

The result in [**VS**], [**K**], and [**GS**] states, in particular, that the orbits for rational polygons are all periodic. In this case, the complement of the limit set is tiled by dynamically invariant convex polygons. Figure 2.1 shows part of the tiling for the kite $K(1/3)$.

This is the simplest tiling[1] we see among all the kites. We have drawn only part of the tiling. The reader can draw more of these figures, and in color, using Billiard King. The existence of these tilings is what motivated me to study outer billiards. I wanted to understand how the tiling changes with the rational parameter and saw that the kites give rise to highly nontrivial figures.

[1]Note that the picture is rotated 90 degrees from the usual normalization.

2.2 SPECIAL ORBITS

Until the last result in this section, the parameter $A = p/q$ is rational. Say that a *special interval* is an open horizontal interval of length $2/q$ centered at a point of the form $(a/q, b)$ with a odd. Here a/q need not be in lowest terms.

Lemma 2.1 *The outer billiards map is entirely defined on any special interval and indeed permutes the special intervals.*

Proof: The four order 2 rotations about the vertices of $K(A)$ send the point (x, y), respectively, to each of the following points.

$$(-2 - x, -y), \qquad (-x, 2 - y), \qquad (-x, -2 - y), \qquad (2A - x, -y). \quad (2.1)$$

The corresponding outer billiards map ψ' is built out of these rotations.

Define

$$\Lambda = 2\mathbf{Z}[A] \times \mathbf{Z}_{\text{odd}}; \qquad \mathbf{Z}[A] = \{mA + n|\ m, n \in \mathbf{Z}\} \qquad (2.2)$$

From Equation 2.1, both Λ and $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ are invariant under ψ'. Therefore the complementary set $\Lambda^c = \mathbf{R} \times \mathbf{Z}_{\text{odd}} - \Lambda$ is also invariant under ψ'. Note that Λ^c is precisely the union of special intervals.

To find the points of $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ where ψ' is not defined, we extend the sides of $K(A)$ and intersect them with $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$. We get 4 families of points.

$$(2n, 2n + 1), \qquad (2n, -2n - 1), \qquad (2An, 2n - 1), \qquad (2An, -2n + 1). \quad (2.3)$$

Here $n \in \mathbf{Z}$. Notice that all these points lie in Λ. Hence ψ' is defined on all points of Λ^c. The first statement of our result now follows from the fact that Λ^c is ψ'-invariant.

For the second statement, note that ψ' is completely defined on any special interval. But ψ' is a piecewise isometric map. By continuity, ψ' is an isometry when restricted to each special interval. But then ψ' must map each special interval to another one. This proves the second statement. □

Remark: For rational kites, the dynamics on $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ is essentially combinatorial. It is just a question of how the special intervals are permuted by the dynamics. Thus we are really dealing with an infinite permutation. Of course, we will sometimes profit from considering this situation geometrically.

.

Lemma 2.2 *Let $A \in (0, 1)$ be arbitrary. Relative to the kite $K(A)$, the entire outer billiards orbit of any point (α, n) is defined provided that $\alpha \notin 2\mathbf{Z}[A]$ and $n \in \mathbf{Z}_{\text{odd}}$.*

Proof: The orbit of the point (α, n) never lands in any of the 4 families of points discussed in the previous result. Hence, at any step in the orbit, both the forward and backward iterates are defined. □

When A is irrational, the set $2\mathbf{Z}[A]$ is a countable dense subset of $\mathbf{R}$. Likewise, $2\mathbf{Z}[A] \times \mathbf{Z}_{\text{odd}}$ is a countable dense set of $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$.

2.3 THE RETURN LEMMA

Let ψ be the square map relative to some kite, as above. As in §1.5, let

$$\Xi = \mathbf{R}_+ \times \{-1, 1\}. \tag{2.4}$$

Lemma 2.3 (Return) *Let $p \in \Xi$. be a point with a well defined outer billiards orbit. Then $\psi^a(p), \psi^{-b}(p) \in \Xi$ for some $a, b > 0$.*

Remark: The main goal of this section is to prove the Return Lemma. The reader interested in the broad picture might want to skip this rather tedious section on the first pass. To accommodate such a reader, we give a quick heuristic explanation of why the Return Lemma is true. The ψ-orbits generally circulate around the kite, skipping at most 2 lines of $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ with each iterate. Being made from 2 consecutive rays, Ξ serves as an impenetrable barrier to the progress of the orbit in both the forward and backward directions.

To prepare for our proof of the Return Lemma, and also for later use in the proof of the Pinwheel Lemma in Part 2, we discuss some structure of the map ψ. For each $p \in \mathbf{R}^2$ at which ψ is well defined, we have $\psi(p) = p + V$ for some vector V that is twice the difference between a pair of vertices of $K(A)$. There are a priori 12 possibilities for V, and the following 10 actually occur.

- $V_1 = -V_5 = (0, 4)$.
- $V_2 = -V_6 = (-2, 2)$.
- $V_3 = -V_7 = (-2 - 2A, 0)$.
- $V_4 = -V_8 = (-2, -2)$.
- $V_4^\sharp = -V_6^\flat = (-2A, 2)$.

When listed in the order $1, 2, 3, 4, 4^\sharp, 5, 6^\flat, 6, 7, 8$, the vectors defined above turn in counterclockwise fashion.

For each index j, there is some region $R_j \subset \mathbf{R}^2 - K(A)$ such that

$$p \in R_j \qquad \Longleftrightarrow \qquad \psi(p) = p + V_j. \tag{2.5}$$

The two regions $R_4^\sharp$ and $R_6^\flat$ are bounded regions. These regions ultimately turn out to be of no importance to us. The remaining regions $R_1, ..., R_8$ are unbounded and play an important role. The 10 regions partition $\mathbf{R}^2 - K(A)$. One can compute this partition by extending the sides of K in pinwheel fashion and then suitably pulling these sides back under the outer billiards map.

We now give a precise but terse description of the partition. For $R_4^\sharp$ and $R_6^\flat$, we list just the vertices of the polygon. The remaining regions are unbounded. The notation $\overrightarrow{q_1}, p_1, ..., p_k, \overrightarrow{q_2}$ indicates the following.

- The two unbounded edges are the rays $\overrightarrow{p_1 q_1}$ and $\overrightarrow{p_k q_2}$.
- $p_2, ..., p_{k-1}$ are any additional intermediate vertices.

Finally, to improve the typesetting, we have set $\alpha = (A-1)^{-1}$.

- R_1: $\overrightarrow{(1,-1)}$, $(1,-2)$, $\overrightarrow{(1,1)}$.
- R_2: $\overrightarrow{(1,1)}$, $(1,-2)$, $(0,-1)$, $\overrightarrow{(A,1)}$.
- R_3: $\overrightarrow{(A,1)}$, $(2A,1)$, $\alpha(2A^2,-1-A)$, $\overrightarrow{(-A,1)}$.
- R_4: $\overrightarrow{(-A,1)}$, $\alpha(2A,A-3)$, $\overrightarrow{(-1,1)}$.
- $R_4^\sharp$: $(A,0)$, $(2A,1)$, $\alpha(2A^2,-1-A)$.
- R_5: $\overrightarrow{(-1,1)}$, $\alpha(2A,A-3)$, $(-A,2)$, $\alpha(2A,3A-1)$, $\overrightarrow{(-1,-1)}$.
- $R_6^\flat$: $(0,1)$, $(-A,2)$, $\alpha(2A,3A-1)$.
- R_6: $\overrightarrow{(-1,-1)}$, $\alpha(2,A+1)$, $\overrightarrow{(-A,-1)}$.
- R_7: $\overrightarrow{(-A,-1)}$, $\alpha(2,A+1)$, $(-2,-1)$, $\overrightarrow{(A,-1)}$.
- R_8: $\overrightarrow{(A,-1)}$, $(-2,-1)$, $(-1,0)$, $\overrightarrow{(1,-1)}$.

Figure 2.2 shows accurately the partition and the vectors for $A = 1/3$. The numbers indicate the regions in an obvious way. The small 4 represents $R_4^\sharp$, for instance. For the vectors, the rule is that that the tail of V_j lies in R_j. The shaded strip is bounded by the lines $y = \pm 1$. Note a certain "kinship" between R_4 and $R_4^\sharp$, and similarly between R_6 and $R_6^\flat$.

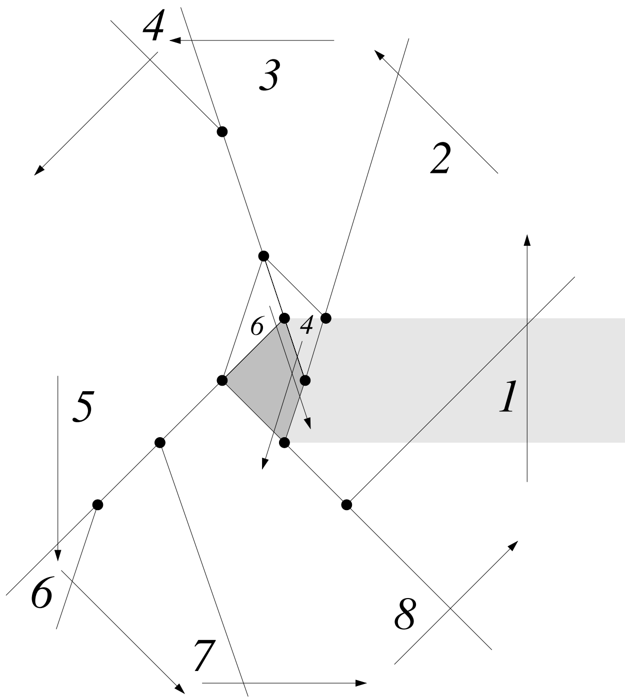

Figure 2.2: The partition for $A = 1/3$.

Figure 2.3 shows the partition for the parameters $A = p/7$ for $p = 1, 2, 3, 4, 5, 6$. The reader can draw the figure for any slice using Billiard King.

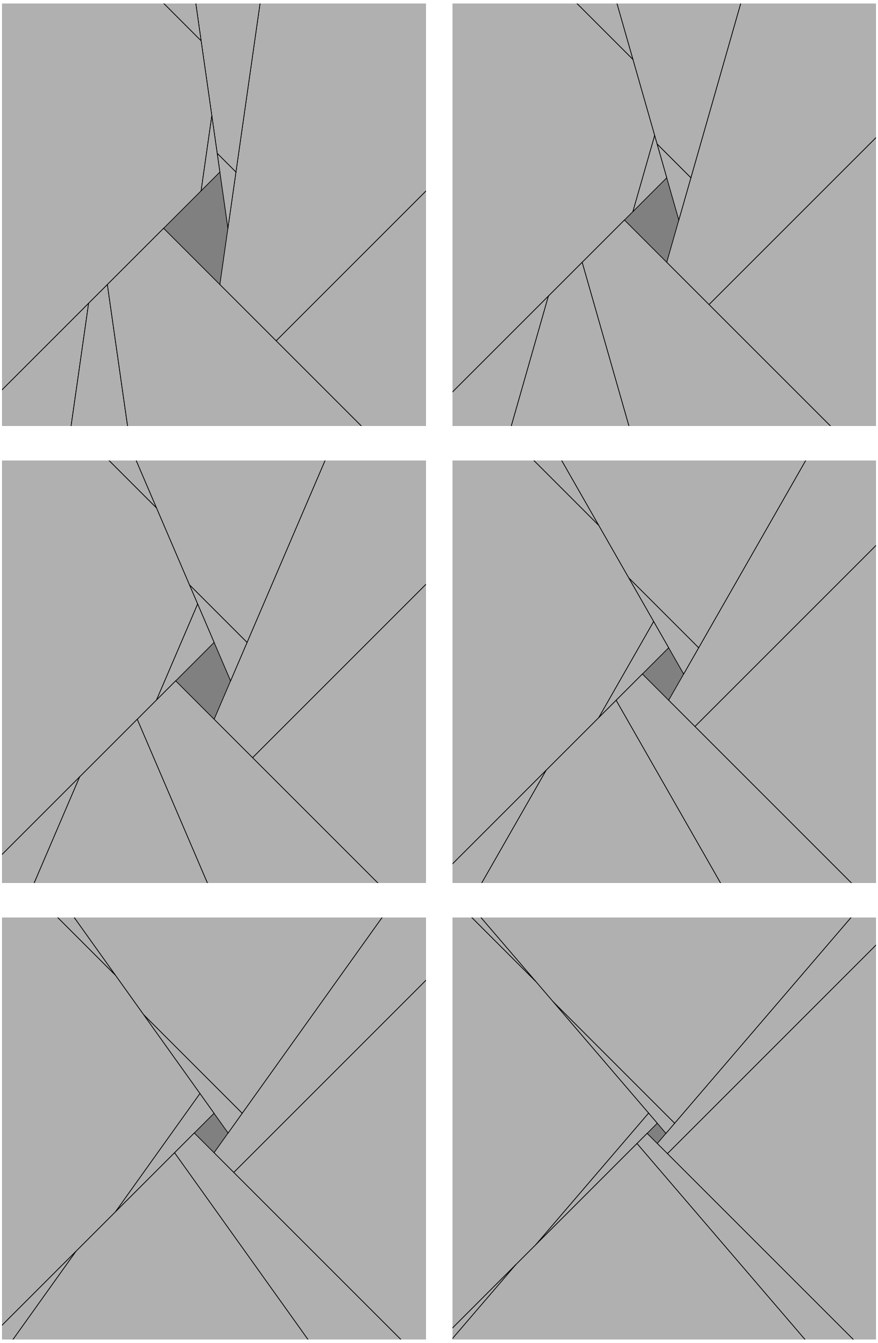

Figure 2.3: The partition for 6 parameters.

We define

$$\widehat{R}_i = R_i + V_i, \qquad S_{ij} = \widehat{R}_i \cap R_j \cap (\mathbf{R} \times \mathbf{Z}_{\mathrm{odd}}). \tag{2.6}$$

We put a 1 in the (ij)th spot of the matrix if there is a parameter A for which $S_{ij} \neq \emptyset$. This means that there is some $p \in R_i \cap (\mathbf{R} \times \mathbf{Z}_{\mathrm{odd}})$ such that $\psi(p) \in R_j$. Both the partition and ψ depend on the parameter, but we omit this from our notation. Note that not all transitions are possible for all parameters. Here is the transition matrix.

$$\begin{array}{c|cccccccccc}
 & R_1 & R_2 & R_3 & R_4 & R_4^\sharp & R_5 & R_6^\flat & R_6 & R_7 & R_8 \\
\widehat{R}_1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\widehat{R}_2 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
\widehat{R}_3 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
\widehat{R}_4 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\widehat{R}_4^\sharp & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\widehat{R}_5 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
\widehat{R}_6^\flat & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\widehat{R}_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
\widehat{R}_7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\widehat{R}_8 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1
\end{array} \tag{2.7}$$

Remark: Though it plays no role in our analysis, we note one pretty symmetry: Reflection in the x-axis swaps $\widehat{R}_i$ and R_j if $i + j = 10$. This works even for the pair $(4^\sharp, 6^\flat)$.

Proof of the Return Lemma: We will consider just the forward orbit. The backward orbit requires the same treatment and indeed follows from symmetry. Given the regions and vectors, the forward orbit of a point cannot stay in one region forever. Starting with a point $z \in \Xi$, we let $i_1 \to i_2 \to \cdots$ denote the sequence of regions encountered by the forward ψ-orbit of z. Let $z_k = (x_k, y_k)$ be the first point in R_{i_k}. Looking at the matrix, we arrive at 3 cases.

Case 1: Suppose $i_k = 1$ for some k. Looking at R_1, we see that $x_k > 0$. The set $\{y \geq 3\}$ is more than 4 units from the region $R_{i_{k-1}}$, and each of the vectors has length at most 4. Hence $y_k \in \{..., -3, -1, 1\}$. As the orbit proceeds, we just keep adding $V_1 = (0, 4)$ until we reach $y_k \in \{-1, 1\}$, and then we are in Ξ.

Case 2: Suppose $i_k = 2$ for some k. The same argument places the constraints on x_k and y_k as in Case 1. Now we also observe that the set $\{y \leq -3\}$ is disjoint from R_2. Hence $y_k \in \{-1, 1\}$. Hence $z_k \in \Xi$.

Case 3: If we never see $i_k \in \{1, 2\}$, then we must have $i_{k-1} = 8$ and $i_k = 4^\sharp$ for some k. We check easily that in this case $z_k \in \Xi$. The argument is similar to that in the previous two cases. $\square$

2.4 THE RETURN MAP

The Return Lemma implies that the *first return map* $\Psi\colon \Xi \to \Xi$ is well defined on any point with an outer billiards orbit. This includes the set

$$(\mathbf{R}_+ - 2\mathbf{Z}[A]) \times \{-1, 1\},$$

as we saw in Lemma 2.2.

Given the nature of the maps in Equation 2.1 comprising ψ, we see that

$$\Psi(p) - (p) \in 2\mathbf{Z}[A] \times \{-2, 0, 2\}.$$

In Part 2, we will prove the main structural result about the first return map, namely, the Master Picture Theorem. As a consequence of the Master Picture Theorem, we get the following result.

$$\Psi(p) - (p) = 2(A\epsilon_1 + \epsilon_2, \epsilon_3), \qquad \epsilon_j \in \{-1, 0, 1\}, \qquad \sum_{j=1}^{3} \epsilon_j \equiv 0 \mod 2. \tag{2.8}$$

The parity result in Equation 2.8 has the following proof. The vectors V_j considered above all have the form

$$(2aA + 2b, 2c), \qquad a + b + c \equiv 0 \mod 2.$$

The vector $\Psi(p) - p$ is some finite sum of these vectors.

We do not have an easy proof for the bound $|\epsilon_j| = 1$, but we can easily give a rough idea. For the reader who skipped the proof of the Return Lemma above, we remark that our explanation here also gives a rough reason why the Return Lemma is true. Consider the forward ψ-orbit of a point of Ξ that is far from the origin. This orbit essentially circulates counterclockwise around the origin, nearly making a giant octagon. Looking at our vectors $V_1, \ldots, V_8$, we see that this near octagon has approximate 4-fold bilateral symmetry. The *return pair* $(\epsilon_1(p), \epsilon_2(p))$ essentially measures the *approximation error* between the true orbit and the closed octagon. There is almost complete cancellation as one goes around this near octagon, and this keeps the return pair uniformly small.

Remarks:

(i) Some version of the first return map is considered in many papers on outer billiards – e.g., [**GS**], [**G**], and [**DF**].

(ii) On a nuts-and-bolts level, this book concerns how the pair $(\epsilon_1(p), \epsilon_2(p))$ depends on $p \in \Xi$. The pair (ϵ_1, ϵ_2) and the parity condition determine ϵ_3. I like to say that this book is really about the infinite accumulation of small errors.

(ii) Reflection in the x-axis conjugates the map ψ to the map ψ^{-1}. Thus, once we understand the orbit of the point $(x, 1)$, we automatically understand the orbit of the point $(x, -1)$. Put another way, the unordered pair of return points $\{\Psi(p), \Psi^{-1}(p)\}$ for $p = (x, \pm 1)$ depends only on x.

2.5 THE ARITHMETIC GRAPH

Fundamental Map: Recall that $\Xi = \mathbf{R}_+ \times \{-1, 1\}$. Given $\alpha \in \mathbf{R}$ and a parameter A, define

$$M = M_{A,\alpha} : \mathbf{Z}^2 \to \mathbf{R} \times \{-1, 1\}$$

by

$$M_{A,\alpha}(m, n) = \Big(2Am + 2m + 2\alpha, (-1)^{m+n+1}\Big). \tag{2.9}$$

The second coordinate of M is either 1 or -1, depending on the parity of $m + n$. This definition is adapted to the parity condition in Equation 2.8. We call M a *fundamental map*. Each choice of α gives a different map.

Main Definition: M is injective when A is irrational and M is injective on any disk of radius q when $A = p/q$. Given $p_1, p_2 \in \mathbf{Z}^2$, we write $p_1 \to p_2$ iff the following hold.

- $\zeta_j = M(p_j) \in \Xi$.
- $\Psi(\zeta_1) = \zeta_2$.
- $\|p_1 - p_2\|$ is as small as possible.

The third condition is relevant only in the rational case. According to Equation 2.8, the choice of p_2 depends uniquely on p_1, in all cases, and $\|p_1 - p_2\| \leq \sqrt{2}$. Our construction gives a directed graph with vertices in $\mathbf{Z}^2$. We call this graph the *arithmetic graph* and denote it by $\widehat{\Gamma}_\alpha(A)$. We usually ignore the isolated vertices of the graph. These correspond to points on which the return map is the identity.

A Convention: When $A = p/q$, any choice of $\alpha \in (0, 2/q)$ gives the same result. This is a consequence of Lemma 2.1. To simplify the formulas, we choose $\alpha = 0_+$, where 0_+ is an infinitesimally small positive number. When we write formulas, we usually take $\alpha = 0$, but we always use the convention that the lattice point (m, n) tracks the orbits just to the right of the points $(2Am + 2n, \pm 1)$. With this convention, we have

$$\widehat{\Gamma}\Big(\frac{p}{q}\Big) = \widehat{\Gamma}_{0_+}\Big(\frac{p}{q}\Big), \qquad M(m, n) = \Big(2\Big(\frac{p}{q}\Big)m + 2n, (-1)^{m+n+1}\Big). \tag{2.10}$$

We say that the *baseline* of $\widehat{\Gamma}(A)$ is the line $M^{-1}(0)$. The baseline is the line of slope $-A$ through a point infinitesimally far beneath the origin. In practice, we think of the baseline as the line of slope $-A$ through the origin.

Translation Symmetry: When p/q is odd, Equation 2.10 gives

$$M(\zeta + V) = M(\zeta), \qquad V = (q, -p), \tag{2.11}$$

for any $\zeta \in \mathbf{Z}^2$. Hence translation by V preserves $\widehat{\Gamma}(p/q)$ as a directed graph. When p/q is even, we have $M(\zeta + V) = R \circ M(\zeta)$, where R is the reflection in the x-axis. The map R conjugates Ψ to Ψ^{-1}. In this case, translation by V preserves $\widehat{\Gamma}(p/q)$ as a graph but reverses the direction.

In Part 3 we will prove the following result.

Theorem 2.4 (Embedding) *Any well defined arithmetic graph is the disjoint union of embedded polygons and bi-infinite embedded polygonal curves.*

Let $\Gamma(p/q)$ denote the component of $\widehat{\Gamma}(p/q)$ that contains the origin. This component corresponds to the fundamental orbit discussed in Theorem 1.7. The component $\Gamma(p/q)$ is never a closed polygon when p/q is odd. This is a consequence of the Room Lemma in Chapter 3. Figure 2.4 shows an example.

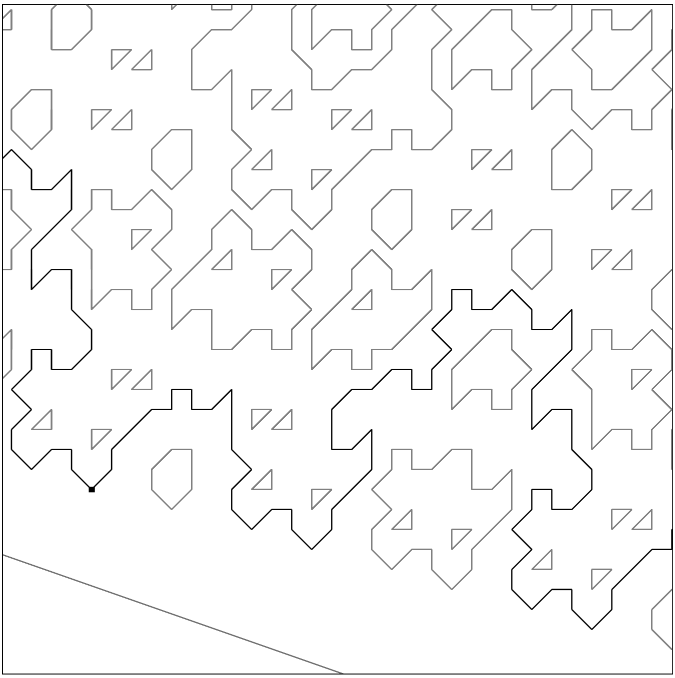

Figure 2.4: Some of $\widehat{\Gamma}(7/25)$, with $\Gamma(7/25)$ in black.

In contrast, we have the following result.

Lemma 2.5 *If p/q is even, then every component of $\widehat{\Gamma}(p/q)$ is a polygon.*

Proof: Suppose that some component β is not a polygon. Since translation by V reverses the direction on $\widehat{\Gamma}$, we have $\beta \neq \beta + V$.

Let $\langle V \rangle \approx \mathbf{Z}$ denote the group generated by integer multiples of V. Let X be the cylinder $\mathbf{R}^2/\langle V \rangle$. Let $\pi : \mathbf{R}^2 \to X$ be the quotient map. By the Embedding Theorem, $\pi(\beta)$ is embedded in X. Since β corresponds to a periodic orbit, $\pi(\beta)$ is a closed loop in X. Since β is not a polygon, $\pi(\beta)$ is nontrivial in the first homology group $H_1(X) = \mathbf{Z} \approx \langle V \rangle$. Because $\pi(\beta)$ is embedded, $\pi(\beta)$ must generate $H_1(X)$. But then $\beta = \beta + V$, a contradiction. $\square$

2.6 LOW VERTICES AND PARITY

Remark: The material in this section is not needed for the proofs of statements 1 and 2 of the Erratic Orbits Theorem.

Let A be any kite parameter. We define the *parity* of a low vertex (m, n) to be the parity of $m + n$. Here we explain the structure of the arithmetic graph at low vertices. Our answer will be given in terms of a kind of phase portrait. Given a point $(x, A) \in (0, 2) \times (0, 1)$, we have

$$\Psi^{\pm 1}(x, -1) = (x, -1) + 2(\epsilon_1^{\pm} A + \epsilon_2^{\pm}, \epsilon_3^{\pm}). \tag{2.12}$$

For the point (x, A) we associate the directed graph

$$(\epsilon_1^-, \epsilon_2^-) \to (0, 0) \to (\epsilon_1^+, \epsilon_2^+).$$

This gives a local picture of the arithmetic at the low vertex (m, n) such that $M_A(m, n) = (x, -1)$. If $M_A(m, n) = (x, 1)$, then we get the local picture by reversing the edges. Figure 2.5 shows the final result. The gray edges in the figure, present for reference, connect $(0, 0)$ to $(0, -1)$. The gray triangle represents the places where the return map is the identity.

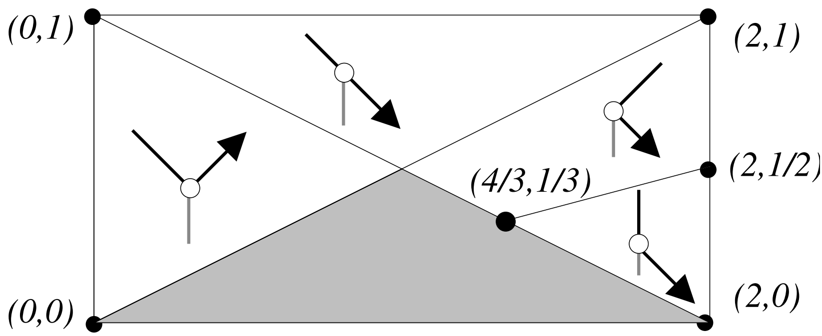

Figure 2.5: Low-vertex phase portrait.

Example: Relative to $A = 1/3$, the vertex $(-7, 3)$ is a low vertex. We compute that

$$M_{1/3}(A) = (4/3 + \alpha, -1).$$

Here α is an infinitesimally small positive number. To see the local picture of the arithmetic graph $\Gamma(1/3)$ at $(-7, 3)$, we observe that the point $p = (4/3 + \alpha, 1/3)$ lies infinitesimally to the right of the point $(4/3, 1/3)$. Hence $(\epsilon_1^-, \epsilon_2^-) = (0, 1)$ and $(\epsilon_1^+, \epsilon_2^+) = (1, -1)$.

In principle, one can derive Figure 2.5 by hand. We will explain how to derive it in §6.8 as a corollary of the Master Picture Theorem.

Lemma 2.6 *No component of* $\widehat{\Gamma}(A)$ *contains low vertices of both parities.*

Proof: Recall that $\widehat{\Gamma}$ is an oriented graph. If v is a nontrivial low vertex of $\widehat{\Gamma}$, we can say whether $\widehat{\Gamma}$ is *left-travelling* or *right-travelling* at v. The definition is this: As we travel along the orientation and pass through v, the line segment connecting v to $v - (0, 1)$ lies on either our left or our right. This gives the name to our definition.

A component of Γ cannot right-travel at one low vertex and left-travel at another. Figure 2.6 shows the problem. The curve γ would create a pocket for itself, and γ could not escape from this pocket because it must stay above the baseline. The low vertices of γ serve as barriers. Travelling into the pocket, γ would have only a finite number of steps before it would have to cross itself. But then we would contradict the Embedding Theorem.

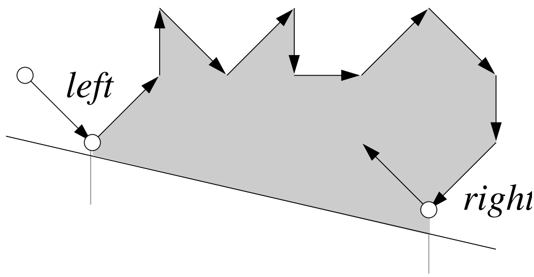

Figure 2.6: γ travels into a pocket.

To finish the proof, we just have to show that a component of $\widehat{\Gamma}$ right-travels at a low vertex v if and only if v has even parity. We will show that a component of $\widehat{\Gamma}$ always right-travels at low vertices of even parity. Let us explain why this suffices. Recall that the fundamental map M maps vertices of even parity to $\mathbf{R}_+ \times \{-1\}$, and vertices of odd parity to $\mathbf{R} \times \{1\}$. Also, recall that reflection in the x-axis conjugates the return map Ψ to Ψ^{-1}. It follows from this symmetry that $\widehat{\Gamma}$ left-travels at all low vertices of odd parity if and only if $\widehat{\Gamma}$ right-travels at all vertices of even parity. But a glance at Figure 2.5 shows that $\widehat{\Gamma}$ right-travels at all vertices of even parity. The gray line segment always lies on the right. $\square$

Corollary 2.7 *Let* $A \in (0, 1)$ *be arbitrary. Suppose that* $\xi_+ \in (0, 2) \times \{1\}$ *and* $\xi_- \in (0, 2) \times \{-1\}$ *have well defined orbits. Then the two orbits* $O_2(\xi_+)$ *and* $O_2(\xi_-)$ *are distinct.*

Proof: Suppose, for the sake of contradiction, that the orbits coincide. Then there is a choice of α such that a component Γ of the arithmetic graph $\Gamma_\alpha(A)$ corresponds to this common orbit. There are low vertices (m_+, n_+) and (m_-, n_-) such that $M_\alpha(m_\pm, n_\pm) = \xi_\pm$. But then (m_+, n_+) and (m_-, n_-) have opposite parity, contradicting the previous result. $\square$

2.7 HAUSDORFF CONVERGENCE

Here we state the basic results that will allow us to take geometric limits of orbits for outer billiards systems with varying parameters. When it comes to taking limits of arithmetic graphs, we will always use the Hausdorff topology.

The Hausdorff Metric and Topology: Given two compact subsets $K_1, K_2 \subset \mathbf{R}^2$, we define $d(K_1, K_2)$ to be the infimal $\epsilon > 0$ such that the set K_1 is contained in the ϵ-tubular neighborhood of the set K_2, and vice versa. The function $d(K_1, K_2)$ is known as the *Hausdorff metric*. A sequence $\{C_n\}$ of closed subsets of $\mathbf{R}^2$ is said to *Hausdorff converge* to $C \subset \mathbf{R}^2$ if $d(C_n \cap K, C \cap K) \to 0$ for every compact subset $K \subset \mathbf{R}^2$. This notion of convergence is called the *Hausdorff topology*.

Remark: In the cases of interest to us, C_n is always an arc of an arithmetic graph that contains $(0, 0)$. In this case, the Hausdorff convergence has a simple meaning. $\{C_n\}$ converges to C if and only if the following property holds true. For any N, there is some N' such that $n > N'$ implies that the first N steps of C_n away from $(0, 0)$ in either direction agree with the corresponding steps of C.

Given a parameter $A \in (0, 1)$ and a point $\zeta \in \Xi$, we say that a pair (A, ζ) is *N-defined* if the first N iterates of the outer billiards map of ζ are defined relative to A in both directions. We let $\Gamma(A, \zeta)$ be as much of the arithmetic graph as is defined. We call $\Gamma(A, \zeta)$ a *partial arithmetic graph*.

Lemma 2.8 (Continuity Principle) *Let $\{\zeta_n\} \in \Xi$ converge to $\zeta \in \Xi$. Let $\{A_n\}$ converge to A. Suppose the orbit of ζ is defined relative to A. Then for any N, there is some N' such that $n > N'$ implies that (ζ_n, A_n) is N-defined. The corresponding sequence $\{\Gamma(A_n, \zeta_n)\}$ of partially defined arithmetic graphs Hausdorff-converges to $\Gamma(A, \zeta)$.*

Proof: Let ψ'_n be the outer billiards map relative to A_n. Let ψ' be the outer billiards map defined relative to A. If $p_n \to p$ and ψ' is defined at p, then ψ'_n is defined at p_n for n sufficiently large and $\psi'_n(p_n) \to \psi(p)$. This follows from the fact that $K(A_n) \to K(A)$ and from the fact that a piecewise isometric map is defined and continuous in open sets. The Continuity Principle now follows from induction. $\square$

In the case when the orbit of ζ_n relative to A_n is already well defined, the partial arithmetic graph is the same as one component of the ordinary arithmetic graph. In this case, we can state the Continuity Principle more simply.

Corollary 2.9 *Let $\{\zeta_n\} \in \Xi$ converge to $\zeta \in \Xi$. Let $\{A_n\}$ converge to A. Suppose the orbit of ζ is defined relative to A and the orbit of ζ_n is defined relative to A_n for all n. Then $\{\Gamma(A_n, \zeta_n)\}$ Hausdorff converges to $\Gamma(A, \zeta)$.*

We will have occasion to use both versions of the Continuity Principle in our proofs.

Remark: The remaining material in this section is not needed for the proofs of statements 1 and 2 of the Erratic Orbits Theorem.

Lemma 2.10 *Let $s \in (0,1)$. If $(\psi')^k(s,1)$ is not defined on $K(A)$ for some $|k| \le N$, then $s = 2Am + 2n$ for some $m, n \in \mathbf{Z}$ such that $|m| \le 2N$.*

Proof: We will consider the case when $k > 0$. The other case is similar. By induction, we may suppose that $t = (t_1, t_2) = (\psi')^{N-1}(s)$ is well defined. Looking at the maps in Equation 2.1, we see inductively that $|t_2| \le 2N$. If $\psi'(t)$ is not defined, then t is one of the points in Equation 2.3 for some $|n| \le N$. Hence

$$t_1 = 2Am' + 2n'; \qquad |m'| \le N.$$

By Equation 2.1 and induction, we have

$$s = 2Am + 2n; \qquad |m| \le N + |m'| \le 2N.$$

This completes the proof □

We think of the next result as a rigidity lemma because it says that certain limits are independent of how we take them.

Lemma 2.11 (Rigidity) *Let A_n be any sequence of parameters converging to the irrational parameter A. Let $\zeta_n \in [0,2] \times \{1\}$ be any sequence of points converging to $(0,1)$. Let $\Gamma(\zeta_n, A)$ be the arithmetic graph of ζ_n relative to A. Then the sequence $\{\Gamma(\zeta_n, A)\}$, if entirely defined, Hausdorff-converges.*

Proof: Let $\epsilon \in (0,1)$ be given. Define

$$\Sigma_\epsilon(A) = \{(s, A')|\ s \in (0,\epsilon),\ \ |A - A'| < \epsilon\}. \tag{2.13}$$

Let $O(s; A')$ denote the outer billiards orbit of $(s,1)$ relative to $K(A')$. Suppose that one of the first N iterates of $O(s, A')$ is not defined. By Lemma 2.10, we have $m, n \in \mathbf{Z}$ such that

$$s = 2A'm - 2n; \qquad |m| \le 2N. \tag{2.14}$$

(We use a minus sign for convenience.) Note that $m \ne 0$. Hence, by Equation 2.14 and the triangle inequality,

$$\left|A - \frac{n}{m}\right| < |A - A'| + \left|A' - \frac{n}{m}\right| < \epsilon + \frac{s}{2|m|} < 2\epsilon. \tag{2.15}$$

This is impossible for ϵ sufficiently small. Hence the first N iterates of $O(s; A')$ in both directions are well defined when $(s, A') \in \Sigma_\epsilon(A)$ and ϵ is sufficiently small.

If all orbits in some interval are defined, then all orbits in that interval have the same combinatorial structure. Hence, for any N, there is some $\epsilon > 0$ such that the combinatorics of the first N iterates, in either direction, of $O(s; A')$ are independent of $(s, A') \in \Sigma_\epsilon(A)$. This lemma now follows from the Return Lemma, which guarantees that, as $N \to \infty$, the number of returns to Ξ tends to ∞ as well. □

This page intentionally left blank

Chapter Three

The Hexagrid Theorem

3.1 THE ARITHMETIC KITE

In this section we describe a certain quadrilateral, which we call the *arithmetic kite*. This object is meant to "live" in the same plane as the arithmetic graph. The diagonals and sides of this quadrilateral define 6 special directions. In the next section we describe a grid made from 6 infinite families of parallel lines based on these 6 directions. Let $A = p/q$. Figure 3.1 accurately shows $\mathcal{K}(A)$ for $A = 1/3$.

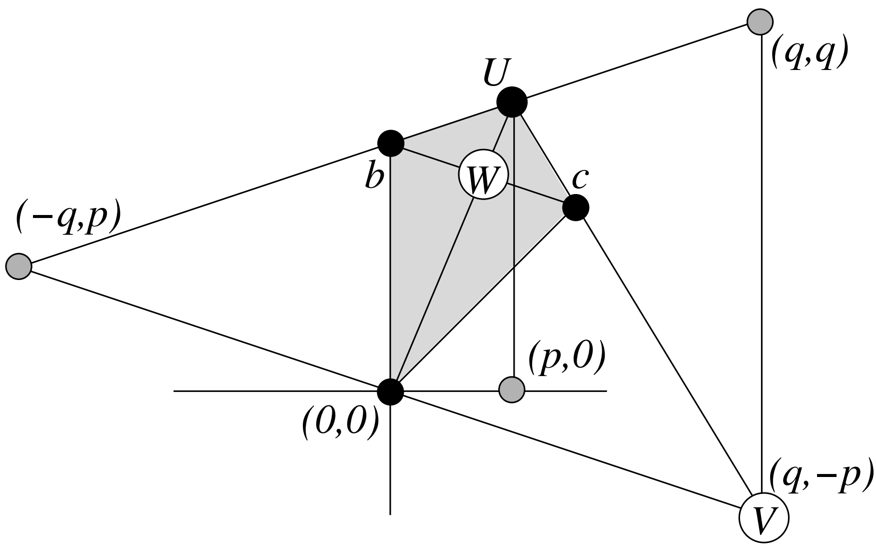

Figure 3.1: The arithmetic kite.

One can construct this figure using straight lines and the given coordinates. The pairs of lines that look parallel are supposed to be parallel. Setting $a = (q, q)$, we have

$$b = \frac{a - V}{2}, \qquad U = Aa + (1 - A)b, \qquad W = \frac{U}{1 + A} = \frac{b + c}{2}. \quad (3.1)$$

The vectors V and W are of special interest to us. We have

$$V = (q, -p), \qquad W = \left(\frac{pq}{p + q}, \frac{pq}{p + q} + \frac{q - p}{2}\right). \quad (3.2)$$

It follows from the rightmost (double) equality in Equation 3.1 that $\mathcal{K}(A)$ is affinely equivalent to $K(A)$.

The *hexagrid* $G(A)$ consists of two interacting grids, which we call the *room grid* $RG(A)$ and the *door grid* $DG(A)$.

Room Grid: When A is an odd rational, $RG(A)$ consists of the lines obtained by extending the diagonals of $\mathcal{K}(A)$ and then taking the orbit under the lattice $\mathbf{Z}[V/2, W]$. These are the black lines in Figure 3.2. In the case when A is an even rational, we make the same definition but use the lattice $\mathbf{Z}[V, 2W]$ instead.

Door Grid: The *door grid* $DG(A)$ is the same for both even and odd rationals. It is obtained by extending the sides of $\mathcal{K}(A)$ and then taking their orbit under the 1-dimensional lattice $\mathbf{Z}[V]$. These are the gray lines in Figure 3.2.

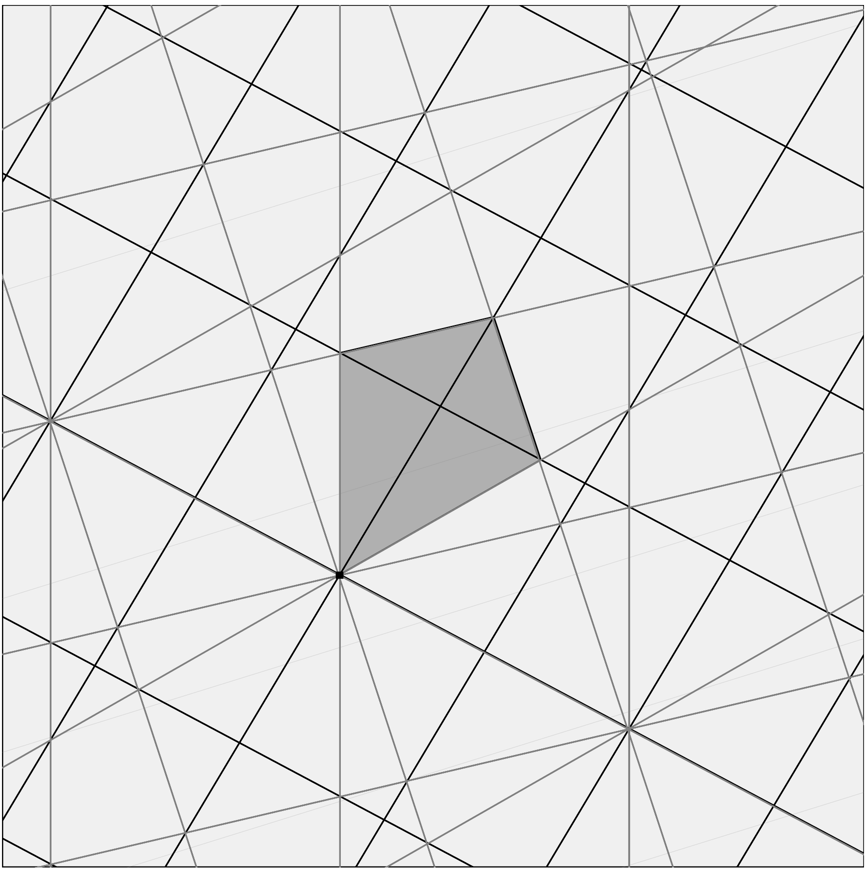

Figure 3.2: $G(25/47)$. and $\mathcal{K}(25/47)$.

Figure 3.2 shows the hexagrid $G(25/47)$ and the arithmetic kite $\mathcal{K}(25/47)$. Billiard King allows the user to draw color versions of such figures for essentially any rational parameter.

3.2 THE HEXAGRID THEOREM

First we will talk informally about the Hexagrid Theorem. In the previous section, we defined two grids, the room grid and the door grid. The Hexagrid Theorem says that these two grids control the large-scale structure of the arithmetic graph. It turns out that the lines of the room grid serve to "confine" the arithmetic graph, in the sense that the arithmetic graph can cross these lines only at very specific locations. The door grid serves to define the locations – i.e., the doors – where the graph can cross the lines of the room grid. Thus the Hexagrid Theorem relates two kinds of objects, *wall crossings* and *doors*. Informally, the Hexagrid Theorem says that the arithmetic graph crosses a wall only at a door. Here are formal definitions.

Rooms and Walls: $RG(A)$ divides $\mathbf{R}^2$ into different connected components which we call *rooms*. Say that a *wall* is the line segment of positive slope that divides two adjacent rooms.

Doors: When p/q is odd, we say that a *door* is a point of intersection between a wall of $RG(A)$ and a line of $DG(A)$. When p/q is even, we have the same definition, except that we exclude intersection points of the form (x, y), where y is a half-integer. Every door is a triple point, and every wall has one door. The first coordinate of a door is always an integer. (See Lemma 15.1.) In exceptional cases – when the second coordinate is also an integer – the door lies in the corner of the room. In this case, we associate the door to both walls containing it. The door $(0, 0)$ has this property.

Crossing Cells: Say that an edge e of $\widehat{\Gamma}$ *crosses a wall* if e intersects a wall at an interior point. Say that a union of two incident edges of Γ *crosses a wall* if the common vertex lies on a wall and the two edges point to opposite sides of the wall. The point $(0, 0)$ has this property. We say that a *crossing cell* is either an edge or a union of two edges that crosses a wall in the manner just described. For instance $(-1, 1) \to (0, 0) \to (1, 1)$ is a crossing cell for any $A \in (0, 1)$.

In Part 3 of the book we will prove the following result. Let $[y]$ denote the floor of y, the greatest integer less than or equal to y.

Theorem 3.1 (Hexagrid) *Let $A \in (0, 1)$ be rational.*

1. *$\widehat{\Gamma}(A)$ never crosses a floor of $RG(A)$. Any edges of $\widehat{\Gamma}(A)$ incident to a vertex contained on a floor rise above that floor (rather than below it.)*
2. *There is a bijection between the set of doors and the set of crossing cells. If y is not an integer, then the crossing cell corresponding to the door (m, y) contains $(m, [y]) \in \mathbf{Z}^2$. If y is an integer, then (x, y) corresponds to 2 doors. One of the corresponding crossing cells contains (x, y), and the other one contains $(x, y - 1)$.*

Figure 3.3 illustrates the Hexagrid Theorem for $p/q = 25/47$. We will explain the shaded parallelogram $R(25/47)$ in the next section. We have shown only the fleeting components in Figure 3.3 – i.e., those components that are not closed polygons. Each persistent component – i.e., those components that are closed polygons – is confined to a single room.

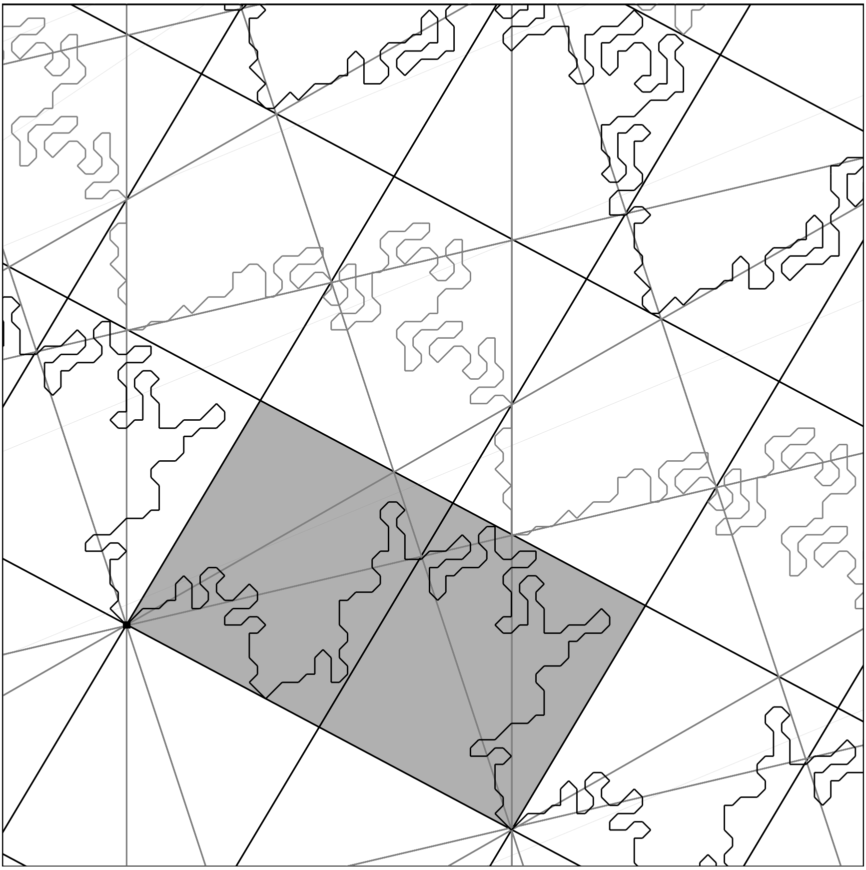

Figure 3.3: $G(25/47)$, $R(25/47)$, and some of $\widehat{\Gamma}(25/47)$.

The figure for the even case looks similar but slightly different. The reader can see much better figures for the Hexagrid Theorem using either Billiard King or our interactive guide to the book. The interactive guide shows only the odd case, but Billiard King shows both the even and odd cases.

The Hexagrid Theorem helps us in two ways. First, the pattern of the doors forces some of the orbits to oscillate on a large scale. Second, the pattern of the walls guarantees that the components do not oscillate too wildly for us to control them. This controlled oscillation will come in handy later on when we discuss period-copying phenomena.

Remark: The Hexagrid Theorem immediately implies that all special orbits on $K(p/q)$ are bounded, and hence periodic.

3.3 THE ROOM LEMMA

The Room Lemma, an easy consequence of the Hexagrid Theorem, is the main result we use to force large oscillations of the fundamental orbit $O(1/q, -1)$.

Let $R(p/q)$ denote the parallelogram whose vertices are

$$(0,0), \qquad V, \qquad W, \qquad V+W. \tag{3.3}$$

Here V and W are as in Equation 3.2. $R(p/q)$ is the shaded parallelogram in Figure 3.3. We also define

$$d_0 = (x, [y]), \qquad x = \frac{p+q}{2}, \qquad y = \frac{q^2-p^2}{4q} \tag{3.4}$$

d_0 lies within 1 vertical unit of the centerline of $R(p/q)$, above the midpoint of the centerline. d_0 is just below the triple point contained in the interior of the shaded parallelogram in Figure 3.3.

Lemma 3.2 (Room) *Let p/q be an odd rational. Then $\Gamma(p/q)$ is an open polygonal curve. One period of $\Gamma(p/q)$ connects $(0,0)$ to d_0 to $(q,-p)$. This period is contained in $R(p/q)$.*

Proof: First of all, for any value of A, it is easy to check that $\Gamma(A)$ contains the arc $(-1,1) \to (0,0) \to (1,1)$. One can see this from the phase portrait shown in Figure 2.6. This is to say that $\Gamma(p/q)$ enters $R(p/q)$ from the left at $(0,0)$. Now $R(p/q)$ is the union of two adjacent rooms R_1 and R_2. Note that $(0,0)$ is the only door on the left wall of R_1, and (x,y) is the only door on the wall separating R_1 and R_2, and $(q,-p)$ is the only door on the right wall of R_2. Here (x,y) is as in Equation 3.4. From the Hexagrid Theorem and the Embedding Theorem, $\Gamma(p/q)$ must connect $(0,0)$ to d_0 to $(q,-p)$. The arithmetic graph $\widehat{\Gamma}(p/q)$ is invariant under translation by $(q,-p)$, and so the whole pattern repeats endlessly to the left and right of $R(p/q)$. Hence $\Gamma(p/q)$ is an open polygonal curve. □

We remark that we did not really need the Embedding Theorem in our proof above.[1] All we require is that $\Gamma(p/q)$ cannot backtrack as we travel from one corner of $R(p/q)$ to the other. Lemma 3.3 below gives a self-contained proof of what we need.

Lemma 3.3 $\Gamma(p/q)$ *has valence* 2 *at every vertex.*

Proof: Recall that Ψ is the first return map to $\mathbf{R}_+ \times \{-1,1\}$. As in our proof of the Room Lemma, $\Gamma(p/q)$ has valence 2 at $(0,0)$. But $\Gamma(p/q)$ describes the forward orbit of $p = (1/q, 1)$ under Ψ. If some vertex of $\Gamma(p/q)$ has valence 1, then Ψ has order 2 when evaluated at the corresponding point. But then Ψ has order 2 when evaluated at v. But then $\Gamma(p/q)$ has valence 1 at $(0,0)$. This is a contradiction. □

[1] I am grateful to Dmitry Dolgopyat and Giovanni Forni for pointing this out to me.

3.4 ORBIT EXCURSIONS

Remark: The material in this section is not needed for the proof of the Erratic Orbits Theorem.

In this section, we prove Theorems 1.7, 1.9, and 1.10.

Let M_1 be the first coordinate of the map M in Equation 2.10. Let $\lambda = 1$ if p/q is odd and $\lambda = 2$ if p/q is even.

Lemma 3.4 *No point of $O_2(1/q) \cap \Xi$ has a first coordinate greater than $\lambda(p+q)$.*

Proof: Let L be the line of the room grid parallel to the baseline that contains the point λW. Here W is as in Equation 3.2. We compute that $M_1(\lambda W) = \lambda(p+q)$. By the Hexagrid Theorem, $\Gamma(p/q)$ lies in the strip bounded by the baseline and L. Looking at Equation 2.10, we see that M_1 is constant on L. Therefore we have the bound $M_1(m,n) \leq \lambda(p+q)$ for any vertex (m,n) of $\Gamma(p/q)$. □

Lemma 3.5 *The first coordinate of some point in $O_2(1/q) \cap \Xi$ exceeds $\lambda(p+q)/2$.*

Proof: To avoid an irritating case in which the calculations yield a bound that is off by 1 unit, we assume that $p > 1$.

In the odd case, $M_1(d_0)$ is the first coordinate of a point of $O_2(1/q,-1) \cap \Xi$, and we compute that $M_1(d_0) > (p+q)/2$. Here d_0 is as in the Room Lemma.

Consider the even case. Let L_0 be the line of the room grid through $(0,0)$ and parallel to the walls. By Lemma 2.5, the component $\Gamma(p/q)$ is a closed polygon. Since $\Gamma(p/q)$ contains the arc $(-1,1) \to (0,0) \to (1,1)$, an arc that has points on either side of L_0, the polygon $\Gamma(p/q)$ must cross L_0 at some point above $(0,0)$ as well. The door on L_0 immediately above $(0,0)$ is within 1 unit of U, the tip of $\mathcal{K}(A)$. See Figure 3.1. By the Hexagrid Theorem, $\Gamma(p/q)$ must cross L_0 within 1 unit of U. Call this crossing point d_0'. We compute that $M(d_0') > p+q$, at least when $p > 1$. But $M_1(d_0')$ is the first coordinate of a point in $O_2(1/q,-1) \cap \Xi$. □

Proof of Theorem 1.7: Lemma 3.5 immediately gives the lower bounds in Theorem 1.7, except in the case $p = 1$. The unimportant case $p = 1$ requires a separate argument which we omit. For the upper bounds, we note that the first coordinates of points in $O_2(1/q,-1) \cap \Xi$ lie in $[0, \lambda(p+q)]$, by Lemma 3.4. The second coordinates belong to the set $\{-1,1\}$. This completes the proof. □

Proof of Theorem 1.9: We will give the proof for odd rationals. The even case is just about the same except for the factor of 2. Suppose that p/q is odd. Since p and q are relatively prime, we can realize any integer as an integer combination of p and q. From this we see that every point of the form s/q, with s odd, lies in the image of M_1. Hence some point of $\mathbf{Z}^2$, above the baseline of $\widehat{\Gamma}(p/q)$, corresponds to the orbit of either $(s/q,1)$ or $(s/q,-1)$. Let the *floor grid* denote the lines of negative slope in the room grid. These lines all have slope $-p/q$. The kth line L_k of the floor grid contains the point

$$\zeta_k = \left(0, \frac{k(p+q)}{2}\right).$$

Modulo translation by V, the point ζ_k is the only lattice point on L_k. Statement 1 of the Hexagrid Theorem says that the edges of Γ incident to ζ_k lie between L_k and L_{k+1} (rather than between L_{k-1} and L_k).

We compute that $M_1(\zeta_k) = k(p+q)$. For all lattice points (m, n) between L_k and L_{k+1}, we therefore have

$$M_1(m, n) \in I_k, \tag{3.5}$$

the interval from Theorem 1.9. This theorem now follows from Equation 3.5, statement 1 of the Hexagrid Theorem, and our remarks about ζ_k. $\square$

Remark: We compare the odd case of Theorem 1.9 to a result in [**K**]. (The even case is similar.) The result in [**K**] is quite general, and so we will specialize it to kites. In this case, a kite is quasirational iff it is rational. The (special case of the) result in [**K**], interpreted in our language, says that every special orbit is contained in one of the intervals J_0, J_1, J_2, ..., where

$$J_a = \bigcup_{i=0}^{p+q-1} I_{ak+i}.$$

The endpoints of the J intervals correspond to *necklace orbits*. A necklace orbit (in our case) is an outer billiards orbit consisting of copies of the kite touching vertex to vertex. Compare Figure 2.1.

Recall that a periodic orbit relative to $K(A)$ is persistent if there exists a nearby and combinatorially identical orbit relative to $K(A')$ for all nearby parameters A'.

Lemma 3.6 *Suppose that $p \in \Xi$ is a periodic point relative to outer billiards on $K(A)$. Then $O_2(p)$ is persistent if and only if the component of the arithmetic graph corresponding to A and p is a closed polygon.*

Proof: Let γ be a the component of interest. By Equation 2.8, we have

$$\Psi^k(p) - p = (2m_k A + 2n_k, 2\epsilon_k), \qquad k = 1, 2, 3, \dots \tag{3.6}$$

Here $m_k, n_k \in \mathbf{Z}$ and $\epsilon_k \in \{-1, 0, 1\}$ and $m_k + n_k + \epsilon_k$ is even. For any given k, the triple (m_k, n_k, ϵ_k) remains the same for small perturbations of the parameter. The point is that a finite amount of combinatorial data determines (m_k, n_k). If γ is a closed polygon, then $(m_k, n_k, \epsilon_k) = (0, 0, 0)$ for some k. But then $\Psi^k(p) = p$ for all parameters near A. Hence $O_2(p)$ is persistent. Conversely, if $O_2(p)$ is persistent then $m_k A' + n_k = 0$ for some k and all A' sufficiently close to A. But this forces $(m_k, n_k) = (0, 0)$. Hence γ is a closed polygon. $\square$

Proof of Theorem 1.10: Lemmas 2.5 and 3.6 combine to prove the even case of Theorem 1.10. Now we establish the odd case. Let p/q be an odd rational. Say that a *suite* is the region between two floors of the room grid. Each suite is partitioned into rooms. Each room has two walls, and each wall has a door in it. From the Hexagrid Theorem, we see that there is an infinite polygonal arc of $\widehat{\Gamma}(p/q)$ that lives in each suite. Let $\Gamma_k(p/q)$ denote the infinite polygonal arc that lives in the kth suite. Here $\Gamma_0(p/q) = \Gamma(p/q)$.

We have just described the infinite family of fleeting components listed in Theorem 1.10. All the other components of $\widehat{\Gamma}(p/q)$ are closed polygons and must be confined to single rooms. The corresponding orbits are persistent, by Lemma 3.6. The already described polygonal arcs use up all the doors.

The point $(m, n) \in \mathbf{Z}^2$ lies on the component of the arithmetic graph corresponding to one of the two orbits $(M(m, n), \pm 1)$. Here M is the fundamental map from Equation 2.9. By the parity result in Equation 2.8, these two points lie on different Ψ-orbits. Therefore each component of $\widehat{\Gamma}$ corresponds to two distinct special orbits. In particular, there are exactly two fleeting orbits U_k^+ and U_k^- contained in the interval I_k, and these correspond to $\Gamma_k(p/q)$. This completes the proof. $\square$

Chapter Four

Period Copying

4.1 INFERIOR AND SUPERIOR SEQUENCES

We discussed inferior and superior sequences in §1.4. Here we say a bit more. Let $p/q \in (0,1)$ be any odd rational. There are unique rationals p_-/q_- and p_+/q_+ such that

$$\frac{p_-}{q_-} < \frac{p}{q} < \frac{p_+}{q_+}, \qquad \max(q_-, q_+) < q, \qquad qp_\pm - pq_\pm = \pm 1. \tag{4.1}$$

See Chapter 17 for more details.

We define the odd rational.

$$\frac{p'}{q'} = \frac{|p_+ - p_-|}{|q_+ - q_-|}, \tag{4.2}$$

where p'/q' is the unique odd rational satisfying the equation

$$q' < q, \qquad |pq' - qp'| = 2. \tag{4.3}$$

We call p'/q' the *inferior predecessor* of p/q, and we write $p'/q' \leftarrow p/q$ or $p/q \to p'/q'$. We can iterate this procedure. Any p/q belongs to a finite chain

$$\frac{1}{1} \leftarrow \frac{p_1}{q_1} \leftarrow \cdots \leftarrow \frac{p_n}{q_n} = \frac{p}{q}. \tag{4.4}$$

Corresponding to this sequence, we define

$$d_k = \mathrm{floor}\left(\frac{q_{k+1}}{2q_k}\right). \tag{4.5}$$

We define the *superior predecessor* of p/q to be p_k/q_k, where k is the largest index such that $d_k \geq 1$. It might happen that the inferior and superior predecessors coincide, and it might not.

For reference, we repeat the example from §1.4. Consider the sequence

$$\mathbf{\frac{1}{1}} \leftarrow \frac{1}{3} \leftarrow \mathbf{\frac{1}{5}} \leftarrow \frac{3}{13} \leftarrow \mathbf{\frac{5}{21}} \leftarrow \frac{13}{55} \leftarrow \mathbf{\frac{21}{89}} \leftarrow \frac{55}{233} \leftarrow \mathbf{\frac{89}{377}} \leftarrow \cdots.$$

$3/13$ has $1/5$ as both a superior and an inferior predecessor. $5/21$ has $3/13$ as an inferior predecessor and $1/5$ as a superior predecessor. The implied limit of this sequence is $\sqrt{5} - 2$, the Penrose kite parameter.

The inferior predecessor construction organizes all the odd rationals into a directed tree of infinite valence. The rational $1/1$ is the terminal node of this tree. The nodes incident to $1/1$ are $1/3$, $3/5$, $5/7$, etc. Figure 4.1 shows part of this tree. The edges are labelled with the d-values from Equation 4.5.

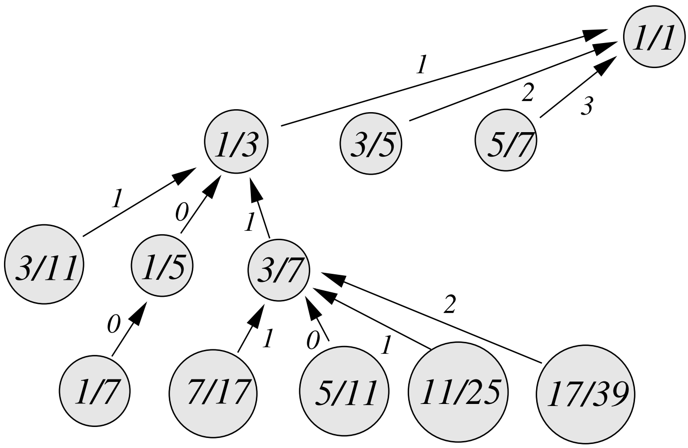

Figure 4.1: Part of the odd tree.

The tree we are drawing has infinite valence at all nodes. With the exception of the top node, $1/1$, all the other nodes have the following structure.

1. There is one incoming arrow labelled 0.

2. There are two incoming arrows labelled k for each $k = 1, 2, 3, \ldots$.

We will establish these results in Part 4 of the book. We will also establish the following result, which identifies certain ends of the tree with the irrationals in $(0, 1)$.

Lemma 4.1 (Superior Sequence) *Let $A \in (0, 1)$ be irrational. There is a unique sequence $\{p_n/q_n\}$ of odd rationals such that*

$$\frac{p_0}{q_0} = \frac{1}{1}, \qquad \frac{p_{n+1}}{q_{n+1}} \to \frac{p_n}{q_n} \quad \forall n, \qquad A = \lim_{n\to\infty} \frac{p_n}{q_n}. \tag{4.6}$$

There are infinitely many indices n such that $2q_n < q_{n+1}$.

We call the sequence $\{p_n/q_n\}$ the *inferior sequence*. We call n a *superior index* if $2q_n < q_{n+1}$. In terms of Equation 4.5, the index n is superior if and only if $d_n \geq 1$. We define the *superior sequence* to be the subsequence indexed by the superior indices. Though there are many inferior and superior sequences containing p_n/q_n, the initial parts of these sequences are determined by p_n/q_n. This comes from the directed tree structure we have already mentioned. The converse result is also true. Any inferior sequence with infinitely many superior terms has an irrational limit. This is a consequence of Lemma 17.4.

Remark: One can also define a similar tree for even rationals. To do this, we modify Equation 4.3 to read $|pq' - qp'| = 1$. For instance, $1/2$ and $2/5$ are related this way. Compare the discussion surrounding Figure 1.5 in the introduction.

4.2 STRONG SEQUENCES

4.2.1 The Main Result

Let A_1 and A_2 be two odd rationals. Let Γ_1 and Γ_2 be the corresponding arithmetic graphs. We fix

$$\epsilon = 1/8. \tag{4.7}$$

This is an arbitrary but convenient choice.

Let $V_1 = (q_1, -p_1)$. Let Γ_1^1 denote the period of Γ_1 connecting $(0,0)$ to V_1 and let Γ_1^{-1} denote the period of Γ_1 connecting $(0,0)$ to $-V_1$. We define

$$\Gamma_1^{1+\epsilon} = \Gamma_1^1 \cup \Big(\Gamma_1 \cap B_{\epsilon q_1}(V_1)\Big), \qquad \Gamma_1^{-1-\epsilon} = \Gamma_1^{-1} \cup \Big(\Gamma_1 \cap B_{\epsilon q_1}(-V_1)\Big). \tag{4.8}$$

We are extending one period of Γ_1 slightly beyond one of its endpoints. Say that a monotone convergent sequence of odd rationals $\{p_n/q_n\}$ is *strong* if it has the following properties.

1. $|A - A_n| < Cq_n^{-2}$ for some universal constant C.
2. If $A_n < A_{n+1}$, then $\Gamma_n^{1+\epsilon} \subset \Gamma_{n+1}^1$.
3. If $A_n > A_{n+1}$, then $\Gamma_n^{-1-\epsilon} \subset \Gamma_{n+1}^{-1}$.

In other words, Γ_{n+1} copies about $1+\epsilon$ periods of Γ_n for every n. As usual, we have set $A_n = p_n/q_n$.

In Part 4 we will prove the following result.

Theorem 4.2 *Any superior sequence has a strong subsequence. In particular, any irrational in* $(0, 1)$ *is the limit of a strong subsequence.*

In the next chapter we will prove that any limit of a strong sequence satisfies the conclusions of the Erratic Orbits Theorem. Thus Theorem 4.2 is one of the key ingredients in the proof of the Erratic Orbits Theorem.

4.2.2 An Easier Result

The proof of Theorem 4.2 is rather involved. It requires all the material in Part 4. It turns out that we can prove a result nearly as strong as the Erratic Orbits Theorem based on the following easier result.

Lemma 4.3 *Let* $A_j = p_j/q_j$ *be odd rationals such that* $|A_1 - A_2| < 1/(2q_1^2)$.

- *If* $A_1 < A_2$, *then* $\Gamma_1^{1+\epsilon} \subset \Gamma_2^1$.
- *If* $A_1 > A_2$, *then* $\Gamma_1^{-1-\epsilon} \subset \Gamma_2^{-1}$.

Here $\epsilon = 1/8$ as above. The proof of Lemma 4.3, given in §17.4 and Chapter 18, requires only a small portion of Part 4.

Let $\Delta_k \subset (0, 1)$ denote the set of irrationals A such that the equation

$$0 < \left| A - \frac{p}{q} \right| < \frac{1}{kq^2}, \qquad p, q \in \mathbf{Z}_{\rm odd} \tag{4.9}$$

holds infinitely often. Lemma 4.3 has the following corollary.

Corollary 4.4 *Every $A \in \Delta_2$ is the limit of a strong sequence.*

Proof: If $A \in \Delta_2$, then there exists a monotone sequence of solutions to Equation 4.9 for $k = 2$. This sequence is strong, by Lemma 4.3. □

Combining the last corollary with our work in the next chapter, we obtain the proof of the Erratic Orbits Theorem for all $A \in \Delta_2$. We close this section with some observations on the size of the sets Δ_k.

Lemma 4.5 *Δ_k has full measure in $(0, 1)$ for any k.*

Proof: Any block of 3 consecutive odd terms $\geq k$ in the continued fraction expansion of A guarantees a solution to Equation 4.9. It follows from the ergodicity of the Gauss map (or the ergodicity of the geodesic flow on the modular surface) that almost every A has infinitely many such blocks. Hence Δ_k has full measure in $(0, 1)$. See [**BKS**] for the relevant ergodic theory. □

It turns out[1] that every irrational in $(0, 1)$ belongs to Δ_1. This result is similar in spirit to Lagrange's famous theorem stating that every irrational A satisfies

$$\left| A - \frac{p}{q} \right| < \frac{1}{\sqrt{5}q^2}$$

infinitely often. Lagrange's theorem does not imply that every irrational lies in Δ_2 because the conditions on Δ_2 involve a parity restriction. In any case, the set Δ_2 seems to be pretty close to all of $(0, 1) - \mathbf{Q}$.

For the interested reader, we sketch the proof of the result we have just mentioned.

Lemma 4.6 $\Delta_1 = (0, 1) - \mathbf{Q}$.

Proof: (Sketch.) Consider the usual horodisk packing in the upper half-plane associated to the modular group. The disk tangent to $\mathbf{R}$ at p/q has diameter $1/q^2$. Remove all horodisks except those based at odd rationals. Dilate each disk (in the Euclidean sense) by a factor of 2 about its tangency point. Observe that the complement of these inflated disks in the hyperbolic plane has infinitely many components. Interpret this result in terms of Δ_1 using the usual connection between the modular horodisk packing and rational approximation. □

Lemma 4.6 plays no role in our proofs, however.

[1] I am grateful to Curt McMullen for pointing this out to me and also for supplying the proof sketched here.

Chapter Five

Proof of the Erratic Orbits Theorem

5.1 PROOF OF STATEMENT 1

In this section we will prove the following result.

Lemma 5.1 *Suppose that A is the limit of a strong sequence $\{A_n\}$. Then statement 1 of the Erratic Orbits Theorem holds for A.*

Statement 1 of the Erratic Orbits Theorem follows from Theorem 4.2 and Lemma 5.1. The reader who is satisfied with the Erratic Orbits Theorem for all $A \in \Delta_2$ can use the much easier Lemma 4.3 in place of Theorem 4.2.

In our proof of Lemma 5.1, we will consider the monotone increasing case. The monotone decreasing case is essentially the same. Let $\epsilon = 1/8$ be as in the definition of strong sequences. Note that our sequence remains strong if we pass to a subsequence. Passing to a subsequence, we arrange that

$$\epsilon q_{n+1} > 10 q_n \tag{5.1}$$

Let $V_n = (q_n, -p_n)$. Define

$$\Gamma_n^2 = \Gamma_n^1 + V_{n+1}, \tag{5.2}$$

Lemma 5.2

$$\Gamma_n^1 \subset \Gamma_{n+1}^1, \qquad \Gamma_n^2 \subset \Gamma_m^1 \qquad \forall m \geq n+2. \tag{5.3}$$

Proof: We have

$$\Gamma_n^{1+\epsilon} \subset \Gamma_{n+1}^1,$$

by definition, and

$$\Gamma_n^1 + V_{n+1} \subset \Gamma_{n+1}$$

because Γ_{n+1} is invariant under translation by V_{n+1}. Our choice of subsequence gives

$$\Gamma_n^1 \subset \Gamma_n^{1+\epsilon} \subset_* B_{10q_n}(0,0) \subset B_{\epsilon q_{n+1}}(0,0) \cap \Gamma_{n+1}. \tag{5.4}$$

The starred containment comes from the Room Lemma. Translating by V_{n+1}, we have

$$\Gamma_n^1 + V_{n+1} \subset B_{\epsilon q_{n+1}}(V_{n+1}) \cap \Gamma_{n+1}^1 \subset \Gamma_{n+1}^{1+\epsilon} \subset \Gamma_{n+2}^1. \tag{5.5}$$

Equation 5.3 follows immediately. $\square$

Figure 5.1 shows schematicaly the sort of binary structure we have set up. In this figure, the notation ij stands for Γ_i^j.

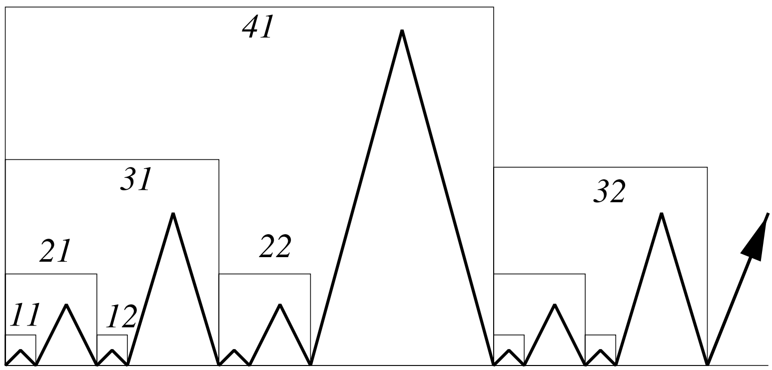

Figure 5.1: Large-scale Cantor set structure.

Figure 5.2 shows a simplified version of Figure 5.1 that retains the structure of interest to us. Below we will analyze this figure carefully.

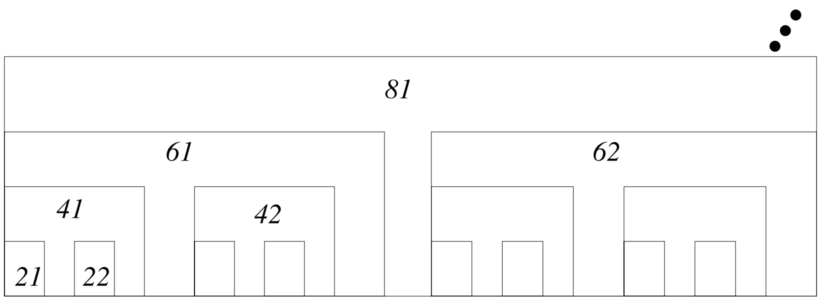

Figure 5.2: Large-scale Cantor set structure simplified.

Corollary 5.3 *The vertex*

$$\omega_n = \omega(\sigma) := \sum_{k=1}^{n-1} \epsilon_k V_{2k+1} \tag{5.6}$$

is a vertex of Γ^1_{2n} *for any binary sequence* $\epsilon_1, \ldots, \epsilon_{n-1}$.

Proof: This follows from Equation 5.3 and induction. □

Let Π denote the set of not-eventually-constant sequences. Given any $\sigma \in \Pi$, we form the sequence of translated baselines and translated graphs

$$L'_n = L_{2n} - \omega_n, \qquad \Gamma'_n = \Gamma^1_{2n} - \omega_n. \tag{5.7}$$

Here ω_n is based on the first $n-1$ terms of σ as in Equation 5.6. The line L_{2n} is the baseline of Γ_{2n}, namely, the line of slope $-A_{2n}$ through the origin.

Lemma 5.4 $\{L'_n\}$ *converges in the Hausdorff topology to a line* L *of slope* $-A$.

Proof: This follows from the fact that there is a uniform bound to the distance from ω_n to L_{2n}. □

Lemma 5.5 $\{\Gamma'_n\}$ *Hausdorff-converges to an open polygonal arc* Γ *that in both directions rises unboundedly far from L and comes arbitrarily close to L.*

Proof: Figure 5.2 shows a pattern of nested rectangles or *boxes*. We now formally define these boxes. Say that the box containing Γ_n^1 is $R_n = R(A_n)$, as in the Room Lemma. We define the 8 smallest boxes in Figure 5.2 as

$$R_2 + \epsilon_1 V_3 + \epsilon_2 V_5 + \epsilon_3 V_7, \qquad \epsilon_j \in \{0, 1\}. \tag{5.8}$$

The larger boxes have similar definitions. We rank each box according to the label of its leftmost translate. The smallest boxes have rank 2; the next-smallest have rank 4; and so on. The following structure emerges.

1. The arc inside a box of rank $2n$ is a translate of Γ_{2n}^1 and has diameter $O(q_{2n})$. This arc contains the bottom corners of the corresponding box and rises up $O(q_{2n})$ units toward the top of the box. This is all a consequence of the Room Lemma and Corollary 5.3.

2. The bottom edge of a box of rank $2n$ lies within $O(1/q_{2n})$ of the bottom edge of the box of rank $2n+2$ that nearly contains it. For the leftmost boxes R_{2n} and R_{2n+2} this result follows from the facts that the bottom edges of these boxes meet at the origin, their slopes differ by $O(1/q_{2n}^2)$, and the shorter edge has length $O(q_{2n})$. Next, since V_{2n+1} is at most $O(1/q_{2n})$ units from the bottom of R_{2n+2}, we get the same result for $R_{2n} + V_{2n+1}$ and R_{2n+2}. The general case now follows from translation.

By construction, the pattern of boxes surrounding ω_n stabilizes when we view any fixed-radius neighborhood of ω_n. More formally, for any R, there is some N such that $m, n > N$ implies that $\Gamma_{2m}^1 \cap B_R(\omega_m)$ is a translate of $\Gamma_{2n}^1 \cap B_R(\omega_n)$. Here we are crucially using the fact that $\sigma \in \Pi$, so that the common pattern of boxes grows both to the left and to the right of the points of interest. Hence the sequence $\{\Gamma'_n\}$ Hausdorff-converges to a limit Γ. The 2 properties above imply that Γ rises unboundedly far from L in both directions.

Now consider the claim about the near approach. Call an arc of Γ'_n *steady* if this same arc also belongs to Γ'_m for $m > n$. By construction, we get the following result. For any k, there is some n such that Γ'_n contains a steady arc of the form $\beta - \omega_n$. Here β is a full period of Γ_k but is contained in Γ_{2n}^1. Some vertex v of β has the form

$$\sum_{j=k}^{n-1} \epsilon_j V_{2j+1}. \tag{5.9}$$

The distance from v to the baseline of Γ_{2n} is $O(1/q_{2k+1})$. But then the distance from $v - \omega_n$ to the baseline of Γ'_n is $O(1/q_{2k+1})$. But $v - \omega_n$ is also a vertex of Γ (by the Hausdorff-convergence) and its distance to the baseline of Γ is also $O(1/q_{2k+1})$. We can choose the arc $\beta - \omega_n$ either to the left or to the right of the origin. Hence both sides of the limit Γ come arbitrarily close to L. $\square$

Now we will recognize Γ as an arithmetic graph corresponding to the parameter A and a certain offset value α. At the same time, L will be the baseline of this graph.

Similar to Equation 2.10, we define

$$M(m,n) = \Big(2Am + 2n, (-1)^{m+n+1}\Big). \tag{5.10}$$

Given $\sigma = \{\epsilon_k\} \in \Pi$, the point

$$\alpha(\sigma) = \bigg(\sum_{k=1}^{\infty} 2\epsilon_k\Big(Aq_{2k+1} - p_{2k+1}\Big), -1\bigg) \tag{5.11}$$

is well defined because the kth term in the series has size $O(1/q_{2k+1})$ and the sequence $\{q_{2k+1}\}$ grows exponentially. The union of such limits, taken over all of Π, contains an uncountable set. Throwing out all points in $2\mathbf{Z}[A] \times \{-1\}$, we still have an uncountable set. Taking σ from this uncountable set, the point $\alpha = \alpha(\sigma)$ we consider has a well defined orbit, by Lemma 2.2.

Lemma 5.6 *Γ is the arithmetic graph of α, and L is the baseline.*

Proof: Let M_{2n} be as in Equation 2.10 for the parameter A_{2n}. Define

$$\alpha_n = M_{2n}(\sigma_n) = \bigg(\sum_{k=1}^{n-1} 2\epsilon_k\Big(A_{2n}q_{2k+1} - p_{2k+1}\Big), -1\bigg). \tag{5.12}$$

There is some constant C such that

$$\Big|(A_{2n}q_{2k+1} - p_{2k+1}) - (Aq_{2k+1} - p_{2k+1})\Big| = q_{2k+1}|A - A_{2n}| < Cq_{2n}^{-1}, \qquad \forall k \le n-1,$$

$$\bigg|\sum_{k=n}^{\infty} 2\epsilon_k(Aq_{2k+1} - p_{2k+1})\bigg| < Cq_{2n}^{-1}.$$

It follows from these estimates and the triangle inequality that

$$|\alpha - \alpha_n| < Cnq_{2n}^{-1}.$$

Hence $\alpha_n \to \alpha$.

By construction, Γ'_n is one period of the arithmetic graph of α_n relative to A_{2n}. The distance that Γ'_n extends from the origin in either direction tends to ∞ with n. By the Continuity Principle, $\{\Gamma'_n\}$ converges to the arithmetic graph of α. But $\{\Gamma'_n\}$ also converges to Γ. Hence Γ is the arithmetic graph of α.

The line $L_{2n} - \omega$ is the baseline for the arithmetic graph Γ'_n that tracks the orbit of α_n. Hence L is the baseline of Γ. $\square$

Combining Lemmas 5.5 and 5.6, we see that α lies on an erratic orbit relative to outer billiards on $K(A)$. But there are uncountably many α to which our argument applies. This proves Lemma 5.1.

5.2 PROOF OF STATEMENT 2

The result we prove in this section shows that statement 1 of the Erratic Orbits Theorem implies statement 2. The reader will see from the next lemma that we do not need the full force of statement 1. We just need the existence of a point sufficiently close to a kite vertex that has a two-sided unbounded orbit.

Lemma 5.7 *Suppose that A is a parameter and $p \in (0,2) \times \{1\}$ has an orbit that is unbounded in both directions. Then all special orbits relative to A are either periodic or unbounded in both directions.*

Proof: We write $p = (2\zeta, 1)$. By hypothesis, $\zeta \in (0,1)$. Suppose that β has an aperiodic orbit that is forward-bounded. (The backward case is similar.) For ease of exposition, we suppose that $\beta \notin 2\mathbf{Z}[A]$, so that all components of the arithmetic graph $\widehat{\Gamma}$ associated to β are well defined. In the case when $\beta \in 2\mathbf{Z}[A]$, we simply apply our argument to a sequence $\{\beta_n\}$ converging to β and invoke the Continuity Principle. Our robust geometric limit argument works the same way with only notational complications.

Let Γ be the component of $\widehat{\Gamma}$ that tracks β. The forward direction Γ_+ remains within a bounded distance of the baseline L of $\widehat{\Gamma}$ and yet is not periodic. Hence Γ_+ travels infinitely far either to the left or to the right. Since L has an irrational slope, we can find a sequence of vertices $\{v_n\}$ of Γ_+ such that the vertical distance from v_n to L converges to $\zeta + N$ for some integer N. Let $w_n = v_n - (0, N)$. Let γ_n be the component of $\widehat{\Gamma}_n$ containing w_n. Note that $M(w_n) \to p$. Here M is as in Equation 2.10.

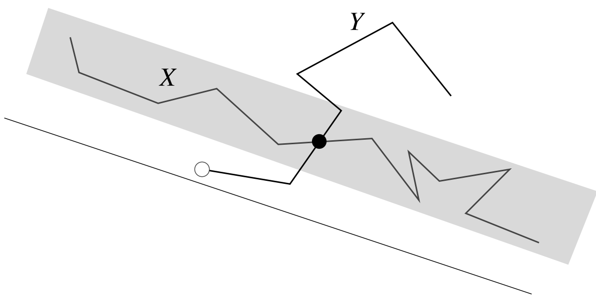

Figure 5.3: The contradiction.

Let T_n be a translation so that $T_n(w_n) = (0,0)$. By compactness, we can choose our sequence so that $\{T_n(\Gamma_+)\}$ converges to an infinite polygonal arc X that remains within a bounded distance of any line parallel to L. By construction, X travels infinitely far both to the left and to the right. At the same time, $\{T_n(\gamma_n)\}$ converges to the arithmetic graph Y of ζ. Here Y starts at $(0,0)$, a point within 1 unit of the baseline $L_\infty = \lim T_n(L)$, and rises unboundedly far from L_∞. Hence Y starts out below X and rises above X, contradicting the Embedding Theorem. Figure 5.3 shows the contradiction. $\square$

5.3 PROOF OF STATEMENT 3

In this section we prove that statement 1 of the Erratic Orbits Theorem implies statement 3. Let A be an irrational parameter for which statement 1 of the Erratic Orbits Theorem holds. Since outer billiards is a piecewise isometry, the set of periodic orbits is open in $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$. We just need to prove that the periodic orbits are dense.

Let $\widehat{\Gamma}$ be an arithmetic graph associated to A such that Γ tracks an erratic orbit. Since A is irrational, we can find a sequence of vertices $\{(m_k, n_k)\}$ of odd parity that converges to the baseline of A. Let γ_k be the component of $\widehat{\Gamma}$ that contains (m_k, n_k). Note that $\gamma_k \neq \Gamma$ because, by Lemma 2.6, Γ contains vertices only of even parity. By the Embedding Theorem, γ_k is trapped underneath Γ. Hence γ_k is a polygon. Let $|\gamma_k|$ denote the maximal distance between a pair of low vertices on γ_k.

Lemma 5.8 *$|\gamma_k| \to \infty$ as $k \to \infty$.*

Proof: By the Rigidity Lemma in §2.7, a very long arc of γ_k, with one endpoint (m_k, n_k), agrees with the Hausdorff limit $\lim_{n\to\infty} \Gamma(p_n/q_n)$. Here $\{p_n/q_n\}$ is an approximating strong sequence. But this limit has vertices within ϵ of the baseline and at least $1/\epsilon$ apart for any $\epsilon > 0$. Our result now follows from Hausdorff continuity. □

Let S_k denote the set of components γ' of $\widehat{\Gamma}$ such that γ' is translation equivalent to γ_k and the corresponding vertices are low. The vertex (m, n) is low if the baseline of $\widehat{\Gamma}$ separates (m, n) and $(m, n-1)$.

Lemma 5.9 *There is some constant N_k so that every point of L is within N_k units of a member of S_k.*

Proof: Say that a lattice point (m, n) is *very low* if it has depth less than $1/100$ (but is still positive.) The polygon γ_k corresponds to a periodic orbit ξ_k. Since ξ_k is periodic, there is an open neighborhood U_k of ξ_k such that all orbits in U_k are combinatorially identical to ξ_k. Let M be a fundamental map associated to $\widehat{\Gamma}$. Then $M^{-1}(U_k)$ is an open strip parallel to L. Since L has an irrational slope, there is some constant N_k so that every point of L is within N_k of some point of $M^{-1}(U_k) \cap \mathbf{Z}^2$. But the components of $\widehat{\Gamma}$ containing these points are translation-equivalent to γ_k. Choosing U_k small enough, we can guarantee that the translations taking γ_k to the other components carry the very low vertices of γ_k to low vertices. □

Given two polygonal components X and Y of $\widehat{\Gamma}$, we write $X \bowtie Y$ if one low vertex of Y lies to the left of X and one low vertex of Y lies to the right of X. See Figure 5.4. In this case, X is trapped underneath Y, by the Embedding Theorem.

Now we pass to a subsequence so that

$$|\gamma_{k+1}| > 10(N_k + |\gamma_k|). \tag{5.13}$$

Equation 5.13 has the following consequence. For any integer N, we can find components γ_j of S_j, for $j = N, ..., 2N$, such $\gamma_N \bowtie \cdots \bowtie \gamma_{2N}$. Let L_N denote the portion of L between the two distinguished low points of γ_N. Let Λ_N denote the set of lattice points within N units of L_N. The set Λ_N is a parallelogram whose base is L_N, a segment whose length tends to ∞ with N. The height of Λ_N tends to ∞ as well.

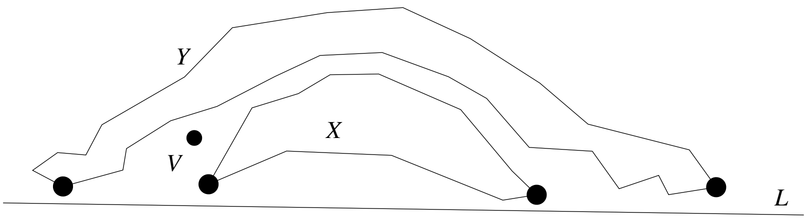

Figure 5.4: One polygon overlying another.

Lemma 5.10 *The set $M(\mathbf{Z}^2 \cap \Lambda_N)$ consists entirely of periodic orbits.*

Proof: Let V be a vertical ray whose x-coordinate is an integer. If V starts out on L_n, then V must travel upward at least N units before escaping from underneath γ_{2N}. This is an application of the pidgeonhole principle. The point is that V must intersect each γ_j, for $j = N, ..., 2N$, in a different lattice point. Hence any point of Λ_N is trapped beneath γ_{2N}. $\square$

Given the facts that both the base and height of Λ_N are growing unboundedly and the fact that A is an irrational parameter, the union $\bigcup_{N=1}^{\infty} M(\Lambda_N \cap \mathbf{Z}^2)$ is dense in $\mathbf{R}_+$. Hence the set of periodic orbits starting in $\mathbf{R}_+ \times \{-1, 1\}$ is dense in the set of all special orbits. Our proof of the Pinwheel Lemma in Part 2 shows that every special orbit eventually lands in $\mathbf{R}_+ \times \{-1, 1\}$. Hence the set of periodic special orbits is dense in $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$.

This page intentionally left blank

Part 2. The Master Picture Theorem

In this part of the book, we will state and prove the Master Picture Theorem. All the auxilliary theorems left over from Part 1 rely on this central result. Here is an overview of the material.

- In Chapter 6, we will state the Master Picture Theorem. Roughly, the Master Picture Theorem says that the structure of the return map Ψ is determined by a pair of maps into a flat 3-torus $\mathbf{R}^3/\Lambda$ together with a partition of $\mathbf{R}^3/\Lambda$ into polyhedra. Here Λ is a certain 3-dimensional lattice that depends on the parameter. We will consider the Master Picture Theorem from several points of view, giving lots of example calculations. The remainder of Part 2 is devoted to the proof of the Master Picture Theorem. The reader who is keen to see the applications can skip directly from Chapter 6 to Part 3.

- In Chapter 7, we will prove the Pinwheel Lemma, a key technical step along the way to the proof of the Master Picture Theorem. The Pinwheel Lemma states that we can factor the return map Ψ into a composition of 8 simpler maps, which we call *strip maps*. A strip map is a very simple map from the plane into an infinite strip.

- In Chapter 8, we prove the Torus Lemma, another key result. The Torus Lemma implies that there exists some partition of the torus into open regions such that the regions determine the structure of the arithmetic graph. The Torus Lemma reduces the Master Picture Theorem to a rough determination of the singular set. The singular set is the (closure of the) set of points in the torus corresponding to points where the return map is not defined.

- In Chapter 9, we verify, with the aid of symbolic manipulation, certain functional identities that arise in connection with the Torus Lemma. These functional identities are the basis for our analysis of the singular set.

- In Chapter 10, we combine the Torus Lemma with the functional identities to prove the Master Picture Theorem.

Billiard King has a module that shows the torus partition and demonstrates the Master Picture Theorem. A separate module on Billiard King shows all the sets involved in the proof of the Pinwheel Lemma. We hope that the material in Chapters 6 and 7 stands on its own, but we strongly recommend that the reader use Billiard King as a guide to this material.

This page intentionally left blank

Chapter Six

The Master Picture Theorem

6.1 COARSE FORMULATION

Recall that $\Xi = \mathbf{R}_+ \times \{-1, 1\}$. We distinguish two special subsets of Ξ.

$$\Xi_+ = \bigcup_{k=0}^{\infty} (2k, 2k+2) \times \{(-1)^k\}, \qquad \Xi_- = \bigcup_{k=1}^{\infty} (2k, 2k+2) \times \{(-1)^{k-1}\}. \tag{6.1}$$

Each set is an infinite disconnected union of open intervals of length 2. The reflection in the x-axis interchanges Ξ_+ and Ξ_-. The union $\Xi_+ \cup \Xi_-$ partitions the set $(\mathbf{R}_+ - 2\mathbf{Z}) \times \{\pm 1\}$.

Define

$$R_A = [0, 1 + A] \times [0, 1 + A] \times [0, 1]\,. \tag{6.2}$$

R_A is a fundamental domain for the action of a certain lattice Λ_A. This lattice is defined by the following matrix.

$$\Lambda_A = \begin{bmatrix} 1 + A & 1 - A & -1 \\ 0 & 1 + A & -1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{Z}^3. \tag{6.3}$$

We mean to say that Λ_A is the $\mathbf{Z}$-span of the column vectors of the above matrix.

We define maps

$$\mu_\pm \colon \Xi_\pm \to R_A \tag{6.4}$$

by the equations

$$\mu_\pm(t, *) = \left(\frac{t-1}{2}, \frac{t+1}{2}, \frac{t}{2}\right) \pm \left(\frac{1}{2}, \frac{1}{2}, 0\right) \quad \text{mod } \Lambda. \tag{6.5}$$

The maps depend on only the first coordinate. In each case, we mean to map t into $\mathbf{R}^3$ and then use the action of Λ_A to move the image into R_A. It might happen that there is not a unique representative in R_A. (There is an issue with boundary points, as is usual with fundamental domains.) However, if $t \not\in 2\mathbf{Z}\,[A]$, this situation does not occur. The maps μ_+ and μ_- are locally affine.

Here is a coarse formulation of the Master Picture Theorem. We will state the entire result in terms of $(+)$, with the understanding that the same statement holds with $(-)$ replacing $(+)$ everywhere. Let $\Psi \colon \Xi \to \Xi$ be the first return map.

Theorem 6.1 *For each parameter A, there is a partition $(\mathcal{P}_A)_+$ of R_A into finitely many convex polyhedra. If Ψ is defined on $\xi_1, \xi_2 \in \Xi_+$ and $\mu_+(\xi_1)$ and $\mu_+(\xi_2)$ lie in the same open polyhedron of $(\mathcal{P}_A)_+$, then $\Psi(\xi_1) - \xi_1 = \Psi(\xi_2) - \xi_2$.*

6.2 THE WALLS OF THE PARTITIONS

In order to make Theorem 6.1 precise, we need to describe the nature of the partitions $(\mathcal{P}_A)_\pm$ and also the rule by which the polyhedron in the partition determines the vector $\Psi(\xi) - \xi$. We will make several passes through the description, adding a bit more detail each time.

The polyhedra of $(\mathcal{P}_A)_\pm$ are cut out by the following 4 families of planes.

- $\{x = t\}$ for $t = 0, A, 1, 1 + A$.
- $\{y = t\}$ for $t = 0, A, 1, 1 + A$.
- $\{z = t\}$ for $t = 0, A, 1 - A, 1$.
- $\{x + y - z = t\}$ for $t = -1 + A, A, 1 + A, 2 + A$.

As a first approximation, we say that the connected components of the complement of the above planes are the polyhedra in the partition. Actually, the best statement is that the polyhedra in the partition are certain convex unions of these components. This is to say that the actual partition into polyhedra is somewhat simpler than what one would get just by taking the complementary regions we are discussing. We will consider the best version at the very end of the chapter.

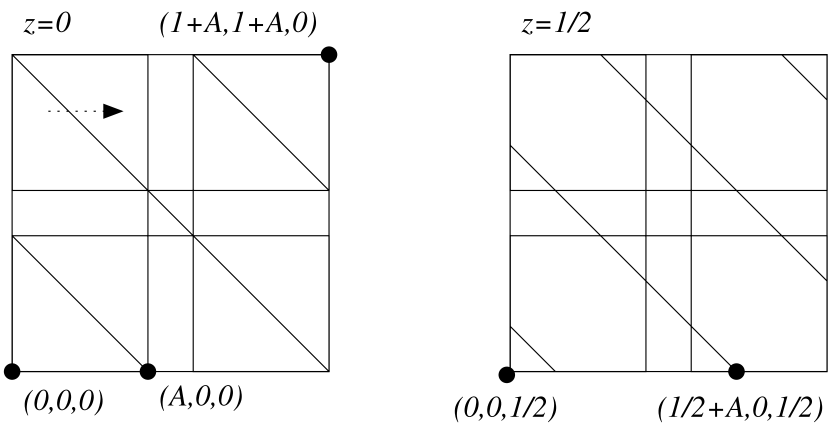

Figure 6.1: Two slices of the partition for $A = 2/3$.

Figure 6.1 shows two slices of the partition for the parameter $A = 2/3$. We have sliced the figure at $z = 0$ and $z = 1/2$ and we have labelled several points just to make the coordinate system more clear. The arrow in indicates the "motion" the diagonal lines would make if we increased the z-coordinate, showing a kind of movie of the partition.

6.3 THE PARTITIONS

For each parameter A we get a solid body R_A partitioned into polyhedra. We can put all these pieces together into a single master picture. We define

$$R = \bigcup_{A\in(0,1)} \big(R_A \times \{A\}\big) \subset \mathbf{R}^4. \tag{6.6}$$

Each 2-plane family discussed above gives rise to a hyperplane family in $\mathbf{R}^4$. These hyperplane families are now all defined over $\mathbf{Z}$ because the variable A is just the 4th coordinate of $\mathbf{R}^4$ in our current scheme. Given that we have two maps μ_+ and μ_-, it is useful for us to consider two identical copies R_+ and R_- of R.

We have a fibration $f\colon \mathbf{R}^4 \to \mathbf{R}^2$ given by

$$f(x, y, z, A) = (z, A). \tag{6.7}$$

This fibration in turn gives a fibration of R over the unit square $B = (0, 1)^2$. Figure 6.1 shows the fiber $f^{-1}(3/2, 1/2)$. The base space B is partitioned into 4 regions, as seen in Figure 6.2.

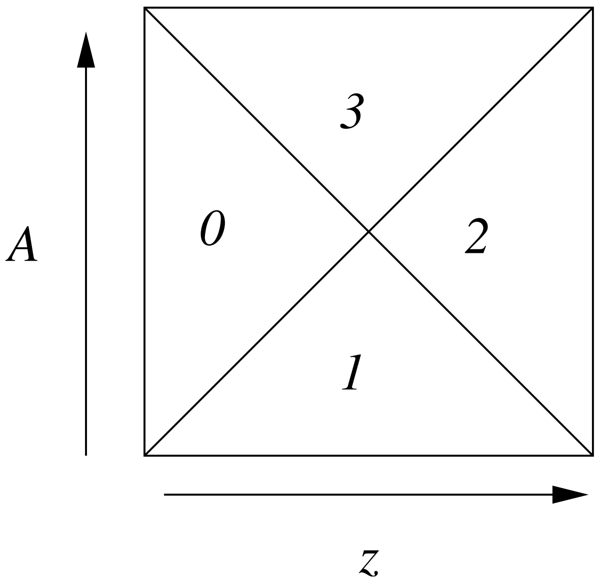

Figure 6.2: The partition of the base space.

All the fibers above the same open region in the base space have the same combinatorial structure. Figure 6.3 shows precisely how the partition assigns the value of the return map. Given a point $\xi \in \Xi_+$, we have a pair of integers $(\epsilon_1^+(\xi), \epsilon_2^+(\xi))$ such that

$$\Psi(\xi) - \xi = 2(\epsilon_1^+, \epsilon_2^+, *). \tag{6.8}$$

The second coordinate, ± 2, is determined by the parity relation in Equation 2.8. Similarily, we have $(\epsilon_1^-, \epsilon_2^-)$ for $\xi \in \Xi_-$.

Figure 6.3 shows a schematic picture of R. For each of the 4 open triangles in the base, we have drawn a cluster of 4 copies of a representative fiber over that triangle. The jth column of each cluster determines the value of $\epsilon_j^{\pm}$. The first row of each cluster determines ϵ_j^{+}, and the second row determines ϵ_j^{-}. Light shading indicates a value of $+1$. Dark shading indicates a value of -1. No shading indicates a value of zero.

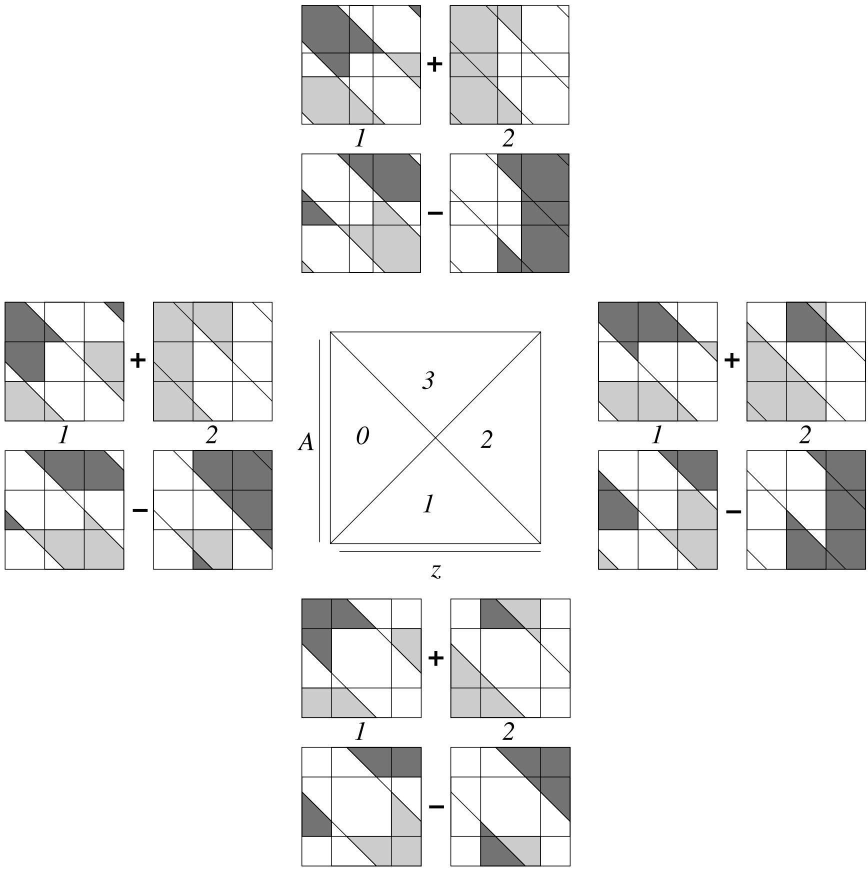

Figure 6.3: The decorated fibers.

Given a generic point $\xi \in \Xi_{\pm}$, the image $\mu_{\pm}(\xi)$ lies in some fiber. We then use the shading scheme to determine $\epsilon_j^{\pm}(\xi)$ for $j = 1, 2$. (See below for examples.) Theorem 6.1, together with the description in this section, constitutes the Master Picture Theorem. In §6.9 we explain with more traditional formulas how to compute these values.

Remark: The hard work in the proof of the Master Picture Theorem is showing that Theorem 6.1 holds with respect to the partition we have defined. Once we know this, a short finite experiment will determine the shading in Figure 6.3.

6.4 A TYPICAL EXAMPLE

Here we will show the Master Picture Theorem in action. We will explain it determines the local structure of the arithmetic graph $\Gamma(3/5)$ at the point $(4, 2)$. Let M be the fundamental map associated to

$$A = 3/5; \qquad A = 1/10.$$

We compute

$$M(4,2) = \big((8)(3/5) + (4) + (1/5), (-1)^{4+2+1}\big) = (9, -1) \in \Xi_-.$$

The point $\mu_-(9, -1)$ determines the forward direction and the point $\mu_+(9, 1)$ determines the backward direction. (Reflection in the x-axis conjugates Ψ to its inverse.)

We compute

$$\mu_+(9,1) = \left(\frac{9}{2}, \frac{11}{2}, \frac{9}{2}\right) \equiv \left(\frac{1}{10}, \frac{3}{2}, \frac{1}{2}\right) \bmod \Lambda,$$

$$\mu_-(9,-1) = \left(\frac{7}{2}, \frac{9}{2}, \frac{9}{2}\right) \equiv \left(\frac{7}{10}, \frac{1}{2}, \frac{1}{2}\right) \bmod \Lambda.$$

In §6.6 we will explain algorithmically how to make these computations. We have $(z, A) = (1/2, 3/5)$. There we need to look at cluster 3, the cluster of fibers above region 3 in the base. Here is the plot of the two points in the relevant fiber. When we look up the regions in Figure 6.3, we find that $(\epsilon_1^+, \epsilon_2^+) = (-1, 1)$ and $(\epsilon_1^-, \epsilon_2^-) = (1, 0)$. The bottom right of Figure 6.4 shows the corresponding local structure for the arithmetic graph.

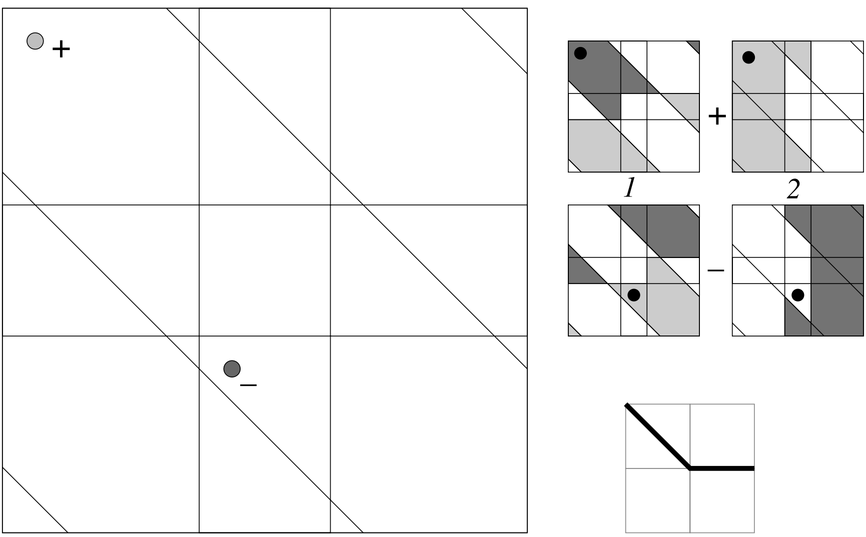

Figure 6.4: Points in the fiber over region 3.

6.5 A SINGULAR EXAMPLE

Sometimes it is an annoyance to deal with the tiny positive constant α that arises in the definition of the fundamental map. In this section we will explain an alternate method for applying the Master Picture Theorem. One situation where this alternate approach proves useful is when we need to deal with the fibers at $z = \alpha$. We much prefer to draw the fibers at $z = 0$ because they do not contain any tiny polygonal regions. All the pieces of the partition can be drawn cleanly. However, in order to make sense of the Master Picture Theorem, we need to slightly redefine how the partition defines the return map.

We define the *lower boundary* of a polyhedron $P \subset \mathbf{R}^3$ as the portion $S \subset \partial P$ such that $x \in S$ implies that $x + \epsilon(1, 1, 1) \in S$ for sufficiently small $\epsilon > 0$. Let $\underline{P}$ denote the union of the interior of P with its lower boundary. When α is sufficiently small, we can set $\alpha = 0$ and determine the return pair using the polyhedra $\underline{P}$ in place of the interior of P, which we used above. In practice, we will use this method when A is rational. In this case, α will always be small enough for our purposes.

We can explain the alternate method in terms of the slices we have drawn above. We redefine the polygonal regions to include their *lower* edges. A lower edge is an edge first encountered by a line of slope 1. Figure 6.5 shows what we have in mind.

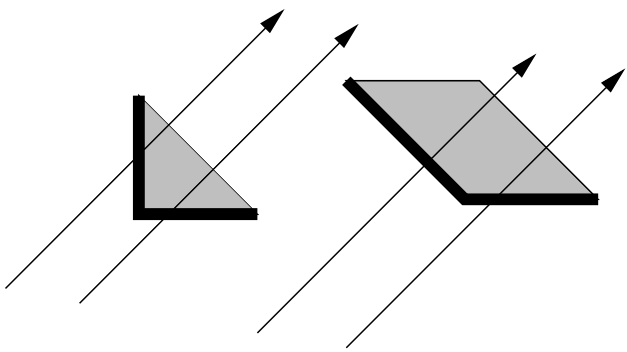

Figure 6.5: Polygons with their lower boundaries included.

We then set $\alpha = 0$ and determine the relevant edges of the arithmetic graph by which *lower-bordered* polygon contains our points. If $z \in \{0, A, 1 - A\}$, then we think of the fiber at z as being the geometric limit of the fibers at $z + \epsilon$ for $\epsilon > 0$. That is, we take a right-sided limit of the figures. When z is not one of these special values, there is no need to do this, for the fiber is completely defined already.

We illustrate our approach with the example $A = 3/5$ and $(m, n) = (0, 8)$. We compute that $t = 8 + \alpha$ in this case. The relevant slices are the ones we get by setting $z = \alpha$. We deal with this by setting $\alpha = 0$ and computing

$$\mu_+(16, 1) = (8, 9, 8) \equiv (4/5, 1, 0) \mod \Lambda$$

$$\mu_-(16, -1) = (7, 8, 8) \equiv (0, 7/5, 0) \mod \Lambda.$$

Figure 6.6 shows the relevant fibers. The bottom right of Figure 6.6 shows the local structure of the arithmetic graph. For instance, $(\epsilon_1^+, \epsilon_2^+) = (0, 1)$.

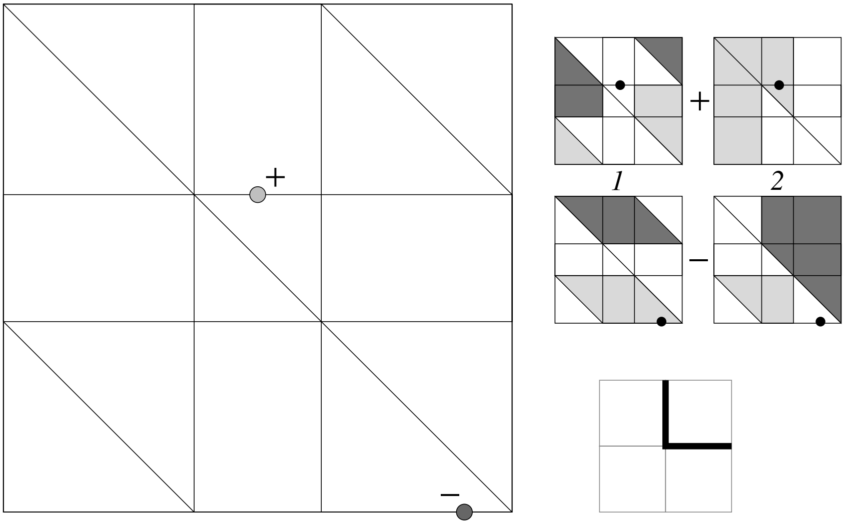

Figure 6.6: Points in the fiber.

The only place where we need to use our special definition of a lower-bordered polygon is for the point in the lower left fiber. This fiber determines the x-coordinate of the edge corresponding to μ_{-}. In this case, we include the point in the lightly shaded parallelogram because the point lies in the lower border of this parallelogram.

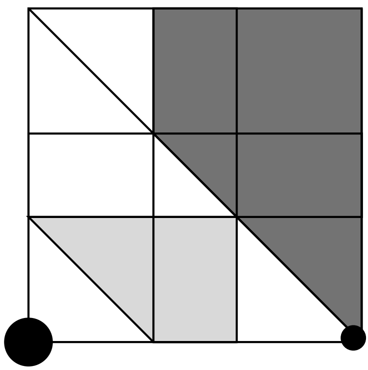

Figure 6.7: An exceptional case.

There is one exception to our construction that requires an explanation. Referring to the lower right fiber, suppose that the bottom point actually is the bottom right vertex as shown in Figure 6.7. In this case, the point is simultaneously the bottom left vertex, and we make the definition using the bottom left vertex. The underlying reason is that a tiny push along the line of slope 1 would move the point into the region on the left. Actually, this case is not really an exception if we think of the left and right hand sides of the fiber as being identified.

6.6 THE REDUCTION ALGORITHM

Let A be a parameter and let α be an offset value. Let M be the fundamental map associated to the pair (A, α), as in Equation 2.9. We define

$$M_+ = \mu_+ \circ M, \qquad M_- = \mu_- \circ \rho \circ M. \tag{6.9}$$

Here $\mu_\pm$ is as in Equation 6.5 and ρ is the reflection in the x-axis. The domain of $\mu_\pm$ is $\Xi_\pm$, the set from Equation 6.1. Note that μ_+ and μ_- depend on only the first coordinate, and this first coordinate is not changed by ρ. The map ρ is present mainly for bookkeeping purposes because $\rho(\Xi_+) = \Xi_-$.

Given a point $p \in \mathbf{Z}^2$, the polyhedron of R_+ containing $M_+(p)$ determines the forward edge of $\widehat{\Gamma}$ incident to p, and the polyhedron of R_- containing $M_-(p)$ determines the backward edge of $\widehat{\Gamma}$ incident to p. Concretely, we have

$$M_+(m,n) = (s, s+1, s) \mod \Lambda,$$

$$M_-(m,n) = (s-1, s, s) \mod \Lambda,$$

$$s = Am + n + \alpha. \tag{6.10}$$

Let $(m,n) \in \mathbf{Z}^2$ be a point above the baseline of $\Gamma_\alpha(A)$. Here we describe how to compute the points

$$\mu_\pm(M_\alpha(m,n)).$$

This algorithm will be important when we prove the Diophantine Lemma in Part 4.

1. Let $z = Am + n + \alpha$.
2. Let $Z = \text{floor}(z)$.
3. Let $y = z + Z$.
4. Let $Y = \text{floor}(y/(1+A))$.
5. Let $x = y - Y(1-A) - 1$.
6. Let $X = \text{floor}(x/(1+A))$.

We then have

$$\mu_-(M_\alpha(m,n)) = \begin{pmatrix} x - (1+A)X \\ y - (1+A)Y \\ z - Z \end{pmatrix}. \tag{6.11}$$

The description of μ_+ is identical except that the third step above is replaced by

$$y = z + Z + 1. \tag{6.12}$$

All this algorithm does is use the lattice Λ_A to move the point (x, y, z) into the fundamental domain R_A.

6.7 THE INTEGRAL STRUCTURE

Let **Aff** denote the group of affine automorphisms of $\mathbf{R}^4$. We define a discrete affine group action $\Lambda \subset \mathbf{Aff}$ on the infinite slab

$$\widehat{R} = \mathbf{R}^3 \times (0, 1). \tag{6.13}$$

The group Λ is generated by the 3 maps $\gamma_1, \gamma_2, \gamma_3$. Here γ_j acts on the first 3 coordinates as translation by the jth column of the matrix Λ_A, and on the 4th coordinate as the identity. We think of the A-variable as the 4th coordinate. $\gamma_1, \gamma_2, \gamma_3$ map the column vector $(x, y, z, A)^t$, respectively, to

$$\begin{bmatrix} x+1+A \\ y \\ z \\ A \end{bmatrix}, \quad \begin{bmatrix} x+1-A \\ y+1+A \\ z \\ A \end{bmatrix}, \quad \begin{bmatrix} x-1 \\ y-1 \\ z+1 \\ A \end{bmatrix}. \tag{6.14}$$

The quotient $\widehat{R}/\Lambda$ is naturally a fiber bundle over $(0, 1)$. Each fiber $(\mathbf{R}^3 \times \{A\})/\Lambda$ is isomorphic to $\mathbf{R}^3/\Lambda_A$. The region R, from Equation 6.6, is a fundamental domain for the action of Λ. Explicitly, the 16 vertices of R are

$$(\epsilon_1, \epsilon_2, \epsilon_3, 0), \qquad (2\epsilon_1, 2\epsilon_2, \epsilon_3, 1), \qquad \epsilon_1, \epsilon_2, \epsilon_3 \in \{0, 1\}. \tag{6.15}$$

Implicit in Figure 6.3 is the statement that the regions R_+ and R_- are partitioned into smaller convex polytopes. The partition here is defined by the 4 families of hyperplanes discussed above. For each pair $(\epsilon_1, \epsilon_2) \in \{-1, 0, 1\}$, let $R_+(\epsilon_1, \epsilon_2)$ denote the closure of the union of regions that assign (ϵ_1, ϵ_2). It turns out that $R_+(\epsilon_1, \epsilon_2)$ is a finite union of convex integral polytopes. There are 14 such polytopes, and they give an integral partition of R_+. Here we list the 14 polytopes. In each case, we list the vertices followed by the pair (ϵ_1, ϵ_2) that the polytope determines.

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\begin{bmatrix}0\\0\\1\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\end{bmatrix}\begin{bmatrix}0\\1\\1\\1\end{bmatrix}\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}\begin{bmatrix}1\\0\\1\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad (1, 1),$$

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\end{bmatrix}\begin{bmatrix}0\\2\\0\\1\end{bmatrix}\begin{bmatrix}0\\2\\1\\1\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}1\\2\\1\\1\end{bmatrix} \quad (-1, 1),$$

$$\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}1\\2\\1\\1\end{bmatrix}\begin{bmatrix}2\\1\\1\\1\end{bmatrix} \quad (-1, -1),$$

$$\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0\\2\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\2\\0\\1\end{bmatrix}\begin{bmatrix}1\\2\\1\\1\end{bmatrix} \quad (0, 1),$$

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}\begin{bmatrix}0\\1\\1\\1\end{bmatrix}\begin{bmatrix}0\\2\\1\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad (0, 1),$$

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\quad (0,1),$$

$$\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\begin{bmatrix}0\\0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\\1\end{bmatrix}\begin{bmatrix}0\\1\\1\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\\1\end{bmatrix}\quad (0,1),$$

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}\begin{bmatrix}0\\2\\1\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}1\\2\\1\\1\end{bmatrix}\quad (-1,0),$$

$$\begin{bmatrix}1\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\2\\0\\1\end{bmatrix}\begin{bmatrix}2\\1\\0\\1\end{bmatrix}\begin{bmatrix}2\\2\\0\\1\end{bmatrix}\begin{bmatrix}2\\2\\1\\1\end{bmatrix}\quad (-1,0),$$

$$\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}1\\2\\1\\1\end{bmatrix}\begin{bmatrix}2\\1\\1\\1\end{bmatrix}\quad (-1,0),$$

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\0\end{bmatrix}\begin{bmatrix}2\\0\\0\\1\end{bmatrix}\begin{bmatrix}2\\1\\0\\1\end{bmatrix}\begin{bmatrix}2\\1\\1\\1\end{bmatrix}\quad (1,0),$$

$$\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}\begin{bmatrix}1\\0\\1\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}2\\0\\1\\1\end{bmatrix}\quad (1,0),$$

$$\begin{bmatrix}1\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\2\\0\\1\end{bmatrix}\begin{bmatrix}1\\2\\1\\1\end{bmatrix}\begin{bmatrix}2\\1\\0\\1\end{bmatrix}\begin{bmatrix}2\\1\\1\\1\end{bmatrix}\begin{bmatrix}2\\2\\1\\1\end{bmatrix}\quad (0,0),$$

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}2\\0\\0\\1\end{bmatrix}\begin{bmatrix}2\\0\\1\\1\end{bmatrix}\begin{bmatrix}2\\1\\1\\1\end{bmatrix}\quad (0,0).$$

Let $\iota\colon R_+ \to R_-$ be given by the map

$$\iota(x,y,z,A) = (1+A-x, 1+A-y, 1-z, A). \tag{6.16}$$

Geometrically, ι is a reflection in the 1-dimensional line. We have the general equation

$$R_-(-\epsilon_1, -\epsilon_2) = \iota(R_+(\epsilon_1, \epsilon_2)). \tag{6.17}$$

Thus the partition of R_- is a mirror image of the partition of R_+. We can use the action of Λ to extend our partitions to give tilings of $\widehat{R}$ by convex integer polytopes. This tiling is our "master picture."

6.8 CALCULATING WITH THE POLYTOPES

We will illustrate a calculation with the polytopes we have listed. Let ι and γ_2 be the maps from Equation 6.7. The region $R_+(0,0)$ consists of two polygons P_1 and P_2. These are the last two listed above. We will show that

$$\iota(P_2) + (1,1,0,0) = \gamma_2(P_2).$$

As above, the coordinates for P_2 are

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}2\\0\\0\\1\end{bmatrix}\begin{bmatrix}2\\0\\1\\1\end{bmatrix}\begin{bmatrix}2\\1\\1\\1\end{bmatrix}.$$

Recall that $\iota(x,y,z,A) = (1+A-x, 1+A-y, 1-z, A)$. For example, $\iota(0,0,0,0) = (1,1,1,0)$. The coordinates for $\iota(P_2)$ are

$$\begin{bmatrix}1\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\2\\1\\1\end{bmatrix}\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}1\\1\\0\\1\end{bmatrix}\begin{bmatrix}0\\2\\1\\1\end{bmatrix}\begin{bmatrix}0\\2\\0\\1\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\end{bmatrix}.$$

The coordinates for $\iota(P_2) + (1,1,0,0)$ are

$$\begin{bmatrix}2\\2\\1\\0\end{bmatrix}\begin{bmatrix}2\\1\\1\\0\end{bmatrix}\begin{bmatrix}2\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\2\\1\\0\end{bmatrix}\begin{bmatrix}2\\3\\1\\1\end{bmatrix}\begin{bmatrix}1\\2\\0\\0\end{bmatrix}\begin{bmatrix}2\\2\\1\\1\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\end{bmatrix}\begin{bmatrix}2\\2\\0\\1\end{bmatrix}\begin{bmatrix}1\\3\\1\\1\end{bmatrix}\begin{bmatrix}1\\3\\0\\1\end{bmatrix}\begin{bmatrix}1\\2\\0\\1\end{bmatrix}.$$

We have $\gamma_2(x,y,z,A) = (x+1-A, y+1+A, z, A)$. For instance, we compute that $\gamma_2(0,0,0,0) = (1,1,0,0)$. The coordinates for $\gamma(P_2)$ are

$$\begin{bmatrix}1\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\2\\0\\0\end{bmatrix}\begin{bmatrix}1\\2\\1\\0\end{bmatrix}\begin{bmatrix}2\\1\\0\\0\end{bmatrix}\begin{bmatrix}1\\2\\0\\1\end{bmatrix}\begin{bmatrix}2\\1\\1\\0\end{bmatrix}\begin{bmatrix}1\\3\\0\\1\end{bmatrix}\begin{bmatrix}2\\2\\1\\0\end{bmatrix}\begin{bmatrix}1\\3\\1\\1\end{bmatrix}\begin{bmatrix}2\\2\\0\\1\end{bmatrix}\begin{bmatrix}2\\2\\1\\1\end{bmatrix}\begin{bmatrix}2\\3\\1\\1\end{bmatrix}.$$

These are the same vectors as listed for $\iota(P_2) + (1,1,0,0)$, but in a different order.

Finally, we illustrate how the general form of the integral partition can justify numerical calculations. Consider the phase portrait described in Figure 2.5. Consider the two rectangles

$$Q_+ = \{(t, t+1, t) |\ t \in (0,1)\} \times [0,1],$$

$$Q_- = \{(t-1, t, t) |\ t \in (0,1)\} \times [0,1].$$

Allow $Q_\pm$ to intersect the polytope $R_\pm$. These intersections partition Q_+ and Q_- into a small finite number of polygons. The partition of $Q_\pm$ tells the behavior of $\Psi^\pm$ on points of $(0,2) \times \{1\}$. By symmetry, the partition of $Q_\mp$ tells the behavior of $\Psi^\pm$ on $(0,2) \times \{-1\}$. The partition of $Q_\pm$ gives us the information needed to build Figure 2.5. Given the simplicity of the partitions involved, we can determine the figure just by plotting (say) 10,000 fairly dense points in the rectangles. This is what we did.

6.9 COMPUTING THE PARTITION

Here we explain how Billiard King implements the Master Picture Theorem. We cannot imagine that a person would want to do this by hand, but it seems worth explaining what the computer actually does.

6.9.1 Step 1

Suppose $(a, b, c) \in R_A$ lies in the range of μ_+ or μ_-. Now we describe how to attach a 5-tuple $(n_0, ..., n_4)$ to (a, b, c).

- Determining n_0:
  - If we are interested in μ_+, then $n_0 = 0$.
  - If we are interested in μ_-, then $n_0 = 1$.
- Determining n_1:
  - If $c < A$ and $c < 1 - A$, then $n_1 = 0$.
  - If $c > A$ and $c < 1 - A$, then $n_1 = 1$.
  - If $c > A$ and $c > 1 - A$, then $n_1 = 2$.
  - If $c < A$ and $c > 1 - A$, then $n_1 = 3$.
- Determining n_2:
  - If $a \in (0, A)$, then $n_2 = 0$.
  - If $a \in (A, 1)$, then $n_2 = 1$.
  - If $a \in (1, 1 + A)$, then $n_2 = 2$.
- Determining n_3:
  - If $b \in (0, A)$, then $n_3 = 0$.
  - If $b \in (A, 1)$, then $n_3 = 1$.
  - If $b \in (1, 1 + A)$, then $n_3 = 2$.
- Determining n_4:
  - Let $t = a + b - c$.
  - Let $n_4 = \text{floor}(t - A)$.

Notice that each 5-tuple $(n_0, ..., n_4)$ corresponds to a (possibly empty) convex polyhedron in R_A. The polyhedron does not depend on n_0. It turns out that this polyhedron is empty unless $n_4 \in \{-2, -1, 0, 1, 2\}$.

6.9.2 Step 2

Let $n = (n_0, ..., n_4)$. We now describe two functions $\epsilon_1(n) \in \{-1, 0, 1\}$ and $\epsilon_2(n) \in \{-1, 0, 1\}$.

Here is the definition of $\epsilon_1(n)$.

- If $n_0 + n_4$ is even, then
  - If $n_2 + n_3 = 4$ or $x_2 < x_3$ set $\epsilon_1(n) = -1$.
- If $n_0 + n_4$ is odd, then
  - If $n_2 + n_3 = 0$ or $x_2 > x_3$, set $\epsilon_1(n) = +1$.
- Otherwise, set $\epsilon_1(n) = 0$.

Here is the definition of $\epsilon_2(n)$.

- If $n_0 = 0$ and $n_1 \in \{3, 0\}$, then
  - If $n_2 = 0$, let $\epsilon_2(n) = 1$.
  - If $n_2 = 1$, and $n_4 \neq 0$ let $\epsilon_2(n) = 1$.
- If $n_0 = 1$ and $n_1 \in \{0, 1\}$, then
  - if $n_2 > 0$ and $n_4 \neq 0$, let $\epsilon_2(n) = -1$.
  - If $n_2 < 2$ and $n_3 = 0$ and $n_4 = 0$, let $\epsilon_2(n) = 1$.
- If $n_0 = 0$ and $n_1 \in \{1, 2\}$, then
  - If $n_2 < 2$ and $n_4 \neq 0$, let $\epsilon_2(n) = 1$.
  - If $n_2 > 0$ and $n_3 = 2$ and $n_4 = 0$, let $\epsilon_2(n) = -1$.
- If $n_0 = 1$ and $n_1 \in \{2, 3\}$, then
  - If $n_2 = 2$, let $\epsilon_2(n) = -1$.
  - If $n_2 = 1$ and $n_4 \neq 0$, let $\epsilon_2(n) = -1$.
- Otherwise, let $\epsilon_2(n) = 0$.

6.9.3 Step 3

Let $A \in (0, 1)$ be any parameter and let $\alpha > 0$ be some parameter such that $\alpha \notin 2\mathbf{Z}[A]$. Given any lattice point (m, n), we perform the following construction.

- Let $(a_\pm, b_\pm, c_\pm) = \mu_\pm(A, m, n)$. See §6.6.
- Let $n_\pm$ be the 5-tuple associated to $(a_\pm, b_\pm, c_\pm)$.
- Let $\epsilon_1^\pm = \epsilon_1(n_\pm)$ and $\epsilon_2^\pm = \epsilon_2(n_\pm)$.

The Master Picture Theorem says that the two edges of $\Gamma_\alpha(m, n)$ incident to (m, n) are $(m, n) + (\epsilon_1^\pm, \epsilon_2^\pm)$.

This page intentionally left blank

Chapter Seven

The Pinwheel Lemma

7.1 THE MAIN RESULT

The Pinwheel Lemma gives a formula for the return map $\Psi: \Xi \to \Xi$ in terms of maps we call *strip maps*. Similar objects are considered in [**GS**] and [**S**].

Consider a pair (Σ, L), where Σ is an infinite planar strip and L is a line transverse to Σ. The pair (L, Σ) determines two vectors V_+ and V_-, each of which points from one boundary component of Σ to the other and is parallel to L. Clearly, $V_- = -V_+$. See Figure 7.1.

For almost every point $p \in \mathbf{R}^2$, there is a unique integer n such that

$$E(p) := p + nV_+ \in \Sigma. \tag{7.1}$$

We call E the strip map defined relative to (Σ, L). The map E is well defined except on a countable collection of parallel and evenly spaced lines.

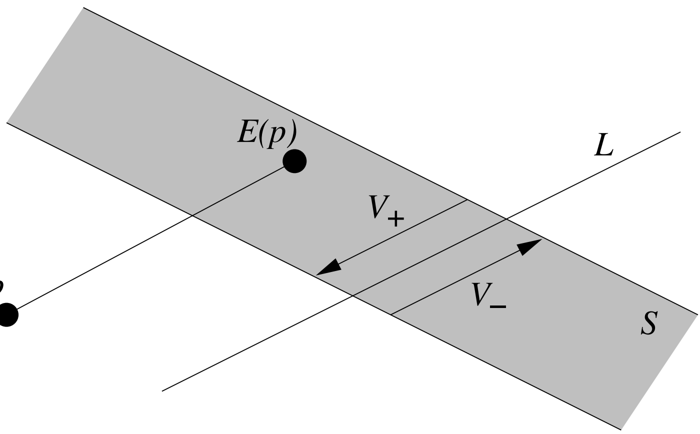

Figure 7.1: A strip map.

Figure 7.2 shows 4 strips $\Sigma_1, ..., \Sigma_4$ we associate to the kite $K(A)$. The labelled points all lie on the x-axis, and we simply give the first coordinate. One edge of each strip contains an edge of $K(A)$. The other edge of the same strip is obtained by reflecting the first edge through the kite vertex that is furthest away from the first edge. Referring to the vectors in §2.3, we associate the vector V_j to Σ_j. We remind the reader that

$$V_1 = (0, 4), \qquad V_2 = (-2, 2), \qquad V_3 = (-2-2A, 0), \qquad V_4 = (-2, -2). \tag{7.2}$$

The corresponding strip map E_j is based on (Σ_j, V_j). To make the notation completely consistent with §2.3, we define

$$\Sigma_{j+4} = \Sigma_j, \qquad V_{j+4} = -V_j, \qquad E_{j+4} = E_j. \tag{7.3}$$

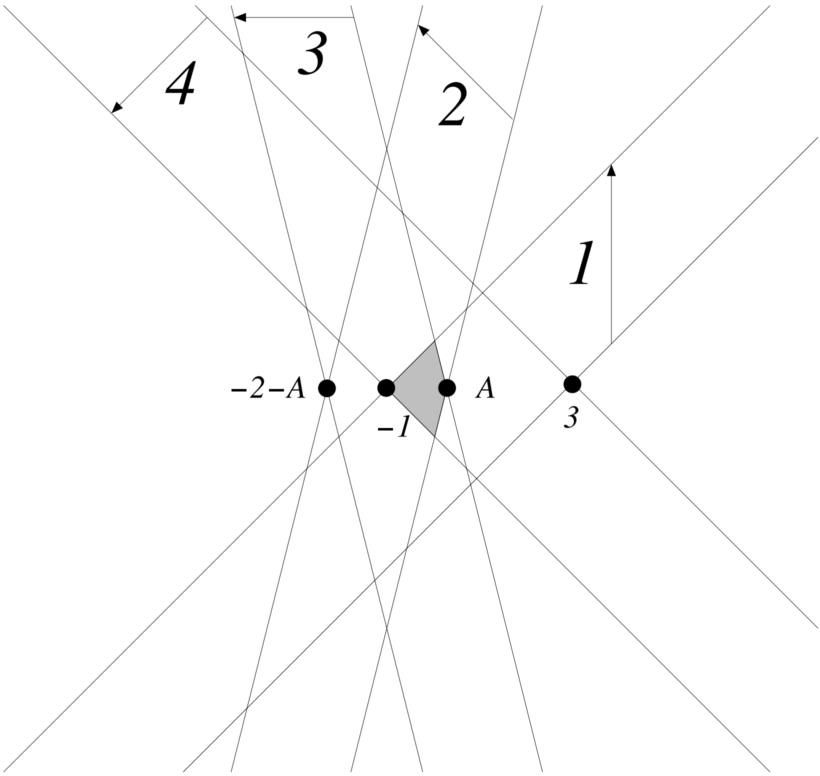

Figure 7.2: The 4 strips for the parameter $A = 1/3$.

To give formulas for the strip maps, we define vectors

$$W_1 = \frac{1}{4}(-1, 1, 3), \qquad W_2 = \frac{1}{2+2A}(-1, A, A),$$

$$W_3 = \frac{1}{2+2A}(-1, -A, A), \qquad W_4 = \frac{1}{4}(-1, -1, 3). \tag{7.4}$$

For a point $p \in \mathbf{R}^2$, we define

$$F_j(p) = W_j \cdot (p_1, p_2, 1). \tag{7.5}$$

$F(j, p)$ measures the position of p relative to the strip Σ_j. This quantity lies in $(0, 1)$ iff p lies in the interior of Σ_j. Letting [ ] denote the floor function, we have

$$E_j(p) = p - \left[F_j(p)\right] V_j. \tag{7.6}$$

We also define a map $\chi\colon \mathbf{R}_+ \times \mathbf{Z}_{\text{odd}} \to \Xi$ by the formula

$$\chi(x, 4n \pm 1) = (x, \pm 1). \tag{7.7}$$

Lemma 7.1 (Pinwheel) Ψ *exists for any point of* Ξ *having a well defined outer billiards orbit. In all cases,* $\Psi = \chi \circ (E_8 \ldots E_1)$.

7.2 DISCUSSION

We call the map in the Pinwheel Lemma the *pinwheel map*. Results like the Pinwheel Lemma seem to be foundational for polygonal outer billiards. Similar ideas appear in [**K**] and [**GS**], for instance. As we will see in the next section, the Pinwheel Lemma is quite easy for points far from the kite. We are forced to consider all points in Ξ because all the unbounded orbits turn out to be erratic; they inevitably come close to the kite.

To prove the Pinwheel Lemma in general, we follow the strategy used for the Return Lemma. We consider all possible sequences of the form

$$i_1 \to i_2 \to i_3 \to \cdots,$$

where R_{i_1}, R_{i_2}, ... denotes the list of successive regions of the partition encountered by the forward ψ-orbit of some point $z_1 \in \Xi$. We let z_j be the first point in the forward orbit in R_{a_j}. Our proof boils down to a case-by-case analysis of the possible sequences. In some cases, the proof relies on some lucky cancellations.

Clearly, something nontrivial must happen to make the Pinwheel Lemma true for all points. Notice that the pinwheel map does not involve the vectors $V_4^{\sharp}$ and $V_6^{\flat}$, and yet these vectors and their corresponding regions are involved in the dynamics. Some kind of lucky cancellation must take place that "edits out" these vectors and regions from the final reckoning. There are two "symmetrically related" lucky accidents, and they are depicted in Figures 7.4 and 7.5 below. The nature of these accidents dictates the order of our proof. First we deal with sequences that do not involve $4^{\sharp}$ and $6^{\flat}$ and then we consider the general case.

As in §2.3, we strongly recommend that the reader use Billiard King to better follow the claims we make here. This program allows the reader to draw all the regions in the partition and their translates, superimposing them as desired over the strips. At the same time, the reader can plot the dynamics of the outer billiards map, checking that all the sets have their advertised properties.

Since the Pinwheel Lemma is a nontrivial result for points near the kite, it seems worth presenting some numerical evidence for the result. Using Billiard King, we compute that the Pinwheel Lemma holds true at the points $(x, \pm 1)$ relative to the parameter A for all

$$A = \frac{1}{256}, ..., \frac{255}{256}, \qquad x = \epsilon + \frac{1}{1024}, ..., \epsilon + \frac{16384}{1024}, \qquad \epsilon = 10^{-6}.$$

The small number ϵ is included to make sure that the outer billiards orbit is actually defined for all the points we sample. This calculation fairly well carpets the "near region" with instances of the truth of the result. While this calculation does not prove anything formally, it serves as a good sanity check that the Pinwheel Lemma is true.

We close this section with a discussion of how the Pinwheel Lemma fits into the proof of the Master Picture Theorem. The Master Picture Theorem really makes a statement about the pinwheel map. The Pinwheel Lemma then translates this statement to a statement about the map Ψ. Thus, if we want to use the Master Picture Theorem to verify a particular statement solely about the pinwheel map, we do not need to know about the truth of the Pinwheel Lemma. This principle will come in handy at the end, saving us some tedious work.

7.3 FAR FROM THE KITE

Here we prove the Pinwheel Lemma for points of Ξ far from K. Logically, the argument we give here is not necessary for our overall proof of the Pinwheel Lemma. However, it is an easy argument, and it serves as a guide for the harder arguments we give in the following sections when we come to the real proof.

Let K' be a large compact set surrounding K. Define

$$S_j = R_j - K', \qquad j = 1, 2, 3, 4, 5, 6, 7, 8. \tag{7.8}$$

K' contains the two compact regions $R_4^\sharp$ and $R_6^\flat$. Figure 7.3 shows how the regions S_j sit with respect to the strips Σ_j. Each S_j shares its unbounded edges with two consecutive strips as shown.

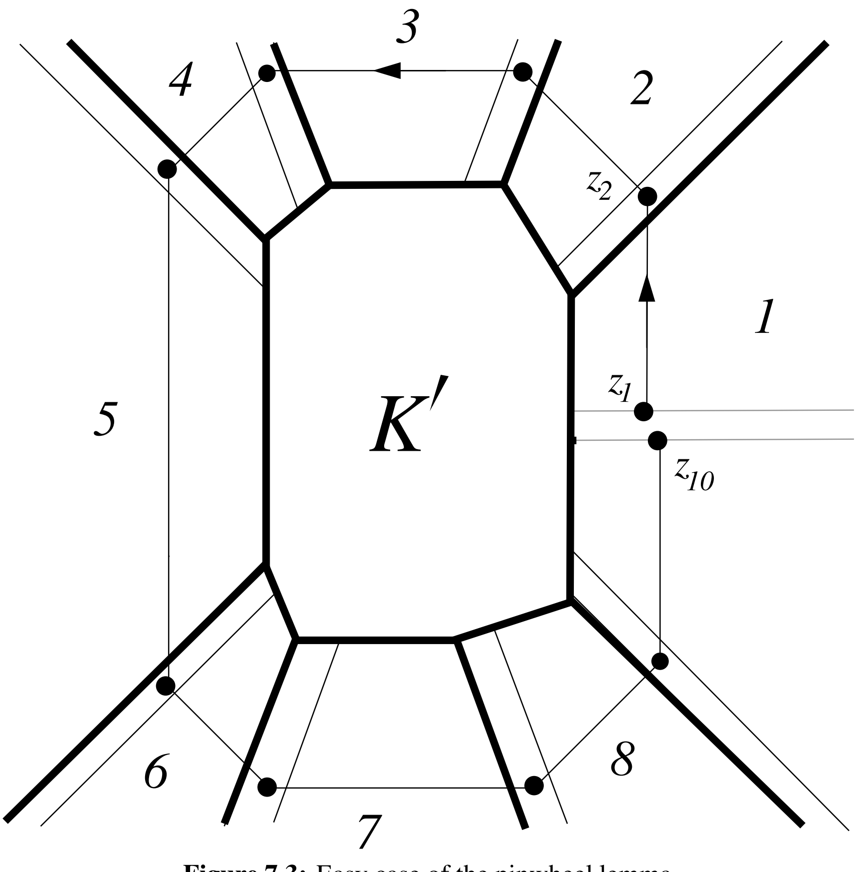

Figure 7.3: Easy case of the pinwheel lemma.

Looking at the figure, we have

$$z_{j+1} = E_j(z_j), \qquad z_{10} = \chi(z_9), \qquad j = 1, ..., 8. \tag{7.9}$$

By induction and Equation 7.8, the point z_{j+1} lies in the forward orbit of z_j for each $j = 1, ..., 8$. But then $z_{10} = \Psi(z_1) = \chi \circ E_8 ... E_1(z_1)$, and we are finished.

7.4 NO SHARPS OR FLATS

Now we turn to the general proof of the Pinwheel Lemma. In this section, we prove the Pinwheel Lemma for sequences that contain neither $4^{\#}$ or $6^{\flat}$. Since $\Xi \subset R_1 \cup R_2 \cup R_4^{\sharp}$, our sequence has the form $i_1 \to \cdots \to i_k$, where $i_1 \in \{1, 2\}$ and $i_k \in \{9, 10\}$. By Equation 2.7, the indices increase, and furthermore they increase by at most 3 each time. We observe, using Billiard King, that

$$0 < k - j < 4, \qquad \Longrightarrow \qquad \widehat{R}_j \cap R_k \subset \Sigma_j \cap \cdots \cap \Sigma_{k-1}. \tag{7.10}$$

Since no sharps and flats are involved, Equation 7.10 implies

$$z_{j+1} = E_{i_j}(z_j). \tag{7.11}$$

We check that $R_2 \cap \Xi \subset \Sigma_1$. Hence, if $i_1 = 2$, we have $E_1(z_1) = z_1$. Therefore, whether $i_1 = 1$ or $i_1 = 2$, Equation 7.11 yields

$$z_2 = E_{i_1} ... E_1(z_1). \tag{7.12}$$

By Equation 7.10, we have

$$z_2 \in \widehat{R}_{i_1} \cap R_{i_2} \subset \Sigma_{i_1} \cap \cdots \cap \Sigma_{i_2-1}, \qquad E_{i_2-1} ... E_{i_1+1}(z_2) = z_2. \tag{7.13}$$

The first equation above implies the second. Combining Equations 7.11–7.13, we have

$$z_3 = E_{i_2}(z_2) = E_{i_2} ... E_{i_1+1}(z_2) = E_{i_2} ... E_1(z_1). \tag{7.14}$$

Repeating the same argument, we have

$$z_4 = E_{i_3} ... E_1(z_1). \tag{7.15}$$

This pattern continues in this way until we arrive at $z_k \in R_9 \cup R_{10}$.

Case 1: Suppose $z_k \in R_9 = R_1$. Then

$$z_k = E_8 ... E_1(z_1). \tag{7.16}$$

The forward iterates of z_k are obtained by repeatedly adding V_1. This is the same as applying the map χ. Hence

$$\Psi(z_1) = \chi(z_k) = \chi \circ E_8 ... E_1(z_1). \tag{7.17}$$

Hence the Pinwheel Lemma holds in this case.

Case 2: Suppose $z_k \in R_{10} = R_2$. Then

$$z_k = E_9 ... E_1(z_1) =^* E_8 ... E_1(z_1) \subset \Xi \tag{7.18}$$

The starred equality comes from the fact that

$$E_8 ... E_1(z_1) \in \Sigma_9 = \Sigma_1, \tag{7.19}$$

by Equation 7.10. Hence $E_9 = E_1$ acts trivially. The containment in Equation 7.18 comes from the same argument we gave in case 2 of the proof of the Return Lemma in §2.3. By Equation 7.18, we have $\chi(z_k) = z_k$, and again the Pinwheel Lemma holds.

7.5 DEALING WITH $4^\sharp$

In this section we will consider sequences that have $4^\#$ in them but not $6^\flat$. In this section, we suppose that $4^\#$ is not the first term. By Equation 2.7, we must have $a \to 3 \to 4^\# \to \cdots$, where $a \in \{1, 2\}$.

Our proof is based on the following items.

1. $R_4^\sharp \subset \Sigma_4 - \Sigma_3$ and $R_4^\sharp + V_3 \in \Sigma_3 - \Sigma_4$.
2. $V_4^\sharp = V_3 - V_4 + V_5$.
3. $\widehat{R}_4^\sharp \cap R_8 \subset \Sigma_5 \cap \Sigma_6 \cap \Sigma_7$.

One can see these at a glance using Billiard King.

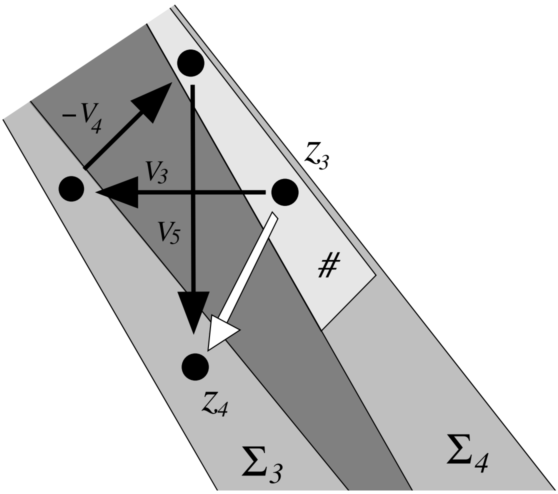

Figure 7.4: The orbit near $R_4^\sharp$.

Consider z_3, the first point in the forward orbit of z_1 that lies in $R_4^\sharp$. (This region is labelled by a # in Figure 7.4.) From Equation 7.10 and from the fact that $3 \to 4^\sharp$, we have

$$z_3 = E_2E_1(z_1) + nV_3, \qquad n \geq 1. \tag{7.20}$$

Item 1 gives

$$E_3E_2E_1(z_1) = E_3(z_3) = z_3 + V_3, \qquad E_4E_3(z_3) = z_3 + V_3 - V_4.$$

Item 2 gives the crucial starred equality in the next equation.

$$z_4 = z_3 + V_4^\sharp =^* z_3 + V_3 - V_4 + V_5 = E_4E_3(z) + V_5 = E_4E_3E_2E_1(z_1) + V_5. \tag{7.21}$$

Equation 2.7 gives $z_4 \in R_5$ or $z_4 \in R_8$. If $z_5 \in E_5$, then

$$z_5 = E_5(z_4) = E_5 ... E_1(z_1). \tag{7.22}$$

If $z_4 \in R_8$, then item 3 gives the starred equality in the following equation.

$$z_5 = E_8(z_4) =^* E_8E_7E_6E_5(z_4) = E_8 ... E_1(z_1). \tag{7.23}$$

In either case, the analysis finishes as in the previous section.

7.6 DEALING WITH $6^\flat$

In this section we will consider sequences that contain $6^\flat$ but not the portion $2 \to 6^\flat$. Our arguments refer mainly to Figure 7.5. By Equation 2.7, we must have $5 \to 6^\flat$. Our argument is based on the following items.

1. $R_6^\flat \subset \Sigma_6 - \Sigma_5$ and $R_6^\flat + V_5 \subset \Sigma_5 - \Sigma_6$.

2. $V_6^\flat = V_5 - V_6 + V_7$.

3. $\widehat{R}_6^\flat \cap R_2 \subset \Sigma_7 \cup \Sigma_8$.

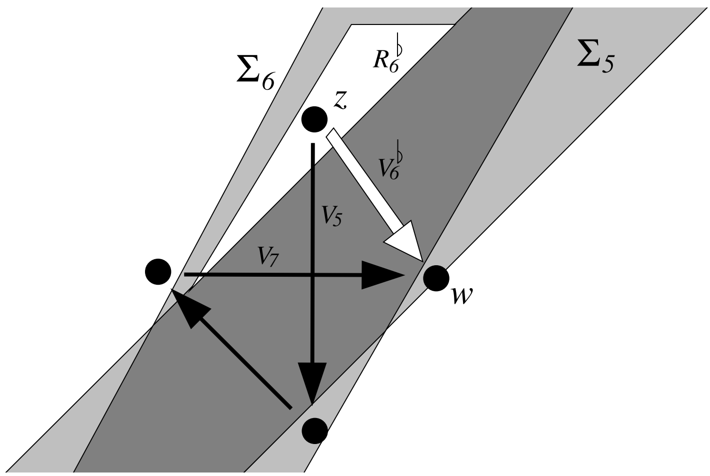

Figure 7.5: The orbit near $R_6^\flat$.

Let z be the first point in the forward orbit of z_1 such that $z \in R_6^\flat$ and let $w = \psi(z)$. From the arguments in the last two sections, we have some $n \geq 1$ such that

$$z = E_4E_3E_2E_1(z_1) + nV_5, \qquad w = z + V_6^\flat \in R_7 \cup R_8 \cup R_2. \tag{7.24}$$

By item 1 above, we have

$$E_5E_4E_3E_2E_1(z_1) = E_5(z) = z + V_5, \qquad E_6E_5E_4E_3E_2E_1(z) = z + V_5 - V_6$$

By item 2, we have

$$w = E_6E_5E_4E_3E_2E_1(z_1) + V_7. \tag{7.25}$$

By Equation 2.7, we have $w \in R_7$ or $w \in R_{10} = R_2$. The first case is just like the first case treated at the end of the last section. In the second case, we have

$$\chi E_8...E_1(z_1) =^1 E_6...W_1(z_1) =^2 w = \Psi(z) \tag{7.26}$$

The first equality comes from item 3. The second equality comes from an argument similar to case 2 at the end of §7.4.

7.7 THE LAST CASES

Now we treat the two cases we have not yet treated.

First, suppose the sequence has the portion $2 \to 6^\flat$. Let w be the orbit point in $R_6^\flat$. We have

$$w \in \widehat{R}_2 \cap R_6^\flat \subset (-2A, 0) \times \{1\}. \tag{7.27}$$

This forces the entire orbit sequence to be $2 \to 6^\flat \to 2$, and

$$z_1 \in (2-2A, 2) \times \{-1\}, \qquad \Psi(z_1) = z_1 - (2-2A, 0). \tag{7.28}$$

Second, suppose the sequence starts with $4^\sharp$. A similar calculation shows that

$$z_1 \in (0, 2A) \times \{1\}, \qquad \Psi(z_1) = z_1 + (2-2A, 0). \tag{7.29}$$

To finish the proof, we just have to compute the pinwheel map on the above intervals and see that it matches Ψ. One can achieve this with the same kind of analysis used in the previous sections. However, we prefer a different method. We can use the formula from the Master Picture Theorem to see that the pinwheel map does the right thing on the above intervals. This is not a circular argument, as we discussed at the end of §7.2.

Chapter Eight

The Torus Lemma

8.1 THE MAIN RESULT

For ease of exposition, we state and prove the (+) halves of our results. The (−) halves have the same formulation and proof.

Let μ_+ be as in Equation 6.5. We write $(\mu_+)_A$ to emphasize the dependence on the parameter A. Let $T^4 = \widehat{R}/\Lambda$, the 4-dimensional quotient discussed in §6.7. Topologically, T^4 is the product of a 3-torus with $(0, 1)$. We now define

$$\mu_+ \colon \Xi_+ \times (0,1) \to T^4$$

by the obvious formula

$$\mu_+(p, A) = ((\mu_+)_A(p), A). \tag{8.1}$$

We are just stacking all these maps together.

Referring to the Pinwheel Lemma, we have $\Psi(p) = \chi \circ E_8 ... E_1(p)$ whenever both maps are defined. Let $p \in \Xi_+$. We set $p_0 = p$ and inductively define

$$p_j = E_j(p_{j-1}) \in \Sigma_j. \tag{8.2}$$

We also define

$$\theta(p) = \min \theta_j(p), \qquad \theta_j(p) = \text{distance}(p_j, \partial \Sigma_j). \tag{8.3}$$

The quantity $\theta(p)$ depends on the parameter A, so we will write $\theta(p, A)$ when we want to be clear about this.

Lemma 8.1 (Torus) *Let $(p, A), (q^*, A^*) \in \Xi_+ \times (0, 1)$. There is some $\eta > 0$, depending only on $\theta(p, A)$ and $\min(A, 1 - A)$, with the following property. Suppose that the pinwheel map is defined at (p, A). Suppose also that $\mu_+(p, A)$ and $\mu_+(q^*, A^*)$ are within η of each other. Then the pinwheel map is defined at (q^*, A^*) and $(\epsilon_1(q^*), \epsilon_2(q^*)) = (\epsilon_1(p), \epsilon_2(p))$.*

Remarks:
(i) In the proof of the Pinwheel Lemma, we started our labelling with z_1, then considered $z_2 = E_1(z_1)$, etc. Here we find it convenient to take $p_j = z_{j+1}$.
(ii) I discovered the Torus Lemma experimentally, but my formal proof owes a considerable intellectual debt to the ideas presented in [**K**] and [**GS**] concerning outer billiards on quasirational polygons. (Compare the remark in the next section.) My proof also owes an intellectual debt to the paper [**T2**], in which S. Tabachnikov describes unpublished work of C. Culter on the existence of periodic orbits for polygonal outer billiards. If all these written sources were not enough, I was also influenced by conversations with John Smillie.

8.2 INPUT FROM THE TORUS MAP

We first prove the Torus Lemma assuming that $A = A^*$. Let $q = q^*$. In this section, we explain the significance of the map μ_+. We introduce the quantities

$$\widehat{\lambda}_j = \lambda_0 \times \cdots \times \lambda_j, \qquad \lambda_j = \frac{\mathrm{area}(\Sigma_{j-1} \cap \Sigma_j)}{\mathrm{area}(\Sigma_j \cap \Sigma_{j+1})}, \qquad j = 1, ..., 7. \tag{8.4}$$

Remark: For a general convex n-gon, one can make the strip construction along the lines of what we have done. The polygon is said to be *quasirational* if all the numbers λ_j are rational. As mentioned in the introduction, the result in [**VS**], [**K**], and [**GS**] is that all outer billiards orbits are bounded relative to quasirational polygons. In hindsight, it is no surprise that these quantities arise in our proof of the Master Picture Theorem.

Let $p = (x, \pm 1)$ and $q = (y, \pm 1)$. We have

$$\mu_+(q) - \mu_+(p) = (t, t, t) \mod \Lambda, \qquad t = \frac{y - x}{2}. \tag{8.5}$$

Lemma 8.2 *For any $\epsilon > 0$, there is a $\delta > 0$ with the following property. If* $\mathrm{dist}(\mu_+(x), \mu_+(y)) < \delta$ *in T^3, then for each k, the quantity $t\widehat{\lambda}_k$ is within ϵ of some integer I_k.*

Proof: We compute

$$\mathrm{area}(\Sigma_0 \cap \Sigma_1) = 8, \qquad \mathrm{area}(\Sigma_1 \cap \Sigma_2) = \frac{8 + 8A}{1 - A},$$

$$\mathrm{area}(\Sigma_2 \cap \Sigma_3) = \frac{2(1 + A)^2}{A}, \qquad \mathrm{area}(\Sigma_3 \cap \Sigma_4) = \frac{8 + 8A}{1 - A}. \tag{8.6}$$

This leads to

$$\widehat{\lambda}_0 = \widehat{\lambda}_4 = 1, \quad \widehat{\lambda}_1 = \widehat{\lambda}_3 = \widehat{\lambda}_5 = \widehat{\lambda}_7 = \frac{1 - A}{1 + A}, \quad \widehat{\lambda}_2 = \widehat{\lambda}_6 = \frac{4A}{(1 + A)^2}. \tag{8.7}$$

The matrix

$$H = \begin{bmatrix} \frac{1}{1+A} & \frac{A-1}{(1+A)^2} & \frac{2A}{(1+A)^2} \\ 0 & \frac{1}{1+A} & \frac{1}{1+A} \\ 0 & 0 & 1 \end{bmatrix} \tag{8.8}$$

conjugates the columns of the matrix defining Λ to the standard basis. Therefore, if $\mu_+(x)$ and $\mu_+(y)$ are close in T^3 then $H(t, t, t)$ is close to a point of $\mathbf{Z}^3$. We compute

$$H(t, t, t) = \left(\frac{4A}{(1 + A)^2}, \frac{2}{1 + A}, 1\right)t = (\widehat{\lambda}_2, \widehat{\lambda}_1 + 1, 1)t. \tag{8.9}$$

Equations 8.7 and 8.9 now finish the proof. $\square$

8.3 PAIRS OF STRIPS

Suppose (S_1, S_2, V_2) is triple, where V_2 is a vector pointing from one corner of $S_1 \cap S_2$ to an opposite corner. Let $p_1 \in S_1$ and $p_2 = E_2(p_1) \in S_2$. Here E_2 is the strip map associated to (S_2, V_2). We define n and α by the equations

$$p_2 - p_1 = nV_2, \qquad \alpha = \frac{\text{area}(B)}{\text{area}(S_1 \cap S_2)}, \qquad \sigma_j = \frac{\|p_j - p'_j\|}{\|V_2\|}. \tag{8.10}$$

All quantities are affine-invariant functions of the quintuple $(S_1, S_2, V_2, p_1, p_2)$.

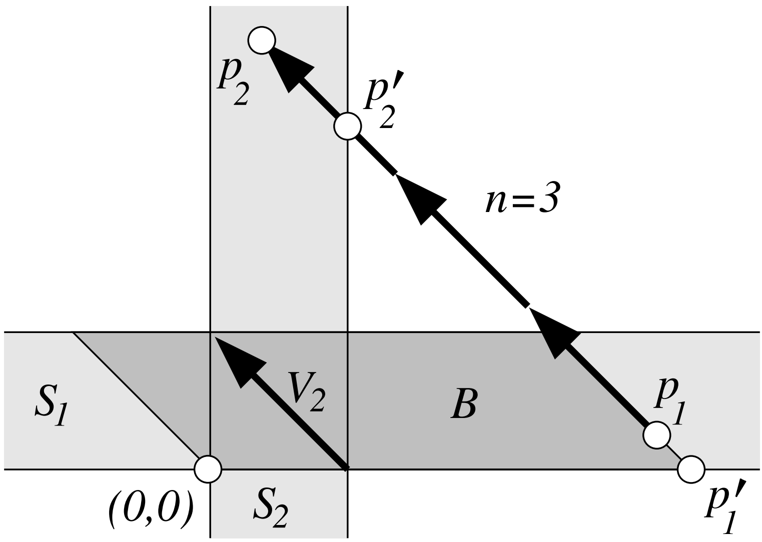

Figure 8.1: Strips and associated objects.

Figure 8.1 shows what we call the *standard pair* of strips, where Σ_j is the strip bounded by the lines $x_j = 0$ and $x_j = 1$. Here we denote points in the plane by (x_1, x_2). To get a better picture of the quantities we have defined, we consider them on the standard pair. We have

- $\alpha = p_{11} + p_{12} = p_{21} + p_{22}$,
- $\sigma_1 = p_{12}$,
- $\sigma_2 = 1 - p_{22}$,
- $n = [p_{11}]$ (the floor of x).

Here p_{ij} is the jth coordinate of p_i. The above equations lead to the following affine-invariant relations. Letting $\langle x \rangle = x - [x]$, the fractional part of x, we have

$$n = [\alpha - \sigma_1], \qquad \sigma_2 = 1 - \langle \alpha - \sigma_1 \rangle. \tag{8.11}$$

Again, the relations in Equation 8.11 hold for any pair of strips.

In our next result, we hold (S_1, S_2, V_2) fixed but compare all the quantities for (p_1, p_2) and another pair (q_1, q_2). Let $n(p) = n(S_1, S_2, V_2, p_1, p_2)$, etc. Also, N stands for an integer.

Lemma 8.3 (Strip) *Let $\epsilon > 0$. There is some $\delta > 0$ with the following property. If*

$$|\sigma(p_1) - \sigma(q_1)| < \delta, \qquad |\alpha(q) - \alpha(p) - N| < \delta,$$

then

$$|\sigma(p_2) - \sigma(q_2)| < \epsilon, \qquad N = n(q) - n(p).$$

The number δ depends on only ϵ and the distance from $\sigma(p_1)$ and $\sigma(p_2)$ to $\{0, 1\}$.

Proof: If δ is small enough, then $\langle \alpha(p) - \sigma(p_1) \rangle$ and $\langle \alpha(q) - \sigma(q_1) \rangle$ are very close and relatively far from 0 or 1. Equation 8.11 now says that $\sigma(p_2)$ and $\sigma(q_2)$ are close. Also, the following two quantities are both near N, while the individual summands are all relatively far from integers.

$$\alpha(q) - \alpha(p), \qquad \big(\alpha(q) - \sigma(q_1)\big) - \big(\alpha(p) - \sigma(p_1)\big).$$

But the second quantity is near the integer $n(q) - n(p)$, by Equation 8.11. □

Suppose now that S_1, S_2, S_3 is a triple of strips and V_2, V_3 is a pair of vectors, such that (S_1, S_2, V_2) and (S_2, S_3, V_3) are as above. Let $p_j \in S_j$, for $j = 1, 2, 3$, be such that $p_2 = E_2(p_1)$ and $p_3 = E_3(p_2)$. For $j = 1, 2$, define

$$\alpha_j = \alpha(S_j, S_{j+1}, V_{j+1}, p_j, p_{j+1}), \qquad \lambda = \frac{\text{area}(S_1 \cap S_2)}{\text{area}(S_2 \cap S_3)}. \tag{8.12}$$

It is convenient to set $\sigma_2 = \sigma(p_2)$.

Lemma 8.4 *There are constants C and D such that $\alpha_2 = \lambda\alpha_1 + C\sigma_2 + D$. The constants C and D depend on the strips.*

Proof: We normalize so that we have the standard pair. Then

$$p_2 = (1 - \sigma_2, \alpha_1 + \sigma_2 - 1). \tag{8.13}$$

There is a unique orientation-preserving affine map T such that $T(S_{j+1}) = S_j$ for $j = 1, 2$, and T carries the line $x_2 = 1$ to the line $x_1 = 0$. Given that $S_1 \cap S_2$ has unit area, we have $\det(T) = \lambda$. Given the description of T, we have

$$T(x_1, x_2) = \begin{pmatrix} a & \lambda \\ -1 & 0 \end{pmatrix} (x_1, x_2) + (b, 1) = (ax_1 + b + \lambda x_2, 1 - x_1). \tag{8.14}$$

Here a and b are constants depending on $S_2 \cap S_3$. Setting $q = T(p_2)$, the relations above give $\alpha = q_1 + q_2$. Hence

$$\alpha_2 = a(1 - \sigma_2) + b + \lambda(\alpha_1 + \sigma_2 - 1) + \sigma_2 = \lambda\alpha_1 + C\sigma_2 + D. \tag{8.15}$$

This completes the proof. □

8.4 SINGLE-PARAMETER PROOF

The Pinwheel Lemma gives a formula for the quantities in Equation 2.8. We have integers $n_0, ..., n_7$ such that

$$p_{j+1} = E_{j+1}(p_j) = p_j + n_j V_{j+1}. \tag{8.16}$$

Compare Figure 7.3. Given the equations

$$V_1 = (0, 4), \qquad V_2 = (-2, 2), \qquad V_3 = (-2 - 2A, 0), \qquad V_4 = (-2, -2), \tag{8.17}$$

we find that

$$\epsilon_1 = n_2 - n_6, \qquad \epsilon_2 = n_1 + n_2 + n_3 - n_5 - n_6 - n_7. \tag{8.18}$$

We are still working under the assumption, in the Torus Lemma, that $A = A^*$. Our main argument relies on Equation 8.18, which gives a formula for the return pairs in terms of the strip maps. We define the point q_j relative to q just as we defined p_j relative to p.

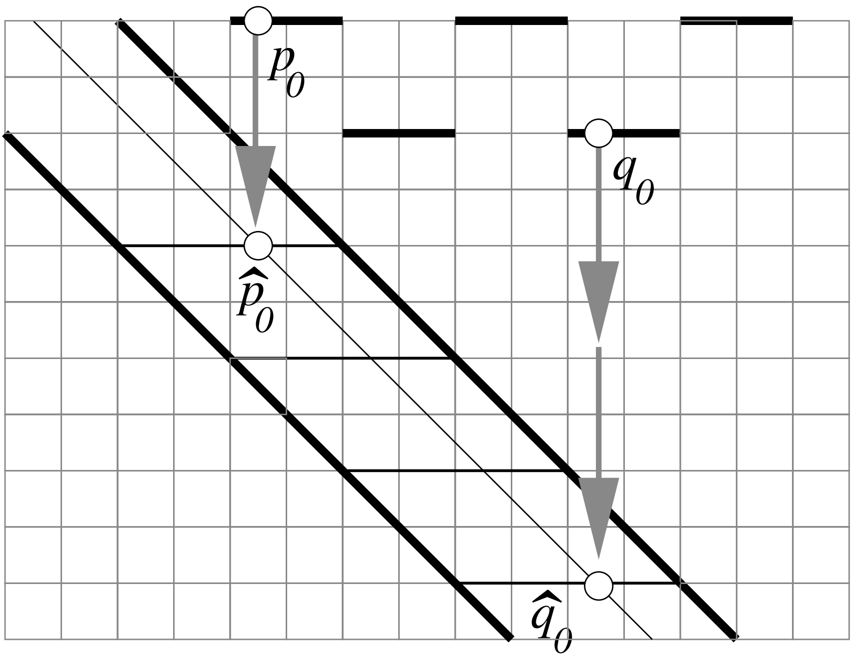

Figure 8.2: The points $\widehat{p}_0$ and $\widehat{q}_0$.

We would like to apply Lemmas 8.2–8.4 inductively. One inconvenience is that p_0 and q_0 do not lie in any of our strips. To remedy this situation, we start with the two points

$$\widehat{p}_0 = E_0(p_0), \qquad \widehat{q}_0 = E_0(q_0). \tag{8.19}$$

See Figure 8.1. We have $\widehat{p}_0, \widehat{q}_0 \in \Sigma_0$. Let t be the near integer from Lemma 8.2. Looking at Figure 8.4, we see that $|\sigma(\widehat{q}_0) - \sigma(\widehat{p}_0)|$ tends to 0 as η tends to 0.

We define

$$\alpha_k(p) = \alpha(\Sigma_k, \Sigma_{k+1}, V_{k+1}, p_k, p_{k+1}) \tag{8.20}$$

It is also convenient to write

$$\sigma_k(p) = \sigma(p_k), \qquad \Delta\sigma_k = \sigma_k(q) - \sigma_k(p). \tag{8.21}$$

For $k = 0$, we use $\widehat{p}_0$ in place of p_0, and $\widehat{q}_0$ in place of q_0, for these formulas.

Remarks:
(i) The functions σ_k play a big role in our overall proof. The next chapter is devoted entirely to obtaining, in a certain sense, closed-form expressions for the functions σ_k. For later reference, we call these functions *strip functions*.
(ii) Our next lemma is stated in a slightly peculiar way because the last-mentioned quantity $n_k(p) - n_k(q)$ is an integer. But that is the whole point of the lemma: Once an integer quantity is sufficiently close to 0, it must actually be 0.

Lemma 8.5 *As $\eta \to 0$, the pairwise differences between the 3 quantities*

$$\alpha_k(q) - \alpha_k(p), \qquad n_k(q) - n_k(p), \qquad t\widehat{\lambda_k}$$

converge to 0 for all k.

Proof: Referring to Figure 8.2, we have

$$\text{area}(\Sigma_0 \cap \Sigma_1) = 8, \qquad \text{area}(B(\widehat{p}_0)) - \text{area}(B(\widehat{q}_0)) = 4y - 4x.$$

This gives $\alpha_0(q) - \alpha_0(p) = t$. Applying Lemma 8.4 inductively, we find that

$$\alpha_k = \alpha_0\widehat{\lambda_k} + \sum_{i=1}^{k} \xi_i\sigma_i + C_k \tag{8.22}$$

for constants $\xi_1, ..., \xi_k$ and C_k that depend analytically on A. Therefore

$$\alpha_k(q) - \alpha_k(p) = t\widehat{\lambda_k} + \sum_{i=1}^{k} \xi_i\,\Delta\sigma_i, \qquad k = 1, ..., 7. \tag{8.23}$$

By Lemma 8.2, the term $t\lambda_k$ is near an integer for all k. By Lemma 8.3 and induction, the remaining terms on the right hand side are near 0. This lemma now follows from Lemma 8.3. $\square$

Combining our last result with Equation 8.7, we see that

$$n_1(q) - n_1(p) = n_3(q) - n_3(p) = n_5(q) - n_5(p) = n_7(q) - n_7(p),$$

$$n_2(q) - n_2(p) = n_6(q) - n_6(p), \tag{8.24}$$

once η is small enough. Given the dependence of constants in Lemma 8.3, the necessary bound on η depends on only $\min(A, 1-A)$ and $\theta(p)$. Equation 8.18 now tells us that $\epsilon_j(p) = \epsilon_j(q)$, for $j = 1, 2$, once η is small enough.

8.5 PROOF IN THE GENERAL CASE

Now we turn to the proof of the Torus Lemma in the general case. Our first result is the key step that allows us to handle pairs of distinct parameters. Once we set up the notation, the proof is almost trivial. Our second result is a variant that will be useful in the next chapter.

Suppose that $(S_1, S_2, V_2, p_1, p_2)$ and $(S_1^*, S_2^*, V_2^*, q_1^*, q_2^*)$ are two quintuples. To fix the picture in our minds, we imagine that (S_1, S_2, V_2) is near (S_1^*, S_2^*, V_2^*), though this is not necessary for the proof of the result to follow. We can define the quantities α, ρ_j, n for each of these quintuples. We place a $*$ by each quantity associated to the second triple.

Lemma 8.6 *Let $\epsilon > 0$. There is some $\delta > 0$ with the following property. If $|\sigma(p_1) - \sigma(q_1^*)| < \delta$ and $|\alpha(q^*) - \alpha(p) - N| < \delta$, then $|\sigma(p_2) - \sigma(q_2^*)| < \epsilon$ and $N = n(q^*) - n(p)$. The number δ depends on only ϵ and the distance from $\sigma(p_1)$ and $\sigma(p_2)$ to* $\{0, 1\}$.

Proof: There is an affine transformation such that $T(X^*) = X$ for each object $X = S_1, S_2, V_2$. We set $q_j = T(q_j^*)$. Then $\alpha(q_1^*) = \alpha(q_1)$, by affine invariance. Likewise for the other quantities. Now we apply Lemma 8.3 to the triple (S_1, S_2, V_2) and the pairs (p_1, p_2) and (q_1, q_2). The conclusion involves quantities with no $*$, but returning the $*$ does not change any of the quantities. □

For use in the next chapter, we state a variant of Lemma 8.6. For this result, we interpret $\langle x \rangle$ as the image of a real number x in $\mathbf{R}/\mathbf{Z}$.

Lemma 8.7 *Let $\epsilon > 0$. There is some $\delta > 0$ with the following property. If $|\sigma(p_1) - \sigma(q_1^*)| < \delta$ and $|\alpha(q^*) - \alpha(p) - N| < \delta$, then the distance from $\langle \sigma(p_2) \rangle$ to $\langle \sigma(q_2^*) \rangle$ in $\mathbf{R}/\mathbf{Z}$ is less than ϵ and $N = n(q^*) - n(p)$. The number δ depends only on ϵ and the distance from $\sigma(p_1)$ and $\sigma(p_2)$ to* $\{0, 1\}$.

Proof: Using the same trick as in Lemma 8.3, we reduce to the single-variable case. In this case, we mainly repeat the proof of Lemma 8.3. If δ is small enough, then $\langle \alpha(p) - \sigma(p_1) \rangle$ and $\langle \alpha(q) - \sigma(q_1) \rangle$ are very close and relatively far from $\langle 0 \rangle$. Equation 8.11 now says that $\langle \sigma(p_2) \rangle$ and $\langle \sigma(q_2) \rangle$ are close in $\mathbf{R}/\mathbf{Z}$. □

In proving the general version of the Torus Lemma, we no longer suppose that $A = A^*$ and we return to the original notation (q^*, A^*) for the second point. In our proof of this result, we attach a $*$ to any quantity that depends on (q^*, A^*). We first need to repeat the analysis from §8.2, this time keeping track of the parameter. Let η be as in the Torus Lemma. We use the "big O" notation.

Lemma 8.8 *There is an integer I_k such that $|\widehat{\alpha_0^* \lambda_k^*} - \alpha_0 \lambda_k - I_k| < O(\eta)$.*

Proof: Let H be the matrix in Equation 8.8. Let $\langle V \rangle$ denote the distance from $V \in \mathbf{R}^3$ to the nearest point in $\mathbf{Z}^3$. Let $p = (x, \pm 1)$ and $q^* = (x^*, \pm 1)$. Recalling the definition of μ_+, the hypotheses in the Torus Lemma imply that the fractional part of

$$H^*\left(\frac{x^*}{2}, \frac{x^*}{2}+1, \frac{x^*}{2}\right) - H\left(\frac{x}{2}, \frac{x}{2}+1, \frac{x}{2}\right) \tag{8.25}$$

has size $O(\eta)$. We compute that $\alpha_0 = x/2 + 1/2$ independent of parameter. Therefore

$$H\left(\frac{x}{2}, \frac{x}{2}+1, \frac{x}{2}\right) = H(\alpha_0, \alpha_0, \alpha_0) + \frac{1}{2}H(-1, 1, -1).$$

The same is true for the starred quantities. Therefore

$$\langle\langle(\widehat{\lambda}_2^*, \widehat{\lambda}_1^* - 1, 1)\alpha_0^* - (\widehat{\lambda}_2, \widehat{\lambda}_1 - 1, 1)\alpha_0\rangle$$

$$= \langle H^*(\alpha_0^*, \alpha_0^*, \alpha_0^*) - H(\alpha_0, \alpha_0, \alpha_0)\rangle < O(\eta) + \|(H^* - H)(-1, 1, -1)\| < O(\eta).$$

The lemma now follows immediately from Equation 8.7. □

The integer I_k of course depends on (p, A) and (q^*, A^*), but in all cases Equation 8.7 gives us

$$I_0 = I_4, \qquad I_1 = I_3 = I_5 = I_7, \qquad I_2 = I_6, \tag{8.26}$$

Lemma 8.9 *As $\eta \to 0$, the pairwise differences between the 3 quantities $\alpha_k^* - \alpha_k$ and $n_k^* - n_k$ and I_k tend to 0 for all k.*

Proof: Here α_k^* stands for $\alpha_k(q^*)$, etc. Equation 8.22 works separately for each parameter. The replacement for Equation 8.23 is

$$\alpha_k^* - \alpha_k = W + X + Y, \qquad W = \alpha_0^*\widehat{\lambda}_k^* - \alpha_0\widehat{\lambda}_k \tag{8.27}$$

$$X = \sum_{i=1}^k \xi_i^*\sigma_i^*(q^*) - \sum_{i=1}^k \xi_i\sigma_i(p) = \sum_{i=1}^k \xi_i\left(\sigma_i^* - \sigma_i\right) + O(|A - A^*|), \tag{8.28}$$

$$Y = \sum_{i=1}^k C_i^* - \sum_{i=1}^k C_i = O(|A - A^*|). \tag{8.29}$$

The estimates on X and Y come from the fact that ξ_i and C_i vary smoothly with A. Putting everything together, we have the following.

$$\alpha_k^* - \alpha_k = \left(\alpha_0^*\widehat{\lambda}_k^* - \alpha_0\lambda_k\right) + \sum_{i=1}^k \xi_i\left(\sigma_i^* - \sigma_i\right) + O(|A - A^*|). \tag{8.30}$$

In light of Lemma 8.8, it suffices to show that $\sigma_i^* - \sigma_i$ tends to 0 as η tends to 0. The same argument as in the single-parameter case works here, with Lemma 8.6 used in place of Lemma 8.3. □

As in the single-parameter case, Equations 8.18 and 8.26 now finish the proof.

Chapter Nine

The Strip Functions

9.1 THE MAIN RESULT

The purpose of this chapter is to understand the functions σ_j that arose in the proof of the Torus Lemma. See Equation 8.21. We continue using the notation from the previous chapter. We call these functions *strip functions*. Let $\langle x \rangle$ denote the fractional part of x. Sometimes we interpret $\langle x \rangle$ as an element of $\mathbf{R}/\mathbf{Z}$.

Let $W_k \subset \Xi_+ \times (0, 1)$ denote the set of points where $E_k \ldots E_1$ is defined but $E_{k+1} E_k \ldots E_1$ is not defined. Let S_k denote the closure of $\mu_+(W_k)$ in R. Here R is as in Equation 6.6. Finally, let

$$W'_k = \bigcup_{j=0}^{k-1} W_j, \qquad S'_k = \bigcup_{j=0}^{k-1} S_j, \qquad k = 1, \ldots, 7. \tag{9.1}$$

The Torus Lemma applies to any point that does not lie in the *singular set*

$$S = S_0 \cup \cdots \cup S_7. \tag{9.2}$$

If $p \in \Xi_+ - W'_k$, then the points $p = p_0, \ldots, p_k$ are defined. Here, as in the previous chapter, $p_j = E_j(p_{j-1})$. The functions $\sigma_1, \ldots, \sigma_k$ and $\alpha_1, \ldots, \alpha_k$ are defined for such a choice of p. Again, σ_j measures the position of p_j in Σ_j relative to $\partial \Sigma_j$. Even if E_{k+1} is not defined on p_k, the equivalence class or p_{k+1} is well defined in the cylinder $\mathbf{R}^2/\langle V_{k+1} \rangle$. The corresponding function $\sigma_{k+1}(q) = \sigma(q_{k+1})$ is well defined as an element of $\mathbf{R}/\mathbf{Z}$.

Let $\pi_j \colon \mathbf{R}^4 \to \mathbf{R}$ be the jth coordinate projection. The following identities refer to the $(+)$ case. We discuss the $(-)$ case at the end of the chapter.

$$\sigma_1 = \left\langle \frac{2 - \pi_3}{2} \right\rangle \circ \mu_+ \qquad \text{on } \Xi_+. \tag{9.3}$$

$$\sigma_2 = \left\langle \frac{1 + A - \pi_2}{1 + A} \right\rangle \circ \mu_+ \qquad \text{on } \Xi_+ - W'_1. \tag{9.4}$$

$$\sigma_3 = \left\langle \frac{1 + A - \pi_1}{1 + A} \right\rangle \circ \mu_+ \qquad \text{on } \Xi_+ - W'_2. \tag{9.5}$$

$$\sigma_4 = \left\langle \frac{1 + A - \pi_1 - \pi_2 + \pi_3}{2} \right\rangle \circ \mu_+ \qquad \text{on } \Xi_+ - W'_3. \tag{9.6}$$

In the next chapter we deduce the Master Picture Theorem from these identities and the Torus Lemma. In this chapter, we establish the identities. Equation 9.3 is true by inspection. The other 3 identities are the nontrivial ones.

9.2 CONTINUOUS EXTENSION

Since the map $\mu_+(\Xi_+ \times (0,1))$ is dense in $R - S_k'$, we define

$$\tilde{\sigma}_j(\tau) := \lim_{n \to \infty} \sigma_j(p_n, A_n), \qquad \tau \in R - S_k'. \tag{9.7}$$

Here (p_n, A_n) is chosen so that all functions are defined and $\mu_+(p_n, A_n) \to \tau$. Note that the sequence $\{p_n\}$ need not converge. So far, we do not know that the limit we take is well defined. However, the next result clears this up.

Lemma 9.1 *The functions $\tilde{\sigma}_1, ..., \tilde{\sigma}_{k+1}$, considered $\mathbf{R}/\mathbf{Z}$-valued functions, are well defined and continuous on $R - S_k'$.*

Proof: For the sake of concreteness, we will give the proof in the case when $k = 2$. This representative case explains the idea. First of all, the continuity follows from the well-definedness. We just have to show that the limit above is always well defined. We need to consider $\tilde{\sigma}_1$, $\tilde{\sigma}_2$, and $\tilde{\sigma}_3$. Our argument is essentially inductive.

Here is the base case. $\tilde{\sigma}_1$ is well defined and continuous on all of R, by Equation 9.3.

Since $S_1' \subset S_2'$, we see that $\tau \in R - S_1'$. Hence τ does not lie in the closure of $\mu_+(W_0)$. Hence there is some $\theta_1 > 0$ such that $\theta_1(p_n, A_n) > \theta_1$ for all sufficiently large n. Note also that there is a positive and uniform lower bound to the quantity $\min(A_n, 1 - A_n)$. Note that

$$\langle \alpha_1(p_n, A_n) \rangle = \langle \pi_3(\mu_+(p_n, A_n)) \rangle.$$

Hence

$$\{\langle \alpha_1(p_n, A_n) \rangle\} \tag{9.8}$$

forms a Cauchy sequence in $\mathbf{R}/\mathbf{Z}$.

Lemma 8.7 now applies uniformly to

$$(p, A) = (p_m, A_m), \qquad (q^*, A^*) = (p_n, A_n)$$

for all sufficiently large pairs (m, n). Since $\{\mu_+(p_n, A_n)\}$ forms a Cauchy sequence in R, Lemma 8.7 implies that $\{\sigma_2(\tau_m, A_m)\}$ forms a Cauchy sequence in $\mathbf{R}/\mathbf{Z}$. Hence $\tilde{\sigma}_2$ is well defined on $R - S_1'$ and continuous.

Since $\tau \in R - S_2'$, we see that τ does not lie in the closure of $\mu_+(W_1)$. Hence there is some $\theta_2 > 0$ such that $\theta_j(p_n, A_n) > \theta_j$ for $j = 1, 2$ and sufficiently large n. As in the proof of the General Torus Lemma, Equation 8.30 now shows that

$$\{\langle \alpha_2(p_n, A_n) \rangle\} \tag{9.9}$$

forms a Cauchy sequence in $\mathbf{R}/\mathbf{Z}$. We now repeat the previous argument to see that $\{\sigma_3(\tau_m, A_m)\}$ forms a Cauchy sequence in $\mathbf{R}/\mathbf{Z}$. Hence $\tilde{\sigma}_3$ is well defined on $R - S_2'$ and continuous. □

Referring to Equations 9.8 and 9.9, we define

$$\beta_k = \langle \alpha_k \rangle \in \mathbf{R}/\mathbf{Z}. \tag{9.10}$$

This function will come in handy in our next result.

9.3 LOCAL AFFINE STRUCTURE

Let $X = R - \partial R \subset \mathbf{R}^4$. Note that X is an open and convex polytope, combinatorially equivalent to the 4-dimensional cube.

Lemma 9.2 *Suppose $X \subset R - S_k'$. Then $\tilde{\sigma}_{k+1}$ is locally affine on X_A.*

Proof: Since $\tilde{\sigma}_{k+1}$ is continuous on X, it suffices to prove this lemma for a dense set of kite parameters A. We can choose A so that $\mu_+(\Xi_+)$ is dense in X_A.

We already know that $\tilde{\sigma}_1, ..., \tilde{\sigma}_{k+1}$ are all defined and continuous on X. We have already remarked that Equation 9.3 is true by direct inspection. As we have already remarked in the previous proof,

$$\beta_0 = \pi_3 \circ \mu_+.$$

Thus we define

$$\tilde{\beta}_0 = \langle \pi_3 \rangle. \tag{9.11}$$

Both $\tilde{\sigma}_0$ and $\tilde{\beta}_0$ are locally affine on X_A.

Let $m \leq k$. The second half of Equation 8.11 tells us that $\tilde{\sigma}_m$ is a locally affine function of $\tilde{\sigma}_{m-1}$ and $\tilde{\beta}_{m-1}$. Below we will prove that $\tilde{\beta}_m$ is defined on X_A and locally affine, provided that $\tilde{\sigma}_1, ..., \tilde{\sigma}_m$ are defined and locally affine on X_A. The lemma follows from this claim and induction.

Now we prove the claim. All the addition below is done in $\mathbf{R}/\mathbf{Z}$. Since $\mu_+(\Xi_+)$ is dense in X_A, we can at least define $\tilde{\beta}_m$ on a dense subset of X_A. Define

$$p = (x, \pm 1), \quad p' = (x', \pm 1), \quad \tau = \mu_+(p), \quad \tau' = \mu_+(p'), \quad t = \frac{x' - x}{2}. \tag{9.12}$$

We choose p and p' so that the pinwheel map is entirely defined.

From Equation 8.23, we have

$$\tilde{\beta}_m(\tau') - \tilde{\beta}_m(\tau) = \langle \widehat{t\lambda_k} \rangle + \sum_{j=1}^{m} \langle \xi_j \times (\tilde{\sigma}_j(\tau') - \tilde{\sigma}_j(\tau)) \rangle. \tag{9.13}$$

Here $\xi_1, ..., \xi_m$ are constants that depend on A. Let H be the matrix in Equation 8.8. We have $H(t,t,t) \equiv H(\tau' - \tau) \bmod \mathbf{Z}^3$ because $(t,t,t) \equiv \tau' - \tau \bmod \Lambda$. Our analysis in §8.2 shows that

$$\langle \widehat{t\lambda_k} \rangle = \langle \pi \circ H(t,t,t) - \epsilon t \rangle = \langle (\pi - \epsilon \pi_3) \circ H(\tau' - \tau) \rangle. \tag{9.14}$$

Here $\epsilon \in \{0, 1\}$ and π is some coordinate projection. The choice of ϵ and π depends on k. We now see that

$$\tilde{\beta}_m(\tau') = \tilde{\beta}_m(\tau) + \langle (\pi + \epsilon_3 \pi) \circ H(\tau' - \tau) \rangle + \sum_{j=1}^{m} \langle \xi_j \times (\tilde{\sigma}_j(\tau') - \tilde{\sigma}_j(\tau)) \rangle. \tag{9.15}$$

The right hand side is everywhere defined and locally affine. Hence we define $\tilde{\beta}_m$ on all of X_A using the right hand side of the last equation. $\square$

Now we come to a subtle point. We have shown that our functions are locally affine when restricted to each A-slice. We would like to remove this caveat and say simply that our functions are locally affine even when A is allowed to vary. The next result makes a weaker statement along these lines. Once we have this result, we will use a bootstrap argument to improve *analytic* to *affine*. Note that the set X, defined above, is an open convex polytope. Thus it makes sense to say that a function is analytic on X. Logically, we could give our overall proof without Lemma 9.3 below. However, Lemma 9.3 is a labor-saving device. The analyticity in Lemma 9.3 allows us to check the identities above on just a fairly small subset of X.

Lemma 9.3 *Suppose $X \subset R - S'_k$. Then σ_{k+1} is analytic on X.*

Proof: The constants ξ_j in Equation 9.13 vary analytically with A. Our argument in Lemma 9.2 therefore shows that σ_{k+1} is an affine function on X_A whose linear part varies analytically with A. We just have to check the additive term. Since X_A is connected, we can compute the additive term of σ_{k+1} at A from a single point. We choose $p = (\epsilon, 1)$, where ϵ is very close to 0. The fact that $A \to \sigma_{k+1}(p, A)$ varies analytically follows from the fact that the strips vary analytically. □

Equations 9.4, 9.5, and 9.6 are formulas for $\tilde{\sigma}_2$, $\tilde{\sigma}_3$, and $\tilde{\sigma}_4$, respectively. Let

$$f_{k+1} = \tilde{\sigma}_{k+1} - \sigma'_{k+1}, \qquad k = 2, 3, 4. \tag{9.16}$$

Here σ'_{k+1} is the right hand side of the identity for $\tilde{\sigma}_{k+1}$. Our goal is to show that $f_{k+1} \equiv \langle 0 \rangle$ for $k = 1, 2, 3$. Call a parameter A *good* if $f_{k+1} \equiv \langle 0 \rangle$ on X_A. Call a subset $S \subset (0, 1)$ *substantial* if S is dense in some open interval of $(0, 1)$.

Lemma 9.4 *$f_{k+1} \equiv 0$ provided that a substantial set of parameters is good.*

Proof: By hypothesis and by continuity, f_{k+1} vanishes on some open subset of X. But the 0-function is the only analytic function that can vanish on an open subset of X. □

In the next section we explain how to verify that a parameter is good. If f_{k+1} were a locally affine map from X_A into $\mathbf{R}$, we would just need to check that $f_{k+1} = 0$ on some tetrahedron on X_A to verify that A is a good parameter. Since the range of f_{k+1} is $\mathbf{R}/\mathbf{Z}$, we have to work a bit harder.

Before we launch into the method, we make one more remark about the details of the verification process. We want to be sure that, at each stage, we can actually apply Lemma 9.3. Here we explain why we can do this. Observe that, in general, we have

$$S_k \subset \tilde{\sigma}_{k+1}^{-1}(\langle 0 \rangle).$$

Given Equation 9.3, we see that $X \subset R - S'_1$. Hence σ_2 is defined on X. Hence σ_2 is analytic on X and locally affine on each X_A. We use these two properties to show that Equation 9.4 is true. But then $X \subset R - S'_2$, etc. So, we will know at each stage of our verification that Lemmas 9.2 and 9.3 apply to the function of interest.

9.4 IRRATIONAL QUINTUPLES

We will give a construction in $\mathbf{R}^3$. When the time comes to use the construction, we will identify X_A as an open subset of a copy of $\mathbf{R}^3$.

Let $\zeta_1, ..., \zeta_5 \in \mathbf{R}^3$ be 5 distinct points. By taking these points 4 at a time, we can compute 5 volumes $v_1, ..., v_5$. Here v_j is the volume of the tetrahedron obtained by omitting the jth point. We say that $(\zeta_1, ..., \zeta_5)$ is an *irrational quintuple* if there is no rational relation

$$\sum_{i=1}^{5} c_j\zeta_j = 0, \qquad c_j \in \mathbf{Q}, \qquad c_1c_2c_3c_4c_5 = 0. \tag{9.17}$$

If we allow all the constants to be nonzero, then there is always a relation.

Lemma 9.5 *Let C be an open convex subset of $\mathbf{R}^3$. Let $f\colon C \to \mathbf{R}/\mathbf{Z}$ be a locally affine function. Suppose that there is an irrational $(\zeta_1, ..., \zeta_5)$ such that $\zeta_j \in C$ and $f(\zeta_j)$ is the same for all j. Then f is constant on C.*

Proof: Since C is simply connected, we can lift f to a locally affine function $F\colon C \to \mathbf{R}$. But then F is affine on C, and we can extend F to be an affine map from $\mathbf{R}^3$ to $\mathbf{R}$. By construction, $F(\zeta_i) - F(\zeta_j) \in \mathbf{Z}$ for all i, j. Adding a constant to F, we can assume that F is linear. There are several cases.

Case 1: If $F(\zeta_j)$ is independent of j, then all the points lie in the same plane. Hence all the volumes are zero. This violates the irrationality condition.

Case 2: Suppose we are not dealing with case 1 and the following is true. For every index j there is a second index k such that $F(\zeta_k) = F(\zeta_j)$. Since there are 5 points total, this means that the set $\{F(\zeta_j)\}$ has a total of only 2 values. But this means that our 5 points lie in a pair of parallel planes $\Pi_1 \cup \Pi_2$, with 2 points in Π_1 and 3 points in Π_2. Let us say that that $\zeta_1, \zeta_2, \zeta_3 \in \Pi_1$ and $\zeta_4, \zeta_5 \in \Pi_2$. But then $v_4 = v_5$, and we violate the irrationality condition.

Case 3: If we are not dealing with the above two cases, then we can relabel so that $F(\zeta_1) \neq F(\zeta_j)$ for $j = 2, 3, 4, 5$. Let

$$\zeta_j' = \zeta_j - \zeta_1.$$

Then $\zeta_1' = (0, 0, 0)$ and $F(\zeta_1') = 0$. But then $F(\zeta_j') \in \mathbf{Z} - \{0\}$ for $j = 2, 3, 4, 5$. Note that $v_j' = v_j$ for all j. For $j = 2, 3, 4, 5$, let

$$\zeta_j'' = \frac{\zeta_j'}{F(\zeta_j')}.$$

Then $v_j''/v_j' \in \mathbf{Q}$ for $j = 2, 3, 4, 5$. Note that $F(\zeta_j'') = 1$ for $j = 2, 3, 4, 5$. Hence there is a plane Π such that $\zeta_j'' \in \Pi$ for $j = 2, 3, 4, 5$.

There is always a rational relation among the areas of the 4 triangles defined by 4 points in the plane. Hence there is a rational relation among $v_2'', v_3'', v_4'', v_5''$. But then there is a rational relation between v_2, v_3, v_4, v_5. This contradicts the irrationality condition. $\square$

9.5 VERIFICATION

We consider the (+) case first and discuss the (−) case at the end. Proceeding somewhat at random, we define

$$\phi_j = \Big(8jA + 1/(2j), 1\Big), \qquad j = 1, 2, 3, 4, 5. \tag{9.18}$$

We check that $\phi_j \in \Xi_+$ for A near $1/2$. Letting

$$\zeta_j = \mu_+(\phi_j), \tag{9.19}$$

we check that

$$f_{k+1}(\zeta_j) = \langle 0 \rangle, \qquad j = 1, 2, 3, 4, 5. \tag{9.20}$$

In the next section, we give an example calculation.

Lemma 9.6 $(\zeta_1, ..., \zeta_5)$ *form an irrational quintuple for a dense set of parameters* A. *In fact this is true for the complement of a countable set of parameters.*

Proof: The 5 volumes associated to our quintuple are as follows.

- $\upsilon_5 = 5/24 - 5A/12 + 5A^2/24$.
- $\upsilon_4 = 71/40 + 19A/20 - 787A^2/120 - 4A^3$.
- $\upsilon_3 = 119/60 + 7A/60 - 89A^2/15 - 4A^3$.
- $\upsilon_2 = -451/240 - 13A/40 + 1349A^2/240 + 4A^3$.
- $\upsilon_1 = -167/80 - 13A/40 + 533A^2/80 + 4A^3$.

If there is an open set of parameters for which the first 4 of these volumes has a rational relation, then there is an infinite set for which the same rational relation holds. Since every formula in sight is algebraic, this means that there must be a single rational relation that holds for all parameters. But then the parametrized curve $A \to (\upsilon_5, \upsilon_4, \upsilon_3, \upsilon_2)$ lies in a proper linear subspace of $\mathbf{R}^4$. We evaluate this curve at $A = 1, 2, 3, 4$ and see that the resulting points are linearly independent in $\mathbf{R}^4$. Hence there is no global rational relation. Hence, for a dense set of parameters, there is no rational relation among the first 4 volumes listed. A similar argument rules out rational relations among any other 4-tuple of these volumes. □

The (−) Case: Equations 9.4 and 9.5 do not change, except that μ_- replaces μ_+ and all the sets are defined relative to Ξ_- and μ_-. Equations 9.3 and 9.6 become

$$\sigma_1 = \left\langle \frac{1 - \pi_3}{2} \right\rangle \circ \mu_- \qquad \text{on } \Xi_-. \tag{9.21}$$

$$\sigma_4 = \left\langle \frac{A - \pi_1 - \pi_2 + \pi_3}{2} \right\rangle \circ \mu_- \qquad \text{on } \Xi_- - S_3'. \tag{9.22}$$

Lemmas 9.2 and 9.3 have the same proof in the (−) case. We use the same method as above, except that we use the points

$$\phi_j + (2, 0); \qquad j = 1, 2, 3, 4, 5. \tag{9.23}$$

These points all lie in Ξ_- for A near $1/2$.

9.6 AN EXAMPLE CALCULATION

Here we work out by hand one of the cases of Equation 9.20. We do the rest of the cases in Mathematica [**W**]. Consider the case $k = 1$ and $j = 1$.

When $A = 1/2$, the length spectrum for ϕ_1 starts out as $(1, 1, 2, 1)$. Hence this remains true for nearby A. Knowing the length spectrum allows us to compute, for instance, that

$$E_2E_1(\phi_1) = \phi_1 + V_1 + V_2 = \left(\frac{-3}{2} + 8A, 7\right) \in \Sigma_2$$

for A near $1/2$. The affine functional

$$(x, y) \to (x, y, 1) \cdot \frac{(-1, A, A)}{2 + 2A} \tag{9.24}$$

takes on the value 0 on the line $x = Ay+A$ and the value 1 on the line $x = Ay-2-A$. These are the two edges of Σ_2. (See §7.1.) Therefore

$$\sigma_2(\phi_1) = \left(\frac{-3}{2} + 8A, 7, 1\right) \cdot \frac{(-1, A, A)}{2 + 2A} = \frac{3}{4 + 4A}.$$

At the same time, we compute that

$$\mu_+(\phi_1) = (1/4)(-7 + 24A, 1 + 4A, -7 + 16A),$$

at least for A near $1/2$. When A is far from $1/2$, this point will not lie in R_A. We then compute

$$\frac{1 + A - \pi_2(\mu_+(\phi_1))}{1 + A} = \frac{3}{4 + 4A}.$$

This shows that $f_2(\zeta_1) = \langle 0 \rangle$ for all A near $1/2$.

This page intentionally left blank

Chapter Ten

Proof of the Master Picture Theorem

10.1 THE MAIN ARGUMENT

First we recall some notation from previous chapters.

- Let S be the singular set defined in Equation 9.2.
- Let $\widehat{S}$ denote the union of hyperplanes listed in Chapter 6.2.
- Let d denote distance on the polytope R.
- Let $\theta(p, A)$ be the quantity from the Torus Lemma in §8.

Below we will establish the following result.

Lemma 10.1 (Hyperplane) *$S \subset \widehat{S}$ and $\theta(p, A) \geq d(\mu_+(p, A), \widehat{S})$.*

The Hyperplane Lemma essentially says that the singular set is small and simple. Before we prove the Hyperplane Lemma, we will finish the proof of the Master Picture Theorem.

Say that a *ball of constancy* in $R - \widehat{S}$ is an open ball B with the following property. If (p_0, A_0) and (p_1, A_1) are two pairs and $\mu_+(p_j, A_k) \in B$ for $j = 0, 1$, then (p_0, A_0) and (p_1, A_1) have the same return pair. Here is a consequence of the Torus Lemma.

Corollary 10.2 *Any point τ of $R - \widehat{S}$ is contained in a ball of constancy.*

Proof: If τ is in the image of μ_+, this result is an immediate consequence of the Torus Lemma. In general, the image $\mu_+(\Xi_+ \times (0, 1))$ is dense in R. Hence we can find a sequence $\{\tau_n\}$ such that $\tau_n \to \tau$ and $\tau_n = \mu_+(p_n, A_n)$. Let $2\theta_0 > 0$ be the distance from τ to S. From the triangle inequality and the second statement of the Hyperplane Lemma,

$$\theta(p_n, A_n) \geq \theta_0 = \theta_1 > 0$$

for large n. By the Torus Lemma, τ_n is the center of a ball B_n of constancy whose radius depends only on θ_0. In particular – and this is really all that matters in our proof – the radius of B_n does not tend to 0. Hence, for n large enough, τ itself is contained in B_n. □

Lemma 10.3 *Let (p_0, A_0) and (p_1, A_1) be two points of $\Xi_+ \times (0,1)$ such that $\mu_+(p_0, A_0)$ and $\mu_+(p_1, A_1)$ lie in the same path-connected component of $R - \widehat{S}$. Then the return pair for (p_0, A_0) equals the return pair for (p_1, A_1).*

Proof: Let $L \subset R - \widehat{S}$ be a path joining points

$$\tau_0 = \mu_+(p_0, A_0), \qquad \tau_1 = \mu_+(p_1, A_1).$$

By compactness, we can cover L by finitely many overlapping balls of constancy. □

Now we just need to see that the Master Picture Theorem holds for one component of the partition of $R - \widehat{S}$. Here is an example calculation that does the job. For each $\alpha = j/16$, for $j = 1, ..., 15$, we plot the image

$$\mu_A(2\alpha + 2n), \qquad n = 1, ..., 2^{15}. \tag{10.1}$$

The image is contained in the slice $z = \alpha$. We see that the Master Picture Theorem holds for all these points. The reader can use Billiard King to plot and inspect millions of points for any desired parameter.

We have really proved only the half of the Master Picture Theorem that deals with Ξ_+ and μ_+. The proof of the half that deals with Ξ_- and μ_- is exactly the same. In particular, both the Torus Lemma and the Hyperplane Lemma hold verbatim in the $(-)$ case. The proof of the Hyperplane Lemma in the $(-)$ case differs only in that the two identities in Equation 9.21 replace Equations 9.3 and 9.6. We omit the details in the $(-)$ case.

10.2 THE FIRST FOUR SINGULAR SETS

The strip function identites make short work of the first four pieces of the singular set.

- Given Equation 9.3,

$$S_0 \subset \{z = 0\} \cup \{z = 1\}. \tag{10.2}$$

- Given Equation 9.4,

$$S_1 \subset \{y = 0\} \cup \{y = 1 + A\}. \tag{10.3}$$

- Given Equation 9.5,

$$S_2 \subset \{x = 0\} \cup \{x = 1 + A\}. \tag{10.4}$$

- Give Equation 9.6,

$$S_3 \subset \{x + y - z = 1 + A\} \cup \{x + y - z = -1 + A\}. \tag{10.5}$$

10.3 SYMMETRY

We use symmetry to deal with the remaining pieces. Suppose we start with a point $p \in \Xi_+$. We define the points $p = p_0, p_1, \ldots$ exactly as in Equation 8.2. However, this time we do not know a priori that all these points are defined. As we proceed in our analysis, we will see that these points are defined for increasingly large values of j. For the purpose of illustration, we will show the case when all points are defined.

Let ρ denote reflection in the x-axis. Then

$$\rho(\Sigma_{9-j}) = \Sigma_j, \qquad q_j = \rho(p_{9-j}), \qquad j = 1, 2, 3, 4. \tag{10.6}$$

Here we use the convention that indices repeat mod 8, as in previous chapters.

In Figure 10.1, the disk in the center is included for artistic purposes, to cover up some messy intersections. In the figure we show the coordinates for the vectors $-V_1$ and $-V_2$ to remind the reader of their values. It is convenient to write $-V_k$ rather than V_k because there are far fewer minus signs involved.

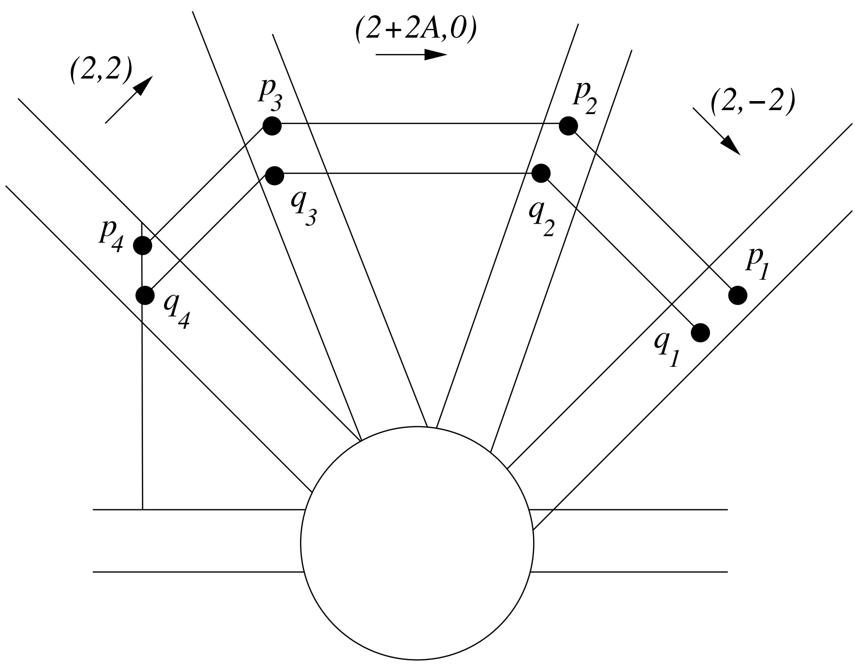

Figure 10.1: Reflected points.

Here is a notion we will use in our estimates. Say that a strip Σ *dominates* a vector V if we can translate V so that it is contained in the interior of the strip. This is equivalent to the condition that we can translate V so that one endpoint of V lies on $\partial\Sigma$ and the other lies in the interior.

10.4 THE REMAINING PIECES

10.4.1 The Set S_4

Suppose $p \in W_4$. Then p_5 and q_4 are defined, and $q_4 \in \partial\Sigma_4$. Given that

$$V_5 = (0, -4)$$

and the y-coordinates of all the points are odd integers, we have

$$p_4 - q_4 = (0, 2) + k(0, 4)$$

for some $k \in \mathbf{Z}$. Given that Σ_4 dominates $p_4 - q_4$, we have $k \in \{-1, 0\}$. Hence $p_4 = q_4 \pm (0, 2)$. If $p_5 \in \partial\Sigma_5$, then $q_4 \in \partial\Sigma_4$. Any vertical line intersects Σ_4 in a segment of length 4. From this we see that p_4 lies on the centerline of Σ_4. That is, $\sigma_4(p) = 1/2$. Given Equation 9.6, we have

$$S_4 \subset \{x + y - z = A\} \cup \{x + y - z = 2 + A\}.$$

10.4.2 The Set S_5

Suppose that $p \in W_5$. Then p_6 and q_3 are defined, and $q_3 \in \partial\Sigma_3$. Given that

$$V_6 = -V_4 = (-2, 2),$$

we see that

$$p_3 - q_3 = \epsilon(0, 2) + k(2, 2), \qquad \epsilon \in \{-1, 1\}, \qquad k \in \mathbf{Z}.$$

The criterion that Σ_3 dominates a vector (x, y) is that $|x + Ay| < 2 + 2A$.

Σ_3 dominates the vector $q_3 - p_3$. If $\epsilon = 1$, then

$$|2k + 2 + 2Ak| < 2 + 2A$$

forces $k \in \{-1, 0\}$. If $\epsilon = -1$, then the condition

$$|2k - 2 + 2Ak| < 2 + 2A$$

forces $k \in \{0, 1\}$. Hence $p_3 - q_3$ is one of the vectors $(\pm 2, 0)$ or $(0, \pm 2)$. Now we have a case-by-case analysis.

Suppose that q_3 lies in the right boundary of Σ_3. Then we have one of the following two conditions.

$$p_3 = q_3 - (2, 0), \qquad p_3 = q_3 + (0, 2).$$

Any horizontal line intersects Σ_3 in a strip of width $2 + 2A$. So, $\sigma_3(p)$ equals either $1/(1 + A)$ or $A/(1 + A)$, depending on whether or not $p_3 = q_3 - (2, 0)$ or $p_3 = q_3 + (0, 2)$. A similar analysis reveals the same two values when q_3 lies on the left boundary of Σ_3. Given Equation 9.5, we have

$$S_5 \subset \{x = A\} \cup \{x = 1\}.$$

10.4.3 The Set S_6

Suppose that $p \in W_6$. Then p_7 and q_2 are defined, and $q_2 \in \partial\Sigma_2$. We have

$$p_2 - q_2 = (p_3 - q_3) + k(2 + 2A, 0). \tag{10.7}$$

The criterion that Σ_3 dominates a vector (x, y) is that $|x - Ay| < 2 + 2A$.

Let $X_1, ..., X_4$ be the possible values for $p_3 - q_3$ as determined in the previous section. Using the values of the vectors X_j and the fact that Σ_2 dominates $p_2 - q_2$, we see that

$$p_2 - q_2 = X_j + \epsilon(2A, 2), \qquad \epsilon \in \{-1, 0, 1\}, \qquad j \in \{1, 2, 3, 4\}. \tag{10.8}$$

Note that the vector $(2A, 2)$ is parallel to the boundary of Σ_2. Hence, for the purpose of computing $\sigma_2(p)$, this vector plays no role. Essentially the same calculation as in the previous section now gives us the same choices for $\sigma_2(p)$ as we had for $\sigma_3(p)$ in the previous section. Given Equation 9.4, we have

$$S_6 \subset \{y = A\} \cup \{y = 1\}.$$

10.4.4 The Set S_7

Suppose that $p \in W_7$. Then p_8 and q_1 are defined, and $q_1 \in \partial\Sigma_1$. We have

$$p_1 - q_1 = (p_2 - q_2) + k(-2, 2). \tag{10.9}$$

Note that the vector $(2, 2)$ is parallel to Σ_1. For the purpose of finding $\sigma_1(p)$, we can do our computation mod $(2, 2)$. For instance, $(2, -2) \equiv (0, 4)$ mod $(2, 2)$. Given Equation 10.8, we have

$$p_1 - q_1 = \epsilon_1(0, 2) + \epsilon_2(2A, 2) + k(0, 4) \mod (2, 2). \tag{10.10}$$

Here $\epsilon_1, \epsilon_2 \in \{-1, 0, 1\}$. Given that any vertical line intersects Σ_1 in a segment of length 4, we see that the only choices for $\sigma_1(p)$ are

$$(k/2) + 2\epsilon A, \qquad \epsilon \in \{-1, 0, 1\}, \qquad k \in \mathbf{Z}.$$

Given Equation 9.3, we see that $S_7 \subset \{z = A\} \cup \{z = 1 - A\}$.

10.5 PROOF OF THE SECOND STATEMENT

Our analysis above establishes the first statement of the Hyperplane Lemma. For the second statement, suppose that $d(\mu_+(p, A), \widehat{S}) = \epsilon$. Given Equations 9.3–9.6, we have

$$\theta_j(p) \geq \epsilon, \qquad j = 1, 2, 3, 4.$$

Given our analysis of the remaining points using symmetry, the same bound holds for $j = 5, 6, 7, 8$. In these cases, $\theta_j(p, A)$ is a linear function of the distance from $\mu_+(p, A)$ to S_{j-1}, and the constant of proportionality is the same as for the index $9 - j$.

This page intentionally left blank

Part 3. Arithmetic Graph Structure Theorems

In this part of the book, we use the Master Picture Theorem to prove most of the structural results for the arithmetic graph that we quoted in Part 1.

- In Chapter 11, we prove the Embedding Theorem.
- In Chapter 12, we prove some results about the symmetries of the arithmetic graph and the hexagrid.
- In Chapter 13, we prove statement 1 of the Hexagrid Theorem, namely, that the arithmetic graph does not cross any floor lines.
- In Chapter 14, we prove a variant of statement 1 of the Hexagrid Theorem. We call the result the Barrier Theorem. Though we do not need this result until Part 6, the proof fits best right after the proof of statement 1 of the Hexagrid Theorem.
- In Chapter 15, we prove statement 2 of the Hexagrid Theorem, namely, that the arithmetic graph crosses the walls only near the doors. The two statements of the Hexagrid Theorem have similar proofs, though statement 2 has a more elaborate proof. We think of the proof of statement 2 of the Hexagrid Theorem as the main event in this part of the book. To make our argument go more smoothly, we defer a technical result, the Intersection Lemma, until the next chapter.
- In Chapter 16, we prove the Intersection Lemma, the technical result left over from the proof given in Chapter 15.

Many of the proofs in this part of the book require us to prove various disjointness results about some 4-dimensional polytopes. We will give short computer-aided proofs of these disjointness results. The proofs involve only a small amount of integer linear algebra. To help make the proofs surveyable, we will include computer images of 2 dimensional slices of our polytopes. These figures, all reproducible on Billiard King, serve as sanity checks for the computer calculations. We will include many figures from Billiard King, but it usually goes without saying that the reader can see much more using the program.

This page intentionally left blank

Chapter Eleven

Proof of the Embedding Theorem

11.1 NO VALENCE 1 VERTICES

Let $\widehat{\Gamma} = \widehat{\Gamma}_\alpha(A)$ be the arithmetic graph for a parameter A and some $\alpha \not\in 2\mathbf{Z}[A]$. The reader will see from our proof that the choice of α is not important. As a first step in the proof of the Embedding Theorem, we show that all nontrivial vertices of $\widehat{\Gamma}$ have valence 2. Dynamically, a vertex of valence 1 corresponds to a point $x \in \Xi$ such that $x \neq \Psi(x) = \Psi^{-1}(x)$.

Let $p \in \mathbf{Z}^2$ be a nontrivial vertex of $\widehat{\Gamma}$. Let q_+ and q_- be the two neighbors of p. We would like to show that $\widehat{\Gamma}$ has valence 2 at p. If this fails, then we must have

$$p \neq q_+ = q_-. \tag{11.1}$$

This means that the maps M_+ and M_- from §6.6 assign the same vector to p. Put another way, this situation occurs iff there is some nontrivial $(\epsilon_1, \epsilon_2) \in \{-1, 0, 1\}$ such that

$$\Lambda R_+(\epsilon_1, \epsilon_2) \cap \big(R_-(\epsilon_1, \epsilon_2) + (1, 1, 0, 0)\big) \neq \emptyset. \tag{11.2}$$

A visual inspection and/or a computer search reveals that at least one of the two sets above is empty unless (ϵ_1, ϵ_2) is one of

$$(1, 1), \qquad (-1, -1), \qquad (1, 0), \qquad (-1, 0). \tag{11.3}$$

It follows from Equation 6.17 that Equation 11.2 holds for (ϵ_1, ϵ_2) if and only if it holds for $(-\epsilon_1, -\epsilon_2)$. Thus we have to deal just with the pairs $(1, 1)$ and $(1, 0)$.

Below we will give a formal argument, based on a small amount of machine computation, that rules out the above kind of intersection. Before we do this, however, we will show some convincing pictures of the relevant sets. As in §6.3, we show (z, A) slices of polytopes in R_+ and R_-. We draw the slices of R_+ with dark shading and the slices of R_- with light shading. Let B_j denote the jth component of the base space B, as in Figure 6.2.

Over the regions B_2 and B_3, at least one of $R_+(1, 1)$ or $R_-(1, 1)$ is empty. Figure 11.1 shows typical slices of

$$\Lambda R_+(1, 1), \qquad \Lambda(R_-(1, 1) + (1, 1, 0, 0))$$

over B_0 and B_1. In all cases, we see that the interiors of the two kinds of pieces are disjoint from each other.

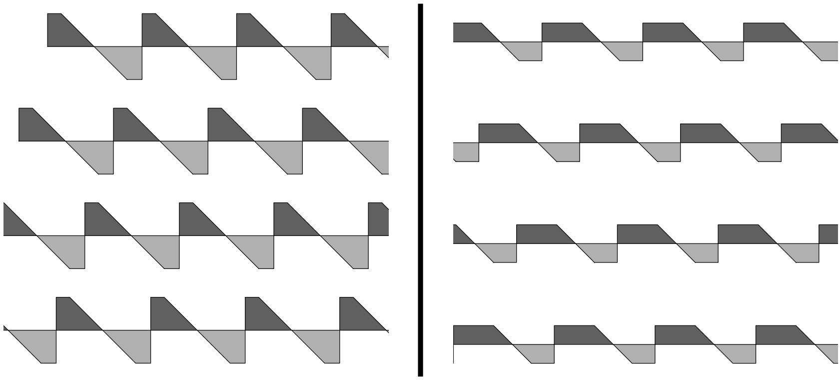

Figure 11.1: Slices of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,1)+(1,1,0,0))$.

Figure 11.2 shows typical slices of

$$\Lambda R_+(1,0), \qquad \Lambda(R_-(1,0)+(1,1,0,0))$$

over each of the regions B_0, B_1, B_2, B_3. We see the same disjoint interiors as above.

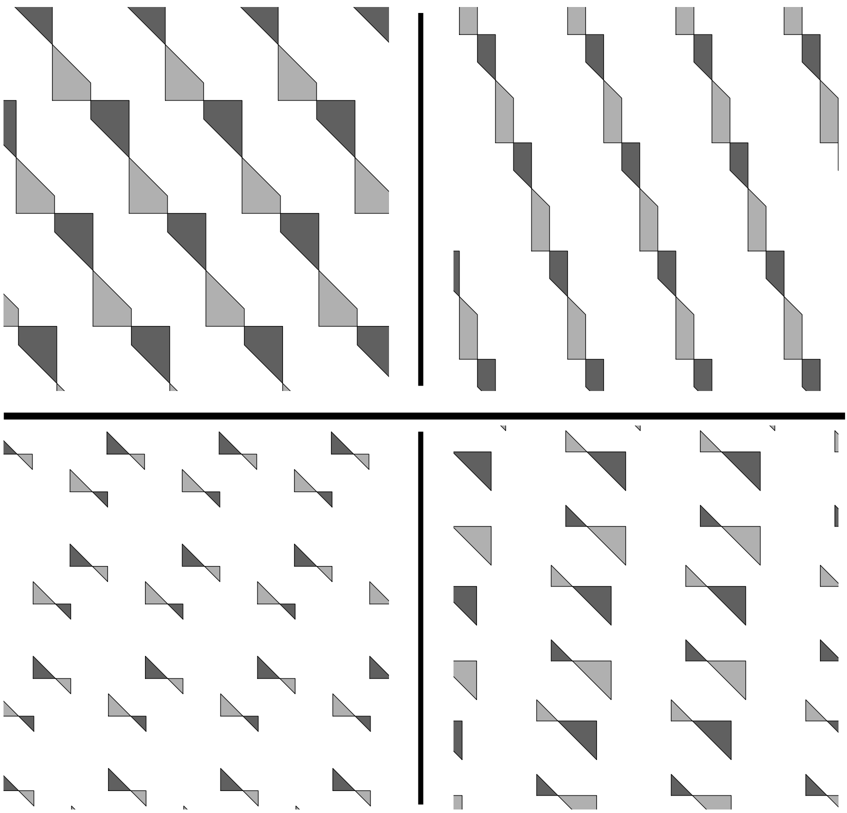

Figure 11.2: Slices $\Lambda R_+(1,0)$ and $\Lambda(R_-(1,0)+(1,1,0,0))$.

Now we give a formal argument. We work in $\mathbf{R}^4$, as discussed in §6.7. All the polytopes of interest are convex integral polytopes. To rule out Equation 11.2, we need to consider all possible pairs (P_1, P_2) of integral convex polytopes such that

$$P_1 \subset \Lambda R_+(\epsilon_1, \epsilon_2), \qquad P_2 \subset R_-(\epsilon_1, \epsilon_2) + (1, 1). \tag{11.4}$$

We hold the second polytope fixed and move the first one around by the action of the entire lattice. At first it looks as if we have an infinite calculation, but actually we will reduce the problem to a finite calculation.

Recall that Λ is generated by the three elements $\gamma_1, \gamma_2, \gamma_3$. Let $\Lambda' \subset \Lambda$ denote the subgroup generated by γ_1 and γ_2. We also define $\Lambda'_{10} \subset \Lambda'$ by the equation

$$\Lambda'_{10} = \{a_1\gamma_1 + a_2\gamma_2 | \, |a_1|, |a_2| \le 10\}. \tag{11.5}$$

Lemma 11.1 *Let* $\gamma \in \Lambda - \Lambda'_{10}$.

$$P_1 = \gamma(Q_1), \qquad Q_1 \subset R_+(\epsilon_1, \epsilon_2), \qquad P_2 \subset R_-(\epsilon_1, \epsilon_2) + (1, 1, 0, 0).$$

Then P_1 *and* P_2 *have disjoint interiors.*

Proof: If $\gamma \notin \Lambda'$, then the third coordinates of points in P_1 lie in $[n, n+1]$ for some integer $n \neq 0$. On the other hand, the third coordinates of points in P_2 lie in $[0, 1]$. Hence P_1 and P_2 have disjoint interiors in this case. This means that we have to worry only about elements of Λ'.

Suppose now that $\gamma \in \Lambda' - \Lambda'_{10}$. In this case, Q_1 is contained in the ball of radius 4 about P_2, but γ moves this ball entirely off itself. □

Now we have a finite problem. Given

$$\gamma \in \Lambda'_{10}, \quad P_1 = \gamma(Q_1), \quad Q_1 \subset R_+(\epsilon_1, \epsilon_2), \quad P_2 \subset R_-(\epsilon_1, \epsilon_2) + (1, 1, 0, 0),$$

we produce a vector

$$w = w(P_1, P_2) \in \{-1, 0, 1\}^4 \tag{11.6}$$

such that

$$\max_{v \in \mathrm{vtx}(P_1)} v \cdot w \le \min_{v \in \mathrm{vtx}(P_2)} v \cdot w. \tag{11.7}$$

This means that a hyperplane separates the interior of P_1 from P_2. In each case we find $v(P_1, P_2)$ by a short computer search and perform the verification using arithmetic with integers.

Remark: It seems rather lucky that we could find such simple hyperplanes separating the polytopes. However, every coordinate of every polytope lies in $\{0, 1, 2\}$, and the relevant pairs of polytopes often have several pairs of vertices in common. This situation makes the existence of the very simple separating hyperplanes less surprising.

11.2 NO CROSSINGS

Given that every nontrivial vertex of $\widehat{\Gamma}$ has valence 2, and also that the edges of $\widehat{\Gamma}$ have length at most $\sqrt{2}$, the only way that $\widehat{\Gamma}$ can fail to be embedded is if there is a situation like the one shown in Figure 11.3.

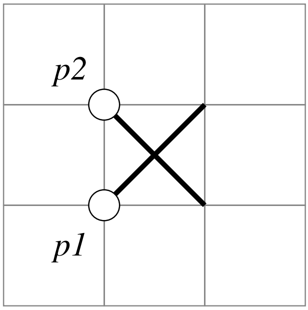

Figure 11.3: Embedding failure.

Let M_+ and M_- be the maps from §6.6. Given the Master Picture Theorem, this situation arises only if

$$M_\pm(p_1) \in R_\pm(1,1), \qquad M_\pm(p_2) \in R_\pm(1,-1) \tag{11.8}$$

This equation implicitly involves 4 cases, depending on the sign choices. Since $p_2 = p_1 + (0,1)$, we have

$$M_\pm(p_2) = M_\pm(p_1) + (1,1,1,0) \bmod \Lambda. \tag{11.9}$$

In particular, the two points $M(p_1)$ and $M(p_2)$ lie in the same fiber of R over the (z, A) square. We see by inspection that no fiber intersects both $R_+(1,1)$ and $R_+(1,-1)$. In light of the nature of the partition, we need to only check 4 fibers. (See the discussion following Figure 6.2.) This rules out the $(+,+)$ case. The same check rules out the $(-,-)$ and $(-,+)$ cases. The only possibility is

$$M_+(p_1) \in R_+(1,1), \qquad M_-(p_2) \in R_-(1,-1). \tag{11.10}$$

Modulo Λ, we have

$$M_-(p_2) \equiv M_-(p_1) + (1,1,1,0) \equiv M_+(p_1) + (0,0,1,0) \equiv M_+(p_1) + (1,1,0,0).$$

In short,

$$M_+(p_1) \equiv M_-(p_2) - (1,1,0,0) \quad \bmod \Lambda. \tag{11.11}$$

Letting $x \in \mathbf{R}^4$ be any representative of $M_+(p_1)$, we see that the orbit Λx intersects both sets

$$R_+(1,1), \qquad R_-(1,-1) - (1,1,0,0).$$

Hence

$$\Lambda R_+(1,1) \cap (R_-(1,-1) - (1,1,0,0)) \neq \emptyset. \tag{11.12}$$

We mean that there is a pair (P_1, P_2) of polytopes, with P_1 in the first set and P_2 in the second set, such that P_1 and P_2 do not have disjoint interiors. We rule out this intersection using exactly the same method as in step 2.

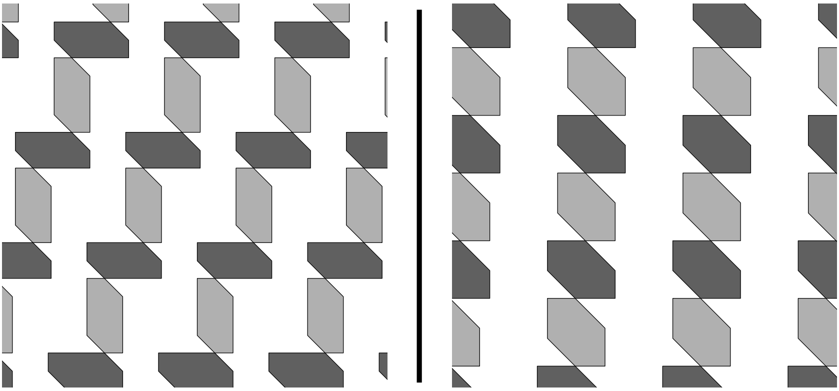

Figure 11.4: Slices of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,-1)-(1,1,0,0))$.

Here is an illustration just like Figures 11.1 and 11.2. Figure 11.4 shows slices of

$$\Lambda R_+(1,1), \qquad \Lambda(R_-(1,-1)-(1,1,0,0))$$

over B_2 and B_3. Over B_0 and B_1, at least one of the sets is empty.

This page intentionally left blank

Chapter Twelve

Extension and Symmetry

12.1 TRANSLATIONAL SYMMETRY

Referring to §6.6, the maps M_+ and M_- are defined on all of $\mathbf{Z}^2$. This gives an extension of the arithmetic graph to all of $\mathbf{Z}^2$. We denote this full extension by $\widehat{\Gamma}$.

Figure 12.1 shows $\widehat{\Gamma}(3/7)$, as well as the hexagrid $G(3/7)$, from §3.1. The bottom of the shaded parallelogram is the baseline. In the rational case, both the arithmetic graph and the hexagrid are invariant under a certain lattice Θ of translations of $\mathbf{Z}^2$. The shaded parallelogram is the fundamental domain for Θ. In this section we give the formulas for the lattice and establish the translational symmetry.

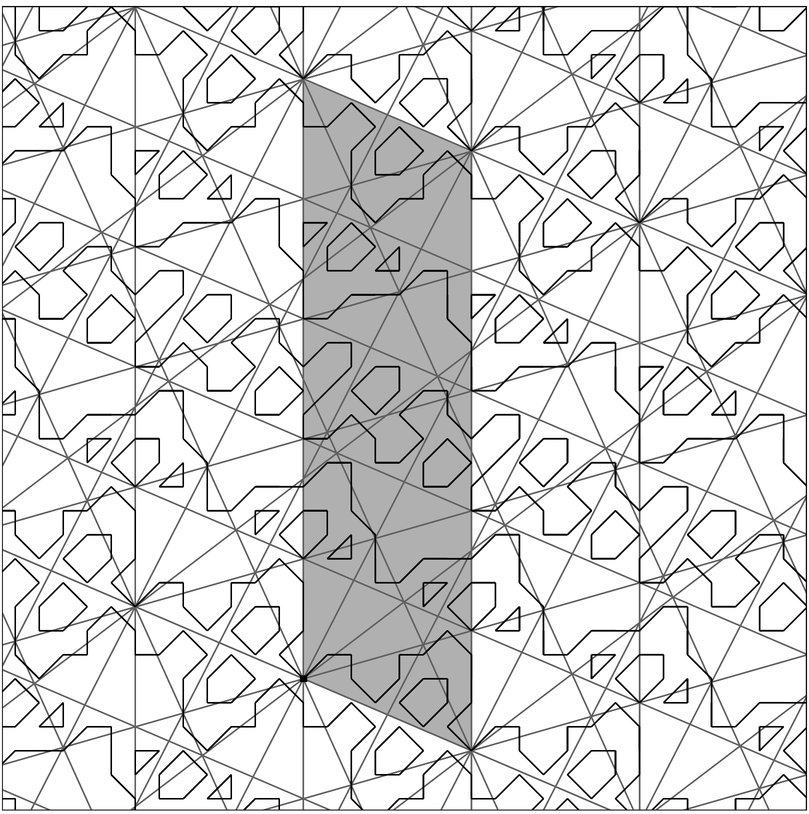

Figure 12.1: $\widehat{\Gamma}(3/7)$ and $G(3/7)$.

Lemma 12.1 *The extended arithmetic graph does not cross the baseline.*

Proof: The arithmetic graph describes the dynamics of the pinwheel map Φ. Note that Φ is generically defined and invertible on $\mathbf{R}_+ \times \{-1, 1\}$. Reflection in the x-axis conjugates Φ to Φ^{-1}. By the Pinwheel Lemma, Φ maps $\mathbf{R}_+ \times \{-1, 1\}$ into itself. By symmetry the same goes for Φ^{-1}. Hence Φ and Φ^{-1} also map $\mathbf{R}_- \times \{-1, 1\}$ into itself. If some edge of $\widehat{\Gamma}$ crosses the baseline, then one of Φ or Φ^{-1} will map a point of $\mathbf{R}_+ \times \{-1, 1\}$ into $\mathbf{R}_- \times \{-1, 1\}$. This is a contradiction. $\square$

Let $\lambda(p/q) = 1$ if p/q is odd and let $\lambda(p/q) = 2$ if p/q is even. Define

$$\Theta = \mathbf{Z}V + \mathbf{Z}V', \qquad V' = \lambda^2\left(0, \frac{(p+q)^2}{4}\right), \qquad \lambda = \lambda(p/q). \tag{12.1}$$

Referring to Figure 12.1, the short edges of the parallelogram are translates of V and the long edges are translates of V'. Thus the shaded parallelogram is a fundamental domain for the action of Θ on $\mathbf{R}^2$.

Lemma 12.2 *The arithmetic graph $\widehat{\Gamma}(p/q)$ is invariant under Θ.*

Proof: We will consider the odd case. The even case is similar. We have already seen that $\widehat{\Gamma}$ is invariant under V. We just have to show invariance for V'. Referring to the notation in §6.6, we have

$$M_\pm(x + V') - M_\pm(x) = (t, t, t) \mod \Lambda, \qquad t = \frac{(p+q)^2}{4}. \tag{12.2}$$

By the Master Picture Theorem, it suffices to prove that $(t, t, t) \in \Lambda$. Setting

$$a = pq, \qquad b = \frac{pq + q^2}{2}, \qquad c = t, \tag{12.3}$$

we express (t, t, t) as an integer combination of vectors in Λ as follows.

$$a\begin{bmatrix} 1+A \\ 0 \\ 0 \end{bmatrix} + b\begin{bmatrix} 1-A \\ 1+A \\ 0 \end{bmatrix} + c\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} t \\ t \\ t \end{bmatrix}. \tag{12.4}$$

This completes the proof. $\square$

Remark: One can probably also see rotational symmetry by looking at Figure 12.1. We will treat this kind of symmetry below.

Our next result deals with the hexagrid and the arithmetic kite $\mathcal{K}(A)$. Both objects are defined in §3.1. Recall that the hexagrid consists of a room grid RG and a door grid DG. Here RG is composed of 2 families of parallel lines and DG is composed of 4 families of parallel lines. The lines of RG are all parallel to the two diagonals of $\mathcal{K}(A)$, and the lines of DG are all parallel to the sides of $\mathcal{K}(A)$. Referring to Figure 12.1, notice that each corner of the shaded parallelogram lies on 6 lines – one per family – of the hexagrid. Our proof of the following result is based on this phenomenon.

Lemma 12.3 *The hexagrid is invariant under the action of* Θ.

Proof: Again, we treat the only odd case. Let $G = G(p/q)$ denote the hexagrid. By construction, G is invariant under translation by V. We just have to show that the same holds for V'. We will show that V' contains 6 lines of G. Translation by V' then maps each family of parallel lines in the hexagrid to itself, and so the whole hexagrid is invariant.

Let W be as in Equation 3.2. For convenience, we repeat the formula for W.

$$W = \left(\frac{pq}{p+q}, \frac{pq}{p+q} + \frac{q-p}{2}\right).$$

We compute that

$$V' = -\frac{p}{2}V + \frac{p+q}{2}W. \tag{12.5}$$

The second coefficient is an integer. Given that the room grid RG is invariant under the lattice $\mathbf{Z}[V/2, W]$, the room grid RG is also invariant under translation by V'. This gives 2 lines L_1 and L_2, one from each family of RG.

Note that the door grid DG is invariant only under $\mathbf{Z}[V]$, so we have to work harder. We need to produce 4 lines of DG that contain V'. Here they are.

- The vertical line L_3 through $(0, 0)$ certainly contains V'. This line extends the bottom left edge of $\mathcal{K}(A)$ and hence belongs to DG.

- Let L_4 be the line containing V' and the point
$$\frac{-(p+q)}{2}V \in \mathbf{Z}[V].$$
We compute that the slope of L_4 coincides with the slope of the top left edge of $\mathcal{K}(A)$. The origin contains a line of DG parallel to the top left edge of $\mathcal{K}(A)$, and hence every point in $\mathbf{Z}[V]$ contains such a line. Hence L_4 belongs to DG.

- Let L_5 be the line containing V' and the point
$$-pV \in \mathbf{Z}[V].$$
We compute that the slope of L_5 coincides with the slope of the bottom right edge of $\mathcal{K}(A)$. The same argument as in the previous case shows that L_5 belongs to DG.

- Let L_6 be the line containing V' and the point
$$\frac{q-p}{2}V \in \mathbf{Z}[V].$$
We compute that the slope of L_6 coincides with the slope of the top right edge of $\mathcal{K}(A)$. The same argument as above shows that L_6 belongs to DG.

The lines $L_1, ..., L_6$ are the desired lines. $\square$

12.2 A CONVERSE RESULT

Here we show that Θ is, in some sense, the maximal group of translational symmetries of the arithmetic graphs. Let $M_{\pm}$ be the map from the Master Picture Theorem. We state our result for the map M_+, but the argument is the same for M_-.

Lemma 12.4 *Let $v_1, v_2 \in \mathbf{Z}^2$. Then $M_+(v_1) \equiv M_+(v_2)$ mod Λ iff $v_1 \equiv v_2$ mod Θ.*

Proof: As usual, $A = p/q$. The proof of Lemma 12.2 shows that $v_1 \equiv v_2$ mod Θ implies $M_+(v_1) \equiv M_+(v_2)$ mod Λ. We must establish the converse. Suppose that $M_+(v_1) \equiv M_+(v_2)$ mod Λ. Let

$$w = v_2 - v_2 = (m, n). \tag{12.6}$$

Our hypothesis implies that

$$(t, t, t) \in \Lambda, \qquad t = Am + n. \tag{12.7}$$

We would like to see that this equation implies that $w \in \Theta$. Recall that Λ is the $\mathbf{Z}$-span of the columns of the matrix in Equation 6.3. The bottom row of this matrix is $(0, 0, 1)$. From this we conclude that $t \in \mathbf{Z}$. Since

$$t = \frac{pm}{q} + n, \tag{12.8}$$

we see that q divides m. But now we can subtract multiples of $V = (q, -p)$ to arrange that $m = 0$. That is, we can assume that $w = (0, n)$. Hence $t = n$. Note that

$$(n, n, n) \equiv (2n, 2n, 0) \mod \Lambda. \tag{12.9}$$

Therefore we have the equation

$$\begin{bmatrix} 2n \\ 2n \end{bmatrix} = a \begin{bmatrix} 1+A \\ 0 \end{bmatrix} + b \begin{bmatrix} 1-A \\ 1+A \end{bmatrix}. \tag{12.10}$$

The solutions are

$$a = \frac{4npq}{(p+q)^2}, \qquad b = \frac{2nq}{p+q}. \tag{12.11}$$

Since p and q are relatively prime, pq is relatively prime to $(p+q)^2$. Since $a \in \mathbf{Z}$, we have that $(p+q)^2$ divides $4n$. Hence

$$n = k\frac{(p+q)^2}{4}, \qquad k \in \mathbf{Z}. \tag{12.12}$$

When p/q is odd, we have $w = kV'$, by Lemma 12.1. When p/q is even, the fact that $n \in \mathbf{Z}$ forces $k = 4k'$ for some $k' \in \mathbf{Z}$. Hence $w = k'V'$ in this case. $\square$

Lemma 12.4 has the following immediate corollary.

Corollary 12.5 *The maps M_+ and M_- from the Master Picture Theorem are well defined and injective on $\mathbf{Z}^2/\Theta$.*

12.3 ROTATIONAL SYMMETRY

Let p/q be an odd rational. Let p_+/q_+ be as in Equation 4.1. Let ι be the rotation

$$\iota(m,n) = V_+ - (m,n), \qquad V_+ = (q_+, -p_+). \tag{12.13}$$

The fixed point of ι is $(1/2)V_+$. This point lies very close to the baseline of $\widehat{\Gamma}(p/q)$. Figure 12.2 shows $\Gamma(7/17)$ centered on this fixed point.

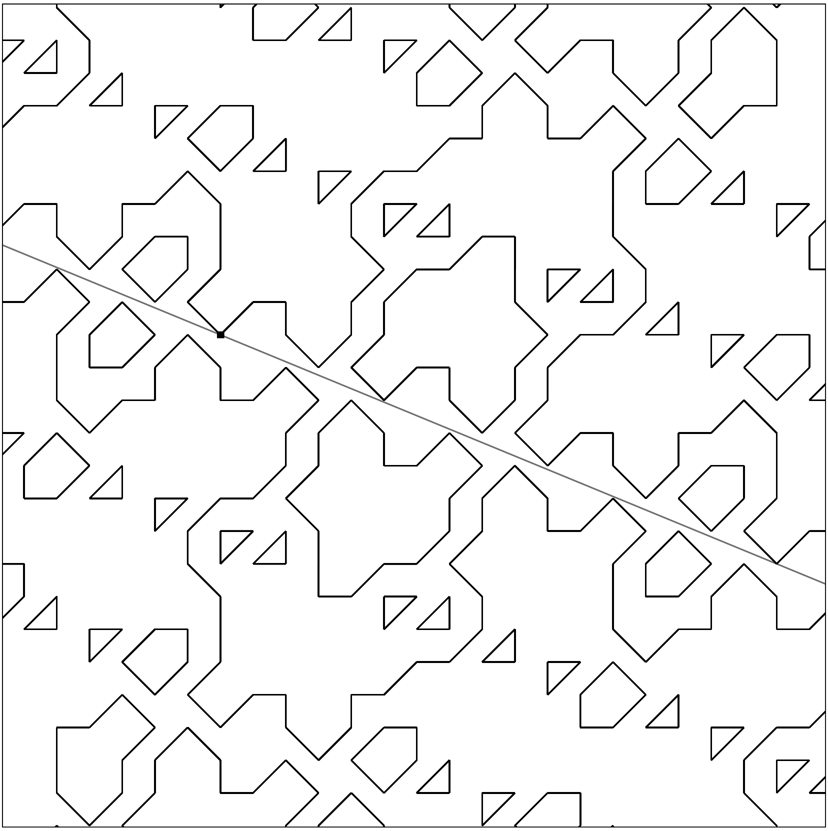

Figure 12.2: $\widehat{\Gamma}(7/17)$ centered on the point $(12, -5)/2$.

Below we prove that $\iota(\widehat{\Gamma}) = \widehat{\Gamma}$, as suggested by Figure 12.2. Combining this result with the translation symmetry above, we see that rotation by π about any of the points

$$\beta + \theta, \qquad \beta = (1/2)V_+, \qquad \theta \in \Theta \tag{12.14}$$

is a symmetry of $\widehat{\Gamma}$.

Remark: In particular, there is an involution swapping $(0,0)$ and $V_+ + kV$ for any $k \in \mathbf{Z}$.

Lemma 12.6 $\iota(\widehat{\Gamma}) = \widehat{\Gamma}$.

Proof: Let M_+ and M_- be as in §6.6. We use the offset value $\alpha = 1/(2q)$. Recall that R_A is the fundamental domain for the action of $\Lambda = \Lambda_A$. Let ρ be reflection through the midpoint of the space R_A. Below we will derive the equations

$$M_+(m,n) = \rho \circ M_-(\iota(m,n)), \qquad M_-(m,n) = \rho \circ M_+(\iota(m,n)). \quad (12.15)$$

Given these equations, we verify by inspection that our partition of R_A is symmetric under ρ and has the labels appropriate to force the type determined by

$$\rho \circ M_+(m,n), \quad \rho \circ M_-(m,n)$$

to be the 180-degree rotation of the type forced by

$$M_-(m,n), \quad M_+(m,n).$$

Indeed, we can determine this with an experiment performed on any rational that is complicated enough such that all regions are sampled.

Now we derive Equation 12.15. We will derive just the first half. The derivation of the second half is entirely similar. We have

$$M_+(m,n) = (t, t+1, t) \mod \Lambda, \qquad t = \frac{pm}{q} + n + \frac{1}{2q}. \quad (12.16)$$

Next, using the fact that $q_+ p - p_+ q = -1$, we have

$$M_-(\iota(m,n)) = (t'-1, t', t') \mod \Lambda,$$

$$t' = \left(\frac{pq_+}{q} - p_+\right) - \left(\frac{pm}{q} + n\right) + \frac{1}{2q} = -\left(\frac{pm}{q} + n\right) - \frac{1}{2q} = -t.$$

In short

$$M_-(\iota(m,n)) = (-t-1, -t, -t) \mod \Lambda. \quad (12.17)$$

We compute easily that $(2+A, A, 1) \in \Lambda$. Hence the points

$$x = (t, t+1, t), \qquad y = (-t-1, -t, -t) + (2+A, A, 1) \quad (12.18)$$

are representatives of the points $M_+(m,n)$ and $M_-(\iota(m,n))$ in $\mathbf{R}^3$. We compute the average.

$$\frac{x+y}{2} = \frac{1}{2}(1+A, 1+A, 1).$$

This is the midpoint of R_A. But then ρ interchanges x and y. Since ρ preserves the elements of Λ, we see that ρ interchanges the full orbits Λx and Λy. But then ρ interchanges $\Lambda x \cap R_A$ with $\Lambda y \cap R_A$. But these two points are $M_+(m,n)$ and $M_-(\iota(m,n))$. This establishes the first half of Equation 12.15. $\square$

12.4 NEAR-BILATERAL SYMMETRY

Our pictures of arithmetic graphs show near-bilateral symmetry. In this section we explain how this arises. Looking at Figure 12.2, we see that there is a natural correspondence between components above the baseline and components below the baseline. Our first result explains this near-bilateral symmetry. There is a second kind of bilateral symmetry that meets the eye in Figure 1.5 or 12.3. After proving our first result, we will explain how this other kind of near-bilateral symmetry arises.

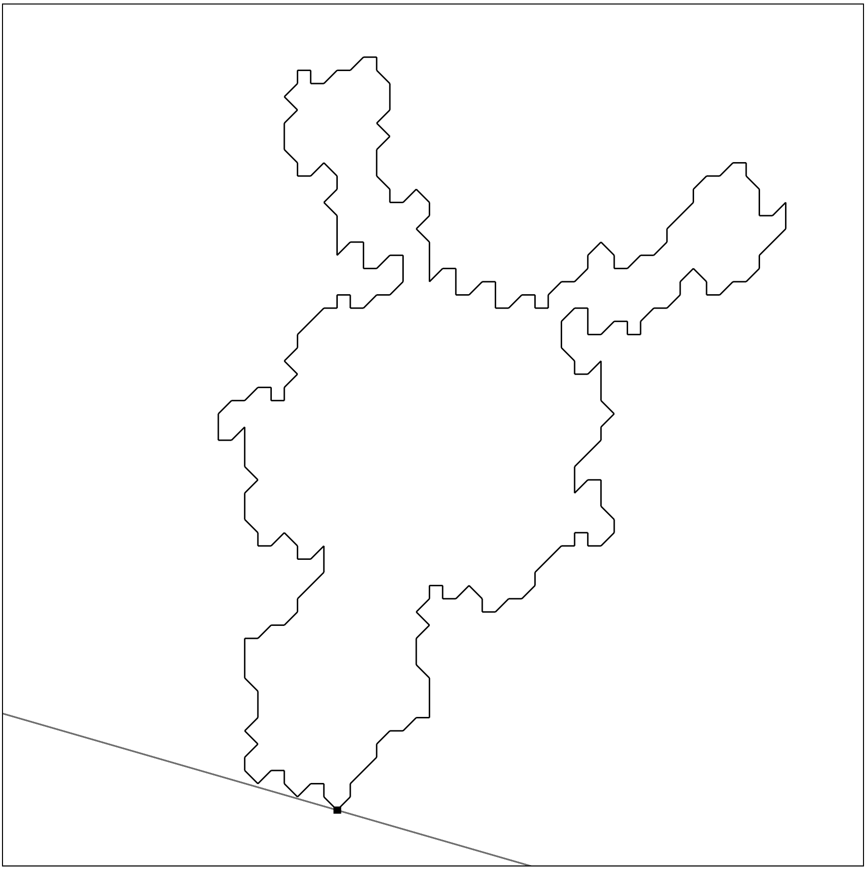

Figure 12.3: $\Gamma(15/52)$.

We say that a map $\mathbf{J}$ from $\widehat{\Gamma}$ to $\widehat{\Gamma}$ is a *combinatorial isomorphism* if $\mathbf{J}$ maps vertices to vertices and edges to edges.

Recall that a low vertex is one that is above the baseline but within 1 vertical unit of it. Say that a *low component* is a component of $\widehat{\Gamma}$ above the baseline that contains a low vertex. Say that this component is *odd* if it contains odd low vertices, and *even* if it contains even low vertices. By Lemma 2.6, this notion is well defined.

Lemma 12.7 *For any rational A, there is a combinatorial isomorphism $\mathbf{J}\colon \widehat{\Gamma} \to \widehat{\Gamma}$ that swaps the components of $\widehat{\gamma}$ above the baseline with the one below the baseline.*

Proof: Let $\Xi_\pm = \mathbf{R}_\pm \times \{-1, 1\}$. Recall that $\Psi\colon \Xi_+ \to \Xi_+$ is the first return map. We can extend Ψ so that it is also the return map from Ξ_- to Ξ_-. We have proved the Return Lemma and the Pinwheel Lemma for the return map to Ξ_+, but essentially the same arguments work with Ξ_- in place of Ξ_+. Thus the portion of $\widehat{\Gamma}$ below the baseline tracks the dynamics of the map $\Psi\colon \Xi_- \to \Xi_-$ just as the portion above the baseline tracks the dynamics of $\Psi\colon \Xi_+ \to \Xi_+$.

Let $\Psi^{1/2}$ be the first return map to $\mathbf{R} \times \{-1, 1\}$. If $p \in \Xi_\pm$, then $\Psi^{1/2}(p) \in \Xi_\mp$. The correspondence $\xi \to \Psi^{1/2}(\xi)$ gives a bijection between Ψ-orbits in Ξ_+ and Ψ-orbits in Ξ_-. The map Ψ is the square of $\Psi^{1/2}$. We define $\mathbf{J}_+(m,n) = (m', n')$, where (m,n) corresponds to ξ and (m', n') corresponds to $\Psi^{1/2}(\xi)$.

We could set $\mathbf{J} = \mathbf{J}_+$ and be finished, but we can somewhat improve the construction. There is a second involution that is just as good as $\mathbf{J}_+$. We can match $\xi \in \Xi_+$ to the point $\Psi^{-1/2}(\xi) \in \Xi_-$. Call this map $\mathbf{J}_-$. Both $\mathbf{J}_+$ and $\mathbf{J}_-$ have the same action on *components*, but they have different actions on individual points.

If γ is a component of $\widehat{\Gamma}$ above the baseline that is not low, we use (say) $\mathbf{J} = \mathbf{J}_+$. For even low components we use $\mathbf{J} = \mathbf{J}_+$. For odd low components, we use $\mathbf{J} = \mathbf{J}_-$. This is our combinatorial isomorphism. □

Lemma 12.7 does not really explain the near-bilateral symmetry we see in Figure 12.3. Here is the explanation. Let ι be the symmetry discussed in the previous section. Then $\iota \circ \mathbf{J}$ permutes the components of $\widehat{\Gamma}$ above the baseline. In particular, $\iota \circ \mathbf{J}$ preserves Γ but reverses its direction. This is the symmetry seen in Figure 12.3.

Now we work out a few more properties of $\mathbf{J}$. Our first result really uses the improved version of $\mathbf{J}$.

Lemma 12.8 *If v is a low vertex, then $\mathbf{J}(v) = v - (0, 1)$.*

Proof: Let M be the fundamental map. Let (m,n) be an even low vertex. Let

$$(x, -1) = M(m,n) \in (0,2) \times \{-1\}.$$

We compute

$$\Psi^{1/2}(x,-1) = \psi^2(x,-1) = (x-2, 1) = M(m, n-1). \tag{12.19}$$

Hence $\mathbf{J}(m,n) = (m, n-1)$. Similarly, if (m,n) is an odd low vertex, then

$$\Psi^{-1/2}(x,1) = \psi^{-2}(x,1) = (x-2,-1) = M(m,n-1). \tag{12.20}$$

Hence $\mathbf{J}(v) = v - (0,1)$ when v is a low vertex. □

We say that $\mathbf{J}$ is *pseudolinear* if there is a linear isomorphism $J\colon \mathbf{R}^2 \to \mathbf{R}^2$ such that J is a bounded distance from $\mathbf{J}$ (in the sup norm.) If J exists, J is unique. We call J the *model* for $\mathbf{J}$. Since we do not need the final result for any purpose, the proof will be a bit sketchy.

Lemma 12.9 $\mathbf{J}$ *is pseudolinear, modelled on the affine map J such that $J(V) = V$ and $J(W) = -W$. Here V and W are as in Equation 3.2.*

Proof: (Sketch) Letting $(x, 1)$ be a point on Ξ_+ about N units from the origin, we roughly trace out the Pinwheel map. First we add some integer multiple of the vector $(0, 4)$, then we add some integer multiple of the vector $(-2, 2)$, etc. When we reach Ξ_- we have a vector of the form

$$x + (2Ac_N + 2d_N, \pm 1).$$

Here (c_N, d_N) depends linearly on N up to a uniformly bounded error. Given a point $v = (m, n)$, we have

$$\mathbf{J}(v) = v + (c_N, d_N), \qquad N = 2Am + 2n. \tag{12.21}$$

This shows that $\mathbf{J}$ is pseudolinear.

Let J be the linear map on which $\mathbf{J}$ is modelled. Given the action of $\mathbf{J}$ on low vertices, we see that $J(V) = V$. To show that $J(W) = -W$, we consider how the pair (c_k, d_k) associated to kW depends on k. Taking the limit as $k \to \infty$, we get an exact formula that shows $J(W) = -W$. We omit the details of this calculation. $\square$

This page intentionally left blank

Chapter Thirteen

Proof of Hexagrid Theorem I

13.1 THE KEY RESULT

The proof of Hexagrid Theorem I is the same in the odd and even cases.

Say that a *floor line* is a negatively sloped line of the floor grid. Floor lines all have slope $-A$. Say that a *floor point* is a point on a floor line. Such a point need not have integer coordinates. The maps M_+ and M_- from §6.6 are constant on floor lines. Thus, if L is a floor line, $M_\pm(L)$ is a single point.

Lemma 13.1 *If p is a floor point, then $M_-(p) \equiv (\beta, 0, 0) \mod \Lambda$ for some $\beta \in \mathbf{R}$.*

Proof: Suppose first that p/q is odd. Since M_- is constant on floor lines, it suffices to consider floor points of the form

$$(0, t), \qquad t = \frac{k(p+q)}{2}, \qquad k \in \mathbf{Z}. \tag{13.1}$$

These points are the intersections of the floor lines with the y-axis. Note that t is an integer because $p+q$ is even.

To compute the image of the point $(0, t)$, we just have to subject the point t to the reduction algorithm from §6.6. The first 4 steps of the algorithm lead to the following result.

1. $z = t$.
2. $Z = \text{floor}(t) = t$ because t is an integer.
3. $y = 2t = k(p+q) = kq(1+A)$.
4. $Y = \text{floor}(y/(1+A)) = kq$.

Hence $z = Z$ and $y = (1+A)Y$. Hence

$$M_-(0, t) = (x - (1+A)X, y - (1+A)Y, z - Z) = (\beta, 0, 0) \tag{13.2}$$

for some number $\beta \in \mathbf{R}$ that depends on A and k.

When p/q is even, the floor grid has a different definition: Only the even floor lines are present in the grid. That is, the number k in Equation 13.1 is an even integer. Hence, for the floor lines in the even case, the number t is an integer. The rest of the proof is the same. $\square$

13.2 A SPECIAL CASE

Say that a floor point is *special* if it lies in $\mathbf{Z}^2$. For instance, (0, 0) is a special floor point. So are the points in Equation 13.1. In this section we will prove statement 1 of the Hexagrid Theorem for special floor points.

Lemma 13.2 *The arithmetic graph rises up above the baseline at a special floor point.*

Proof: Let v be a special floor point. By Lemma 13.1, we have $M_{\pm}(v) \equiv (\beta, 0, 0)$ mod Λ. In particular, $M_{\pm}(v)$ lies in the kind of singular fiber that we considered in §6.5. The fiber we mean is $\{z = 0\}$. The slices as shown in Figure 6.3 determine the nature of the edges of the arithmetic graph, although the slices currently of interest to us are not shown there. We are interested in following the method discussed in §6.5, where we set $\alpha = 0$ and consider the singular situation. The points $M_{-}(\zeta_k)$ and $M_{+}(\zeta_k)$ both lie in the $(0, A)$ slices of the partitions. Figure 13.1 does for these slices what Figure 6.3 does for the generic slice. The point $M_{-}(\zeta_k)$ always lies along the bottom edge of the fiber, and the point $M_{+}(\zeta_k)$ just above the edge contained in the line $y = 1$. The relevant edges are highlighted.

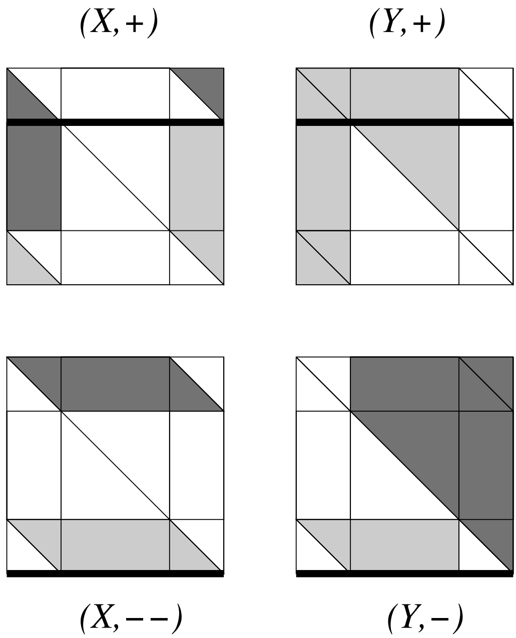

Figure 13.1: The $(0, A)$ slices.

From this figure we can see that the only edges emanating from ζ_k are those corresponding to the pairs

$$(0, 1), \qquad (1, 0), \qquad (1, 1), \qquad (-1, 1).$$

All of these edges point into the half-plane above the relevant floor line. This is what we wanted to establish.

13.3 PLANES AND STRIPS

We say that an edge e of the arithmetic graph is a *crossing cell* if e crosses the arithmetic graph in an interior point. If statement 1 of the Hexagrid Theorem fails, then a crossing cell must exist. One vertex of a crossing cell lies above a floor line and one vertex lies below. We shall be interested in the above-lying vertex. Call this vertex the *top vertex* of the crossing cell.

For each pair $(\epsilon_1, \epsilon_2) \in \{-1, 0, 1\}^2$, let $\Sigma(\epsilon_1, \epsilon_2) \subset \mathbf{R}^2$ denote the set of points (m, n) such that some floor line separates (m, n) from $(m, n) + (\epsilon_1, \epsilon_2)$. The set $\Sigma(\epsilon_1, \epsilon_2)$ is a countable union of open infinite strips, one per floor line. Depending on the choice of (ϵ_1, ϵ_2), the strips lie either above the floor lines or below them. We shall be interested in the above-lying strips. These strips correspond to the pairs

$$(-1, 0), \qquad (-1, -1), \qquad (0, -1), \qquad (1, -1). \tag{13.3}$$

Lemma 13.3 *Let c be a crossing cell and let v be the top vertex of c. Then we have $v \in \Sigma(\epsilon_1, \epsilon_2)$ for one of the choices in Equation 13.3.*

Proof: This is a tautology. □

Now we switch gears and talk about the situation in $\mathbf{R}^3$. Let $\Pi_- \subset \mathbf{R}^3$ denote the plane given by $z = y$. Equivalently, Π_- is the plane through the origin generated by the vectors $(1, 0, 0)$ and $(1, 1, 1)$. Let $\Pi_-(0) \subset \Pi_-$ denote the line through the origin parallel to $(1, 0, 0)$. Define

$$\Pi_+ = \Pi_- + (1, 1, 0), \qquad \Pi_+(0) = \Pi_-(0) + (1, 1, 0). \tag{13.4}$$

Let $\Pi(\lambda) \subset \Pi_\pm$ denote the strip bounded by the two lines

$$\Pi_\pm(0), \qquad \Pi_\pm(0) + \lambda(1, 1, 1). \tag{13.5}$$

We take the strips to be open in $\Pi_\pm$, and we always take $\lambda > 0$. We define

$$\lambda(\epsilon_1, \epsilon_2) = -(A\epsilon_1 + \epsilon_2). \tag{13.6}$$

Lemma 13.4 *Let $\lambda = \lambda(\epsilon_1, \epsilon_2)$. Suppose that $(m, n) \in \Sigma(\epsilon_1, \epsilon_2)$. Then*

$$M_\pm(m, n) \in \Pi_\pm(\lambda).$$

Proof: We consider the case of M_- and the pair $(-1, 0)$. In this case, $\lambda(-1, 0) = A$. The other cases have essentially the same proof. If $(m, n) \in \Sigma(-1, 0)$, then there is some x such that (x, n) lies on a floor line and $0 < m - x < 1$. Given the definition of M_-, there is some $0 < s < A$ such that

$$M_-(m, n) - M_-(m, x) = (s, s, s).$$

By Lemma 13.1,

$$M_-(m, n) = M_-(x, n) + (s, s, s) \equiv (\beta, 0, 0) + s(1, 1, 1) \mod \Lambda.$$

This completes the proof. □

13.4 THE END OF THE PROOF

Let $\mathcal{R}_\pm$ be the polyhedron partition from the Master Picture Theorem associated to A. For each pair (ϵ_1, ϵ_2) above, let $R_\pm(\epsilon_1, \epsilon_2; A)$ denote the finite union of polyhedra corresponding to the pair (ϵ_1, ϵ_2). In our next result, ΛR denotes the orbit of R under the lattice $\Lambda = \Lambda_A$ from the Master Picture Theorem.

Lemma 13.5 *The following is true for either choice of sign, for any parameter A, and for any (ϵ_1, ϵ_2) in Equation 13.3.*

$$\Pi_\pm(\epsilon_1, \epsilon_2; A) \cap \Lambda R_\pm(\epsilon_1, \epsilon_2; A) = \emptyset.$$

Proof: Our notation above emphasizes the dependence on the parameter A. We check the disjointness for all parameters at the same time. Let

$$\Pi_\pm(\epsilon_1, \epsilon_2) = \bigcup_{A \in (0,1)} \Big(\Pi_\pm(\epsilon_1, \epsilon_2; A) \times \{A\}\Big). \tag{13.7}$$

Let $\Pi^*(\ldots)$ denote the portion of $\Pi(\ldots)$ between the hyperplanes $\{x = 0\}$ and $\{x = 2\}$. The element γ_1 from Equation 6.14 preserves both $\Pi(\ldots)$ and the tiling. Also, since γ translates by at most 2 units in the x-direction, $\Pi^*(\ldots)$ contains a fundamental domain for the action of γ on $\Pi(\ldots)$. Hence, to establish our result, it suffices to establish

$$\Pi^*_\pm(\epsilon_1, \epsilon_2) \cap \Lambda R_\pm(\epsilon_1, \epsilon_2) \tag{13.8}$$

for all relevant choices. Here $R_\pm(\epsilon_1, \epsilon_2)$ is one of the convex integral polytopes described in §6.9. The set $\Pi^*_\pm(\epsilon_1, \epsilon_2) \subset \Pi$ is the interior of a convex integral polyhedron in $\mathbf{R}^4$. In $(-)$ cases, the vertices of this polyhedron are (perhaps redundantly)

$$\begin{bmatrix}0\\0\\0\\0\end{bmatrix}\begin{bmatrix}2\\0\\0\\0\end{bmatrix}\begin{bmatrix}0\\-\epsilon_2\\-\epsilon_2\\0\end{bmatrix}\begin{bmatrix}2\\-\epsilon_2\\-\epsilon_2\\0\end{bmatrix}\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\begin{bmatrix}2\\0\\0\\1\end{bmatrix}\begin{bmatrix}0\\-\epsilon_1-\epsilon_2\\-\epsilon_1-\epsilon_2\\1\end{bmatrix}\begin{bmatrix}2\\-\epsilon_1-\epsilon_2\\-\epsilon_1-\epsilon_2\\1\end{bmatrix}$$

Using a method just like that in §11.1, we check Equation 13.8 for all relevant choices. $\square$

Suppose that statement 1 of the Hexagrid Theorem fails for some parameter A. Then there is some crossing cell c. By the Master Picture Theorem, one of the two maps $M_\pm$ (say M_+) is such that

$$M_+(v) \in \mathcal{R}(\epsilon_1, \epsilon_2; A), \tag{13.9}$$

where (ϵ_1, ϵ_2) is one of the pairs from Equation 13.3. By Lemma 13.4, we have

$$M_+(v) \subset \Pi_+(\epsilon_1, \epsilon_2; A). \tag{13.10}$$

But these last two equations together contradict Lemma 13.5. This contradiction establishes statement 1 of the Hexagrid Theorem.

13.5 A VISUAL TOUR

Our computational proof of Lemma 13.5 does not really give a feel for what is going on. Here we illustrate the result with images taken from Billiard King. To draw figures, we will identify the planes $\Pi_{\pm}$ with $\mathbf{R}^2$ using the projection

$$\pi(x, y, z) = (x, z). \tag{13.11}$$

In fact, this simple projection will be our constant companion for the rest of this part of the book. All our constructions depend on the parameter A, but we sometimes omit A from our notation.

Under the identification, the sets

$$\pi(R_{\pm}(\epsilon_1, \epsilon_2; A) \cap \Pi) \tag{13.12}$$

are rectangles whose sides are parallel to the coordinate axes! Our proof of Lemma 14.3 in the next chapter justifies this claim.

The coordinates of the rectangle vertices are small rational combinations of 1 and A and can easily be determined by inspection. The whole figure is invariant under translation by $(1 + A, 0)$. The thick line on the left corresponds to $\Pi_-(0)$, the black dot is $(A, 0)$, and the white dot is $(1 + 2A, 0)$.

The unlabelled rectangles in Figure 13.2 show one period of the tiling of the strip $\Pi(1 + 2A)$ for the parameter $A = 1/3$. The shaded and labelled rectangles to the right of the partition give the shading scheme. For instance, the dark left rectangle corresponds to $R_-(-1, -1)$. The white rectangles have various labels that do not matter to us. The line corresponding to the label of (ϵ_1, ϵ_2) indicates the placement of the top edge of the strip $\Pi_-(\lambda(\epsilon_1, \epsilon_2))$. In each case, the relevant strip lies below the relevant shaded piece of the partition.

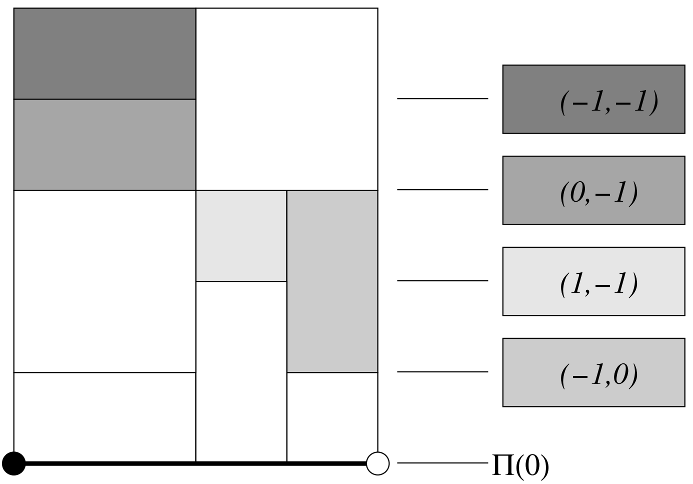

Figure 13.2: The $(-)$ case for $A = 1/3$.

Figure 13.3, taken from Billiard King, shows the partitions of the strip $\Pi_-(2)$ for several parameters. We show somewhat more of the tiling than in Figure 13.2. One can match part of the top right of Figure 13.3 with Figure 13.2.

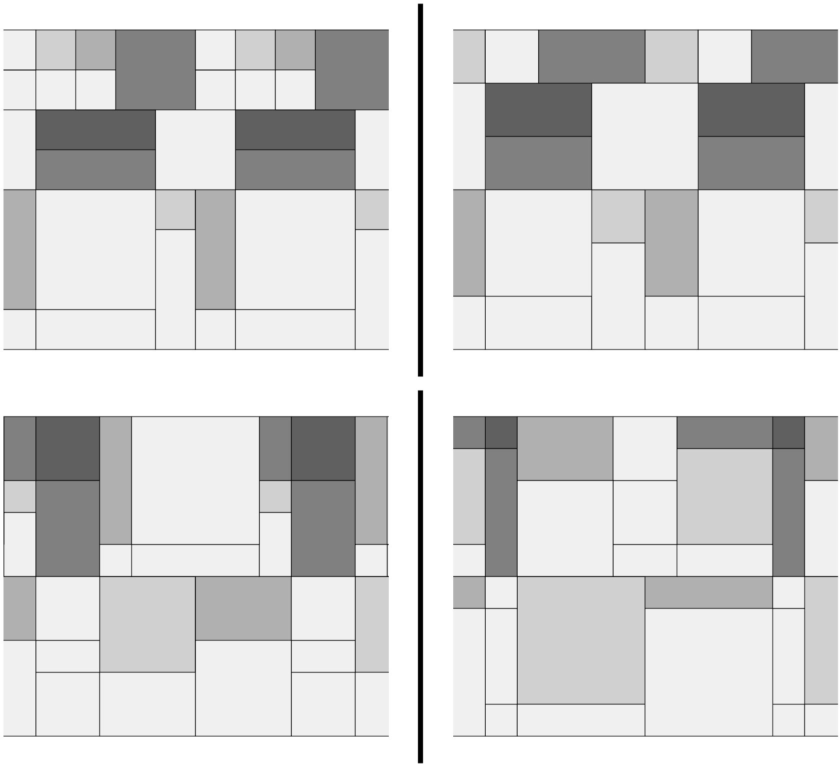

Figure 13.3: The $(-)$ case for $A = 1/4, 1/3, 3/5, 4/5$.

Figures 13.4 and 13.5 show the same thing for the $(+)$ case. Here the black dot is $(0, 0)$ and the white dot is $(1 + A, 0)$. These figures are not as interesting. Only the levels $(-1, -1)$ and $(-1, 0)$ play a role, and there are no close calls.

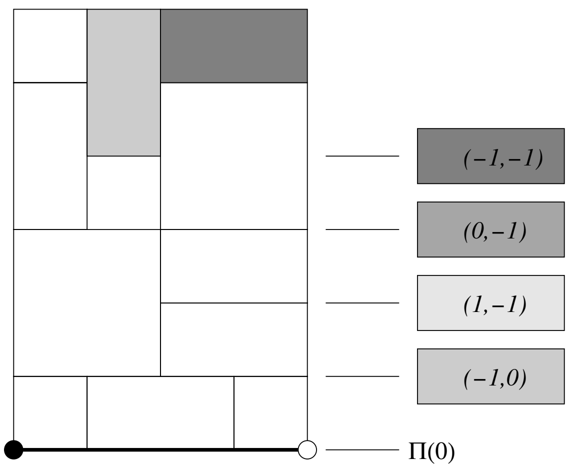

Figure 13.4: The $(+)$ case for $A = 1/3$.

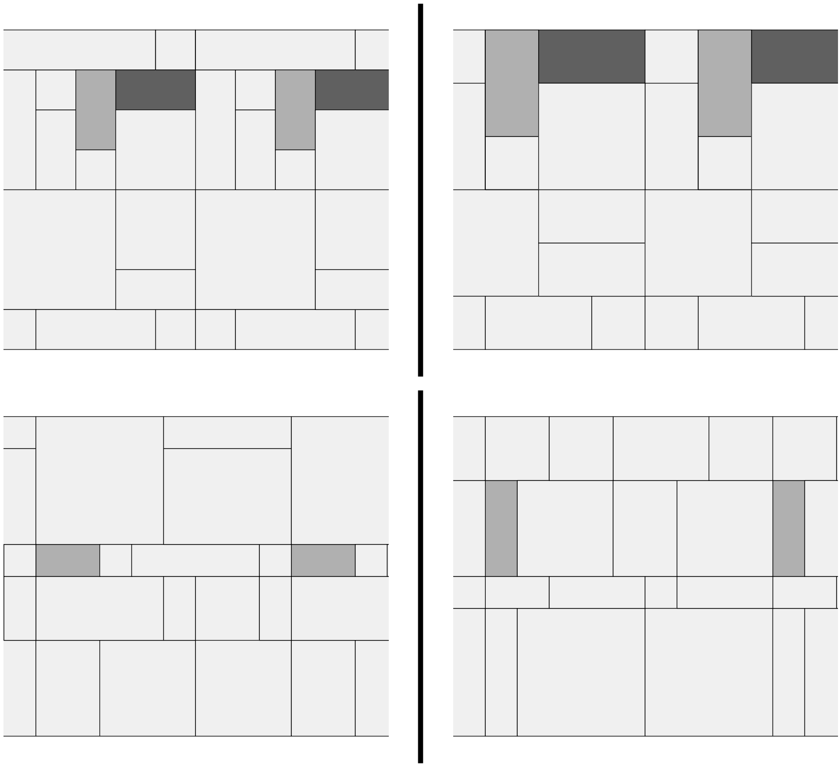

Figure 13.5: The $(+)$ case for $A = 1/4, 1/3, 3/5, 4/5$.

This page intentionally left blank

Chapter Fourteen

The Barrier Theorem

14.1 THE RESULT

Let p/q be an even rational. Let $V = (q, -p)$. Referring to Equation 4.1, one of the two rationals $p_{\pm}/q_{\pm}$ is even and one is odd. Let p'/q' denote whichever of these rationals is odd. (We allow the case when $p'/q' = 1/1$.) We call p'/q' the *odd predecessor* of p/q. We say that the *barrier* is the line parallel to V that contains the point

$$\left(0, \frac{p'+q'}{2}\right). \tag{14.1}$$

Theorem 14.1 (Barrier) *Let e be an edge of $\widehat{\Gamma}(p/q)$ that crosses the barrier line. Then there is some $k \in \mathbf{Z}$ such that the translate $e + kV$ is an edge of $\Gamma(p/q)$. Moreover, there are only two such edges modulo translation by $\mathbf{Z}[V]$.*

We will not need the Barrier Theorem until Part 6 of the book. The reader who is interested in only the Erratic Orbits Theorem can skip this chapter. The reason that we prove the Barrier Theorem here is that the proof involves a modification of the argument we gave in the last chapter. Also, our proof of statement 2 of the Hexagrid Theorem uses some of the ideas we present first in the proof of the Barrier Theorem. Compare §16.5.

The interested reader can observe, using Billiard King, that the Barrier Theorem and the Hexagrid Theorem are specially related: The arithmetic graph always crosses the barrier line within 1 unit of a line from the door grid. We will not establish this fact because we do not need it for any purpose.

We have stated the precise version of the Barrier Theorem that we need for our applications, but the Barrier Theorem is really part of a more robust general theorem. If we replace A' by some parameter A^* that is close to A in the sense of Diophantine approximation, then we get the general result that the corresponding "barrier line" is not frequently crossed by $\widehat{\Gamma}$. The basic reason is that Λ^* serves as a kind of memory of the Hexagrid Theorem for the parameter A^*. The two graphs $\widehat{\Gamma}$ and $\widehat{\Gamma}^*$ mainly agree along Λ^*, and the only crossings take place at the few mismatches in the graphs.

Figure 14.1 illustrates the Barrier Theorem for the parameter $A = 12/47$. The bottom straight line in the figure is the baseline. The top straight line is the barrier. The black component is $\Gamma(12/47)$. The reader can see other parameters using Billiard King.

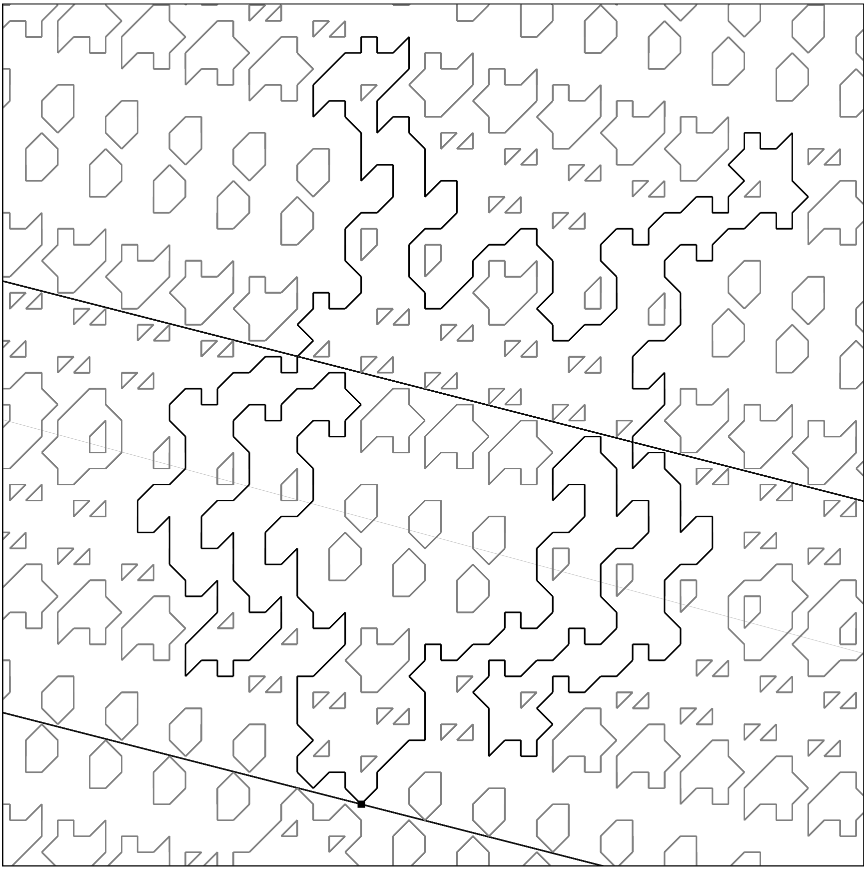

Figure 14.1: Components of $\widehat{\Gamma}(12/47)$ and the barrier.

Before we give the formal proof of the Barrier Theorem, we indicate the main idea. In the previous chapter we saw that M_- mapped the points of $\mathbf{Z}^2$ of interest to us, namely those contained in the strip

$$\Sigma(\epsilon_1, \epsilon_2),$$

into a strip

$$\Pi_-\big(\lambda(\epsilon_1, \epsilon_2)\big) \subset \mathbf{R}^3. \tag{14.2}$$

We then showed that

$$\Pi_-\big(\Lambda(\epsilon_1, \epsilon_2)\big) \cap R_-(\epsilon_1, \epsilon_2) = \emptyset \tag{14.3}$$

for the relevant pairs. We did the same thing for $(+)$ in place of $(-)$. In all, we had 8 cases to consider.

For the Barrier Theorem, we have a similar setup. This time, however, the strips we get are slight translates of those in Equation 14.2. The small translation causes the intersection in Equation 14.3 to be nonempty but quite small. The tiny intersections give rise to the crossings we see in the Barrier Theorem. The main point is to bound the number of potential new crossings.

14.2 THE IMAGE OF THE BARRIER LINE

Let Λ be the barrier line. Here we prove an analog of Lemma 13.1 from the previous chapter. There is one result for $A' > A$ and one result for $A' < A$. We will concentrate on the case $A' > A$. At the end of the chapter, we will deal with the other case.

Lemma 14.2 *Suppose that $A' > A$. There is some real β such that*

$$M_-(\Lambda) = \Big(\beta, 1/q, 0\Big). \tag{14.4}$$

Proof: The key fact here is that

$$q'(A' - A) = 1/q. \tag{14.5}$$

Since Λ is parallel to the baseline, M_- is constant on Λ. Hence we just have to compute

$$M_-(0, t'), \qquad t' = \frac{p' + q'}{2}.$$

To compute the image of the point $(0, t')$, we just have to subject the point t' to the reduction algorithm from §6.6. The first 4 steps of the algorithm lead to the following result.

1. $z = t'$.
2. $Z = \text{floor}(t') = t'$ because t' is an integer.
3. $y = 2t = p' + q' = q'(1 + A') = q'(1 + A) + q'(A' - A) = q'(1 + A) + (1/q)$.
4. $Y = \text{floor}(y/(1 + A)) = q'$.

Hence $z = Z$ and $y = (1 + A)Y + (1/q)$. Hence

$$M_-(0, t) = (x - (1 + A)X, y - (1 + A)Y, z - Z) = (\beta, 1/q, 0) \tag{14.6}$$

for some number $\beta \in \mathbf{R}$ that depends on A and k. □

For any relevant set $X \subset \mathbf{R}^3$, we define

$$X' = X + (0, 1/q, 0). \tag{14.7}$$

We define the strips $\Sigma(\epsilon_1, \epsilon_2)$ exactly as in the previous chapter, except that we use the barrier line Λ as the bottom of the strips rather than the floor lines. We are just translating the strips. Now that we know Lemma 14.2, the same argument as in the previous chapter shows that

$$M_\pm\Big(\Sigma(\epsilon_1, \epsilon_2)\Big) = \Pi'_\pm\big(\lambda(\epsilon_1, \epsilon_2)\big). \tag{14.8}$$

We draw figures using the projection map

$$\pi(x, y, z) = (x, y), \tag{14.9}$$

just as in the previous chapter. Note that $\pi(X') = \pi(X)$. Therefore the composition $\pi \circ M_\pm$ maps $\Sigma(\epsilon_1, \epsilon_2)$ to precisely the same planar set as in the previous chapter. Even though the domains have changed, the ranges have not.

What changes is the projection of the intersection of $\Pi'_\pm$ with the partition $R_\pm$. That is, there is a difference between the two planar patterns of rectangles:

$$\pi(\Pi_\pm \cap R_\pm), \qquad \pi(\Pi'_\pm \cap R_\pm). \tag{14.10}$$

Say that the planes cutting out the partition of $R_\pm$ are the *partition planes*. These planes belong to 4 families and are listed in §6.2. The following result explains how the rectangle pattern changes. Incidentally, this result explains why we really do get a pattern of rectangles.

Lemma 14.3 *Let W be a partition plane. Then the two lines $\pi(W \cap \Pi_\pm)$ and $\pi(W \cap \Pi'_\pm)$ either coincide or are exactly $1/q$ apart in the plane.*

Proof: The result depends on only the normals of the planes involved and not on the (initial) positions. Thus we can work with Π_- and with 4 planes through the origin, each parallel to one of the partition planes in the 4 different families. For ease of notation, let $\Pi = \Pi_-$ and let $s = 1/q$. Here are the 4 cases.

- Let $W = \{z = 0\}$. The map $X \to X'$ preserves W. Therefore we have $\Pi' \cap W = (\Pi \cap W)'$. But then $\pi(\Pi' \cap W) = \pi(\Pi \cap W)$. We remark that $\Pi \cap W$ is the line through the origin parallel to $(1, 0, 0)$. Hence $\pi(\Pi \cap W)$ is a horizontal line.

- Let $W = \{z = 0\}$. The map $X \to X'$ preserves W, and the same argument works as in the previous case. We remark that $\Pi \cap W$ is the line through the origin parallel to $(0, 0, 1)$. Hence $\pi(\Pi \cap W)$ is a vertical line.

- Let $W = \{y = 0\}$. In this case, $W \cap \Pi$ is the x-axis and $W \cap \Pi'$ is parallel to the x-axis but contains the point
$$(0, 0, -s) = (0, s, 0) - s(1, 1, 1) + s(1, 0, 0).$$
In this case, $\Pi \cap W$ and $\Pi' \cap W$ are exactly s units apart and the map π is an isometry. The images under π are parallel horizontal lines exactly s units apart from each other.

- Let $W = \{x + y - z = 0\}$. In this case, we compute that $W \cap \Pi$ and $W \cap \Pi'$ are the lines given by the parametric equations
$$t(0, 1, 1), \qquad (-s, s, 0) + t(0, 1, 1).$$
The corresponding lines $\pi(W \cap \Pi)$ and $\pi(W \cap \Pi')$ are parallel vertical lines exactly s units apart.

This completes the proof. $\square$

14.3 AN EXAMPLE

We consider the parameter $A = 4/15$. We consider the plane Π_- and its corresponding translate Π'_-. Here we illustrate Lemma 14.3 in action.

Figure 14.2 shows one period for the parameter $4/15$. The left hand side of Figure 14.2 shows $\pi(\Pi_- \cap R_-)$, and the right hand side shows $\pi(\Pi'_- \cap R_-)$.

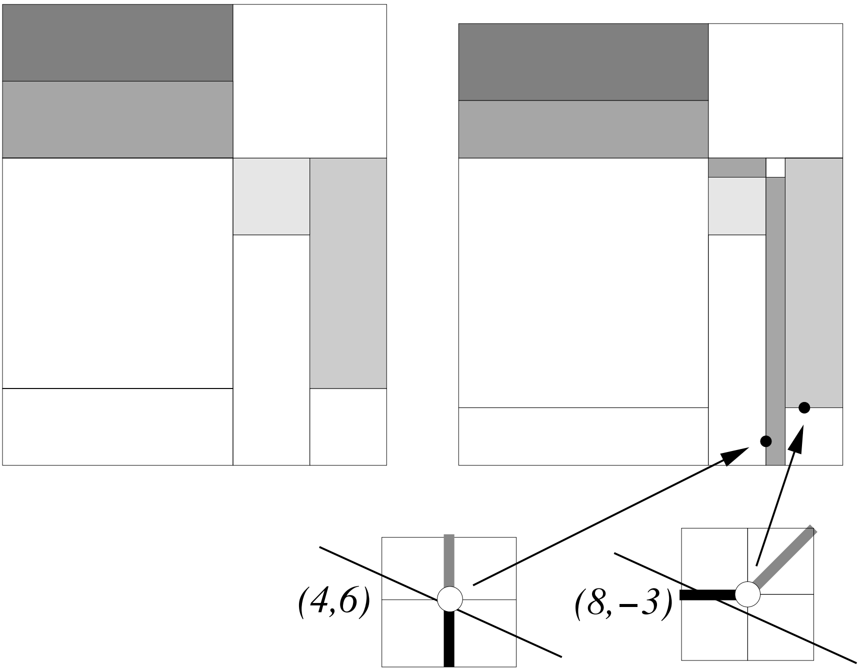

Figure 14.2: The slices Π_- and Π'_-.

Comparing the right hand side with the left hand side, we notice several changes. First, 3 new regions have become visible. Two of these regions are long and thin, and one of them is a little square. The common width of these regions is $1/15$. Second, some of the other regions have slightly changed their positions. In all cases when an edge moves, the offset is by $1/15$, as predicted by Lemma 14.3.

We compute that the two relevant crossings occur at the points $(4, 6)$ and $(8, -3)$. Figure 14.2 illustrates the locations of the points $M_-(4, 6)$ and $M_-(8, -3)$ and the corresponding crossings of the barrier that arise from these images. The tall thin region, which gets labelled $(0, -1)$, causes a downward crossing at $(4, 6)$. The leftmost shaded region, which is labelled $(-1, 0)$, has shifted downward slightly so as to meet $M_-(8, -3)$ and cause a leftward crossing. Were we to analyze the figure relative to the parameter $A' = 3/11$, these offending points would be assigned noncrossing edges.

14.4 BOUNDING THE NEW CROSSINGS

In the new setting, our analysis for statement 1 of the Hexagrid Theorem does not *completely succeed* because of the emergence of the new regions and the slight perturbations of the existing regions. Let us analyze the failures. Referring to the right side of Figure 14.2, the images of the relevant vertices all lie on a diagonal line of slope 1. This line starts somewhere on the bottom edge of the tiled rectangle and wraps around when it hits the right edge. Considered mod $1 + A$, the difference in the x-coordinates between successive points is $1/q$.

The bottom of each modified rectangle is at most $1/q$ units lower than the original. Since the original rectangle was disjoint from the relevant strip, the modified rectangle intersects only the top $1/q$ rim of the same strip. Thus each modified rectangle gives rise to at most one new crossing. The horizontal lines bounding a new region come from partition planes in different families. Looking at the cases in the proof of Lemma 14.3, we see that one of these lines moves and one does not. Thus a new region has width exactly $1/q$. Likewise, a new region has height exactly $1/q$. Therefore each new region gives rise to at most 1 crossing.

Looking carefully at which shaded regions actually move down when Π_- is replaced by Π'_-, we arrive at the 4 shaded regions shown in Figure 14.3. Here is a trick to get down to 2.

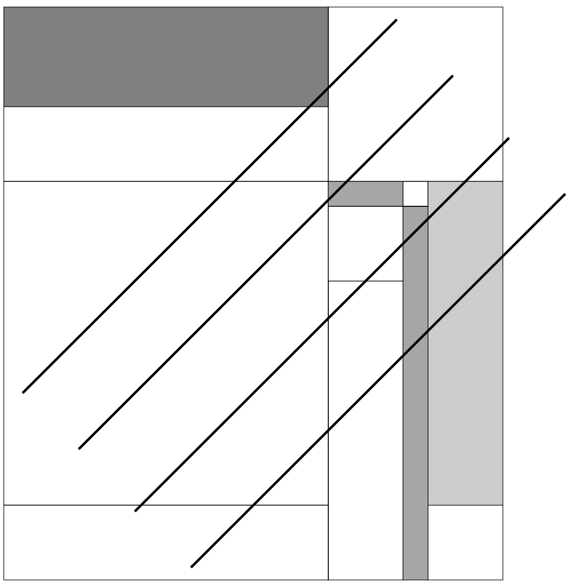

Figure 14.3: Some of the shaded rectangles.

Note that any diagonal line intersects at most 2 of the 4 relevant rectangles. Therefore what seems like 4 potential crossings is just 2. The argument works much the same for other parameters. Figure 14.4 shows the picture for 3 other parameters. In each row, the left hand side shows the slice corresponding to statement 1 of the Hexagrid Theorem, and the right hand side shows the perturbed slice we are interested in here.

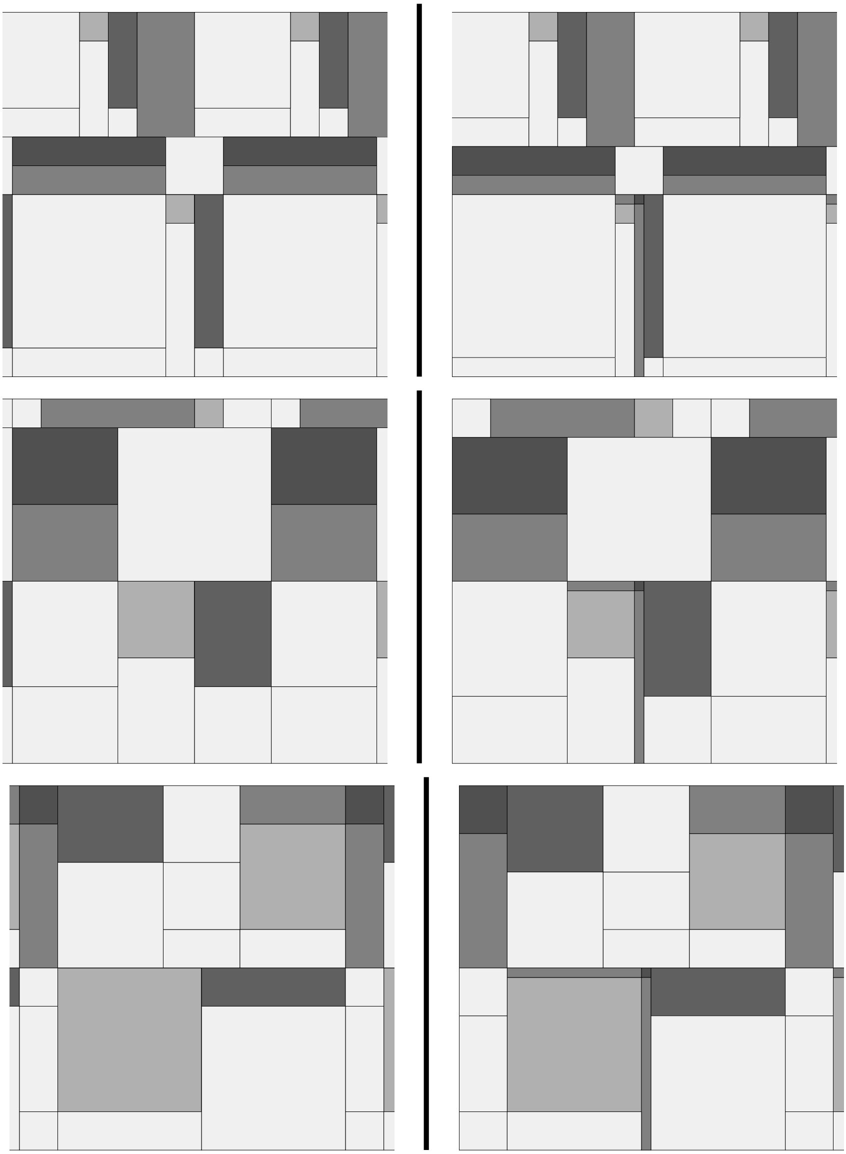

Figure 14.4: The $(-)$ figure for $A = 3/19, 8/19, 15/19$.

The figure for Π_+ is easier to analyze. Recall from the proof of the Hexagrid Theorem that all the relevant rectangles were well above the range of the corresponding vertices. See Figures 13.3 and 13.4. Thus we only have to worry about the emergence of new rectangles. The only new rectangle to emerge within range is a rectangle labelled $(-1, 0)$ that emerges at the very bottom. Hence there is at most 1 crossing. See Figure 14.5.

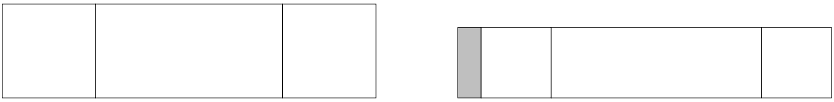

Figure 14.5: The bottom row of Π_+ and Π'_+.

All in all, there are at most 3 barrier crossings within one period. Also, the number of barrier crossings is even because every component is a polygon. Hence there are exactly 2 barrier crossings. The major components do cross the barrier, and hence this accounts for the 2 crossings.

14.5 THE OTHER CASE

An analysis similar to thte one above takes care of the case when $A' < A$. However, we will take a different approach based on symmetry. Let Λ_+ denote the barrier line. There is nothing special about the fact that Λ_+ lies above the baseline. We could consider the corresponding line Λ_- below the baseline. Here Λ_- is parallel to V and contains

$$P_- = \left(0, \frac{-p' + q'}{2}\right). \tag{14.11}$$

Actually, to do things exactly right, we think of Λ_+ and Λ_- lying infinitesimally near, but below, the lines we have defined. Thus, in particular, P_- lies above Λ_-.

We compute that

$$M(P_-) = \left(\beta, 1/q, 0\right)$$

for some $\beta \in \mathbf{R}$. Thus, by considering Λ_- in place of Λ_+, we have returned to the case already analyzed. But now we can apply the rotational symmetry ι considered in §12.3. Assuming that $\iota(\Lambda_-) = \Lambda_+$, the result for Λ_+ follows from the result for Λ_-.

It is not quite true that $\iota(\Lambda_-) = \Lambda_+$. In fact, $\iota(\Lambda_-)$ is parallel to Λ_+ and exactly $1/q$ vertical units beneath Λ_-. Thus we have actually proved the Barrier Theorem for a barrier that is lower by a tiny bit. This result suffices for all purposes.

To obtain the stated result right on the nose, we note that P_- is the only point adversely affected: $\iota(P_-)$ lies beneath Λ_+, whereas P_- lies on Λ_-. However, recall that we consider these lines to be infinitesimally beneath the lines through integer points. Thus, as mentioned above, P_- lies above Λ_-. So, even though $\iota(\Lambda_-) \neq \Lambda_+$, all the relevant lattice points lie on the correct sides.

This completes the proof of the Barrier Theorem.

Chapter Fifteen

Proof of Hexagrid Theorem II

We will prove statement 2 of the Hexagrid Theorem for odd rationals. The even case has an essentially identical proof. Here we remark on one small difference. Call a point in $\mathbf{R}^2$ *bad* if it has the form (m, y), where y is a half-integer. According to statement 3 of Lemma 15.1 below, a door cannot be a bad point in the odd case. In the even case, we simply declare that a door cannot be a bad point. See the definition in Chapter 3. Having ruled out the bad points in both cases, our proof is practically independent of parity.

15.1 THE STRUCTURE OF THE DOORS

Our proof of statement 2 of the Hexagrid Theorem requires a careful analysis of the doors. In this first section, we will establish a technical result about the doors. Say that a *wall line* is a line of positive slope in the room grid. These lines are all parallel to the vector W, from Equation 3.2. Recall that Θ is the lattice, from §12.1. We distinguish two special kinds of points in $\mathbf{R}^2$.

- *Type 1:* $(aq, b/p)$, with $a, b \in \mathbf{Z}$.
- *Type 2:* $(ap, b/q)$, with $a, b \in \mathbf{Z}$.

A point could have both types. Here is our structural result.

Lemma 15.1 *The following are true.*

1. *Any two wall lines are equivalent mod* Θ.
2. *The only points of* $\mathbf{Z}^2$ *lying on wall lines are elements of* Θ.
3. *Every door on* L_0 *has type 1 or type 2 (or both).*

Proof: Statement 1: Recall that Θ is generated by V and V', the vectors from Equation 12.1. Modulo translation by $\mathbf{Z}[V]$, any wall line is equivalent to L_0 or L_1. We just need to show that these two wall lines are equivalent to each other mod Θ. We check explicitly that the following equation holds.

$$V' + \frac{p+1}{2}V \in \Theta \cap L_1.$$

Hence addition by some vector in Θ carries L_0 to L_1. Hence L_0 and L_1 are equivalent mod Θ.

Statement 2: By statement 1, it suffices to consider the case when $(m,n) \in L_0$. Any point in L_0 is a real multiple of W. Such a point has the form

$$sW = \frac{s}{2(p+q)}(2pq, (p+q)^2 - 2p^2). \tag{15.1}$$

In order for this point to lie in $\mathbf{Z}^2$, the first coordinate must be an integer. Hence

$$s = \frac{k(p+q)}{pq}, \qquad k \in \mathbf{Z}. \tag{15.2}$$

Hence

$$n = k \times \frac{pq + (q^2 - p^2)/2}{pq} \in \mathbf{Z}. \tag{15.3}$$

Since p and q are relatively prime, the numerator and denominator of the rational on the right side of Equation 15.3 are relatively prime. Hence pq divides k. Hence (m,n) is an integer multiple of the point

$$(p+q)W = \left(pq, \frac{(p+q)^2}{2} - p^2\right) = 2V' + pV \in \Theta.$$

Here V and V' are the vectors generating Θ, as in Equation 12.1.

Statement 3: Let $\mathcal{K}$ denote the arithmetic kite associated to the parameter. Call a line in the door grid *top* if it is parallel to one of the top two edges of $\mathcal{K}$. Call a line in the door grid *bottom* if it is parallel to one of the bottom two edges of $\mathcal{K}$. Call a door *top* if it lies on a top door line, and *bottom* if it lies on a bottom door line.

Our argument crucially uses the point U in Figure 3.1. We have

$$U = \left(p, \frac{q^2 - p^2 + 2pq}{2q}\right). \tag{15.4}$$

The bottom doors are evenly spaced on L_0. Two consecutive ones are

$$(0,0), \qquad \frac{q}{p}U = \left(q, \frac{q^2 - p^2 + 2pq}{2p}\right) = \left(q, \frac{b}{p}\right). \tag{15.5}$$

Every bottom door on L_0 is a multiple of the nontrivial one we have listed. Hence every bottom door has type 1.

The top doors are evenly spaced on L_0. Two consecutive ones are

$$(0,0), \qquad U = (p, b/q). \tag{15.6}$$

Every top door on L_0 is an integer multiple of the nontrivial one we have listed. Hence such doors have type 2. □

Remark: In the even case, statement 1 of Lemma 15.1 has a trivial proof: Any two wall lines are equivalent modulo translations by integer multiples of V.

15.2 ORDINARY CROSSING CELLS

The bijection between crossing cells and doors described in statement 2 of the Hexagrid Theorem commutes with the action of the symmetry group Θ. The point is that Θ preserves both the hexagrid and the arithmetic graph. Hence, by statement 1 of Lemma 15.1, it suffices to consider those crossing cells that cross L_0.

We first deal with two trivial cases. Recall that the point $(0, 0)$ gives rise to two doors. One of the doors, which we denote $(0, 0)_+$, is attached to the wall above $(0, 0)$. The other door, which we denote $(0, 0)_-$, is attached to the wall below $(0, 0)$. Any door lying in Θ is equivalent to one of these, by statement 2 of Lemma 15.1. One of the crossing cells has vertices $(-1, 1)$, $(0, 0)$, and $(1, 1)$. We associate $(0, 0)_+$ to this crossing cell. Another crossing cell has vertices $(0, -1)$ and $(-1, 0)$. We associate the door $(0, 0)_-$ to this cell.

Henceforth we consider crossing cells that cross L_0 but are not equivalent mod Θ to either of the ones we have just described. We call these remaining crossing cells *ordinary cells*. Given an ordinary cell c, let v_c denote the vertex of c that lies to the right of L_0. (The first statement of the next lemma justifies the existence of v_c.)

Lemma 15.2 *An ordinary cell c has a single edge that crosses L_0 in its interior. Moreover, $v_c + (\epsilon_1, \epsilon_2) \notin L_0$ for any choice of $(\epsilon_1, \epsilon_2) \in \{-1, 0, 1\}^2$.*

Proof: Let c be a crossing cell. If an edge of c fails to cross L_0 at an interior point, then a vertex of c lies on L_0. But then $c \equiv (0, 0)$ mod Θ, by statement 2 of Lemma 15.1. If $v_c + (\epsilon_1, \epsilon_2) \in L_0$, then $v_c \equiv (-\epsilon_1, -\epsilon_2)$ mod Θ, by statement 2 of Lemma 15.1. Hence $(-\epsilon_1, -\epsilon_2)$ is the right vertex of a crossing cell. This happens for $(1, 1)$ and $(0, -1)$, but these are the special crossing cells we have already handled. The only point in L_0 within reach of either $(1, -1)$ or $(1, 0)$ is $(0, 0)$, and we already know that $(0, 0)$ does not connect to these points. The remaining 4 choices lie to the left of L_0. This rules out all cases. $\square$

Now we describe the bijection between ordinary crossing cells and doors. Below we will prove the following result.

Lemma 15.3 (Separation) *Let c be an ordinary cell and let v_c be the right vertex of c. Then L_0 separates v_c from $v_c + (0, 1)$.*

We write $v_c = (m, n)$. Let $\theta \in (n, n + 1)$ be the point such that $(m, \theta) \in L_0$. We define

$$\Upsilon(c) = (n, \theta). \tag{15.7}$$

Lemma 15.4 (Door) *Let v be an ordinary crossing cell. Then $\Upsilon(c)$ is a door.*

The map $c \to \Upsilon(c)$ is certainly injective. To finish our proof of the Hexagrid Theorem, we will prove the following result.

Lemma 15.5 (Surjection) *The map Υ is a surjective map from the set of ordinary crossing cells to the set of doors on L_0 that do not lie in Θ.*

15.3 NEW MAPS

The key to our proof is to use variants of the maps M_+ and M_- from Equation 6.6. Let Λ be the lattice from the Master Picture Theorem. We will produce maps Δ_+ and Δ_-, which have, mod Λ, the same action as M_+ and M_- on $\mathbf{Z}^2$. However, the action of $\Delta_\pm$ on all of $\mathbf{R}^2$ is quite different from the action of $M_\pm$ on $\mathbf{R}^2$.

Now we give the definition. Let $A \in (0, 1)$ be any parameter. Define

$$\Pi = \{x + y = A\}. \tag{15.8}$$

The plane Π plays the same role in the proof of Hexagrid Theorem II that the similarly named plane plays in the proof of Hexagrid Theorem I.

For $(m, n) \in \mathbf{R}^2$, we define

$$\Delta_+(m, n) = (x, y, z),$$

$$x = 2A(1 - m + n) - m, \qquad y = A - x, \qquad z = Am. \tag{15.9}$$

We also define

$$\Delta_-(m, n) = \Delta_+(m, n) + (-A, A, 0). \tag{15.10}$$

Note that $\Delta_\pm(m, n) \in \Pi$. Indeed, $\Delta_\pm$ is an affine isomorphism from $\mathbf{R}^2$ onto Π. We found the maps $\Delta_\pm$ after considerable trial and error.

Lemma 15.6 *Suppose that $(m, n) \in \mathbf{Z}^2$. Then $\Delta_\pm(m, n)$ and $M_\pm(m, n)$ are equivalent mod Λ.*

Proof: Let v_1, v_2, v_3 be the three columns of the matrix defining Λ. So,

$$v_1 = (1 + A, 0, 0), \qquad v_2 = (1 - A, 1 + A, 0), \qquad v_3 = (-1, -1, 1).$$

Let

$$c_1 = -1 + 2m, \qquad c_2 = 1 - m + 2n, \qquad c_3 = n.$$

We compute directly that

$$M_+(m, n) - \Delta_+(m, n) = c_1 v_1 + c_2 v_2 + c_3 v_3.$$

$$M_-(m, n) - \Delta_-(m, n) = c_1 v_1 + (c_2 - 1) v_2 + c_3 v_3.$$

This completes the proof. □

We introduce the vector

$$\zeta = (-A, A, 1) \in \Lambda. \tag{15.11}$$

Referring to the proof of our last result, we have $\zeta = v_2 + v_3$. This explains why $\zeta \in \Lambda$. Note that Π is invariant under translation by ζ.

Below we will specialize to the case when $A = p/q$ is an odd rational. Also, we will extend $\Delta_\pm$ so that it acts linearly on $\mathbf{R}^2$. Now we will see the difference between $\Delta_\pm$ and $M_\pm$. We will see that $\Delta_\pm$ is specially adapted to the wall lines.

Let L_0 denote the wall line through the origin.

Lemma 15.7 *$\Delta_{\pm}(L_0)$ is parallel to ζ and contains $(-2A, A, 0)$.*

Proof: We refer to the points in Figure 3.1. The points W and $(0, 0)$ both lie on L_0. We compute

$$\Delta_+(W) - \Delta_+((0,0)) = \frac{p^2}{p+q}\zeta.$$

Hence $\Delta_+(L_0)$ is parallel to ζ. We compute that $\Delta_+(0,0) = (2A, -A, 0)$. □

We introduce the notation $\Pi(x)$ to denote the line in Π that is parallel to ζ and contains the point $(x, A - x, 0)$. For instance,

$$\Delta_+(0,0) \in \Pi(2A), \qquad \Delta_-(0,0) \in \Pi(A). \tag{15.12}$$

Let $\Pi(r, s)$ denote the infinite strip bounded by the lines $\Pi(r)$ and $\Pi(s)$.

For each pair of indices $(\epsilon_1, \epsilon_2) \in \{-1, 0, 1\}^2$, we let $\Sigma(\epsilon_1, \epsilon_2)$ denote the set of points (m, n) such that L_0 separates (m, n) from $(m + \epsilon_1, n + \epsilon_2)$. We care only about the integer points in $\Sigma(\epsilon_1, \epsilon_2)$, but our definition allows $(m, n) \in \mathbf{R}^2 - \mathbf{Z}^2$ as well. Note that $\Sigma(\epsilon_1, \epsilon_2)$ is an infinite strip whose left boundary is L_0. Now we define constants

1. $\lambda(0, 1) = 2A$,
2. $\lambda(-1, -1) = 1 + 0A - A^2$,
3. $\lambda(-1, 0) \quad = 1 + 2A - A^2$,
4. $\lambda(-1, +1) = 1 + 4A - A^2$.

We have included $0A = 0$ above to make the pattern more clear.

Lemma 15.8 *Let (ϵ_1, ϵ_2) be any of the 4 pairs listed above. Let $\lambda = \lambda(\epsilon_1, \epsilon_2)$. Then*

$$\Delta_+(\Sigma(\epsilon_1, \epsilon_2)) = \Pi(2A-\lambda, 2A), \qquad \Delta_-(\Sigma(\epsilon_1, \epsilon_2)) = \Pi(A-\lambda, A). \tag{15.13}$$

Proof: Given that $\Delta_- = \Delta_+ + (-A, A, 0)$, it suffices to establish the first equation. In light of Lemma 15.7 and the fact that Δ_+ is an affine isomorphism from $\mathbf{R}^2$ to Π, it suffices to check what happens to a single point on the right boundary component of $\Sigma(\epsilon_1, \epsilon_2)$. Indeed, in all cases, we can chose the point $(-\epsilon_1, \epsilon_2)$. We compute

1. $\Delta_+(0, -1) = (0, A, 0) \in \Pi(0) = \Pi(2A - \lambda(0, 1))$.
2. $\Delta_+(1, 1) = (-1+2A, 1-A, A) \in \Pi(A^2+2A-1) = \Pi(2A-\lambda(-1, -1))$.
3. $\Delta_+(1, 0) = (-1, 1 + A, A) \in^* \Pi(1 - A^2) = \Pi(2A - \lambda(-1, 0))$.
4. $\Delta_+(1, -1) = (-1-2A, 1+3A, A) \in \Pi(A^2-2A-1) = \Pi(2A-\lambda(-1, 1))$.

This completes the proof. □

15.4 INTERSECTION RESULTS

Here we describe some intersection results that we prove in the next chapter.

For ease of notation, we define

$$\Pi_+((\epsilon_1, \epsilon_2; A)) = \Pi(2A - \lambda(\epsilon_1, \epsilon_2), 2A), \tag{15.14}$$

$$\Pi_-((\epsilon_1, \epsilon_2; A)) = \Pi(A - \lambda(\epsilon_1, \epsilon_2), A). \tag{15.15}$$

To be precise, we take $\Pi_\pm((\epsilon_1, \epsilon_2))$ to be the interior (relative to Π) of the strip. These strips correspond to those in Lemma 15.8.

As usual, ΛR denotes the orbit of R under the lattice Λ. In the next result, X^o denotes the interior of X. We prove the following result in Chapter 16.

Lemma 15.9 (Intersection) *The following hold for all* $A \in (0, 1)$.

1. *For each pair* (ϵ_1, ϵ_2) *from Lemma 15.8,*

$$\Pi_\pm((\epsilon_1, \epsilon_2; A)) \cap \Lambda R^o_\pm(\epsilon_1, \epsilon_2; A) \equiv (0, 0).$$

2. *Let* (ϵ_1, ϵ_2) *be either* $(-1, -1)$ *or* $(-1, 1)$*. Then*

$$\Pi_\pm((\epsilon_1, \epsilon_2; A)) \cap \Lambda \underline{R}_\pm(\epsilon_1, \epsilon_2; A) \subset \partial\Pi_\pm((0, 1)).$$

3. *Let* (ϵ_1, ϵ_2) *be either* $(-1, 0)$ *or* $(0, 1)$*. Then*

$$\Pi_\pm((\epsilon_1, \epsilon_2; A)) \cap \Lambda \underline{R}_\pm(\epsilon_1, \epsilon_2; A) \subset \Pi^o_\pm(0, 1).$$

Remark: Let Π_{old} denote the plane we considered in the proof of Hexagrid Theorem I. By construction, the vector $(1, 1, 1)$ is contained in Π_{old}. Thus, when we use the method of §6.5 to implement the Master Picture Theorem, we need only look at how Π_{old} intersects the *interiors* of the polyhedra in the partitions. On the other hand, $(1, 1, 1)$ is not contained in the plane $\Pi_{\text{new}} = \Pi$. It turns out that Π does intersect the lower boundaries of some of the polyhedra in the partition, and this creates the crossings. In other words, case 3 of the Intersection Lemma is nontrivial.

Proof of the Separation Lemma: Suppose c is an ordinary crossing cell. Let $v = v_c$ be the right vertex. Suppose that the left vertex is $v + (\epsilon_1, \epsilon_2)$. There is some choice of sign (say $+$) such that

$$\Delta_+(v) \in \Pi_+((\epsilon_1, \epsilon_2; A)) \cap \Lambda \underline{R}_+(\epsilon_1, \epsilon_2; A). \tag{15.16}$$

The first containment comes from Lemma 15.8. The second containment comes from the Master Picture Theorem. Applied directly, the Master Picture Theorem refers to the maps $M_\pm$, but Lemma 15.6 lets us replace $M_\pm$ with $\Delta_\pm$.

The intersection in Equation 15.16 is empty in case 1 of the Intersection Lemma. By Lemma 15.2, we have $v \notin \Sigma_\pm(0, 1)$. Hence $\Delta_+(v) \notin \partial\Pi_+(0, 1)$. Hence case 2 of the Intersection Lemma does not apply here. We must have case 3. By case 3, we have $v \in \Pi_\pm(0, 1)$. But then, by Lemma 15.8, we have $v \in \text{interior}(\Sigma(0, 1))$. $\square$

To prove the Door Lemma and the Surjection Lemma, we need to describe how $\Pi_\pm((0,1))$ intersects $\underline{R}_\pm(-1,0)$ and $\underline{R}_\pm(0,1)$. The plane $\Pi = \{x+y=A\}$ is transverse to all the planes listed in §6.2. Hence Π does not share any faces with the polyhedra in the partition. We find the edges by inspecting the partition. We see the figure by plotting the intersection of the partition with the slightly perturbed plane.

$$\Pi + (s,s,s) \tag{15.17}$$

When ϵ is small, we see some very thin rectangles. Taking the limit as $s \to 0$, we find the edges. See §16.5 for detailed figures.

To show the final answer, we will use the projection

$$\pi(x,y,z) = (x,z). \tag{15.18}$$

Once again, π maps all intersections to rectangles having horizontal and vertical sides. We have

$$\pi(\zeta) = \pi(-A,A,1) = (-A,1). \tag{15.19}$$

Thus, translation by the vector $(-A,1)$ identifies the top points and the bottom points in Figures 15.1 and 15.2. These figures are meant to be infinite, and invariant under translation by $(-A,1)$. We show just one period.

We give two labels to the vertices in Figures 15.1 and 15.2. The label (x,y) denotes the coordinates of the vertex. The label $((\epsilon_1,\epsilon_2))$ pair associated to the point. We also label the lines by $((\epsilon_1,\epsilon_2))$. If a set X is labelled by $((\epsilon_1,\epsilon_2))$ on the left hand side, it means that

$$\Delta_+(p) \in X \qquad \Longrightarrow \qquad x \in \underline{R}_+(\epsilon_1,\epsilon_2). \tag{15.20}$$

The labels on the right hand side have the same interpretation, with $(-)$ replacing $(+)$. The gray vertices correspond to $\Delta_\pm(0,0)$. The white dots are labelled $((0,0))$.

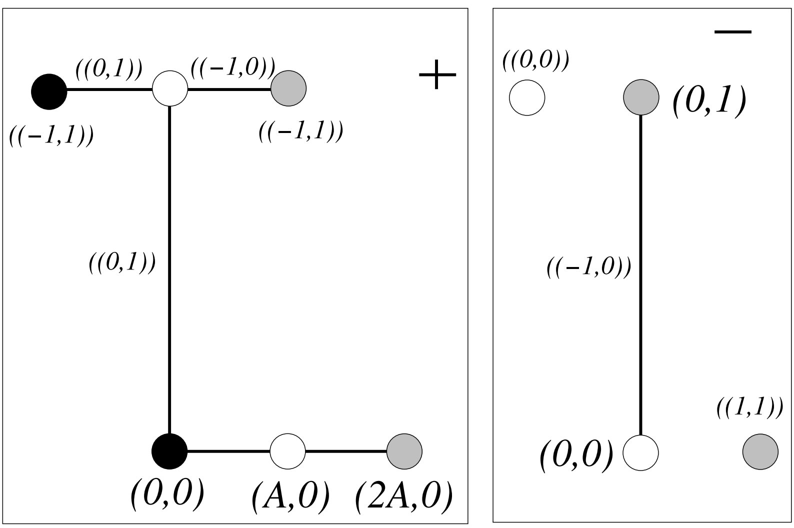

Figure 15.1: The edge intersections for $A = 1/3$.

Figure 15.2 shows the result of superimposing the left and right hand sides of Figure 15.1.

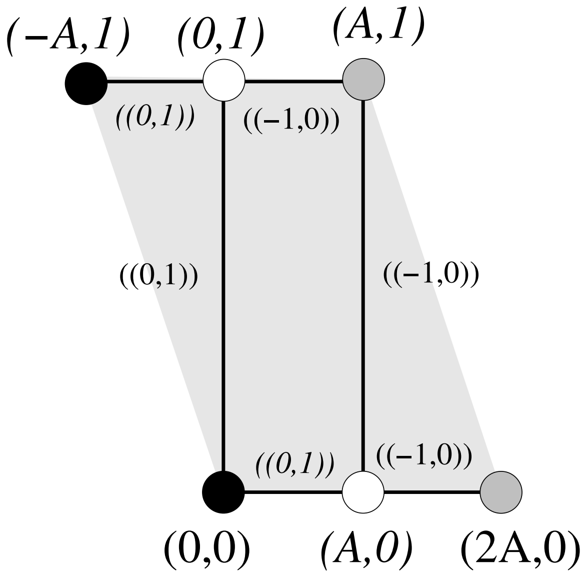

Figure 15.2: Superimposed figures.

Lemma 15.10 *Let c be an ordinary crossing cell. Let v_c be the right vertex of c. Then $\pi \circ \Delta_+(v_c)$ lies in one of the labelled segments of Figure 15.2.*

Proof: Our proof starts out exactly as in the Separation Lemma, and we use the notation there. From the Separation Lemma, we conclude that $v \in \Sigma(0,1)$. Let us suppose first that, as in the proof of the Separation Lemma, the choice of sign is $(+)$, so that

$$\Delta_+(v) \in \Pi_+((0,1)) \cap \underline{R}_+((\epsilon_1, \epsilon_2)). \tag{15.21}$$

Then $\Delta_+(v)$ must lie in one of the open segments on the left hand side of Figure 15.2. The black and gray dots correspond to the special crossing cells we have already analyzed, and the white dot is labelled $((0,0))$.

Now suppose that the choice of sign is $(-)$. Then

$$\Delta_-(v) \in \Pi_-((0,1)) \cap \underline{R}_-((\epsilon_1, \epsilon_2)). \tag{15.22}$$

We get all the same conclusions for Δ_- in place of Δ_+, using the right hand side of Figure 15.1 instead of the left hand side. Hence $\Delta_-(v)$ lies in the vertical segment in the right hand side of Figure 15.1. However, since

$$\pi \circ \Delta_-(v) = \pi \circ \Delta_+(v) - (A, 0),$$

this means that $\Delta_+(v)$ lies on the right hand vertical segment of Figure 15.2. $\square$

15.5 THE END OF THE PROOF

Proof of the Door Lemma: Let $v = (m, n)$ be the right vertex of an ordinary crossing cell. Let $\Upsilon(c) = (m, \theta)$. Here n is the floor of θ. Let

$$\Delta_+(v) = (x, y, z). \tag{15.23}$$

There are two cases to consider. Suppose $\pi \circ \Delta_+(v)$ lies in one of the open horizontal segments of Figure 15.2. Then

$$(x, y, z) \equiv (t, A - t, 0) \mod \Lambda, \qquad t \in (0, A) \cup (A, 2A). \tag{15.24}$$

By Equation 15.24, the third coordinate of $\Delta_+(v)$ is an integer. By the definition of Δ_+, we have $Am = pm/q \in \mathbf{Z}$. Hence q divides m. Hence $v = (kq, n)$ for some $k \in \mathbf{Z}$. Hence v lies in the intersection of L_0 with a door line. Hence v is a door.

Suppose $\pi \circ \Delta_+(v)$ lies in a vertical segment in Figure 15.2. Looking at the positions of the vertical line segments in Figure 15.2, we have

$$x = kA, \qquad k \in \mathbf{Z}. \tag{15.25}$$

From the definition of Δ_+, we have

$$2(1 - m + n) - \frac{m}{A} = \frac{x}{A} \in \mathbf{Z}. \tag{15.26}$$

Hence $m/A \in \mathbf{Z}$. Hence $m = kp$. But then the first coordinate of $\Upsilon(c)$ coincides with the first coordinate of a door on L_0, by statement 3 of Lemma 15.1. Since $\Upsilon(c) \in L_0$, we now see that $\Upsilon(c)$ is a door. □

Proof of the Surjection Lemma: We would like to see that each door actually arises in our construction above. There are two cases.

Type 1: By statement 3 of Lemma 15.1, each type 1 door has the form $(aq, b/p)$, where $a \in \mathbf{Z}$ and b/p is not a half-integer. Let n be the floor of (b/p), let $v = (aq, n)$, andet $(x, y, z) = \Delta_+(v)$. We will show that v is the right vertex of an ordinary crossing cell.

Since the first coordinate of v has the form aq, we have $x \in \mathbf{Z}$. Since $v \in \Sigma(0, 1)$, we have $\Delta_+(v) \in \Pi(0, 2A)$. Hence Equation 15.24 holds. We rule out the case that $t = A$ because b/p is not a half-integer. Hence $\Delta_+(v)$ lands in a horizontal strip in Figure 15.2. Hence one of the edges of $\widehat{\Gamma}$ emanating from v is either $(0, 1)$ or $(-1, 0)$. This edge crosses L_0 because $v \in \Sigma(0, 1) \subset \Sigma(-1, 0)$. Hence v is the right vertex of a crossing cell.

Type 2: By symmetry, it suffices to consider the type 2 doors on L_0. By statement 3 of Lemma 15.1, such a door has the form $(ap, b/q)$. Let $v = (ap, n)$, as in the first case. With the same notation as above,

$$x = 2A(1 - ap - n) - aqA = a'A \tag{15.27}$$

for some $a' \in \mathbf{Z}$. Also, $\Delta_+(v) \in \Pi(0, 2A)$. Hence $\Delta_+(v)$ lands in one of the vertical strips of Figure 15.2. The same argument as in the previous case finishes the proof. □

15.6 THE PATTERN OF CROSSING CELLS

Our proof is finished (modulo the Intersection Lemma), but we would like to say more about the beautiful order of the crossing cells. We present these final details without proof. They can be gleaned from what we have said above. First of all, there are two crossing cells consisting of edges of slope ± 1. These crossing cells correspond to the black and gray corner dots in Figure 15.2. These are the trivial cases we ruled out first.

The remaining crossing cells correspond to the labelled open segments in Figure 15.2. There are exactly $p+q$ crossing cells mod Θ. These cells are indexed by the value of $\theta - n$. The possible numbers are

$$0, \frac{1}{p}, \ldots, \frac{p-1}{p}, \frac{1}{q}, \ldots, \frac{q-1}{q}, 1.$$

Excluding 0 and 1, we have the ordinary crossing cells. We can enhance Figure 15.2 by locating the images of these crossing cells. Figure 15.3 shows the pattern for $p/q = 3/5$. The general case is similar. The lines inside the dots show the nature of the crossing cell. The dashed grid lines in the figure are present to delineate the structure.

One can think of the index values in the following way. Sweep across the plane from right to left by moving a line of slope $-5/3$ parallel to itself. (The diagonal line in Figure 15.3 is one such line.) The indices are ordered according to how the moving line encounters the vertices. The lines we are using correspond to the lines in Π that are parallel to the vector ζ.

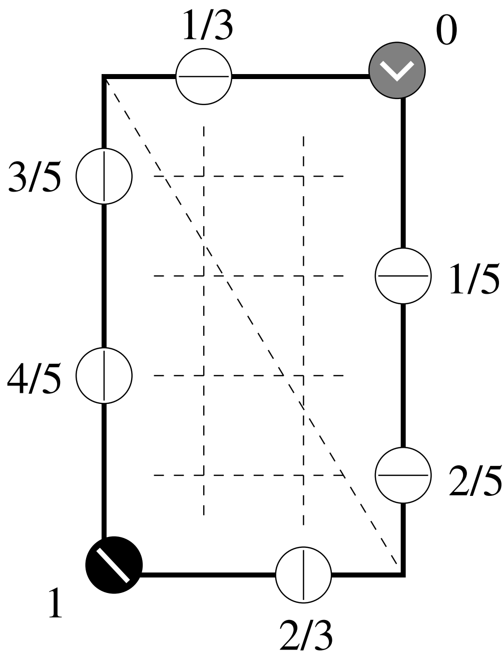

Figure 15.3: The pattern of crossing cells.

Chapter Sixteen

Proof of the Intersection Lemma

16.1 DISCUSSION OF THE PROOF

One way to prove the Intersection Lemma is just by inspection. One can play with Billiard King and see that the result is true. Given the simple nature of the partitions involved, a falsehood in the lemma would be easily detectable by a small amount of experimentation.

Rather than just appeal to experimentation, we will explain a proof that involves finding the intersection patterns of finitely many convex lattice polytopes in $\mathbf{R}^4$. The proof we give is similar to that presented in previous chapters. Previously, e.g., in §11.1, our method was straightforward. Here there is a technical complication that we need to address. This chapter is really about dealing with this complication.

Let $X(\ldots; A) \subset \mathbf{R}^3$ denote some subset of $\mathbf{R}^3$ that depends on the parameter A. For such an object, we define

$$X(\ldots) = \bigcup_A \big(X(\ldots; A) \times \{A\}\big). \tag{16.1}$$

For instance, the sets $R_\pm(\epsilon_1, \epsilon_2)$ are exactly the convex polytopes from §6.7.

Let $S \subset \mathbf{R}^3$ denote the infinite slab bounded by the planes $\{z = 0\}$ and $\{z = 1\}$. Let

$$\Sigma^*_\pm(\epsilon_1, \epsilon_2; A) = \Sigma_\pm(\epsilon_1, \epsilon_2; A) \cap S. \tag{16.2}$$

We include the boundary pieces $\Sigma(\ldots) \cap \partial S$. Thus we are including the tops and bottoms of the parallelogram but not its sides. Figure 15.2 provides a good impression of what this parallelogram looks like.

The set $\Sigma^*_\pm(\epsilon_1, \epsilon_2)$ is contained in a hyperplane of $\mathbf{R}^3$. Unfortunately, this set is not a polyhedron. For instance, the vertices vary quadratically with the parameter. Thus our method breaks down: We cannot control $\Sigma^*_\pm(\epsilon_1, \epsilon_2)$ just by its vertices in $\mathbf{R}^4$.

The trick is to cover $\Sigma(\ldots; A)$ by 2 quadrilaterals $Q_1(\ldots; A)$ and $Q_2(\ldots; A)$ whose vertices vary linearly with the parameter. The linear variation in itself is not enough to guarantee that the corresponding unions $Q_1(\ldots)$ and $Q_2(\ldots)$ are convex, but it turns out that these unions are indeed convex integral polyhedra. When we use $Q_1(\ldots)$ and $Q_2(\ldots)$ in place of $\Sigma(\ldots)$, we create no new interesections – at least not in the interiors. Thus], we prove the Intersection Lemma for these larger sets by the same method we used in §11.1. When we are finished we interpret the results in terms of the original sets.

16.2 COVERING PARALLELOGRAMS

16.2.1 Two Methods

As a first step in making the quadrilateral covering, we describe an entirely planar construction in which we cover a planar parallelogram by 2 rectangles. After we set up the construction, we will relate it to the Intersection Lemma. The only nod we give to the Intersection Lemma in this subsection is that we insist on working in the xz-plane. This is the range of the projection π we used in the last chapter.

Let $A \in (0, 1)$. All the parallelograms we consider have the following properties.

- Their bottom side lies in the line $\{z = 0\}$.
- Their top side lies in the line $\{z = 1\}$.
- Their other sides have slope $-A$.

All the rectangles we consider always have their sides parallel to the coordinate axes.

Figure 16.1 shows a very simple method for covering P with 2 rectangles. The gray dot in this figure has second coordinate A. It seems easier just to amalgamate these rectangles into a single one, but we prefer to always cover the parallelograms with 2 rectangles. This allows us to have more uniform notation.

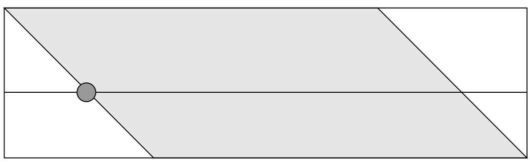

Figure 16.1: Covering a parallelogram with a rectangle.

Figure 16.2 shows a different covering of P with 2 rectangles.

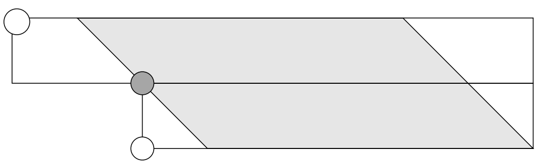

Figure 16.2: Covering a parallelogram with 2 rectangles.

Our geometric construction is determined by the following information.

1. The gray dot lies on the left edge of P. The z (meaning second) coordinate of this dot is A.
2. The line connecting the 2 white dots is parallel to the sides of P.

We will give 4 examples of these constructions in action. We continue working with the parameter A. The reader will recognize the constants from Lemma 15.8 and its proof. Let $P(r, s)$ denote the parallelogram, as above, such that the bottom vertices are $(r, 0)$ and $(s, 0)$.

16.2.2 Example 1

Consider the paralleogram $P(0, 2A)$. Using the first method, we cover $P(0, 2A)$ with rectangles Q_1 and Q_2. The vertices of Q_1 are

$$(0,0), \qquad (0, A), \qquad (2A, 0), \qquad (2A, A). \tag{16.3}$$

The vertices of Q_2 are

$$(0, A), \qquad (0, 1), \qquad (2A, A), \qquad (2A, 1). \tag{16.4}$$

Compare item 1 in the proof of Lemma 15.8.

16.2.3 Example 2

Let $\lambda = \lambda(-1, -1) = 1 - A^2$. We cover $P(2A - \lambda, 2A)$ with 2 rectangles Q_1 and Q_2 using the method above. The coordinates of Q_1 are

$$(-1 + 2A, 0), \qquad (-1 + 2A, A), \qquad (2A, 0), \qquad (2A, A). \tag{16.5}$$

The coordinates of Q_2 are

$$(-1 + A, A), \qquad (-1 + A, 1), \qquad (2A, A), \qquad (2A, 1). \tag{16.6}$$

Compare item 2 in the proof of Lemma 15.8. Note that the coordinates of parallelogram P vary quadratically with A, whereas the coordinates of the rectangles vary linearly.

16.2.4 Example 3

Let $\lambda = \lambda(-1, 0) = 1 + 2A - A^2$. We cover $P(2A - \lambda, 2A)$ with 2 rectangles Q_1 and Q_2 using the method above. The coordinates of Q_1 are

$$(-1, 0), \qquad (-1, A), \qquad (2A, 0), \qquad (2A, A). \tag{16.7}$$

The coordinates of Q_2 are

$$(-1 - A, A), \qquad (-1 - A, 1), \qquad (2A, A), \qquad (2A, 1). \tag{16.8}$$

Compare item 3 in the proof of Lemma 15.8.

16.2.5 Example 4

Let $\lambda = \lambda(-1, 1) = 1 + 4A - A^2$. We cover $P(2A - \lambda, 2A)$ with 2 rectangles Q_1 and Q_2 using the method above. The coordinates of Q_1 are

$$(-1 - 2A, 0), \qquad (-1 - 2A, A), \qquad (2A, 0), \qquad (2A, A). \tag{16.9}$$

The coordinates of Q_2 are

$$(-1 - 3A, A), \qquad (-1 - 3A, 1), \qquad (2A, A), \qquad (2A, 1). \tag{16.10}$$

Compare item 4 in the proof of Lemma 15.8.

16.3 PROOF OF STATEMENT 1

The projection $\pi(x, y, z) = (x, z)$ is an isomorphism from the plane Π to the xz-plane. The inverse map is given by

$$\pi^{-1}(x, z) = (x, A - x, z). \tag{16.11}$$

For any pair (ϵ_1, ϵ_2) considered in the previous section, we define $Q_{j,+}(\epsilon_1, \epsilon_2; A)$ to be the inverse image of the relevant version of Q_j constructed above.

Example: The vertices of $Q_{2,+}(-1, -1; A)$ are

$$(-1 - 3A, 4A + 1, A), \quad (-1 - 3A, 4A + 1, 1), \quad (2A, -A, A), \quad (2A, -A, 1).$$

Once we make this construction, we have

$$\Sigma_+(\epsilon_1, \epsilon_2; A) \subset \bigcup_{j=1}^{2} Q_{j,+}(\epsilon_1, \epsilon_2; A). \tag{16.12}$$

To find the covering for $\Sigma_-(\ldots)$ we simply add the vector $(-A, A, 0)$ to all the coordinates.

We can easily work out the vertices of the corresponding 4-dimensional polytopes. We just compute the vertices at $A = 0$ and at $A = 1$ and take the convex hull. Thus the vertices of $Q_{2,+}(-1, -1)$ are

$$\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} -4 \\ 5 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

Working out the remaining 7 polytopes for the $(+)$ case is similar. Once we have these, we find the polytopes for the $(-)$ case by adding $(-1, 1, 0, 0)$ to all the vertices. These polytopes are stored in Billiard King.

We use the same method as in §11.1 to show that the 2 polytopes

$$Q_{j,\pm}(\epsilon_1, \epsilon_2), \qquad \gamma\left(R_\pm(\epsilon_1, \epsilon_2)\right)$$

have disjoint interiors for all $\gamma \in \Lambda$ and all possible choices. This time we need to use vectors in $\{-2, -1, 0, 1, 1\}^4$ to separate out the polytopes. This shows that the 2 regions

$$\Sigma^*_\pm(\epsilon_1, \epsilon_2), \qquad \gamma\left(R_\pm(\epsilon_1, \epsilon_2)\right)$$

have disjoint interiors for all choices.

There is one last detail to check. Recall that $S \subset \mathbf{R}^3$ is the slab between the planes $\{z = 0\}$ and $\{z = 1\}$. We still have the a priori possibility that the 2 sets

$$\Sigma(\ldots; A)^* \cap \partial S, \qquad R(\ldots; A)$$

are not disjoint for some A and some set of choices. In this case, a point in the interior of the infinite strip $\Sigma(\ldots; A)$ lies in the interior of $R(\ldots; A)$. But then some point in the interior of $\Sigma^*(\ldots; A)$ also lies in the interior of $R(\ldots; A)$. We have already ruled this out.

The proof we have given hides the pretty relationships between the various sets. The reader can get a better feel for why the Intersection Lemma is true using the hexagrid demo in Billiard King.

Here we show some representative images from this demo. We consider the pair $(-1,-1)$ in the $(-)$ case. Figure 16.3 shows the parallelogram $\Sigma^*_-(-1,-1;A)$ and the tiling $\mathcal{R}_- \cap \Sigma$. The slanting lines are part of the parallelogram and so are the top and bottom of the figure. The top is the line $\{z = 1\}$, and the bottom is the line $\{z = 0\}$. We use the usual planar projection to draw the figures. The rectangles $R_-(-1,-1;A)$ are darkly shaded. The rest of the tiling is lightly shaded. Notice the exact fit.

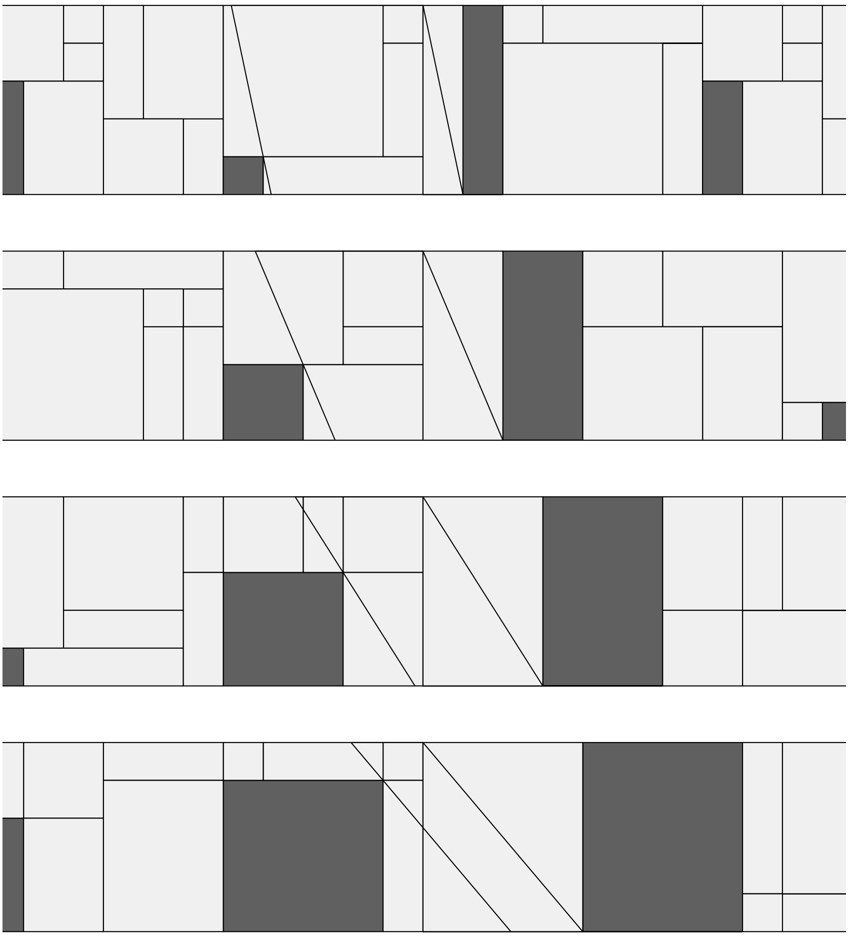

Figure 16.3: $\Sigma^*_-(-1,-1;A)$ and $\mathcal{R}_-(A)$ for $A = p/5$. Here $p = 1, 2, 3, 4$.

The picture is similar for other parameters and other choices of an (ϵ_1, ϵ_2) pair.

16.4 PROOF OF STATEMENT 2

Recall that

$$\Pi_A = \{x + y = A\} \subset \mathbf{R}^3$$

for each parameter. Here we write Π_A to emphasize the dependence on A. The hyperplane

$$\Pi = \bigcup_A \big(\Pi_A \times \{A\}\big) \tag{16.13}$$

is perpendicular to $(1, 1, 0, -1)$.

Say that a vector in $\mathbf{R}^4$ is *positive* if it lies on the same side of Π as the vector $(1, 1, 1, 0)$. Say that a convex integral polytope in $\mathbf{R}^4$ is *semipositive* if all of its vertices either lie on Π or else are positive.

Lemma 16.1 *Let P_A be a polyhedron in the orbit $\Lambda \underline{R}_{\pm}(\epsilon_1, \epsilon_2; A)$. Let P be the corresponding polytope. If*

$$\Pi_{\pm}(\epsilon_1, \epsilon_2; A) \cap \underline{P}_A \neq \emptyset, \tag{16.14}$$

then P is semipositive.

Proof: Let $X_A = \Pi_{\pm}(\epsilon_1, \epsilon_2; A)$. By statement 1 of the Intersection Lemma, X_A is disjoint from the interior of P_A. However, X_A is not disjoint from ∂P_A. Moreover, X_A is an open set in Π_A. From these properties, we see that P_A cannot have vertices on both sides of Π_A. Let $x_A \in \underline{P}_A \cap X_A$. By definition $x_A + (s, s, s) \in P_A$ for small s. Let $x = x_A \times \{A\}$. Then $x \in \partial P$ and $x + (s, s, s, 0) \in P$. This is possible only if P has some positive vertices. □

To finish the proof, it is just a matter of listing the semipositive polytopes and examining the vertices that lie on Π. As in §11.1, it suffices to examine a large but finite part of the orbit. Recall that Λ is generated by the three elements $\gamma_1, \gamma_2, \gamma_3$. Let $\Lambda_{10} \subset \Lambda$ be the set

$$\Lambda_{10} = \{a_1\gamma_1 + a_2\gamma_2 + a_3\gamma_3 | \; |a_1|, |a_2|, |a_3| \leq 10\}. \tag{16.15}$$

An argument similar to that in Lemma 11.1 shows that any intersection of the kind in Equation 16.14 for $P \in \Lambda\mathcal{R}$ is equivalent mod Λ to an intersection with $P \in \Lambda_{10}\mathcal{R}$.

Examining all the vertices of these finitely many polytopes, we find that the intersection points of

$$\Pi_+(-1, 1; A) \cap \Lambda_{10}\underline{R}_+(-1, 1; A)$$

are all equivalent mod Λ to $(0, A, 0) \in \partial\Pi_+(0, 2A)$, and moreover that there are no intersection points in the other cases. This establishes statement 2 of the Intersection Lemma.

16.5 PROOF OF STATEMENT 3

We prove statement 3 by the same method that we used for statement 2. Inspecting the vertices, we find exactly the pattern shown in Figure 15.2. Rather than dwell on this calculation, we show some figures from Billiard King. Define

$$\Pi_A^{(k)} = \Pi_A + 2^{-k}(1, 1, 1). \tag{16.16}$$

This is a slightly perturbed plane.

In Figure 16.4, we fix the parameter $A = 1/3$ and we plot the intersection of $\Pi^{(k)}$ with the tiling for $k = 3, 4, 5, 6$. The lightly shaded rectangles correspond to the label $(0, 1)$. The darkly shaded rectangles correspond to the label $(-1, 0)$. The figure evidently converges to what we have on the left hand side of Figure 15.1. The right hand side of Figure 15.1 is similar.

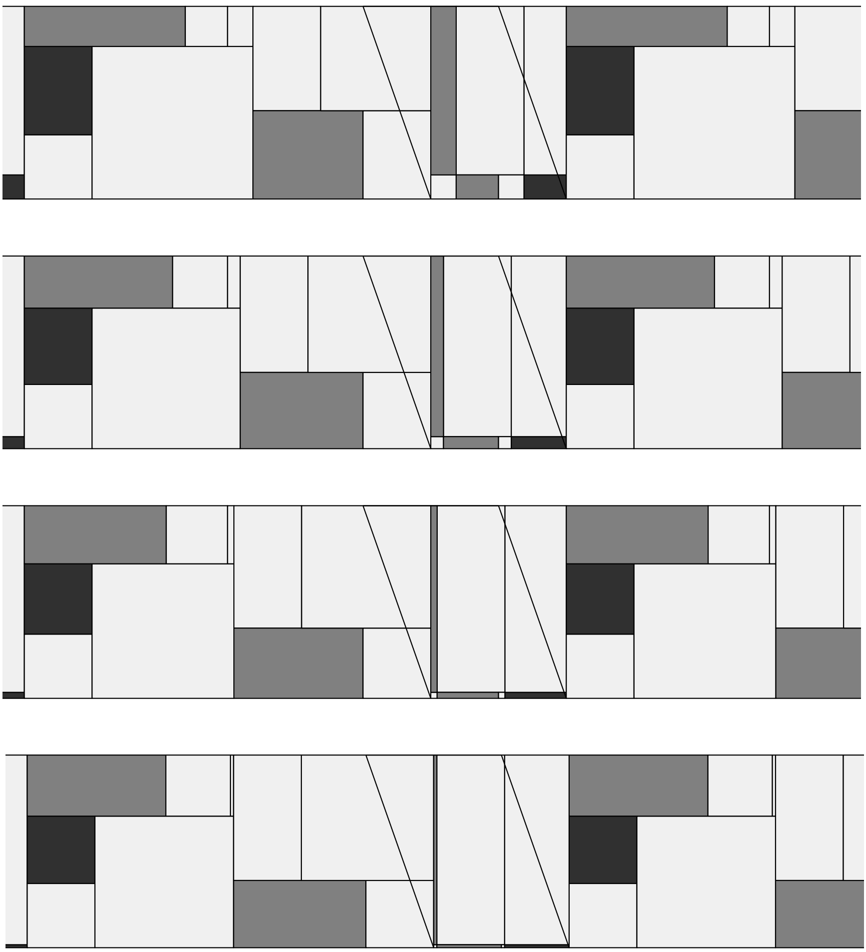

Figure 16.4: Perturbed slices.

In Figure 16.5, we keep $k = 5$ and show the parameters $A = p/5$ for $p = 1, 2, 3, 4$. The detail outside the parallelogram, though interesting, is irrelevant for our purposes.

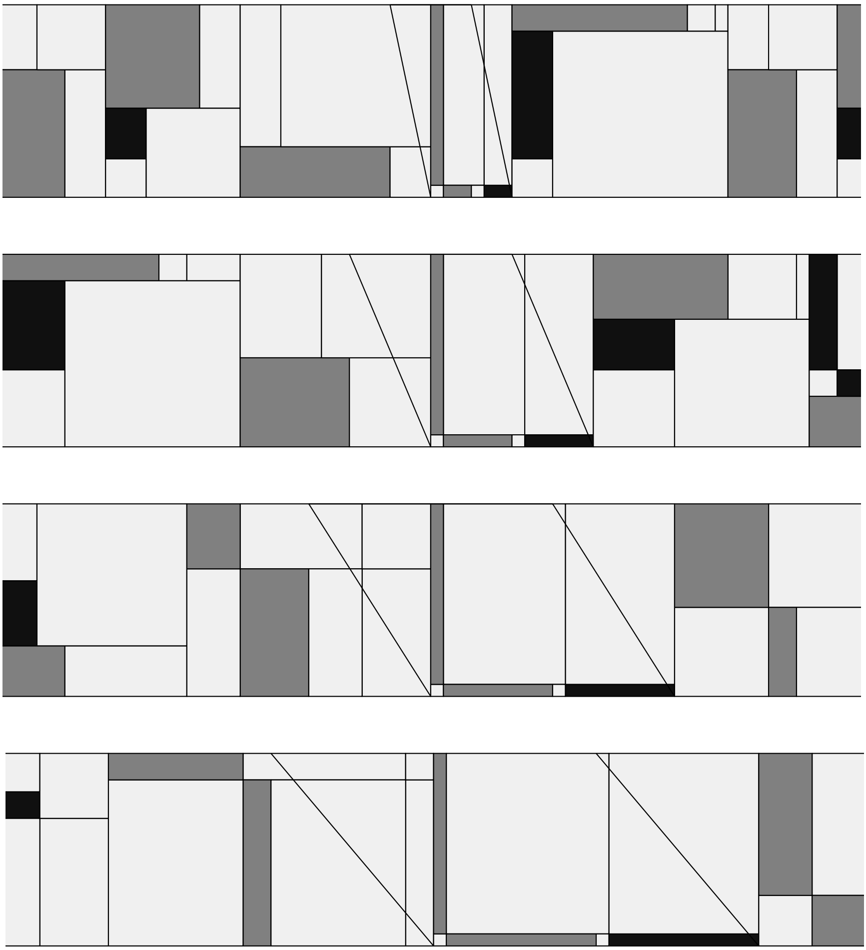

Figure 16.5: Perturbed slices.

Part 4. Period-Copying Theorems

In this part of the book, we will establish some results on period-copying. Our efforts culminate in the proof of Theorem 4.2, the final result needed for the proof of the Erratic Orbits Theorem. In Parts 5 and 6 we will use some of the other results we prove in this part.

- In Chapter 17, we prove some results about Diophantine approximation. There are two main topics. The first is an analysis of the inferior and superior sequences from Chapter 4, including a proof of the Superior Sequence Lemma. The second is the analysis of a device we call the *Diophantine constant*. We introduce the Diophantine constant in §17.4 and it plays an important role in our subsequent results. The reader interested only in Lemma 4.3 can skip everything in this chapter except §17.4.

- In Chapter 18, we prove the Diophantine Lemma. This result is the source of most of our period-copying results. As a quick application, we use the Diophantine Lemma to prove Lemma 4.3, the final ingredient in the proof of the Erratic Orbits Theorem for almost every parameter. The reader who is satisfied with the Erratic Orbits Theorem for almost every parameter can stop reading the book after this chapter.

- In Chapter 19, we state and prove the Decomposition Theorem. This theorem is an enhancement of the Room Lemma in §3.3. Our proof of the Decomposition Theorem is somewhat more tedious than we would like, but it turns out that Theorem 4.2 requires only a part of the Decomposition Theorem that is easier to prove. When the time comes, we will indicate what is necessary and what is not. We do need the full Decomposition Theorem for our work in Parts 5 and 6, however.

- In Chapter 20, we prove Theorem 4.2 by combining the Diophantine Lemma and the Decomposition Theorem.

This page intentionally left blank

Chapter Seventeen

Diophantine Approximation

17.1 EXISTENCE OF THE INFERIOR SEQUENCE

We will describe a hyperbolic geometry construction of the inferior sequence defined in §4.1. Our proof is similar to that for ordinary continued fractions. See [**BKS**]. Also, see [**Be**] for background on hyperbolic geometry, and [**Da**] for the classic theory of continued fractions.

Our model for the hyperbolic plane is the upper half-plane $\mathbf{H}^2 \subset \mathbf{C}$. The group $SL_2(\mathbf{R})$ acts isometrically by linear fractional transformations. The geodesics are vertical rays or semicircles centered on $\mathbf{R}$. The *Farey graph* is a tiling of $\mathbf{H}^2$ by ideal triangles. We join p_1/q_1 and p_2/q_2 by a geodesic iff $|p_1q_2 - p_2q_1| = 1$. The resulting graph divides the hyperbolic plane into an infinite symmetric union of ideal geodesic triangles. The Farey graph is one of the most beautiful constructions in mathematics. Figure 17.1 shows some of the edges of the Farey graph. The vertical lines in Figure 17.1 represent geodesics connecting 0 and 1 to ∞.

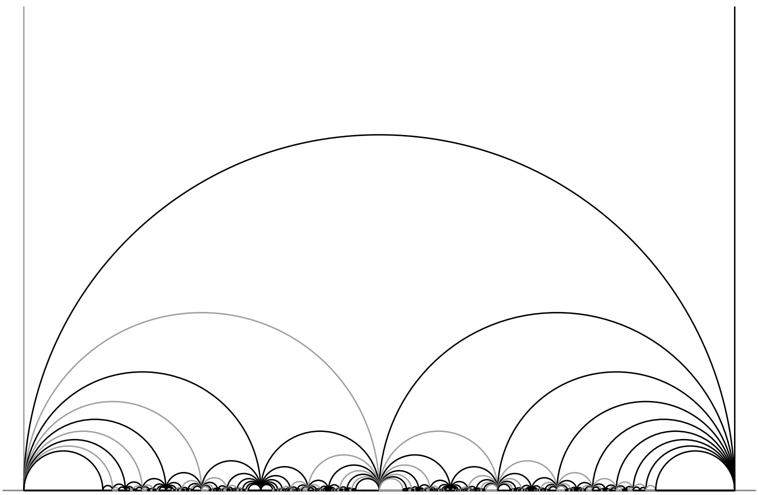

Figure 17.1: The Farey graph.

We modify the Farey graph by erasing all the geodesics that connect even fractions to each other. In Figure 17.1 these geodesics are shown in gray. The remaining edges partition $\mathbf{H}^2$ into an infinite union of ideal squares. The $(2, \infty, \infty)$-triangle group mentioned in Theorem 1.5 is the full isometry group of the Farey graph that respects the shadings in Figure 17.1.

We say that a *basic square* is one of these squares that has all vertices in the interval (0, 1). Each basic square has two opposing vertices that are labelled by positive odd rationals p_1/q_1 and p_2/q_2. These odd rationals satisfy

$$|p_1q_2 - p_2q_1| = 2. \tag{17.1}$$

Ordering so that $q_1 < q_2$, we call p_1/q_1 the *head* of the square, and p_2/q_2 the *tail* of the square. We draw an arrow in each odd square that points from the tail to the head, as in $p_1/q_1 \leftarrow p_2/q_2$. We call the odd square *right-biased* if the rightmost vertex is an odd rational, and *left-biased* if the leftmost vertex is an odd rational. Figure 17.2 shows a prototypical right-biased ideal square.

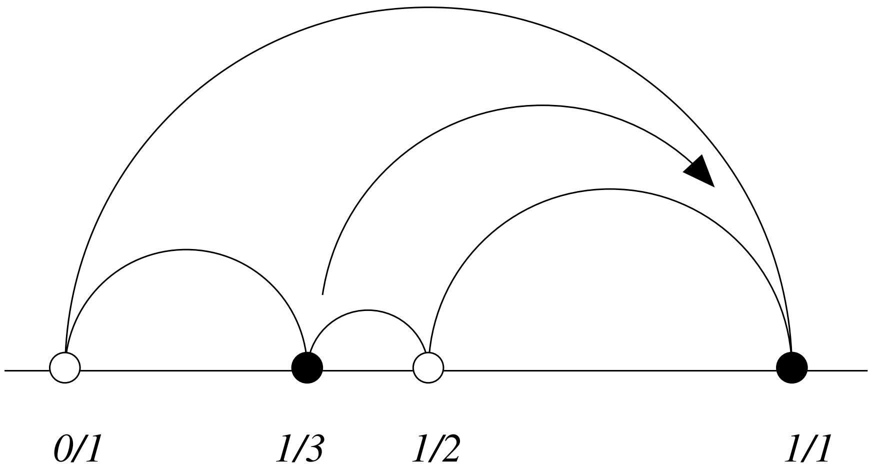

Figure 17.2: A right-biased ideal square.

The general form of a right-biased square is

$$\circ\frac{a_1}{b_1}, \qquad \bullet\frac{2a_1+a_2}{2b_1+b_2}, \qquad \circ\frac{a_1+a_2}{b_1+b_2}, \qquad \bullet\frac{a_2}{b_2}. \tag{17.2}$$

The general form of a left-biased square is

$$\bullet\frac{a_1}{b_1}, \qquad \circ\frac{a_1+a_2}{b_1+b_2}, \qquad \bullet\frac{a_1+2a_2}{b_1+2b_2}, \qquad \circ\frac{a_2}{b_2}. \tag{17.3}$$

The rightmost vertex in a right-biased square is the head. The leftmost vertex in a left-biased square is the head.

For an irrational parameter A, we simply drop the vertical line down from ∞ to A and record the sequence of basic squares we encounter. To form the inferior sequence, we list the heads of the encountered squares and weed out repeaters. Every time we encounter a new rational on our list, this rational and its predecessor are the two odd vertices of an ideal square. The nesting properties of the squares guarantee convergence.

17.2 STRUCTURE OF THE INFERIOR SEQUENCE

Now suppose that $\{p_n/q_n\}$ is the inferior sequence approximating A. Referring to Equation 4.1, we write $A_n = p_n/q_n$ and $(A_n)_\pm = (p_n)_\pm/(q_n)_\pm$. We have

$$(A_n)_- < A_n < (A_n)_+, \tag{17.4}$$

and these numbers form 3 vertices of an ideal square. A_n is the tail of the square.

Lemma 17.1 *The following are true for all indices m.*

1. *Let $N > m$. Then $A_{m-1} < A_m$ iff $A_{m-1} < A_N$.*
2. *If $A_{m-1} < A_m$, then $(q_m)_- = q_{m-1} + (q_m)_+$.*
3. *If $A_{m-1} > A_m$, then $(q_m)_+ = q_{m-1} + (q_m)_-$.*
4. *Either $A_m < A < (A_m)_+$ or $(A_m)_- < A < A_m$.*

Proof: Statement 1 follows from the nesting properties of the ideal squares encountered by the vertical geodesic γ as it converges to A.

For statement 2, note that $A_{m-1} < A_m$ iff these two rationals participate in a left-biased basic square, which happens iff $(q_m)_+ < (q_m)_-$. By definition,

$$q_{m-1} = |(q_m)_- - (q_m)_+|.$$

When $(q_m)_+ < (q_m)_-$, we can simply remove the absolute value symbol and solve for $(q_m)_-$. statement 3 is similar.

For statement 4, we will consider the case when $A_m < A_{m-1}$. The other case is similar. At some point, γ encounters the basic square with vertices

$$(A_m)_- < A_m < (A_m)_+ < A_{m-1}.$$

If $A_{m+1} < A_m$, then γ exits S between $(A_m)_-$ and A_m. So,

$$(A_m)_- < A < A_m.$$

If $A_{n+1} > A_m$, then γ exits S to the right of A_m. If γ exits S to the right of $(A_m)_+$, then γ next encounters a basic square S' with vertices

$$(A_m)_+ < O < E < A_{m-1},$$

where O and E are odd and even rationals, respectively. But then A_m would not be the term in the sequence after A_{m-1}. The term after A_{m-1} would lie in the interval $\big[O, A_{m-1})$. This is a contradiction. □

Let $[x]$ denote the floor of x. Let d_n be as in Equation 4.5. That is,

$$d_n = \left[\frac{q_{n+1}}{2q_n}\right], \qquad n = 0, 1, 2, 3...$$

Relatedly, define

$$\delta_n = \left[\frac{q_{n+1}}{q_n}\right], \qquad n = 0, 1, 2, 3... \tag{17.5}$$

Now we come to our main technical result about inferior sequences. This result is similar to results one sees for the successive terms of continued fraction approximants. See [**Da**]. Before we give the result, we make several clarifying remarks about it.

Remarks:
(i) In the result below, the notation $A_{m-1} < A_m > A_{m+1}$ means that $A_{m-1} < A_m$ and $A_m > A_{m+1}$, and similarly for the other lines.
(ii) There is a basic symmetry in the result below. If we swap all inequalities, then the signs $(+)$ and $(-)$ all switch. This symmetry swaps cases 1 and 3 and likewise swaps cases 2 and 4.
(iii) The same results hold for p in place of q. We used q just for notational convenience.

Lemma 17.2 *The following are true for any index* $m \geq 1$.

1. *If* $A_{m-1} < A_m < A_{m+1}$, *then*
   - δ_m *is odd,*
   - $(q_m)_+ < (q_m)_-$,
   - $(q_{m+1})_+ = d_m q_m + (q_m)_+$,
   - $(q_{m+1})_- = (d_m + 1)q_m + (q_m)_+$.
2. *If* $A_{m-1} > A_m < A_{m+1}$, *then*
   - δ_m *is even,*
   - $(q_m)_- < (q_m)_+$,
   - $(q_{m+1})_+ = d_m q_m - (q_m)_-$,
   - $(q_{m+1})_- = d_m q_m + (q_m)_+$.
3. *If* $A_{m-1} > A_m > A_{m+1}$, *then*
   - δ_m *is odd,*
   - $(q_m)_- < (q_m)_+$,
   - $(q_{m+1})_+ = (d_m + 1)q_m + (q_m)_-$,
   - $(q_{m+1})_- = d_m q_m + (q_m)_-$.
4. *If* $A_{m-1} < A_m > A_{m+1}$, *then*
   - δ_m *is even,*
   - $(q_m)_+ < (q_m)_-$,
   - $(q_{m+1})_+ = d_m q_m + (q_m)_-$,
   - $(q_{m+1})_- = d_m q_m - (q_m)_+$.

Proof: Cases 3 and 4 follow from cases 1 and 2 by symmetry. We will consider case 1 in detail, and case 2 briefly at the end.

In case 1, the vertical geodesic γ to A passes through the basic square S with vertices

$$A_{m-1} < (A_m)_- < A_m < (A_m)_+.$$

Since $A_n < A_{m+1}$, the geodesic γ next crosses through the geodesic α_m connecting A_m to $(A_m)_+$. Following this, γ encounters the basic squares S'_k for $k = 0, 1, 2, ...$ until it crosses a geodesic that does not have A_m as a left endpoint. By Equation 17.3 and induction, we get the following list of vertices for the square S'_k.

$$\frac{p_m}{q_m} < \frac{(k+1)p_m + (p_m)_+}{(k+1)q_m + (q_m)_+} < \frac{(2k+1)p_m + 2(p_m)_+}{(2k+1)q_m + 2(q_m)_+} < \frac{kp_m + (p_m)_+}{kq_m + (q_m)_+}. \quad (17.6)$$

Here S'_k is a left-biased square. But then there is some k such that

$$\frac{p_{m+1}}{q_{m+1}} = \frac{(2k+1)p_m + 2(p_m)_+}{(2k+1)q_m + 2(q_m)_+}, \qquad \frac{(p_{m+1})_+}{(q_{m+1})_+} = \frac{kp_m + (p_m)_+}{kq_m + (q_m)_+}. \quad (17.7)$$

Since $(q_m)_+ < (q_m)_-$, we have

$$2(q_m)_+ < q_m. \quad (17.8)$$

But then we have

$$\frac{p_{m+1}}{q_{m+1}} - \frac{p_m}{q_m} = \frac{2}{(2k+1)q_m^2 + 2q_m(q_m)_+} \in \left(\frac{2}{(2k+2)q_m^2}, \frac{2}{(2k+1)q_m}\right). \quad (17.9)$$

Hence

$$\delta_m = (2k+1) \equiv 1 \mod 2.$$

Here $k = d_m$. This takes care of the second implication. Equation 17.7 is the formula for $(q_{m+1})_+$. Lemma 17.1 now gives the formula for $(q_{m+1})_-$.

In case 2, the vertical geodesic γ again encounters the basic square S. This time γ exits S through the geodesic joining $(A_m)_-$ to A_m. This fact follows from the inequality

$$A_m > A_{m-1} > (A_m)_-,$$

a result of Lemma 17.1. Following this, γ encounters the basic squares S''_k, for $k = 0, 1, 2, ...$ until it crosses a geodesic that does not have A_m as a right endpoint. The coordinates for the vertices of S''_k are just like those in Equation 17.7, except that all the terms have been reversed and each $(\cdot)_+$ is switched to $(\cdot)_-$. The rest of the proof is similar. □

Remark: An important corollary of Lemma 17.2 is that either of the following data determines the inferior sequence uniquely.

- The sequence $\{\delta_n\}$.
- The sequence $\{d_n\}$ and the sequence $\{\sigma_n\}$, where σ_n is the sign of $A_{n+1} - A_n$.

The sequence $\{d_n\}$ in itself does not have enough information to determine the inferior sequence uniquely.

17.3 EXISTENCE OF THE SUPERIOR SEQUENCE

The following result completes the proof of the Superior Sequence Lemma.

Lemma 17.3 $d_m \geq 1$ *infinitely often.*

Proof: We can sort the indices of the sequence into 4 types, depending on which case holds in Lemma 17.2. If this lemma is false, then n eventually has odd type. But it is impossible for n to have type 1 and for $n+1$ to have type 3. Hence n eventually has constant type, say type 1. (The type 3 case has a similar treatment.) Looking at the formula in case 1 of Lemma 17.2, we see that the sequence $\{(q_n)_+\}$ eventually is constant. But then

$$r = \lim_{n\to\infty} \frac{(q_n)_+ p_n}{q_n}$$

exists. Since

$$(q_n)_+ p_n \equiv -1 \mod q_n,$$

$q_n \to \infty$, we must have $r \in \mathbf{Z}$. But then $\lim p_n/q_n \in \mathbf{Q}$, and we have a contradiction. □

Lemma 17.4 *If* $d_m \geq 1$, *then*

$$\left|\frac{p_N}{q_N} - \frac{p_m}{q_m}\right| < \frac{2}{d_m q_m^2} \quad \forall N > m, \qquad \left|A - \frac{p_m}{q_m}\right| \leq \frac{2}{d_m q_m^2}.$$

Proof: The first conclusion implies the second. We will consider the case when $A_m < A_{m+1}$. By Lemma 17.1, we have

$$|A_N - A_m| \leq |(A_{m+1})_+ - A_m| = \frac{1}{q_m (q_{m+1})_+}. \tag{17.10}$$

If m is an index of type 1, then

$$(q_{m+1})_+ = d_m q_m + (q_m)_+ > d_m q_m. \tag{17.11}$$

If m is an index of type 2, then Lemma 17.2 tells us that

$$(q_{m+1})_+ = (q_{m+1})_- - q_m = d_m q_m + (q_m)_+ - q_m > \left(d_m - \frac{1}{2}\right) q_m \geq \frac{1}{2} d_m q_m. \tag{17.12}$$

Combining Equations 17.10–17.12 we obtain the result. □

Remark: The superior sequence has Diophantine approximation properties similar to those of the sequence of continued fraction approximants. While these two sequences are related, they are generally not the same. For one thing, the superior sequence involves only odd rationals. We can, for example, certainly find irrationals whose sequence of continued fraction approximants consists of only even rationals. In this case, the two sequences are forced to be different.

17.4 THE DIOPHANTINE CONSTANT

17.4.1 Basic Definition

We have two odd rationals $A_1 = p_1/q_1$ and $A_2 = p_2/q_2$. We define the real number $a = a(A_1, A_2)$ by the formula

$$\left|\frac{p_1}{q_1} - \frac{p_2}{q_2}\right| = \frac{2}{aq_1^2}. \tag{17.13}$$

We call (A_1, A_2) *admissible* if $a(A_1, A_2) > 1$.

Define

$$\lambda_1 = \frac{(q_1)_+}{q_1} \in (0, 1). \tag{17.14}$$

If $A_1 < A_2$, we define

$$\Omega = \text{floor}\big((a/2) - \lambda_1\big) + 1 + \lambda_1. \tag{17.15}$$

If $A_1 > A_2$, we define

$$\Omega = \text{floor}\big((a/2) + \lambda_1\big) + 1 - \lambda_1. \tag{17.16}$$

Remark: The only fact relevant for Lemma 4.3 is that $a > 4$ implies that $\Omega > 2$. The reader who cares mainly about Lemma 4.3 can skip the rest of this chapter.

17.4.2 Meaning of the Constant

Let $[x]$ denote the floor of x. We say that an integer μ is *good* if

$$[\mu A_1] = [\mu A_2]. \tag{17.17}$$

Our next result is meant to apply when (A_1, A_2) is admissible. Also, we consider the case where $A_1 < A_2$.

Lemma 17.5 (Goodness) *If $\mu \in (-q_1, \Omega q_1) \cap \mathbf{Z}$, then μ is a good integer.*

We will prove this result in two steps.

Lemma 17.6 *If $\mu \in (-q_1, 0)$, then μ is good.*

Proof: Since q_1 is odd, we have unique integers j and M such that

$$\mu A_1 = M + (j/q_1), \qquad |j| < q_1/2 \tag{17.18}$$

By hypotheses, $a > 1$. Hence

$$|A_2 - A_1| < 2/q_1^2 \tag{17.19}$$

in all cases. If this result is false, then there is some integer N such that

$$\mu A_2 < N \leq \mu A_1. \tag{17.20}$$

Referring to Equation 17.18, we have

$$\frac{|j|}{q_1} < \mu A_1 - N \leq \mu A_1 - \mu A_2 < \frac{2|\mu|}{q_1^2} < \frac{2}{q_1}. \tag{17.21}$$

If $j = 0$, then q_1 divides μ, which is impossible. Hence $|j| = 1$. If $j = -1$, then μA_1 is $1/q_1$ less than an integer. Hence $\mu A_1 - N \geq (q_1 - 1)/q_1$. This is false, so we must have $j = 1$.

From the definition of λ_1, we have the following implication.

$$\mu \in (-q_1, 0) \quad \text{and} \quad \mu p_1 \equiv 1 \mod q_1 \quad \Longrightarrow \quad \mu = -\lambda_1 q_1. \tag{17.22}$$

Equation 17.18 implies

$$\frac{\mu p_1}{q_1} - \frac{1}{q_1} \in \mathbf{Z}.$$

But then $\mu p_1 \equiv 1 \mod q_1$. Equation 17.22 now tells us that $\mu = -\lambda_1 q_1$. Hence $|\mu| < q_1/2$. But now Equation 17.21 is twice as strong and gives $|j| = 0$. This is a contradiction. □

Lemma 17.7 *If $\mu \in (0, \Omega q_1)$, then μ is good.*

Proof: We observe that $\Omega < a$, by Equation 17.15. If this result is false, then there is some integer N such that $\mu A_1 < N \leq \mu A_2$. If $\mu A_2 = N$, then q_2 divides μ. But then

$$\mu \geq q_2 \geq a q_1 > \Omega q_1.$$

This is a contradiction. Hence

$$\mu A_1 < N < \mu A_2. \tag{17.23}$$

Referring to Equation 17.18, we have

$$\frac{|j|}{q_1} \leq N - \mu A_1 < \mu(A_2 - A_1) = \frac{2\mu}{a q_1^2} < \frac{2}{q_1}. \tag{17.24}$$

Suppose that $j \in \{0, 1\}$ in Equation 17.18. Then

$$1 - \frac{1}{q_1} \leq N - \mu A_1 \leq \mu A_2 - \mu A_1 < \frac{1}{q_1},$$

a contradiction. Hence $j = -1$. Hence $\mu > a q_1/2$.

Since $j = -1$, Equation 17.18 now tells us that $\mu p_1 + 1 \equiv 0 \mod q_1$. But then

$$\mu = k q_1 + (q_1)_+ \tag{17.25}$$

for some $k \in \mathbf{Z}$. On the other hand, from Equation 17.15 and the fact that $\mu < \Omega q_1$, we have

$$\mu < k' q_1 + (q_1)_+, \qquad k' = \Big(\text{floor}\big((a/2) - \lambda_1\big) + 1\Big). \tag{17.26}$$

Comparing the last two equations, we have $k \leq k' - 1$. Hence

$$k \leq \Big(\text{floor}\big((a/2) - \lambda_1\big)\Big). \tag{17.27}$$

Therefore

$$\mu \leq \Big(\text{floor}\big((a/2) - \lambda_1\big)\Big) q_1 + \lambda_1 q_1 \leq a q_1/2.$$

But we have already shown that $\mu > a q_1/2$. This is a contradiction. □

17.5 A STRUCTURAL RESULT

Now we will explain how the Diophantine constant interacts with the inferior sequence we defined above. Let $A = p/q$ be an odd rational. We say that A' is a *near predecessor* of A if A' precedes A in the inferior sequence but does not precede the superior predecessor of A. The inferior and superior predecessors of A are the two extreme examples of near predecessors of A. Here is a nice characterization of the Diophantine constant for these pairs of rationals.

Lemma 17.8 *If A' is a near predecessor of A, then the following are true.*

1. *If $A' < A$, then $\Omega q' = q' + q_+$.*
2. *If $A' > A$, then $\Omega q' = q' + q_-$.*

Proof: There is a finite chain

$$A' = A_1 \leftarrow \cdots \leftarrow A_m = A. \tag{17.28}$$

Referring to Equation 4.5, we have

$$d_1 \geq 0, \qquad d_2 = \cdots = d_{m-1} = 0.$$

By Lemma 17.1, $A_1 < A_2$ iff $A' < A$. We will consider the case when $A_1 < A_2$. The other case is similar. Recall that

$$A - A' = \frac{2}{a(q')^2}, \qquad \Omega = \text{floor}\left(\frac{a}{2} - \lambda\right) + 1 + \lambda, \qquad \lambda = \frac{q'_+}{q'}. \tag{17.29}$$

Hence

$$\Omega q' = q'(N+1) + q'_+, \qquad N = \text{floor}\big((a/2) - \lambda\big). \tag{17.30}$$

There are two cases to consider, depending on whether δ_1 is odd or even. Here δ_1 is as in Equation 17.5. If δ_n is odd, then we have case 1 of Lemma 17.2. In this case, we will show below that $d_1 = N$. By case 1 of Lemma 17.2, we have

$$(q_2)_+ = d_1 q_1 + (q_1)_+ = N q_1 + (q_1)_+. \tag{17.31}$$

If δ_n is even, then we show below that $d_1 = N + 1$. By case 2 of Lemma 17.2, we have

$$(q_2)_+ = d_1 q_1 - (q_1)_- = (d_1 - 1) q_1 + q_+ = N q_1 + (q_1)_+. \tag{17.32}$$

We obtain the same result in both cases.

Repeated applications of Lemma 17.2, case 1, give us

$$\begin{gathered} q_+ = (q_m)_+ = \ldots = (q_2)_+ = \\ Nq' + q'_+ = \\ (N+1)q' - q' + q'_+ = \\ \Omega q' - q'. \end{gathered}$$

Rearranging this gives statement 1. □

We have some unfinished business from the previous result. As above, we define

$$\lambda = \frac{q'_+}{q'}, \qquad N = \text{floor}\left(\frac{a}{2} - \lambda\right) \tag{17.33}$$

Also, the sequences $\{\delta_n\}$ and $\{d_n\}$ are as in Lemma 17.2.

Lemma 17.9 *If $A_1 < A_2$ and δ_1 is odd, then $d_1 = N$.*

Proof: Rearranging the basic definition of $a(A', A)$ and using $A' = A_1$ and $A = A_m$ in Equation 17.28, we have

$$\frac{a}{2} = \frac{1}{q_1^2|A_1 - A_m|}.$$

By Lemma 17.1 and monotonicity, we have

$$\frac{1}{q_1^2|A_1 - (A_2)_+|} < \frac{a}{2} < \frac{1}{q_1^2|A_1 - A_2|}. \tag{17.34}$$

After some basic algebra, we have

$$d_1 + \lambda_1 =^* \frac{(q_2)_+}{q_1} < \frac{a}{2} < \frac{q_2}{2q_1}. \tag{17.35}$$

The starred inequality is case 1 of Lemma 17.2. The lower bound gives us

$$d_1 < (a/2) - \lambda_1 \tag{17.36}$$

Here λ_1 is the same as λ in Equation 17.29. Since $d_1 \in \mathbf{Z}$, we obtain $d_1 \leq N$. On the other hand, the upper bound gives us

$$N = \text{floor}\left(\frac{a}{2} - \lambda_1\right) \leq \text{floor}\left(\frac{q_2}{2q_1} - \lambda_1\right) \leq d_1. \tag{17.37}$$

In short, $N \leq d_1$. Combining the two halves gives $N = d_1$. □

Lemma 17.10 *If $A_1 < A_2$ and δ_1 is even, then $d_1 = N + 1$.*

Proof: The proof is very similar to that for the other case. Here we mention the 2 changes. The first change is that $(d_1 - 1) + \lambda_1$ occurs on the left hand side of Equation 17.35, by case 2 of Lemma 17.2. This gives us

$$d_1 \leq N + 1.$$

The second change occurs on the right hand side of Equation 17.37. By case 2 of Lemma 17.2, we know that floor(q_2/q_1) is even. Hence $q_2/(2q_1)$ has a fractional part less than $1/2$. But, also by case 2 of Lemma 17.2, λ_1 has a fractional part greater than $1/2$. Hence

$$\text{floor}\left(\frac{q_1}{2q_1} - \lambda_1\right) = \text{floor}\left(\frac{q_1}{2q_1}\right) - 1 \leq d_1 - 1.$$

This gives us the bound $N \leq d_1 - 1$, or $N + 1 \leq d_1$. Putting the two halves together, we have $d_1 = N + 1$. □

Chapter Eighteen

The Diophantine Lemma

18.1 THREE LINEAR FUNCTIONALS

Let p/q be an odd rational.

Consider the following linear functionals.

$$F(m,n) = \left(\frac{p}{q}, 1\right) \cdot (m,n). \tag{18.1}$$

$$G(m,n) = \left(\frac{q-p}{p+q}, \frac{-2q}{p+q}\right) \cdot (m,n). \tag{18.2}$$

$$H(m,n) = \left(\frac{-p^2+4pq+q^2}{(p+q)^2}, \frac{2q(q-p)}{(p+q)^2}\right) \cdot (m,n). \tag{18.3}$$

We have $F = (1/2)M$, where M is the fundamental map from Equation 2.10. We can understand G and H by evaluating them on a basis.

$$H(V) = G(V) = q; \qquad H(W) = -G(W) = \frac{q^2}{p+q}. \tag{18.4}$$

Here $V = (q, -p)$ and W are the vectors from Equation 3.2. We can also understand G by evaluating on a simpler basis.

$$G(q,-p) = q; \qquad G(-1,-1) = 1. \tag{18.5}$$

We can also (further) relate G and H to the hexagrid in Chapter 3. A direct calculation establishes the following result.

Lemma 18.1 *The fibers of G are parallel to the top left edge of the arithmetic kite. The fibers of H are parallel to the top right edge of the arithmetic kite. Also, $\|\nabla G\| \leq 3$ and $\|\nabla H\| \leq 3$.*

Here ∇ is the gradient.

Given any interval I, define

$$\Delta(I) = \{(m,n) |\ G(m,n), H(m,n) \in I\} \cap \{(m,n) |\ F(m,n) \geq 0\}. \tag{18.6}$$

This set is a triangle whose bottom edge is the baseline of $\Gamma(p/q)$.

18.2 THE MAIN RESULT

Lemma 18.2 (Diophantine) *Let (A_1, A_2) be an admissible pair of odd rationals.*

1. *If $A_1 < A_2$, let $I = [-q_1 + 2, \Omega q_1 - 2]$.*

2. *If $A_1 > A_2$, let $I = [-\Omega q_1 + 2, q_1 - 2]$.*

Then $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ agree on $\Delta_1(I) \cup \Delta_2(I)$.

Figure 18.1 illustrates our result for $A_1 = 7/25$ and $A_2 = 11/39$. We have plotted the arithmetic graphs for both parameters and then superimposed them. The "lines" that stick out in the figure are the places where the graphs disagree. These "lines" are essentially parallel to the lines of the hexagrid for either graph. (For the two graphs, the respective hexagrid lines are nearly parallel to each other on account of the nearness of the two rationals involved.) The shaded region is $\Delta_1(-q_1, \Omega q_1)$, a set very slightly larger than $\Delta_1(I)$. The sets $\Delta_1(I)$ and $\Delta_2(I)$ are almost identical.

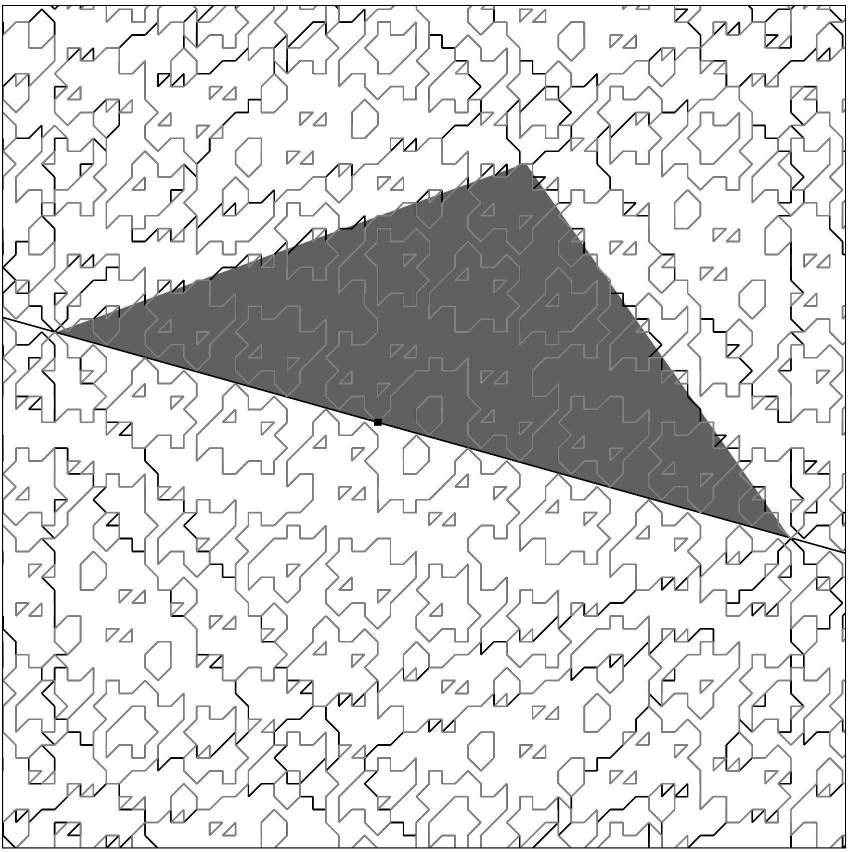

Figure 18.1: The Diophantine Lemma for 7/25 and 11/39.

18.3 A QUICK APPLICATION

Here we use the Diophantine Lemma to prove Lemma 4.3. This completes our proof of the Erratic Orbits Theorem for almost every parameter, as we indicated in Part 1. The reader who is satisfied with this result can stop reading the book at the end of this chapter.

We will prove Lemma 4.3 when $A_1 < A_2$. The other case is similar. By hypothesis, we have

$$a(A_1, A_2) > 4. \tag{18.7}$$

From Equation 17.15, we get

$$\Omega > 2. \tag{18.8}$$

Let $R_1 = R(A_1)$ be the parallelogram from the Room Lemma. Let

$$u = W_1, \qquad w = V_1 + W_1 \tag{18.9}$$

denote the top left and right vertices of R_1. We compute

$$G_1(u) = -\frac{q_1^2}{p_1 + q_1} > -q_1 + 2, \qquad H_1(w) = \frac{q_1^2}{p_1 + q_1} + q_1 < \Omega q_1 - 2. \tag{18.10}$$

The inequalities hold once p_1 is sufficiently large. Given the description of the fibers of G, we have

$$G(u) \le G(v) \le H(v) \le H(w), \qquad \forall v \in R_1. \tag{18.11}$$

The middle inequality uses the fact that $F(v) \ge 0$. In short, we have made the extremal calculations. This calculation shows that $v \in \Delta_1(I)$ for all $v \in R_1$. The Diophantine Lemma now shows that Γ_1 and Γ_2 agree in R_1.

When v lies in the bottom edge of R_1, we have

$$G_1(v), H_1(v) \in [0, q_1]. \tag{18.12}$$

Given the gradient bounds $\|\nabla G_1\| \le 3$ and $\|\nabla H_1\| \le 3$, we see that

$$G_1(v), H_1(v) \in [-q_1 + 2, \Omega q_1 + 2], \tag{18.13}$$

provided that v is within $q_1/4$ of the bottom edge of R_1. Hence Γ_1 and Γ_2 agree in the $q_1/4$ neighborhood of the bottom edge of R_1.

By the Room Lemma, $\Gamma_1^1 \subset R_1$. Hence $\Gamma_1^1 \subset \Gamma_2$. The calculation involving the bottom edge of R_1 shows that $\Gamma_1^{1+\epsilon} \subset \Gamma_2$ for $\epsilon = 1/4$. Since the right endpoint of Γ_2^1 is far to the right of any point on $\Gamma_1^{1+\epsilon}$, we have $\Gamma_1^{1+\epsilon} \subset \Gamma_2^1$, as desired.

Remark: We proved Lemma 4.3 for $\epsilon = 1/4$ rather than $\epsilon = 1/8$, which is what we originally claimed. We do not care about the value of ϵ as long as it is positive.

18.4 PROOF OF THE DIOPHANTINE LEMMA

We will establish the case when $A_1 < A_2$. The other case has a nearly identical proof. Recall that an integer μ is good if

$$[\mu A_1] = [\mu A_2] \tag{18.14}$$

We call μ 1-*good* if $\mu + \epsilon$ is good for all $\epsilon \in \{-1, 0, 1\}$. We can subject a lattice point (m, n) to the reduction algorithm in §6.6. For $\theta \in \{1, 2\}$, we perform the algorithm relative to the parameter A_θ. This produces integers X_θ, Y_θ, and Z_θ. Below we prove the following result.

Lemma 18.3 (Agreement) *Suppose, for at least one choice of* $\theta \in \{1, 2\}$*, that the following numbers are all* 1*-good.*

- m
- $m - X_\theta$
- $m - Y_\theta$
- $m + Y_\theta - X_\theta$.

Then $\widehat{\Gamma}_1$ *and* $\widehat{\Gamma}_2$ *agree at* (m, n).

Next, we prove the following result.

Lemma 18.4 (Good Integer) *If* $(m, n) \in \Delta_1(I) \cup \Delta_2(I)$*, then the integers in the Agreement Lemma all lie in* $(-q_1 + 1, \Omega q_1 - 1)$ *for at least* 1 *choice of* $\theta \in \{1, 2\}$.

By the Goodness Lemma in §17.4, all the numbers in the Agreement Lemma are 1-good. The Diophantine Lemma now follows immediately.

Remarks:
(i) As one can see in Figure 18.1, the Diophantine Lemma also works for points below the baseline. One can give a proof for points below the baseline that is nearly identical to the proof we give for points above the baseline. We have stated only the "above" case because the restriction makes our argument a bit easier and this is the only case we need for applications. In light of the symmetry results we established in §12.3 and §12.4, the fact that the result holds symmetrically above and below the baseline should not be surprising.
(ii) As one can see from Figure 18.1, the Diophantine Lemma is quite sharp. We think that the sharp version runs as follows. The two arithmetic graphs agree at any point in $\Delta_1(-q_1, \Omega q_1)$ that is not adjacent to a point that lies outside $\Delta_1(-q_1, \Omega q_1)$. The slight fudging of the boundaries is an artifact of our proof. Our proof of the Decomposition Theorem in Chapter 19 would go easier if we had the sharp version of the Diophantine Lemma at our disposal, but the result we prove here is the best we can do.

18.5 PROOF OF THE AGREEMENT LEMMA

Lemma 18.5 *Let $\mu, \nu, N_j \in \mathbf{Z}$ and*

$$N_j = \left[\frac{\mu A_j + \nu}{1 + A_j}\right].$$

Suppose there is some $\theta \in \{1, 2\}$ such that both $\mu - N_\theta$ and $\mu - N_\theta + 1$ are good. Then $N_1 = N_2$.

Proof: Here [ ] is the floor function, as above. For the sake of contradiction, assume without loss of generality. that $N_1 < N_2$. Then

$$\mu A_1 + \nu < N_2(A_1 + 1), \qquad (\mu - N_2)A_1 < N_2 - \nu$$

$$N_2(A_2 + 1) \leq \mu A_2 + \nu, \qquad N_2 - \nu \leq (\mu - N_2)A_2.$$

The first equation implies the second in each case. The second items imply that $\mu - N_2$ is not good. On the other hand, we have

$$\mu A_1 + \nu < (N_1 + 1)(A_1 + 1), \qquad A_1(m - N_1 + 1) < N_1 + 1 - n.$$

$$(N_1 + 1)(1 + A_2) \leq \mu A_2 + \nu, \qquad A_2(m - N_2 + 1) \geq N_1 - 1 + n.$$

The first equation implies the second in each case. The second items imply that $\mu - N_1 + 1$ is not good. Now we have a contradiction. □

Corollary 18.6 *Referring to the Agreement Lemma, $(X_1, Y_1, Z_1) = (X_2, Y_2, Z_2)$.*

Proof: We apply the reduction algorithm from §6.6. We focus on the $(-)$ case, indicating the small differences for the $(+)$ case as we go along.

1. Let $z_j = A_j m + n$.

2. Let $Z_j = \text{floor}(z_j)$. Since m is good, we have $Z_1 = Z_2$. Call this common integer Z.

3. $y_j = z_j + Z_j = z_j + Z$. Hence $y_j = mA_j + n'$ for some $n' \in \mathbf{Z}$. [We have $y_j = z_j + Z + 1$ in the $(+)$ case.]

4. Recall that $Y_j = \text{floor}(y_j/(1 + A))$. To see that $Y_1 = Y_2$ we apply Lemma 18.5 to $(\mu, \nu, N_j) = (m, n', Y_j)$. Here we use the fact that $m - Y_\theta$ and $m - Y_\theta + 1$ are good. We set $Y = Y_1 = Y_2$. [We apply Lemma 18.5 to $(\mu, \nu, N_j) = (m, n' + 1, Y_j)$ in the $(+)$ case.]

5. Let $x_j = y_j - Y(1 - A_j) - 1$. Hence $x_j = (m + Y)A_j + n''$.

6. Recall that $X_j = \text{floor}(x_j/(1 + A))$. To see that $X_1 = X_2$, we apply Lemma 18.5 to $(\mu, \nu, N_j) = (m + Y, n'', X_j)$. Here we use the fact that $m + Y - X_\theta$ and $m + Y - X_\theta + 1$ are good integers. □

In the next result, all quantities except A_1 and A_2 are integers.

Lemma 18.7 *If $\mu - dN - \epsilon_1$ is good, then the statement*
$$(\mu A_j + \nu) - N(dA_j + 1) < \epsilon_1 A_j + \epsilon_2$$
is true or false independent of $j = 1, 2$.

Proof: Assume without loss of generality. that the statement is true for $j = 1$ and false for $j = 2$. Then
$$(\mu - dN - \epsilon_1)A_1 < \epsilon_2 + N - \nu \le (\mu - dN - \epsilon_1)A_2,$$
a contradiction. □

Let M_+ and M_- be as in §6.6. By the Master Picture Theorem, it suffices to show that the two images $M_+(m, n)$ and $M_-(m, n)$ land in the same polyhedra for both A_1 and A_2. We have already seen that the basic integers (X, Y, Z) are the same relative to both parameters. Here we recall the planes from §6.2.

- $\mathcal{Z}$, the union $\{z = 0\} \cup \{z = A\} \cup \{z = 1 - A\} \cup \{z = 1\}$.
- $\mathcal{Y}$, the union $\{y = 0\} \cup \{y = A\} \cup \{y = 1\} \cup \{y = 1 + A\}$.
- $\mathcal{X}$, the union $\{x = 0\} \cup \{x = A\} \cup \{x = 1\} \cup \{x = 1 + A\}$.
- $\mathcal{T}$, the union $\{x + y - z = A + j\}$ for $j = -2, 1, 0, 2, 1$.

Letting $\mathcal{S}$ stand for one of these partitions, we say that $\mathcal{S}$ is *good* if, for both sign choices and both parameters, the points $M_\pm(m, n)$ land in the same component of $R_\pm - \mathcal{S}$. Here we set $R_\pm = \mathbf{R}^3/\Lambda$, the domain of the maps $M_\pm$. By the Master Picture Theorem, Γ_1 and Γ_2 agree at (m, n), provided all the partitions are good. The proof works the same for the $(+)$ and the $(-)$ cases.

- For $\mathcal{Z}$, we apply Lemma 18.7 to $(\mu, \nu, d, N) = (m, n, 0, Z)$ to show that the statement $z_j - Z < \epsilon_1 A_j + \epsilon_2$ is truly independent of j for $\epsilon_1 \in \{-1, 0, 1\}$ and $\epsilon_2 \in \{0, 1\}$. The relevant good integers are $m - 1$ and m and $m + 1$.
- For $\mathcal{Y}$, we apply Lemma 18.7 to $(\mu, \nu, d, N) = (m, n', 1, Y)$ to show that the statement $z_j - Z < \epsilon_1 A_j + \epsilon_2$ is truly independent of j for $\epsilon_1 \in \{0, 1\}$ and $\epsilon_2 \in \{0, 1\}$. The relevant good integers are $m - Y$ and $m - Y - 1$.
- For $\mathcal{X}$, we apply Lemma 18.7 to $(\mu, \nu, d, N) = (m + Y, n'', 1, X)$. The relevant good integers are $m + Y - X$ and $m + Y - X - 1$.
- For $\mathcal{T}$, we define
$$\sigma_j = (x_j - X(1 + A_j)) + (y_j - Y(1 + A_j)) - (z_j - Z).$$
We have $\sigma_j = (m - X)A_j + n'''$ for some $n''' \in \mathbf{Z}$. Let $h \in \mathbf{Z}$ be arbitrary. To see that the statement $\sigma_j < A_j + h$ is truly independent of j, we apply Lemma 18.7 to $(\mu, \nu, d, N) = (m - X, n''', 1, 0)$. The relevant good integer is $m - X - 1$.

Remark: Our proof does not use the fact that $m - X + 1$ is a good integer. This technical detail is relevant for Lemma 18.10.

18.6 PROOF OF THE GOOD INTEGER LEMMA

We will assume that $(m, n) \in \Delta_\theta(I)$, for one of the two choices $\theta \in \{1, 2\}$. Here I is as in the Diophantine Lemma. Our proof works the same for $\theta = 1$ and $\theta = 2$. We set $p = p_\theta$ and $q = q_\theta$, etc.

We will show that all the integers that arise in our proof of Lemma 18.3 lie in $(-q_1, \Omega q_1)$. These integers have the form $N + \epsilon$ for $\epsilon \in \{-1, 0, 1\}$. We will show, for all relevant integers (except one), that $N \in J := (-q_1 + 1, \Omega q_1 - 1)$. For the exceptional case, see the remark following Lemma 18.10.

Lemma 18.8 $m \in J$.

Proof: We have $z = Am + n \geq 0$. We compute

$$-q_1 + 2 \leq G(m, n) = m - \frac{2z}{1 + A} \leq m. \tag{18.15}$$

$$\Omega q_1 - 2 \geq H(m, n) = m + \frac{2z(1 - A)}{(1 + A)^2} \geq m. \tag{18.16}$$

These inequalities establish that $m \in J$. □

Lemma 18.9 $m - Y \in J$.

Proof: We have $Y \geq 0$. Hence $m - Y \leq m \leq \Omega q_1 - 2$. We just need the lower bound and worry about the lower bound on $m - Y$. We first deal with the algorithm in §6.6 for the $(-)$ case. Let $G = G(m, n)$. We have $y = z + Z \leq 2z$. By the definition of Y, we have

$$Y \leq \frac{y}{1 + A} \leq \frac{2z}{1 + A}, \qquad Y < \frac{2z}{1 + A}. \tag{18.17}$$

At least one of the first two inequalities is sharp. This gives us the second inequality. Now we know that

$$m - Y > m - \frac{2z}{1 + A} = G \geq -q_1 + 2. \tag{18.18}$$

The last equality comes from Equation 18.15. In the $(+)$ case, we add 1 to Y, giving $m - Y > -q_1 + 1$. □

Lemma 18.10 $m - X \in J \cup \{\Omega q_1 - 1\}$.

Proof: The condition that $F(m, n) \geq 0$ implies that $y \geq Y \geq 0$. Hence

$$x = y - Y(1 - A) - 1 \in [-1, y - 1]. \tag{18.19}$$

Hence $X \in [-1, Y - 1]$. Hence

$$m - X \in [m - Y + 1, m + 1] \subset J \cup \{\Omega q_1 - 1\},$$

by the two previous results. □

Remark: As we remarked at the end of the proof of Lemma 18.3, the integer $m - X + 1$ does not arise in our proof of Lemma 18.3. The relevant integers $m - X$ and $m - X - 1$ are good, by the result above.

Lemma 18.11 $m + Y - X \in J$.

Proof: Our proof works the same in the $(+)$ and $(-)$ cases. Lemma 18.10 gives us $Y - X \geq 0$. Hence

$$m + Y - X \geq m > -q_1 + 1.$$

This takes care of the lower bound. Now we treat the upper bound. We have

$$Y = \text{floor}\left(\frac{y}{1+A}\right) \leq \frac{y}{1+A}, \qquad 1 + X = \text{floor}\left(1 + \frac{x}{1+A}\right) \geq \frac{x}{1+A}.$$

Hence

$$\begin{aligned} Y - X - 1 &\leq \\ \frac{y - x}{1+A} &=^1 \\ Y\frac{1-A}{1+A} + \frac{1}{1+A} &<^* \\ 2z\frac{1-A}{(1+A)^2} + \frac{1}{1+A} &=^2 \\ H - m + \frac{1}{1+A} &< \\ H - m + 1. \end{aligned}$$

The first equality comes from Equation 18.19. The second equality comes from Equation 18.16. The starred inequality comes from the upper bound in Equation 18.17. Adding m to both sides, we have

$$m + Y - X < H + 1 \leq \Omega q_1 - 1.$$

This completes the proof. □

Chapter Nineteen

The Decomposition Theorem

19.1 THE MAIN RESULT

The Room Lemma confines one period of $\Gamma(p/q)$ to a certain parallelogram $R(p/q)$ when p/q is odd. In this section we explain a sharper result, along the same lines, that confines one period of $\Gamma(p/q)$ to a union of two parallelograms. The reader might want to glance at Figure 19.1 before reading the definitions that follow.

Given an odd rational $A = p/q$, we construct the even rationals $A_\pm = p_\pm/q_\pm$. We let A' be the inferior predecessor of A, and we let A^* be the superior predecessor, as in §4.1. For each rational, we use Equation 3.2 to construct the corresponding V and W vectors. For instance, $V_+ = (q_+, -p_-)$ and $V_* = (q_*, -p_*)$. Now we define the following lines.

- L_0^- is the line parallel to V and containing W.
- L_1^- is the line parallel to V and containing W^*.
- L^- is the line parallel to V through $(0, 0)$.
- L_0^+ is the line parallel to W through $(0, 0)$.
- If $q_+ > q_-$, then L_1^+ is the line parallel to W through $-V_-$.
- If $q_+ < q_-$, then L_1^+ is the line parallel to W through $+V_+$.
- If $q_+ > q_-$, then L_2^+ is the line parallel to W through $+V_+$.
- If $q_+ < q_-$, then L_2^+ is the line parallel to W through $-V_-$.

The lines with the $(-)$ superscript have negative slope, and the lines with the $(+)$ superscript have positive slope. All the $(-)$ lines are parallel to each other, and all the $(+)$ lines are parallel to each other. Now we define the following parallelograms:

- R_1 is the parallelogram bounded by L^- and L_1^- and L_0^+ and L_1^+.
- R_2 is the parallelogram bounded by L^- and L_0^- and L_0^+ and L_2^+.

The parallelogram R_2 is the larger of the two parallelograms. It is both wider and taller. Note that translation by V carries the leftmost edge of $R_1 \cup R_2$ to the rightmost edge.

Here is the main result of this chapter.

Theorem 19.1 (Decomposition) *$R_1 \cup R_2$ contains a period of Γ.*

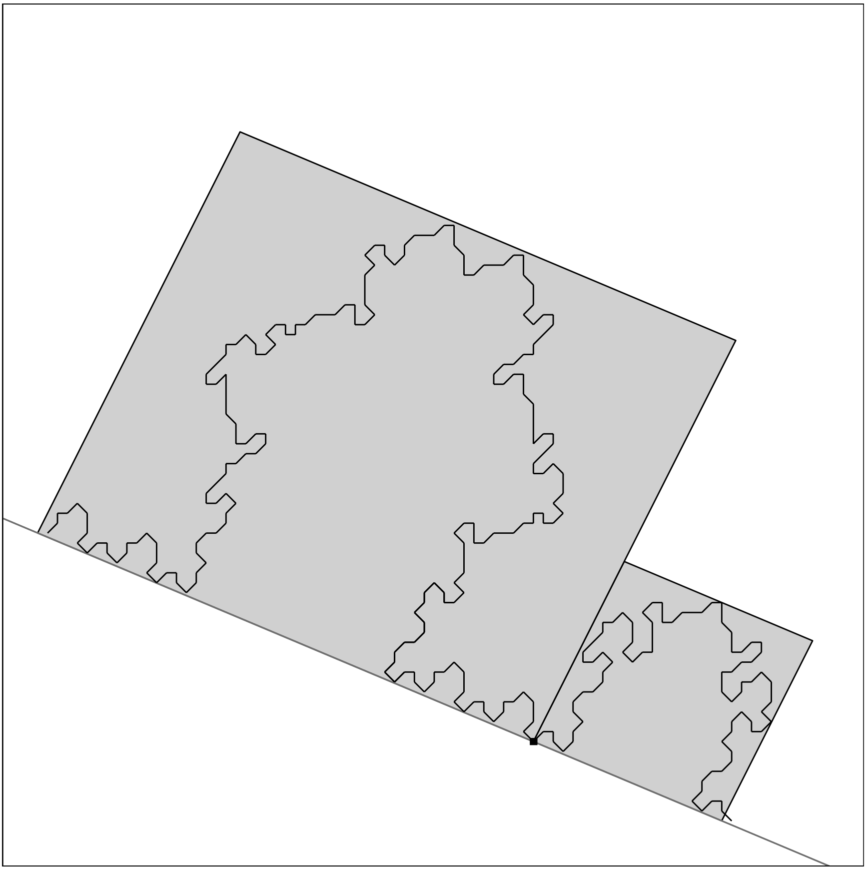

Figure 19.1: $\Gamma(29/69)$ and $R_1(29/69)$ and $R_2(29/69)$.

Figure 19.1 shows the example $A = 29/69$. In this case,

$$A_- = 21/50, \qquad A_+ = 8/19, \qquad A' = A^* = 13/31.$$

Since $q_+ < q_-$, the smaller R_1 lies to the right of the origin. The ratio between the heights of the two parallelograms is $q^*/q = 31/69$. The ratio between the widths is $q_+/q_- = 19/50$.

We would like to point out two features of this figure.

- The containment is very efficient. Notice that we cannot lower the tops of the parallelograms at all and still contain the polygonal arc.
- The arcs $\Gamma \cap R_1$ and $\Gamma \cap R_2$ have approximate bilateral symmetry. This is another indication that the decomposition is somehow canonical. The results in §12.4 explain this near-bilateral symmetry.

The interested reader can see the same phenomena for any other smallish odd rational using Billiard King.

19.2 A COMPARISON

The Room Lemma has two purposes. One purpose is to show that the graph $\Gamma(p/q)$ rises up $O(q)$ units away from the baseline. The second purpose is to confine the graph $\Gamma(p/q)$ to a small region in the plane. As we saw in the proof of Lemma 4.3, such a confinement result is necessary if we want to use the Diophantine Lemma. The Diophantine Lemma shows that a pair of arithmetic graphs agree in a certain region, and we must know that the portions of the graphs of interest to us actually lie in these regions.

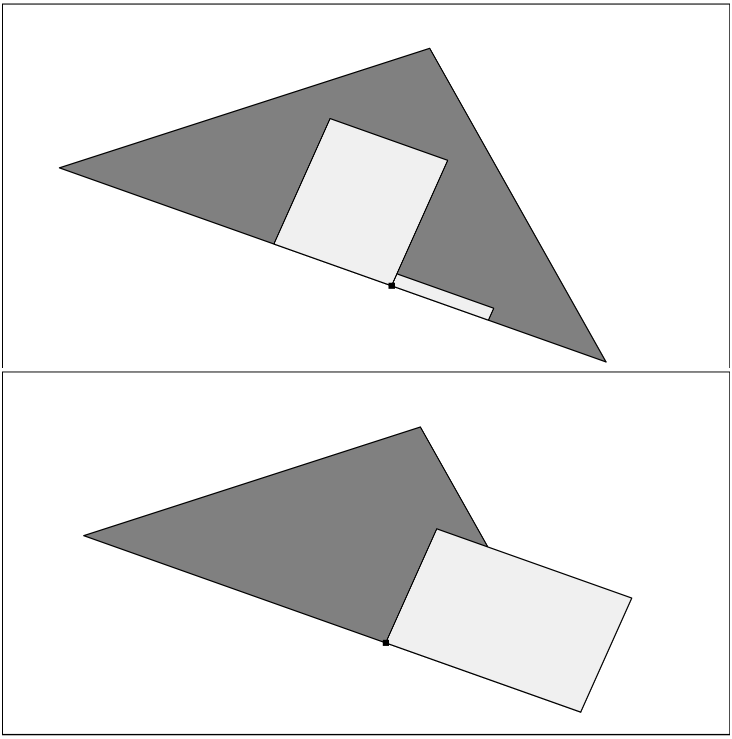

Figure 19.2: Two results compared.

It turns out that the Room Lemma is not a sufficiently strong result to give us the period copying we need in the general case. Figure 19.2 illustrates what we are talking about. Both parts show the region Δ from the Diophantine Lemma corresponding to the pair of rationals $11/31 \leftarrow 23/65$. The top also shows the region from the Decomposition Theorem. This region lies entirely inside Δ. Thus, from the top part, we conclude that $\Gamma(23/65)$ copies the same period of $\Gamma(11/31)$. The bottom part shows the room $R(11/31)$. From this figure we cannot conclude that $\Gamma(23/65)$ copies a full period of $\Gamma(11/31)$. At the same time, the translate $R(11/31) - V(11/31)$ that would lie just to the left of $R(11/31)$ also sticks out of Δ. Thus, from the bottom part, we cannot conclude that $\Gamma(23/65)$ copies *any* period of $\Gamma(11/31)$.

19.3 A CROSSING LEMMA

Now we begin the proof of the Decomposition Theorem. For ease of exposition, we treat the case when $q_- < q_+$. The other case has essentially the same proof. Recall that $v \in \mathbf{Z}^2$ is a low vertex if the baseline separates v from $v - (0, 1)$.

Lemma 19.2 (Crossing) Γ *crosses each of* L_1^+ *and* L_2^+ *only once and at a low vertex.*

Proof: Figure 19.3 illustrates our proof. Let L denote the line of slope $-A$ through the origin – i.e., the baseline. Σ_+ (respectively, Σ_-) is the infinite strip bounded by L and the first ceiling line above (respectively, below) L. By Theorem 1.10, there is one infinite component of $\widehat{\Gamma}$ in $\Sigma_\pm$. We call this component $\Gamma_\pm$. Here $\Gamma_+ = \Gamma$ is the component of interest to us.

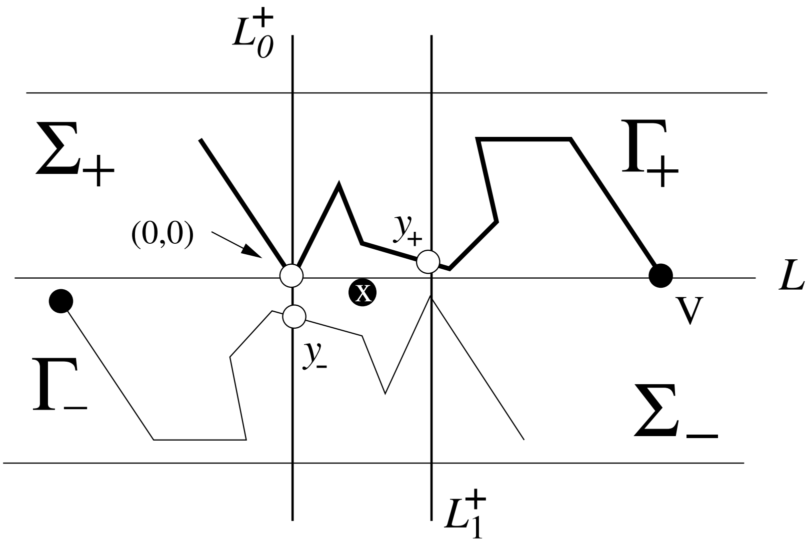

Figure 19.3: Applying rotational symmetry.

The point $x = (1/2)V_+$ is the fixed point of $\imath$, the rotation from Equation 12.13. We have

$$\imath(L_0^+) = L_1^+, \qquad \imath(\Gamma_-) = \Gamma_+, \qquad \imath(L) \downarrow L. \tag{19.1}$$

The last piece of notation means that $\imath(L)$ lies (very slightly) beneath L.

By the Hexagrid Theorem, $(0, 0)$ is the door corresponding to the point where Γ_+ crosses L_0^+ and *also* to the point y_- where Γ_- crosses L_0^+. This point is the intersection of L_0^+ with the edge connecting $(0, -1)$ to $(-1, 0)$. The image $y_+ = \imath(y_-) \in \imath(L_0^+) = L^+$ is the only point where $\imath(\Gamma_-) = \Gamma_+$ crosses L_+. This point is less than 1 unit from L because $\imath(L)$ lies beneath L. Hence $\Gamma = \Gamma_+$ crosses L^+ only once, within 1 unit of L. Since $L_1^+ = L_2^+ \pm V$ and Γ is invariant under translation by V, it suffices to prove the result for one of the lines, as we have finished. $\square$

19.4 MOST OF THE PARAMETERS

Let $A = p/q$ be an odd rational and let $A' = p'/q'$ be the superior predecessor. For Theorem 4.2, all we need is the following result.

Corollary 19.3 *The Decomposition Theorem holds if* $\min(p', q')$ *is sufficiently large.*

In this section we will prove the following explicit version of Corollary 19.3.

Lemma 19.4 *The Decomposition Theorem holds as long as* $p' \geq 3$ *and* $q' \geq 7$.

We will prove Lemma 19.4 through a series of smaller results. By the Crossing Lemma, we can divide a period of Γ into the union of two connected arcs. One of the these lies in what we call R_0 and the other lies in R_2. Each arc connects points near the bottoms of the boxes and otherwise does not cross the boundaries. Figure 19.4 is a schematic figure. Here R_0 is the union of the two shaded regions. Our main goal is to show that $\Gamma \cap R_0 \subset R_1$.

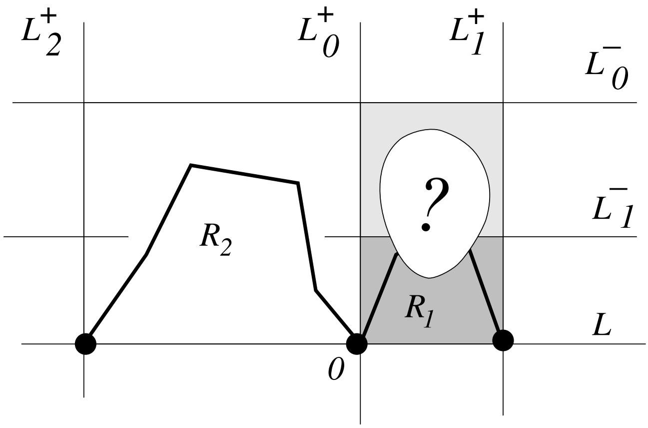

Figure 19.4: Dividing Γ^1 into two arcs.

Let $A' = p'/q'$ denote the superior predecessor of A. Let $\Omega = \Omega(A', A)$. We consider the case when $A' < A$.

Lemma 19.5 *The second coordinate of any point in* R_1 *lies in* $(0, \Omega q_1' - 1)$.

Proof: By convexity, it suffices to consider the vertices of R_1. The bottom vertices of R_1 have first coordinates 0 and q_+, whereas $\Omega q' = q_+ + q'$. This takes care of the bottom vertices. Let $u = (u_1, u_2)$ be the top left vertex of R_1. Since R_1 is a parallelogram, we can finish the proof by showing that $u_1 \in (0, q' - 1)$. Let $y = (p' + q')/2 \leq q' - 1$. Note that u lies on a line of slope in $(1, \infty)$ through the origin. Since the top edge of R_1 has negative slope and contains $(0, y)$, we have $u_2 < y$. Hence $u_1 < y$ as well. $\square$

Lemma 19.6 *Let A' denote the superior predecessor of A. Suppose that $A' \neq 1/1$. Then $\Gamma' \cap R_0 \subset R_1$.*

Proof: Let $\gamma = \Gamma' \cap R_0$. Since γ starts out in R_1 (at the origin), we just need to see that γ never crosses the top edge of R_1. The top edge of R_1 is contained in the line $\lambda = L_1^-$ of slope $-A$ though the point $X = (0, (p'+q')/2)$. By the Room Lemma, γ does not cross the (nearly identical) line $\lambda' = (L_0^-)'$ of slope $-A'$ through X.

If γ crosses the top edge of R_1, then there is a lattice point (m, n) between λ and λ' and within 1 unit of R_1. But then

$$\text{floor}(Am) \neq \text{floor}(A'm), \qquad m \in (-1, q' + q_+) = (-q', \Omega q'). \tag{19.2}$$

The second equation comes from our previous result. Our last equations contradict Lemma 17.5. □

Corollary 19.7 *Suppose that Γ and Γ' agree in R_1. Then the Decomposition Theorem holds for A.*

Proof: Let us trace $\Gamma \cap R_0$ from left to right, starting at $(0, 0)$. By hypothesis, this arc does not cross the top of R_1 until it leaves R_0. Once $\Gamma \cap R_0$ leaves R_0 from the right, it never reenters. This is a consequence of Lemma 19.2. □

By Corollary 19.7, it suffices to prove that Γ' and Γ agree in R_1.

Lemma 19.8 *$\Gamma' \cap R_1$ and $\Gamma \cap R_1$ have the same outermost edges.*

Proof: The leftmost edge of both arcs is the edge connecting $(0, 0)$ to $(1, 1)$. Looking at the proof of Lemma 19.2, we see that the rightmost edge e of $\Gamma \cap R_0$ connects $V_+ + (0, 1)$ to $V_+ + (1, 0)$. Here $V_+ = (q_+, -p_+)$. Applying Lemma 19.2 to Γ', we see that some edge e' of Γ' connects $V'_+ + (0, 1)$ to $V'_+ + (1, 0)$. But repeated applications of case 1 or case 2 of Lemma 17.2 tell us that $V_+ = V'_+ + kV'$ for some $k \in Z$. Since Γ' is invariant under translation by V', we see that e is also an edge of Γ'. □

Mismatch Principle: Lemma 19.8 has the following corollary. If Γ' and Γ fail to agree in R_1, then there are 2 adjacent vertices of $\Gamma' \cap R_1$ where the two arithmetic graphs $\widehat{\Gamma}$ and $\widehat{\Gamma}'$ do not agree. One can see this by tracing the 2 curves from left to right, starting at the origin. Once we get the first mismatch on Γ' the arc Γ has veered off, and the next vertex on Γ' is also a mismatch.

In our analysis below, we will treat the case when $A' < A$. The other case is similar. The bottom right vertex of R_1 lies on a line of slope in $(1, \infty)$ that contains the point V_+. The point V_+ has the same first coordinate as the very nearby point

$$\tilde{V}_+ = \frac{q_+}{q} V. \tag{19.3}$$

Indeed, the 2 points differ by exactly $1/q$. Let $\tilde{R}_1$ denote the slightly smaller parallelogram whose vertices are

$$(0,0), \qquad u, \qquad \tilde{V}_+, \qquad \tilde{w} = u + \tilde{V}_+. \tag{19.4}$$

If the Decomposition Theorem fails for A, then at least one of the adjacent vertices of mismatch will lie in $\tilde{R}_1$. (There are not 2 adjacent vertices between the nearly identical right edges of R_1 and $\tilde{R}_1$.)

As in the previous chapter, it suffices to make the extremal calculations

$$G(u) \geq -q' + 2, \qquad H(\tilde{w}) \leq \Omega q' - 2 = q' + q_+ - 2. \tag{19.5}$$

The Diophantine Lemma then finishes the proof.

We first need to locate u. There is some r such that $v_1 = rW$. Letting M be the map from Equation 2.10, relative to the parameter A, we have

$$M(v_1) = M(rW) = p' + q'.$$

Solving for r gives

$$v_1 = \left(\frac{p' + q'}{p + q}\right) W. \tag{19.6}$$

We compute

$$\begin{aligned} G(u) &= \frac{p' + q'}{p + q} G(W) \\ &= -\frac{p' + q'}{p + q} \times \frac{q^2}{p + q} \\ &= \frac{-(1 + A')q'}{(1 + A)^2} \\ &> \frac{-q'}{1 + A'}. \end{aligned} \tag{19.7}$$

$$\begin{aligned} H(\tilde{w}) &= H(u) + (q_+/q) H(V) \\ &= \frac{(1 + A')q'}{(1 + A)^2} + q_+ \\ &< \frac{q'}{1 + A'} + q_+. \end{aligned} \tag{19.8}$$

The last inequality in each case uses the fact that $0 < A' < A$. Notice the great similarity between these two calculations. One can ultimately trace this symmetry back to the affine symmetry of the arithmetic kite $\mathcal{K}(A)$ defined in Chapter 3.

The conditions in Equation 19.5 are simultaneously met, provided

$$\frac{-q'}{1 + A'} \geq -q' + 2, \qquad \left(\iff \quad \frac{1}{p'} + \frac{1}{q'} \leq \frac{1}{2}\right). \tag{19.9}$$

The equation on the right is equivalent to the one on the left. We easily see that it holds as long as $p' \geq 3$ and $q' \geq 7$.

In the next two sections we will make a more detailed study of the few exceptions to Lemma 19.3. The reader mainly interested in the Erratic Orbits Theorem can stop reading here.

19.5 THE EXCEPTIONAL CASES

19.5.1 Case 1

We use the notation from the previous section. We assume first that $A' \neq 1/1$ is one of the rationals not covered by Theorem 19.3. Our argument uses the linear functionals G' and H' associated to A' in place of the linear functionals G and H used above. Before we begin our argument, we warn the reader that G' is not the derivative of G. We will denote the partial derivatives of G' by $\partial_x G'$ and $\partial_y G'$.

Lemma 19.9 $G'(v) \geq -q' + 2$ *for all* $v \in R_1$.

Proof: We have to worry only about points near the top left corner of R_1. Such points lie on the first period of Γ' to the right of the origin. Call this period β'. When $A' \in \{3/5, 3/7, 5/7\}$, we check this result explicitly for every point on β'. When $A' = 1/q'$, we note that $\partial_x G' > 0$ and $\partial_y G' < 0$. We also note that all points in R_1 have positive first and second coordinates of at most $(q'-1)/2$. Thus the point that minimizes G' is $v = (1, (q'-1)/2)$. We compute

$$G'(v) + q - 2 = \frac{q'-3}{q'+1} \geq 0.$$

The extreme case occurs when $q' = 3$. □

H' is tougher to analyze because the points of interest to us are near the top right corner of R_1, and this corner can vary drastically with the choice of A. We will use rotational symmetry to bring the points of interest back into view, so to speak. Let ι be the isometric involution that swaps $(0,0)$ and V_+. Repeated applications of Lemma 17.2 show that $V_+ = V'_+ + kV'$ for some $d \in \mathbf{Z}$. Hence ι is a symmetry of $\widehat{\Gamma}'$. See the remark following Equation 12.14.

The infinite arc $\iota(\Gamma')$ is the open component of $\widehat{\Gamma}'$ that lies just beneath the baseline. One period of $\iota(\Gamma')$ connects $(0,-1)$ to $(q', -p'-1)$. Let us denote this period by β'. Compare the proof of Lemma 19.2. The points of R_1 near the top right corner correspond to points on β'. To evaluate H' on the points near the top right corner of R_1, we evaluate H' on points of β' and then relate the results.

Lemma 19.10 *For any* $v \in \mathbf{R}^2$, *we have*

$$|H'(v) + H'(\iota(v)) - q_+| < 2/q'.$$

Proof: Since H' is a linear functional, it suffices to prove the result for $v = (0,0)$. In this case, we must demonstrate that $|H'(V_+) - q_+| < 2/(q')$. We have already remarked that $V_+ = V'_+ + kV'$. Hence $q_+ = q'_+ + kq'$. From Lemma 18.1, we have $H'(kV') = kq'$. Hence the equality is equivalent to

$$|H'(V'_+) - q'_+| < 2/q'. \tag{19.10}$$

The point V'_+ lies on the same vertical line as the point $u' = (q'_+/q')V'$ and exactly $1/q'$ units away. Equation 19.10 now follows from the next 3 facts.

$$H'(u') = q'_+, \qquad |\partial_y H'| < 2, \qquad \|u' - V'_+\| = 1/q'. \tag{19.11}$$

The first fact comes from Lemma 18.1. The second fact is an easy calculus exercise. The third fact, already mentioned, is an easy exercise in algebra that uses $|q'p'_+ - p'q'_+| = 1$. □

The bound

$$H'(v) \leq \Omega q' - 2 = q' + q_+ - 2$$

fails only for points very near the top right vertex of R_1. Any such point has the form $\iota(v)$ for some $v \in \beta'$. Thus, to establish the above bound, it suffices to prove that

$$H'(v) \geq -q' + 2 + 2/q'. \tag{19.12}$$

This inequality can fail for very small choices of q'. However, from the Mismatch Principle, the inequality must fail for at least 2 vertices on β', and this does not happen.

We check all cases with $q' \leq 7$ by hand. This leaves only $A' = 1/q'$ for $q' \geq 9$. Reasoning as we did in Lemma 19.9, we see that the extreme point is $v = (0, (1-q)/2)$. We compute

$$H'(v) - \left(-q' + 2 + \frac{1}{q'}\right) = \frac{2(q'^2 - 2q' - 1)}{(1+q')^2} - \frac{2}{q'} > 0. \tag{19.13}$$

The last equation is an easy exercise in calculus. This completes our proof of the Decomposition Theorem for all parameters A such that $A' \neq 1/1$.

19.5.2 Case 2

Now we deal with the case when $A' = 1/1$ is the superior predecessor of A. We have the following structure.

$$\frac{1}{1} \leftarrow A_1 = \frac{2k-1}{2k+1} \leftarrow \cdots \leftarrow A_m = 1. \tag{19.14}$$

Here $k \geq 1$. For instance, when $A = 17/21$, we have $1/1 \leftarrow 9/11 \leftarrow 17/21$. Figure 19.5 shows $\Gamma(17/21)$. In this case $\Gamma \cap R_1$ is the line segment connecting $(0,0)$ to $(-5,5) = (-k,k)$. We will establish this structure in general.

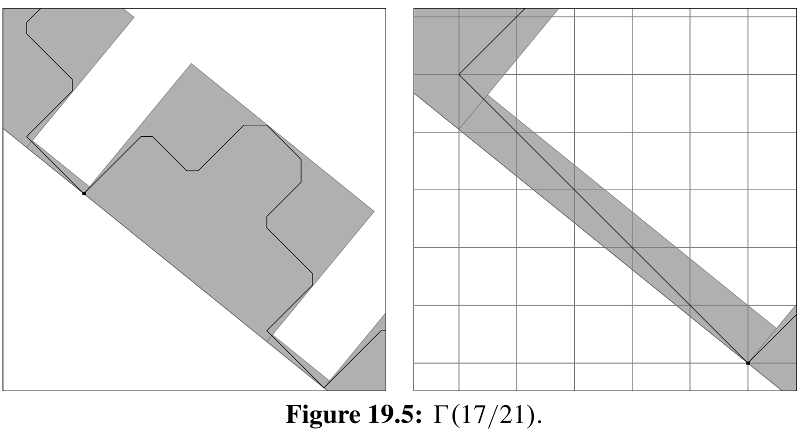

Figure 19.5: $\Gamma(17/21)$.

R_1 is the very short and squat parallelogram on the far left side of Figure 19.5. This time R_1 lies to the left of the origin. The right side of Figure 19.5 shows a closeup of this parallogram superimposed on the integer grid. The left side of R_1 lies in L_1^+. Repeated applications of Lemma 17.2 show that $(-k, k-1) \in L_1^+$. The right side of R_0 lies in L_0^+, the parallel line through the origin. The top of R_1 contains $(0, 1)$ and is parallel to the baseline.

Let $\gamma = \Gamma \cap R_0$. The rightmost vertex of γ is $(0, 0)$, and the rightmost edge of γ connects $(0, 0)$ to $(-1, 1)$. Compare the proof of the Room Lemma.

Lemma 19.11 *The leftmost edge of* γ *connects* $(-k, k)$ *to* $(-k+1, k-1)$.

Proof: By Lemma 19.2, there is a unique edge e of Γ that crosses L_1^+. Looking at the proof of Lemma 19.2, we see $e = \iota(e')$, where e' connects $(0, -1)$ to $(-1, 0)$ and ι is the order 2 rotation about the point

$$\left(\frac{-q_-}{2}, \frac{p_-}{2}\right) = \left(\frac{-k}{2}, \frac{k-1}{2}\right). \tag{19.15}$$

From this, we conclude that e connects $(-k, k)$ to $(-k+1, k-1)$. The leftmost edge of γ crosses L_1^+. This edge must be e. □

Lemma 19.12 *The line segment* γ' *connecting* $(0, 0)$ *to* $(-k, k)$ *lies beneath* L_0^-. *Hence* $\gamma' \cap R_0 \subset R_1$.

Proof: Letting $F(m, n) = Am + n$, we have $F(0, 1) = 1$. Hence $F(x) = 1$ for all $x \in L_0^-$. On the other hand, we compute that $F(0, 0) = 0$ and $F(-k, k) = 2k/(2k+1) < 1$. By convexity, $F(y) < 1$ for all $y \in \gamma'$. □

To finish our proof, we just have to show that $\gamma' = \gamma$. The first and last edges of γ and γ' agree, and these edges are $\pm(1, -1)$, with the sign depending on which way we orient the curves. Let $p_j = (-j, j)$ for $j = 2, ..., k-1$. By Lemma 17.1, we have

$$(A_1)_- = \left(\frac{k-1}{k}\right) < A < \left(\frac{k}{k+1}\right) = (A_1)_+, \qquad \frac{1}{k+1} < 1 - A < \frac{1}{k}. \tag{19.16}$$

The first equation implies the second. We compute

$$M_+(p_j) = (x_j, y_j, z_j) = j(1-A, 1-A, 1-A) + (0, 1, 0) \mod \Lambda. \tag{19.17}$$

Equation 19.16 combines with the fact that $j \in \{1, ..., k-1\}$ to give

$$x_j = z_j \in [1-A, A), \qquad x_j + y_j - z_j = y_j \in (1, 1+A) \subset (A, 1+A). \tag{19.18}$$

We check that these inequalities always specify the edge $(-1, 1)$. Hence γ' and γ are both line segments. Hence $\gamma = \gamma'$. This takes care of the case when $A' = 1/1$.

Our proof of the Decomposition Theorem is complete.

Chapter Twenty

Existence of Strong Sequences

In this chapter, we prove Theorem 4.2. For the sake of efficiency, our proof will be essentially algebraic. However, a clear geometric picture underlies our constructions. We discussed this geometric picture in §19.2. The reader might want to reread that section before going through the proof here. Also, the reader might want to review the proof we gave of Lemma 4.3 in §18.3. Our proof here is similar to the one given there.

20.1 STEP 1

Let A be any irrational parameter. Let $\{p_n/q_n\}$ denote the superior sequence associated to A. Let $\mathcal{S}$ be a monotone subsequence of the superior sequence. We will treat the case when $\mathcal{S}$ is monotone increasing. The other case is entirely similar.

By Corollary 19.3, we can cut off finitely many terms of $\mathcal{S}$, leaving a sequence for which the Decomposition Theorem always holds. This is what we will do.

For any odd rational p/q, let $R^*(p/q)$ denote the rectangle with vertices

$$-\frac{V}{2}, \qquad -\frac{V+W}{2}, \qquad \frac{V+W}{2}, \qquad \frac{V}{2}. \tag{20.1}$$

Here V and W are as in Equation 3.2. The parallelogram R^* is just as wide as R but half as tall. Also, the bottom edge of R^* is centered on the origin.

Lemma 20.1 *If $A_1 \leftarrow A_2$ and p_1 is sufficiently large, then Γ_1 and Γ_2 agree in A_1. Moreover, Γ_1 and Γ_2 agree in the $q_1/8$ neighborhood of the bottom edge of R_1^*.*

Proof: The proof works the same way regardless of the sign of $A_1 - A_2$. The main point is that $\Omega > 1$. Note that (A_1, A_2) is admissible. We use the linear functionals G_1 and H_1 associated to A_1. Let

$$u = \frac{-V+W}{2}, \qquad w = \frac{V+W}{2}$$

denote the top left and right vertices of R_1^*, respectively. We compute

$$-G_1(u) = H_1(w) = \frac{q_1(p_1+2q_1)}{2p_1+2q_1} < q_1 - 2 < \Omega q_1 - 2. \tag{20.2}$$

The same argument as in Lemma 4.3 now finishes the proof. □

Remark: We have not yet used the Decomposition Theorem.

20.2 STEP 2

Now we are really going to use the Decomposition Theorem, as discussed in §19.2.

Lemma 20.2 *Suppose that $A_1 < A_2$ and A_1 is the superior predecessor of A_2. If A_1 has sufficiently large complexity, then $\Gamma_1^{1+\epsilon} \subset \Gamma_2^1$.*

Proof: If $\Omega > 2$, we have the same proof as in Lemma 4.3. Equation 17.15 does not allow $\Omega = 2$. We just need to consider the case $\Omega < 2$. By Equation 17.15, we must have

$$\text{floor}(a/2 - \lambda_1) = 0. \tag{20.3}$$

Since $a > 1$, we must have

$$\lambda_1 > 1/2. \tag{20.4}$$

Since $\lambda_1 = (q_1)_+/q_1$ and $q = q_+ + q_-$, we must have

$$(q_1)_- < (q_1)_+. \tag{20.5}$$

This seemingly minor fact is crucial to our argument.

Let $R(A_1)$ denote the parallelogram from the Room Lemma. In contrast, let $R_1(A_1)$ and $R_2(A_2)$ denote the smaller parallelograms from the Decomposition Theorem. Since $(q_1)_- < (q_-)_+$, we see that $R_2(A_1)$ lies to the left of $R_1(A_1)$. By the Decomposition Theorem,

$$\Gamma_1 \cap R(A_1) \subset R_2(A_1) \cup (R_1(A_1) + V_1) \tag{20.6}$$

Figure 20.1 is a schematic picture.

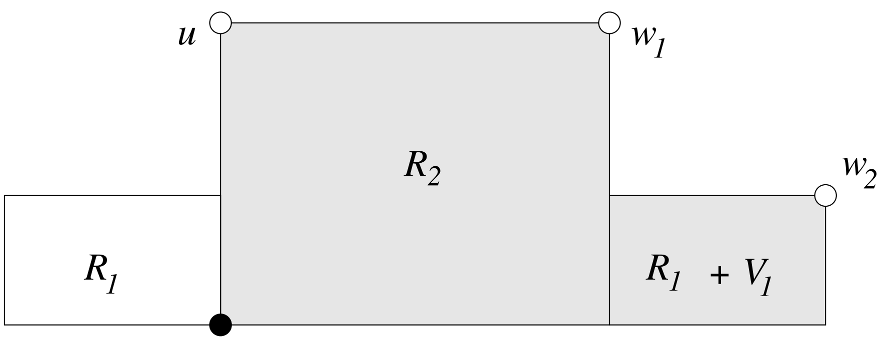

Figure 20.1: $R_2(A_1)$ and $R_1(A_1) + V_1$.

The vertices shown in Figure 20.1 are

$$u = W_1, \qquad w_1 \approx W_1 + \lambda_1 V_1, \qquad w_2 \approx V_1 + \mu W_1. \tag{20.7}$$

Here $\mu = q_0/q_1 < 1/2$, where $A_0 = p_0/q_0$ is the superior predecessor of A_1. Also, $\lambda_1 = (q_1)_+/(q_1)$, as in Equation 17.15.

The approximation sign means that the distance between the two points is at most 1 unit. For instance, w_1 is the intersection of the line parallel to W_1 and containing

V_+ with the line parallel to V_1 and containing W_1. The point V_+ is $O(q_1^{-2})$ of the point $\lambda_1 V_1$. Hence w_1 is within $O(q_1^{-2})$ of $W_1 + \lambda_1 V_1$. The argument for w_2 is similar.

As in the proof of Lemma 4.3, we have $G_1(u) > -q_1 + 2$ once p_1 is large. The computations for $H_1(w_1)$ and $H_1(w_2)$ are the interesting ones. Case 1 of Lemma 17.2 gives $(q_2)_+ \geq (q_1)_+$. Hence, for p_1 sufficiently large, we have the following inequalities.

$$\begin{aligned} 2 + H_1(w_1) &\leq \\ \Big(2 + \|\nabla H\|\Big) + H_1(W_1) + \lambda_1 H_1(V_1) &\leq \\ 5 + \frac{q_1^2}{p_1 + q_1} + (q_1)_+ &< \\ q_1 + (q_1)_+ &\leq \\ q_1 + (q_2)_+ &= \\ \Omega q_1. \end{aligned} \tag{20.8}$$

Here we use the bound $\|\nabla H\| \leq 3$. We have already remarked that $(q_2)_+ \geq (q_1)_+$. We also know that $(q_1)_+ > q_1/2$. Hence

$$\Omega q_1 = q_1 + (q_2)_+ > (3/2)q_1. \tag{20.9}$$

For p_1 large, we have

$$\begin{aligned} 2 + H_1(w_2) &\leq \\ \Big(2 + \|\nabla H\|\Big) + H_1(V_1) + \mu H_1(W_1) &< \\ 5 + H_1(V_1) + (1/2)H_1(W_1) &= \\ 5 + q_1 + \frac{q_1^2}{2(p_1 + q_1)} &< \\ (3/2)q_1 &< \\ \Omega q_1. \end{aligned} \tag{20.10}$$

These arguments show that $v \in \Delta_1(I)$ for all $v \in \Gamma_1^1$. The rest of the proof is just like the proof of Lemma 4.3. □

20.3 STEP 3

Suppose $A_1' < A_2'$ are two consecutive terms in $\mathcal{S}$ when we have a finite chain

$$A_1' = A_1 \leftarrow A_2 \leftarrow \cdots \leftarrow A_n = A_2', \qquad A_1 < A_n, \qquad q_2 > 2q_1. \tag{20.11}$$

The following result finishes the proof of Theorem 4.2.

Lemma 20.3 $\Gamma_1^{n+\epsilon} \subset \Gamma_n^1$.

Proof: We will change our notation slightly from the previous result. We let $R_1 = R(A_1)$ denote the parallelogram from the Room Lemma. Likewise, we let

$R_k^* = R^*(A_k)$ denote the parallelogram from Theorem 20.1. For any parallelogram R_k, let XR_k denote the union of R with the points within $q_k/8$ units from the bottom edge of R_k. Likewise, define XR_k^*.

Since $A_1 < A_n$, we have $A_1 < A_2$ by Lemma 17.1. We now have

$$\Gamma_1^{1+\epsilon} \subset \Gamma_1 \cap XR_1 \subset \Gamma_2. \tag{20.12}$$

The first containment comes from the Room Lemma and the definition of $\Gamma_1^{1+\epsilon}$. The second containment is Theorem 20.2. Theorem 20.1 gives us

$$\Gamma_k \cap XR_k^* \subset \Gamma_{k+1}, \qquad k = 2, ..., n-1. \tag{20.13}$$

Let us compare R_1 and R_k^* for $k \geq 2$.

1. The sides of R_1 have length $O(q_1)$.
2. The slope of each side of R_1 is within $O(q_1^{-2})$ of the slope of the corresponding side of R_k^*. This comes from Lemma 17.4.
3. Each side of R_1 is less than half as long as the corresponding side of R_k^*. This follows from the first two facts and from the fact that $2q_1 < q_2 \leq q_k$. Indeed, the quantity $q_2 - 2q_1$ tends to ∞ with the complexity of A_1.

These properties give us

$$XR_1 \subset XR_k^*, \qquad k = 2, ..., n-1. \tag{20.14}$$

Figure 20.2 is a schematic picture.

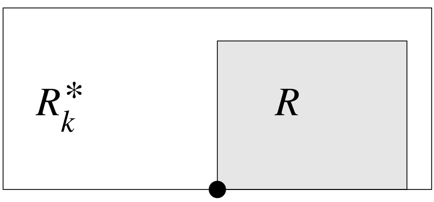

Figure 20.2: R_1 and R_k^* for any $k \geq 2$.

We already know that $\Gamma_1 \cap XR_1 \subset \Gamma_2$. Suppose $\Gamma_1 \cap XR_1 \subset \Gamma_k$ for some $k \geq 2$. Then

$$\Gamma_1 \cap XR_1 \subset \Gamma_k \cap XR_1 \subset \Gamma_k \cap XR_k^* \subset \Gamma_{k+1}. \tag{20.15}$$

Hence, by induction, $\Gamma_1^{1+\epsilon} \subset \Gamma_n$. The right endpoint of Γ_n^1 lies far to the right of any point on $\Gamma_1^{1+\epsilon}$. Hence $\Gamma_1^{1+\epsilon} \subset \Gamma_n^1$. □

This completes the proof of Theorem 4.2. Our proof of the Erratic Orbits Theorem is finished as well.

Part 5. The Comet Theorem

In this part of the book, we prove the Comet Theorem and its corollaries. As we did in Part 1, we defer the proofs of many of the auxilliary results. In Part 6, we take care of all the remaining details.

- In Chapter 21, we prove some further results about the inferior and superior sequences. We list the basic results in the first section and then spend the rest of the chapter proving these results.
- In Chapter 22, we prove Theorem 1.8. We also build a rough model for the way the orbit $O_2(1/q_n, -1)$ returns to the interval $I = [0, 2] \times \{-1\}$. Our work here depends on two technical results, the Copy Theorem and the Pivot Theorem, which we establish in Part 6.
- In Chapter 24, we prove the Comet Theorem, modulo some technical details that we handle in Part 6.
- In Chapter 24, we deduce a number of dynamical consequences of the Comet Theorem, including minimality of the set of unbounded orbits. We also define the cusped solenoids and explain how the time-one map of their geodesic flow models the outer billiards dynamics.
- In Chapter 25, we analyze the structure of the Cantor set C_A from the Comet Theorem. This chapter has a number of geometric results, such as a formula for $\dim(C_A)$ when A is a quadratic irrational.

This page intentionally left blank

Chapter Twenty-One

Structure of the Inferior and Superior Sequences

21.1 THE RESULTS

Let $\{p_n/q_n\}$ be the inferior sequence associated to an irrational parameter $A \in (0, 1)$ and let $\{d_n\}$ be the sequence obtained from Equation 4.5. We call $\{d_n\}$ the *inferior renormalization sequence*. We call the subsequence of $\{d_n\}$ corresponding to the superior terms the *superior renormalization sequence* or just the *renormalization sequence*. Referring to the inferior sequences, we have $d_n = 0$ if and only if n is not a superior term. In this case, we call n an *inferior term*. So, the renormalization sequence is created from the inferior renormalization sequence simply by deleting all the 0s.

For any odd rational $p/q \in (0, 1)$, define

$$p^* = \min(p_-, p_+); \qquad q^* = \min(q_-, q_+). \tag{21.1}$$

Here p^*/q^* is one of the rationals $p_\pm/q_\pm$. It is convenient to define

$$\frac{p_0^*}{q_0^*} = \frac{1}{0}. \tag{21.2}$$

Given the superior sequence $\{p_n/q_n\}$, we define

$$\lambda_n = |Aq_n - p_n|; \qquad \lambda_n^* = |Aq_n^* - p_n^*|; \tag{21.3}$$

Note that

$$\lambda_0^* = 1. \tag{21.4}$$

For the purpose of making a clean statement, we define $\lambda_{-1} = +\infty$. All our results are meant to apply to the superior sequence for indices $n \geq 0$.

$$d_n\lambda_n < 2q_n^{-1}, \tag{21.5}$$

$$q_{2n} > (5/4)^n D_{2n}, \tag{21.6}$$

$$\sum_{k=n}^{\infty} d_k\lambda_k = \lambda_n^* < \lambda_{n-1}. \tag{21.7}$$

Note that Equation 21.5 is an immediate consequence of Lemma 17.4. The rest of the chapter is devoted to proving Equations 21.6 and 21.7.

21.2 THE GROWTH OF DENOMINATORS

Here we establish some terminology.

- Referring to Equation 17.5, we call $\{\delta_n\}$ the *enhanced inferior renormalization sequence* (*EIRS*).
- We call the subsequence corresponding to the superior indices the *enhanced renormalization sequence*.

The reason for the terminology is that we can determine the inferior renormalization sequence from the EIRS, but not vice versa.

Say that a parameter A is *superior to* a parameter A' if the EIRS for A' is obtained by inserting some 1s into the EIRS for A. For instance, $\sqrt{5}-2$ has EIRS

$$3, 1, 2, 1, 2, ...,$$

and $\sqrt{2}-1$ has EIRS sequence

$$3, 2, 2, 2, ...$$

Hence $\sqrt{2}-1$ is superior to $\sqrt{5}-2$.

Lemma 21.1 *Suppose that A is superior to A'. Then $q_n \leq q'_n$ for all n.*

Proof: The EIRS determines the inferior sequence. We have $p_0 = \delta_0 - 2$ and $q_0 = \delta_0$. Then, by Lemma 17.2, each $(q_{n+1})_\pm$ is a nonnegative integer linear combination of $(q_n)_\pm$, and the coefficients are determined by $\{\delta_n\}$. Call this the *positivity property*.

Consider the operation of inserting a 1 into the mth position in the EIRS for A and recomputing $\{A_n\}$. Call this new sequence the A^*-sequence. We have

$$(q^*_{m+1})_\pm \geq (q_m)_\pm.$$

By induction, and the positivity property, we have

$$(q^*_{n+1})_\pm \geq (q_n)_\pm.$$

Now let us delete the $(m+1)$th term from the A^*-sequence. Call the new sequence the A'-sequence. We have $q'_n \geq q_n$ for all n. Our result now follows from induction. $\square$

Call A *superior* if the corresponding inferior sequence has no inferior terms. That is, the EIRS has no 1s in it. For instance, $\sqrt{2}-1$ is a superior parameter. If we want to find a lower bound on the growth of denominators, it suffices to consider only the superior parameters. Equation 21.6 follows from induction and our next lemma.

Lemma 21.2 *Suppose that A_1, A_2, A_3 are 3 consecutive terms in the superior sequence. Let d_1, d_2, d_3 be the corresponding terms of the renormalization sequence. Then $q_3 > (5/4)(d_1+1)(d_2+1)q_1$.*

Proof: It suffices to assume that A is a superior parameter, so that A_1, A_2, A_3 are (also) 3 consecutive terms in the inferior sequence.

First of all, the estimates

$$q_{n+1} > 2d_n q_n, \qquad q_{n+1} > \delta_n q_n \tag{21.8}$$

follow directly from the definitions. Our notation is as in Lemma 17.2. Now we have 3 cases.

Case 1: Suppose that $\min(d_1, d_2) \geq 2$. Then

$$q_3 > 4d_1 d_2 q_1 > (4/3)(d_1+1)(d_2+1)q_1. \tag{21.9}$$

Case 2: Now suppose that $d_1 = d_2 = 1$ and $\min(\delta_1, \delta_2) \geq 3$. Then

$$q_3 > 6q_1 = (3/2)(d_1+1)(d_2+1)q_1. \tag{21.10}$$

Case 3: Suppose finally that $d_1 = d_2 = 1$ and $\delta_1 = \delta_2 = 2$. We will deal with the case when $A_1 < A_2$. The other case is similar. In this case, we must have

$$A_0 > A_1 < A_2 > A_3, \tag{21.11}$$

by Lemma 17.2.

By case 2 of Lemma 17.2,

$$(q_2)_- = q_1 + (q_1)_+, \qquad (q_2)_+ = (q_1)_+. \tag{21.12}$$

By case 4 of Lemma 17.2,

$$(q_3)_+ = q_2 + (q_2)_-, \qquad (q_3)_- = (q_2)_-. \tag{21.13}$$

Hence

$$q_3 = (q_3)_+ + (q_3)_- = q_2 + 2(q_2)_- =^* q_2 + 2q_1 + 2(q_2)_+. \tag{21.14}$$

The starred equality comes from Lemma 17.1 since $A_1 < A_2$.

Since $A_0 > A_1$, Lemma 17.2 says that

$$2(q_1)_+ > (q_1)_+ + (q_1)_- = q_1. \tag{21.15}$$

Combining Equations 21.12, 21.14, and 21.15, we have

$$q_3 = q_2 + 2q_1 + 2(q_1)_+ > q_2 + 3q_1 > 5q_1. \tag{21.16}$$

Hence

$$q_3 > (5/4)(d_1+1)(d_2+1)q_1. \tag{21.17}$$

This completes our proof. □

21.3 THE IDENTITIES

We first verify the identity in Equation 21.7. In this identity, we sum over the superior indices. However, notice that we get the same answer when we sum over all indices. The point is that $d_n = 0$ when n is an inferior index. So, for our derivation, we work with the inferior sequence. Let $\{p_n/q_n\}$ be the inferior sequence associated to A. Define

$$\Delta(n, N) = |p_N q_n - q_N p_n|, \qquad \Delta^*(n, N) = |p_N q_n^* - q_N p_n^*|, \qquad N \geq n. \tag{21.18}$$

Lemma 21.3 $\Delta^*(n, N) - \Delta^*(n+1, N) = d_n \Delta(n, N)$.

Proof: The quantities relevant to the case $n = 0$ are

$$A_0 = \frac{1}{1}, \qquad A_0^* = \frac{1}{0}, \qquad A_1^* = \frac{d_0 - 1}{d_0} < A_1 = \frac{2d_0 - 1}{2d_0 + 1}.$$

In this case, a simple calculation checks the formula directly.

Now suppose $n \geq 1$. We suppose that $A_{n-1} < A_n$. The other case requires a similar treatment. Let r stand for either p or q. There are two cases, depending on whether the index n has type 1 or type 4.

Case 1: When n has type 1, Lemma 17.2 gives

$$r_n^* = (r_n)_+, \qquad r_{n+1}^* = (r_{n+1})_+, \qquad r_n^* = d_n r_n - r_{n+1}^*. \tag{21.19}$$

We have $\Delta^*(n, N) = |a_1 - a_2|$, where

$$a_1 = d_n p_N q_n - d_n q_N p_n = d_n \Delta(n, N),$$

$$a_2 = p_N q_{n+1}^* - q_N p_{n+1}^* = -\Delta^*(n+1, N). \tag{21.20}$$

The sign for a_1 is correct because $A_N > A_n$. The sign for a_2 is correct because, by Lemma 17.1, we have $A_N < (A_{n+1})_+ = A_{n+1}^*$. The identity in this lemma follows immediately.

Case 2: When n has type 4, Lemma 17.2 gives

$$r_n^* = (r_n)_+, \qquad r_{n+1}^* = (r_{n+1})_-, \qquad r_n^* = d_n q_n - r_{n+1}^*.$$

Hence $\Delta^*(n, N) = |a_1 + a_2'|$, where $a_2' = -a_2$. The sign changes for a_2' because $A_N > (A_{n+1})_- = A_{n+1}^*$. In this case, we get the same identity. □

Dividing the equation in Lemma 21.3 by q_N, we have

$$|A_N p_n^* - q_n^*| - |A_N p_{n+1}^* - q_{n+1}^*| = d_n |A_N p_n - q_n|. \tag{21.21}$$

Taking the limit as $N \to \infty$, we get

$$\lambda_n^* - \lambda_{n+1}^* = d_n \lambda_n. \tag{21.22}$$

Summing this equation from $n+1$ to ∞ gives the equality in Equation 21.7.

Now we will verify the inequality in Equation 21.7.

Lemma 21.4 $\lambda_{n+1}^* < \lambda_n$.

Proof: There are two cases to consider, depending on whether $A_n < A$ or $A_n > A$. We will consider the case when $A_n < A$. The other case requires a similar treatment. By Lemma 17.1, we have $A_n < A_{n+1}$. Therefore, by Lemma 17.2 (applied to $m = n+1$), we have $(q_{n+1})_+ < (q_{n+1})_-$. But this means that $A_{n+1}^* = (A_{n+1})_+$. By Lemma 17.1, we have

$$A_n < A < A_{n+1}^*. \tag{21.23}$$

Given the above ordering, we have

$$\lambda_n = |Aq_n - p_n| = Aq_n - p_n$$

and

$$\lambda_{n+1}^* = |Aq_{n+1}^* - p_{n+1}^*| = p_{n+1}^* - Aq_{n+1}^*.$$

Hence

$$\lambda_n - \lambda_{n+1}^* = A(q_n + q_{n+1}^*) - (p_n + p_{n+1}^*). \tag{21.24}$$

But

$$q_n + q_{n+1}^* = q_n + (q_{n+1})_+ = \Big((q_{n+1})_- - (q_{n+1})_+\Big) - (q_{n+1})_+ = (q_{n+1})_-.$$

Likewise,

$$p_n + p_{n+1}^* = (p_{n+1})_-.$$

Combining these identities with Equation 21.24, we have

$$\lambda_n - \lambda_{n+1}^* = A(q_{n+1})_- - (p_{n+1})_- = (q_{n+1})_-(A - (A_{n+1})_-) > 0.$$

This completes the proof. $\square$

This page intentionally left blank

Chapter Twenty-Two

The Fundamental Orbit

22.1 MAIN RESULTS

We will assume that $p/q = p_n/q_n$, the nth term in a superior sequence. We call $O_2(1/q_n, -1)$ the *fundamental orbit*. Let C_n denote the set from Theorem 1.8. Let

$$C_n' = O_2(1/q_n, -1) \cap I, \qquad I = [0,2] \times \{-1\}. \tag{22.1}$$

Theorem 1.8 says that $C_n = C_n'$. In this chapter, we will prove Theorem 1.8 and establish some geometric results about how the orbits return to C_n.

After we prove Theorem 1.8, we will establish a coarse model for how the points of $O_2(1/q_n)$ return to C_n. Statement 2 of the Comet Theorem is s kind of geometric limit of the Discrete Theorem, and statement 3 of the Comet Theorem is the "geometric limit" of the coarse model we build here.

Let Π_n denote the truncation of the space defined in Equation 1.7. Let

$$\chi \colon \Pi_n \to C_n \tag{22.2}$$

denote the mapping that is implicit in the statement of Theorem 1.8. There is an ordering on Π_n such that $\chi(\kappa)$ returns to $\chi(\kappa_+)$, where κ_+ is the successor of κ in the ordering. We will describe this ordering.

Here we will define two natural orderings on the sequence space Π_n associated to p_n/q_n. Let $\{d_n\}$ be the renormalization sequence.

Reverse Lexicographic Ordering: Given two finite sequences $\{a_i\}$ and $\{b_i\}$ of the same length, let k be the largest index, where $a_k \neq b_k$. We define $\{a_i\} \prec' \{b_i\}$ if $a_k < b_k$, and $\{b_i\} \prec' \{a_i\}$ if $a_k > b_k$. This ordering is known as the *reverse lexicographic* ordering.

Twist Automorphism: Given a sequence $\kappa = \{k_i\} \in \Pi_n$, we define $\tilde{k}_i = k_i$ if $A_i < A_n$, and $\tilde{k}_i = d_i - k_i$ if $A_i > A_n$. We define $\tilde{\kappa} = \{\tilde{k}_i\}$. The map $\kappa \to \tilde{\kappa}$ is an involution on Π_n. We call this involution the *twist involution*.

Twirl Ordering: Any ordering on Π_n gives an ordering on C_n via the formula in Theorem 1.8. Now we describe the ordering that comes from the first return map. Given two sequences $\kappa_1, \kappa_2 \in \Pi_n$, we define $\kappa_1 \prec \kappa_2$ if and only if $\tilde{\kappa}_1 \prec' \tilde{\kappa}_2$. We call the ordering determined by $\prec$ the *twirl ordering*. We think of the word twirl as a kind of acronym for *twisted reverse lexicographic*. We will give an example below.

Lemma 22.1 *When C_n is equipped with the twirl order, each element of C_n except the last returns to its immediate successor, and the last element of C_n returns to the first.*

Our last goal in this chapter is to understand $O_2(1/q_n, -1)$ far away from I. Let $h_1(\kappa)$ denote the maximum distance the forward Ψ-orbit of $\chi(\kappa)$ gets from the kite vertex (0, 1) before returning as $\chi(\kappa_+)$. Let $h_2(\kappa)$ denote the number of iterates it takes before the forward Ψ-orbit of $\chi(\kappa)$ returns as $\chi(\kappa_+)$.

Let $\mathrm{ind}(\kappa)$ be the largest index k such that the sequences corresponding to κ and κ_+ differ in the kth position. Here $\mathrm{ind}(\kappa) \in \{0, ..., n-1\}$. Finally, we define $\mathrm{ind}(\kappa) = n$ if κ is the last element of Π_n.

Lemma 22.2 *Let* $m = \mathrm{ind}(\kappa)$. *Then*

$$q_m/2 - 4 < h_1(\kappa) < 2q_m + 4, \qquad\qquad h_2(\kappa) < 5q_m^2.$$

Example: The table below encodes the example from the introduction.

$$\frac{p_0}{q_0} = \frac{1}{1} > \frac{1}{3} < \frac{5}{13} > \frac{19}{49} = \frac{p_3}{q_3}.$$

The first 3 columns indicate the sequences. The 4th column indicates the first coordinate of $49\chi(\kappa)$. The first point of C_3 is $(65/49, -1)$. The 5th column shows $(m) = \mathrm{ind}(\kappa)$. The last column shows q_m.

1	0	1	$\longrightarrow$	65	(0)	1
0	0	1	$\longrightarrow$	5	(1)	3
1	1	1	$\longrightarrow$	81	(0)	1
0	1	1	$\longrightarrow$	21	(1)	3
1	2	1	$\longrightarrow$	97	(0)	1
0	2	1	$\longrightarrow$	37	(2)	13
1	0	0	$\longrightarrow$	61	(0)	1
0	0	0	$\longrightarrow$	1	(1)	3
1	1	0	$\longrightarrow$	77	(0)	1
0	1	0	$\longrightarrow$	17	(1)	3
1	2	0	$\longrightarrow$	93	(0)	1
0	2	0	$\longrightarrow$	33	(3)	49

For instance, the the Ψ orbit of $37/49$ wanders between

$$13/2 - 4 = 5/2$$

and

$$2 * 13 + 4 = 30$$

units away before returning to $61/49$ in less than $5 \times (13^2)$ steps. The results in the table are not very inspiring. A larger table would show more dramatic results.

22.2 THE COPY AND PIVOT THEOREMS

Here we describe the technical results that we will establish in Part 6.

Relative to the parameter A, we associate a sequence of *pairs of points* in $\mathbf{Z}^2$. We call these points the *pivot points*. We make the construction relative to the inferior sequence.

Define $E_0^{\pm} = (0,0)$ and $V_n = (q_n, -p_n)$. Define

$$A_n < A_{n+1} \implies E_{n+1}^- = E_n^-, \qquad E_{n+1}^+ = E_n^+ + d_n V_n. \tag{22.3}$$

$$A_n > A_{n+1} \implies E_{n+1}^- = E_n^- - d_n V_n, \qquad E_{n+1}^+ = E_n^+. \tag{22.4}$$

We have set $A_n = p_n/q_n$. Here is an example.

$$\frac{1}{1} \overset{>}{\leftarrow} \frac{3}{5} \overset{>}{\leftarrow} \frac{17}{29} \overset{<}{\leftarrow} \frac{37}{63} \overset{<}{\leftarrow} \frac{57}{97} \overset{>}{\leftarrow} \frac{379}{645}.$$

The inferior renormalization sequence is $2, 2, 1, 0, 3$. We compute

- $E_1^+ = E_0^+ = (0,0)$.
- $E_2^+ = E_1^+ = (0,0)$.
- $E_3^+ = E_2^+ + 1(29, -17)$.
- $E_4^+ = E_3^+ + 0(97, -57) = (29, -17)$.
- $E^+(379/645) = E_5^+ = E_4^+$.
- $E_1^- = E_0^- - 2(1, -1) = (-2, 2)$.
- $E_2^- = E_1^- - 2(5, -3) = (-12, 8)$.
- $E_3^- = E_2^- = (-12, 8)$.
- $E_4^- = E_3^- = (-12, 8)$.
- $E^-(379/645) = E_5^- = E_4^- - 3(97, -57) = (-303, 197)$.

This procedure gives an inductive way to define the pivot points for a pair of odd rationals. We define the *pivot arc* $P\Gamma$ of Γ to be the arc whose endpoints are E^+ and E^-. It turns out that the pivot arc is well defined – this is something we will prove simultaneously with our Copy Theorem below. This is to say that E^+ and E^- are both vertices of Γ. In Part 6 we prove the following result.

Theorem 22.3 (Copy) *If $A_1 \leftarrow A_2$, then $P\Gamma_2 \subset \Gamma_1$.*

Figure 22.1 illustrates the Copy Theorem. The first 3 frames are of $\Gamma(57/97)$, drawn in black and $P\Gamma(57/97)$ drawn in gray. The last frame shows several periods of $\Gamma(17/29)$.

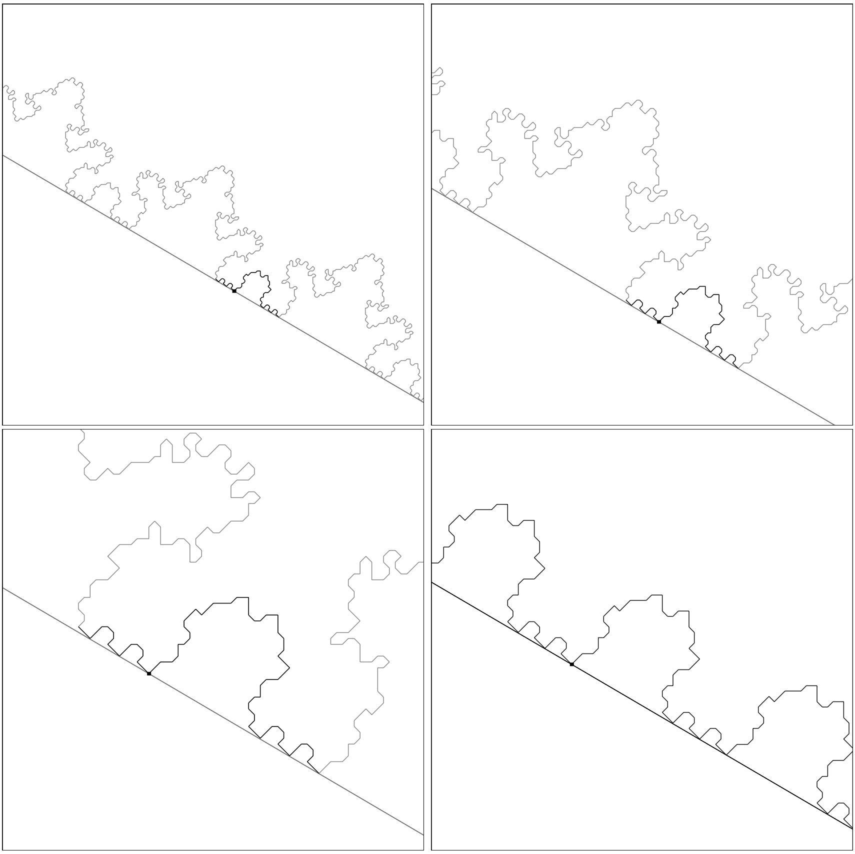

Figure 22.1: $P\Gamma(57/97) \subset \Gamma(17/29)$.

Now we turn to the statement of the Pivot Theorem. Given an odd rational parameter $A = p/q$, let V be the vector from Equation 3.2. Let $\mathbf{Z}[V]$ denote the group of integer multiples of $V = (q, -p)$. In Part 6 we prove the following result.

Theorem 22.4 (Pivot) *Every low vertex of* Γ *is equivalent mod* $\mathbf{Z}[V]$ *to a vertex of* $P\Gamma$. *That is,* $P\Gamma$ *contains one period's worth of low vertices on* Γ.

The Pivot Theorem makes a dramatic statement. Another way to state this theorem is to say that there are no low vertices on the complementary arc $\gamma - P\Gamma$. Here γ is the arc just to the right of $P\Gamma$ such that $P\Gamma \cup \gamma$ is one full period of Γ. A glance at Figure 22.1 will make this clear. We will prove the Pivot Theorem in Part 6. We will also prove the following easy estimate. See §26.2.

Lemma 22.5

$$-\frac{q}{2} < \pi_1(E^-) < \pi_1(E^+) < \frac{q}{2}.$$

22.3 HALF OF THE RESULT

We will prove that $C_n \subset C_n'$. This is almost an immediate consequence of the Copy Theorem.

For convenience, we recall the definition of C_n. Let $\mu_i = |p_n q_i - q_n p_i|$.

$$C_n = \bigcup_{\kappa \in \Pi_n} \Big(X_n(\kappa), -1 \Big), \qquad X_n(\kappa) = \frac{1}{q_n}\bigg(1 + \sum_{i=0}^{n-1} 2k_i \mu_i\bigg). \tag{22.5}$$

It is convenient to write

$$\tilde{V}_k = \text{sign}(A_{k+1} - A_k) V_k = \pm(q_k, -p_k). \tag{22.6}$$

When $1/1 \leftarrow A$, the pivot arc $P\Gamma(A)$ contains the points

$$k\tilde{V}_0, \qquad k = 0, ..., d_1, \qquad \tilde{V}_0 = (-1, 1). \tag{22.7}$$

This is a consequence of the argument in §19.5.

In general, suppose $A_1 \leftarrow A_2$ are two parameters. Then, by construction, the pivot arc $P\Gamma_2$ contains all points

$$v + k\tilde{V}_1 \qquad k \in \{0, ..., d\}, \qquad d = \text{floor}(q_2/2q_1). \tag{22.8}$$

Here v is any vertex of $P\Gamma_1$. It now follows from induction that $P\Gamma_n$ contains all points of the form

$$\sum_{j=0}^{n-1} k_j \tilde{V}_j, \qquad k_j \in \{0, ..., d_j\}. \tag{22.9}$$

Let M denote the map from Equation 2.10. Usually we take M so that $M(0,0) = 0$, but for the proof here, we adjust so that

$$M(0,0) = (1/q_n, -1). \tag{22.10}$$

(This makes no difference; see the discussion surrounding the definition of M in §2.5.) Call a lattice point *even* if the sum of its coordinates is even. Note that $\tilde{V}_j$ is even for all j. Hence all points in Equation 22.9 are even. The images of these points under M have their second coordinate equal to -1. We just have to worry about the first coordinate. We have

$$M(\tilde{V}_j) = \frac{1}{q_n} + 2|Aq_j - p_j| = \frac{1}{q_n} + \frac{1}{q_n} 2|p_n q_j - q_n p_j|. \tag{22.11}$$

The absolute value in the equation comes from the sign choice in the definition of $\tilde{V}_j$.

It now follows from the affine nature of M and from the definition of C_n that

$$C_n \subset O_2(1/q_n, -1). \tag{22.12}$$

It follows from the case $n = 0$ of Equation 21.7 that $C_n \subset [0, 2] \times \{-1\}$.

22.4 THE INHERITANCE OF LOW VERTICES

The rest of Theorem 1.8 follows from the Pivot Theorem and from what we have done by applying the information contained in the Pivot Theorem to what we did in the previous section. To make the argument work, we first need to deal with a tedious technical detail, which we take care of in this section.

Let $A_1 \leftarrow A_2$ be two odd rationals. As usual, we have

$$d_1 = \text{floor}(q_2/2q_1). \tag{22.13}$$

Let v_1 be a vertex on the pivot arc $P\Gamma_1$. Define

$$v_2 = v_1 + k\tilde{V}_1, \qquad k \in \{0, ..., d_1\}. \tag{22.14}$$

Here we mean to choose some arbitrary k. The argument we give will work for any choice. Notice that, as k ranges over all possibilities, we are considering exactly the same vertices as in §22.3. Now we want to take a close look at these vertices. Here is the main result of this section.

Lemma 22.6 *v_1 is low with respect to A_1 iff v_2 is low with respect to A_2.*

Proof: There are two cases to consider, depending on whether $A_1 < A_2$ or $A_2 < A_1$. We will consider the former case. The latter case has essentially the same treatment. In our case, we have $\tilde{V}_1 = V_1$. Let $E_j^{\pm}$ be the pivot points for Γ_j. Say that a vertex is *high* if it is not low.

We will first suppose that v_1 is low with respect to A_1 and that v_2 is high with respect to A_2. This will lead to a contradiction. We write $v_j = (m_j, n_j)$. Let M_j be the fundamental map from Equation 2.10. Since v_1 is low and v_2 is high, we have

$$2A_1m_1 + 2n_1 + \frac{1}{q_1} = M_1(v_1) \leq 2 - \frac{1}{q_1},$$

$$2A_2m_2 + 2n_2 + \frac{1}{q_2} = M_2(v_2) \geq 2 + \frac{1}{q_2}.$$

Rearranging terms,

$$2\left(\frac{p_2}{q_2}m_2 + n_2\right) - 2\left(\frac{p_1}{q_1}m_1 + n_1\right) \geq \frac{2}{q_1}. \tag{22.15}$$

Plugging in the relations $m_2 = m_1 + kq_1$ and $n_2 = n_1 - kp_1$ and simplifying, we have

$$\frac{(m_1 + kq_1)(p_2q_1 - p_1q_2)}{q_1q_2} \geq \frac{1}{q_1}. \tag{22.16}$$

Since $A_1 \leftarrow A_2$ and $A_1 < A_2$, we have

$$p_2q_1 - p_1q_2 = 2. \tag{22.17}$$

Hence

$$m_1 + kq_1 \geq q_2/2. \tag{22.18}$$

Combining Equations 22.3 and 22.18, we have

$$E_1^+(A_2) = E_1^+(A_1) + d_1 q_1 \geq^* m_1 + kq_1 > q_2/2 >^* E_1^+(A_2).$$

This is a contradiction. The first starred inequality comes from the Pivot Theorem and the fact that $k \leq d_1$. The second starred inequality comes from Corollary 22.5.

Now we will suppose that v_1 is high with respect to A_1 and v_2 is low with respect to A_2. This also leads to a contradiction. Let M_1 denote the first coordinate of the fundamental map relative to the parameter A_1, adjusted so that $M_1(0, 0) = 1/q_1$. That is,

$$M_1(m, n) = 2A_1 m + 2n + (1/q_1). \tag{22.19}$$

Since v_1 is high, we have the following dichotomy.

$$M_1(v) \geq 2 + \frac{1}{q_1}, \qquad M_1(v) > 2 + \frac{1}{q_1} \implies M_1(v) \geq 2 + \frac{3}{q_1}. \tag{22.20}$$

We will consider these two cases in turn.

Case 1: If $M_1(v_1) = 2 + 1/q_1$, then

$$M_1(m_1, n_1 - 1) = 1/q_1 = M_1(0, 0).$$

But then

$$(m_1, n_1 - 1) = jV_1$$

for some integer j. But then $|m_1| \geq q_1$. Since $v_1 \in P\Gamma_1$, this contradicts Corollary 22.5. Hence

$$v_1 = (0, 1), \qquad v_2 = kV_1 + (0, 1).$$

If v_2 is low, then

$$0 = 2k(A_1 q_1 - p_1) < 2k(A_2 q_1 - p_1) = M_2(v_2) - M_2(0, 1) \leq 0.$$

This is a contradiction. The first inequality comes from $A_1 < A_2$.

Case 2: If $M_1(v_1) \geq 2 + 3/q$, then the same reasoning as in Equations 22.15–22.17 (but with signs reversed) leads to

$$m_1 + kq_1 < -3q_2. \tag{22.21}$$

But then

$$\frac{-q_2}{2} < \frac{-q_1}{2} <^* m_1 \leq m_1 + kq_1 < -3q_2.$$

The starred inequality comes from Corollary 22.5. Again we have a contradiction, this time by a wide margin. □

22.5 THE OTHER HALF OF THE RESULT

Now we finish the proof of Theorem 1.8.

We revisit the construction in §22.3 and show that actually $C_n = C'_n$. Let Λ_n denote the set of low vertices of $P\Gamma_n$. By the Pivot Theorem, every low vertex on $P\Gamma_n$ is equivalent to a point of Λ_n mod $\mathbf{Z}[V_n]$.

Lemma 22.7 *For any $n \geq 0$, we have*

$$\Lambda_{n+1} = \bigcup_{k=0}^{d_n} (\Lambda_n + k\tilde{V}_n).$$

Proof: Induction. For $n = 0$ we have

$$E_1^- = (-d_0, d_0), \qquad E_1^+ = (0, 0).$$

In this case, the right hand side of the equation precisely describes the set of points on the line segment joining the pivot points. The case $n = 0$ therefore follows directly from the Pivot Theorem.

Let Λ'_{n+1} denote the right hand side of the main equation. Since Γ_n is invariant under translation by V_n, every vertex of Λ'_{n+1} is low with respect to A_n. Hence, by Lemma 22.6, every vertex of Λ'_{n+1} is low with respect to A_{n+1}. Combining this fact with Equation 22.8, we see that

$$\Lambda'_{n+1} \subset \Lambda_{n+1}. \tag{22.22}$$

By Lemma 22.6 again, every $v \in \Lambda_{n+1}$ is also low with respect to A_n. Hence

$$v = v' + k\tilde{V}_n, \qquad k \in \mathbf{Z} \tag{22.23}$$

for some $v' \in \Lambda_n$. If $k \notin \{0, ..., d_n\}$, then v lies either to the left of the left pivot point of Γ_{n+1} or to the right of the right pivot point of Γ_{n+1}. Hence $k \in \{0, ..., d_n\}$. This proves that

$$\Lambda_{n+1} \subset \Lambda'_{n+1}. \tag{22.24}$$

Combining the two facts completes the induction step. □

We proved Lemma 22.7 with respect to the inferior sequence. However, notice that, if $d_n = 1$, then $\Lambda_{n+1} = \Lambda_n$. Thus we get precisely the same result for consecutive terms in the superior sequence. We have shown that $v \in \Gamma_n$ is low if and only if $v \in \Lambda_n$ mod $\mathbf{Z}[V]$. But then

$$O_2(1/q_n, -1) \cap I = M(\Lambda_n), \qquad I = [0, 2] \times \{-1\}. \tag{22.25}$$

Here M is the fundamental map. Recognizing Λ_n as the set from Equation 22.9, we get precisely the equality in Theorem 1.8.

There is one last detail. One might worry that M maps some points of Λ_n to points on $[0, 2] \times \{1\}$ (rather than to $[0, 2] \times \{-1\}$). However, all points in Λ_n have even parity. Hence this does not happen.

This completes the proof of Theorem 1.8.

22.6 THE COMBINATORIAL MODEL

Here we prove Lemmas 22.1 and 22.2.

22.6.1 Combinatorics of the Return Map

Let Σ_n denote the union of all points in Equation 22.9. We have

$$M(\Sigma_n) = C_n. \tag{22.26}$$

The ordering on Σ_n determines the ordering of the return dynamics to C_n. We set $\Sigma_0 = \{(0,0)\}$ for convenience. We can determine the ordering on Σ_{n+1} from the ordering on Σ_n and the sign of $A_{n+1} - A_n$. When $A_n < A_{n+1}$, we can write the relation

$$\Sigma_n + kV_n \prec \Sigma_n + (k+1)V_n, \qquad k = 0, ..., (d_n - 1) \tag{22.27}$$

to denote that each point in the left hand set precedes each point in the right hand set. Within each set, the ordering does not change. When $A_n > A_{n+1}$, we can write the relation

$$\Sigma_n - (k+1)V_n \prec \Sigma_n - kV_n, \qquad k = 0, ..., (d_n - 1). \tag{22.28}$$

Lemma 22.1 follows from these facts and induction.

22.6.2 Geometry of the Return Map

Let β_n denote the arc of $P\Gamma_n$, chosen so that $P\Gamma_n \cup \beta_n$ is one period of $P\Gamma_n$. Let L_n be the line of slope $-A_n$ through the origin.

Lemma 22.8 *No point of β_m lies more than q_m vertical units away from L_m, and some point of β_m lies at least $q_m/4$ vertical units away from L_m.*

Proof: By the Room Lemma, $\beta_m \subset R(A_m)$. The upper bound follows immediately from this containment. For the lower bound, recall from the Room Lemma that $P\Gamma_m$ crosses the centerline L of $R(A_m)$ once, and this crossing point lies at least $(p_m + q_m)/4 > q_m/4$ vertical units from L_m. By Lemma 22.5 and symmetry, the left endpoint of β_m lies to the left of L and the right endpoint of β_m lies to the right of L. Hence β_m contains the crossing point we have mentioned. For an alternative argument, we note that no point on the pivot arc crosses the line parallel to the floor and ceiling of $R(A_m)$ and halfway between them, whereas the crossing point lies above this midline. $\square$

Notice that the line L_n replaces the line L_m in the next lemma.

Lemma 22.9 *Let $m \le n$ and $q_m > 10$. Then some point of β_m lies at least $q_m/4 - 1$ vertical units from L_n. Moreover, no point of β_m lies more than $q_m + 1$ vertical units away from L_n.*

Proof: Some point v of β_m is at least q_m vertical units from L_m, by the previous result. From Lemma 17.4, we have

$$|A_m - A_n| < 2/(q_m^2). \tag{22.29}$$

On the other hand, by the Room Lemma and by construction, $P\Gamma_m$ is contained in two consecutive translates of $R(A_m)$, one of which is $R(A_m)$ itself. Hence $P\Gamma_m$ lies entirely inside the ball B of radius $4q_m$ about the origin. By Equation 22.29, the Hausdorff distance between the seqments $L_m \cap B$ and $L_n \cap B$ is less than 1 once $m > 10$. By construction, the vertical line segment starting at v and dropping down $q_m - 1$ units is disjoint from $L_n \cap B$. But this segment is disjoint from $L_n - B$ as well. Hence v is at least $q_m/2 - 1$ vertical units from L_n. The upper bound has a similar proof. □

Lemma 22.10 *β_m has length at most $5q_m^2$.*

Proof: β_m is contained in one period of $P\Gamma_m$. Hence it suffices to bound the length of any one period of $P\Gamma_m$. By the Room Lemma, one such period is contained in $R(A_m)$. We compute easily that the area of $R(A_m)$ is much less than $5q_m^2$. Hence there are fewer than $5q_m^2$ vertices in $R(A_m)$. Hence the length of one period of $P\Gamma_m$ is less than $5q_m^2$. □

Suppose now that κ and κ_+ are two consecutive points of Σ_n. By this we mean that there is a portion of $P\Gamma_n$ connecting κ to κ_+ when it is oriented from left to right.

We want to understand the arc of $P\Gamma_n$ that joins κ and κ_+. Suppose that $\text{ind}(\kappa) = m$. It follows from induction and from the Copy Theorem that there is some translation T such that $T(\kappa)$ and $T(\kappa_+)$ are the endpoints of the arc β_m. The arc joining κ to κ_+ has the same length as β_m, and this length is less than $5q_m^2$. This gives us the estimate for h_2.

Now we deal with h_1. We check the result by hand for $q_n < 10$. So, suppose that $q_n > 10$. All the vertices κ, κ_+, $T(\kappa)$, and $T(\kappa_+)$ lie within 1 vertical unit of the baseline L_n. We know that the vertical distance from some point of β_m to L_n is at least $q_m/2 - 1$. Hence the vertical distance from some point on $T(\beta_m)$ to L_n is at least $q_m/2 - 2$. Similarly, the vertical distance from any point of β_m to L_n is at most $q_m + 2$. If two points in $\mathbf{Z}^2$ have vertical distance d, then the images of these points under the fundamental map M_n have horizontal distance $2d$. In short, the fundamental map doubles the relevant distances. This fact gives us the estimate for h_1.

This completes the proof of Lemma 22.2. □

Remark: We have tried to give fairly precise estimates in our arguments, but actually we do not use these estimates for any purpose.

22.7 THE EVEN CASE

Here we discuss Theorem 1.8 in the even case. For each even rational $A_1 \in (0,1)$, there is a unique odd rational A_2 such that (in terms of Equation 4.1) $A_1 = (A_2)_\pm$ and $q_2 < 2q_1$. In Lemma 27.2, we will show that Γ_1 (a closed polygon) contains a copy of $P\Gamma_2$, and all low vertices of Γ_1 lie on this arc. From this fact, we see that

$$O(1/q_1, -1) = M_1(\Sigma_1), \tag{22.30}$$

just as in the odd case. Here M_1 is the fundamental map defined relative to the parameter A_1, and Σ_1 is the set of low vertices on $P\Gamma_1$.

Note that $\Sigma_1 = \Sigma_2$, where Σ_2 is the set of low vertices on $P\Gamma_2$. The only difference between the two sets $M_1(\Sigma_1)$ and $M_2(\Sigma_2)$ is the difference in the maps M_1 and M_2. Now we explain the precise form of Theorem 1.8 that this structure entails.

Switching notation, let A be an even rational. One of the two rationals $A_\pm$ from Equation 4.1 is odd, and we call this rational A'. We can find the initial part of a superior sequence $\{A_k\}$ such that $A' = A_{n-1}$. We set $A = A_n$ even though A does not belong to this sequence. Referring to Theorem 1.8, we define Π_n exactly as in the odd case but for one detail. In case $2q' > q$, we simply ignore the nth factor of Π_n. That is, we treat q' as an inferior term. Once changed in this way, Theorem 1.8 holds in the even case and has a proof that follows the odd case word for word.

Here we give an example. Let $A_1 = 12/31$. Then $A_2 = 19/49$, exactly is in the introduction. We have $n = 3$, and the sequence is

$$\frac{p_0}{q_0} = \frac{1}{1}, \frac{1}{3}, \frac{5}{13}, \frac{12}{31} = \frac{p_3}{q_3}.$$

All terms are superior, so this is also the superior sequence. The renormalization sequence is $1, 2, 1$. The μ sequence is $19, 5, 1$. The first coordinates of the 12 points of $O_2(1/49) \cap I$ are given by

$$\bigcup_{k_0=0}^{1} \bigcup_{k_1=0}^{2} \bigcup_{k_2=0}^{1} \frac{2(19k_0 + 5k_1 + 1k_2) + 1}{31}.$$

Writing these numbers in a suggestive way, the union above works out to

$$\frac{1}{31} \times \big(1 \quad 3 \qquad 11 \quad 13 \qquad 21 \quad 23 \qquad\qquad\qquad 39 \quad 41 \qquad 49 \quad 51 \qquad 59 \quad 61\big).$$

Chapter Twenty-Three

The Comet Theorem

23.1 STATEMENT 1

We fix an irrational parameter $A \in (0,1)$. Let $\{A_n\}$ be the superior sequence approximating A. Let $\widehat{\Gamma}_n$ be the arithmetic graph corresponding to A_n. We say that a vertex v of $\widehat{\Gamma}_n$ is *D-low* if v is within D vertical units of the baseline of $\widehat{\Gamma}_n$.

Note that the low vertices considered in the previous chapter are 1-low vertices. These vertices play a special role in our arguments. The fundamental map M from Equation 2.10 maps the 1-low vertices into the interval

$$J = (0,2) \times \{-1,1\}. \tag{23.1}$$

In Part 6 we prove the following result.

Theorem 23.1 (Low Vertex) *Fix N_0. There are constants N_1 and N_2 with the following property. If v_n is an N_0-low vertex contained in a component of $\widehat{\Gamma}_n$ having diameter at least N_1, then there is an arc of $\widehat{\Gamma}_n$ that has length at most N_2 and connects v_n to a 1-low vertex. The constants N_1 and N_2 depend only on A and N_0.*

Now we will deduce statement 1 of the Comet Theorem. Looking at Figure 1.2, we see that

$$J = I \cup (\psi')^3(I), \tag{23.2}$$

where ψ' is the outer billiards map. Hence it suffices to prove statement 1 of the Comet Theorem with J in place of I. Since $\psi = (\psi')^2$, it suffices to prove the result with J in place of I and ψ in place of ψ'. This is what we will do.

Fix $N > 0$. The constants $N_0, N_1, ...$ depend only on A and N. Recall that $\Xi = \mathbf{R}_+ \times \{-1,1\}$ and that Ψ is the first return map to Ξ. Recall also that U_A is the union of the unbounded special orbits.

Corollary 23.2 *If $\xi \in \Xi \cap U_A$ and $\|\xi\| < N$ then $\Psi^k(\xi) \in J$ for some $|k| < N_2$.*

Proof: The arithmetic graph $\widehat{\Gamma}_n$ tracks the orbits of the special intervals defined in §2.2. For each n we choose some special interval I_n whose closure contains ξ. Typically the choice is unique, but when ξ lies in the boundary of a special interval, there are two choices and we pick one arbitrarily.

Let v_n be the vertex of $\widehat{\Gamma}_n$ corresponding to I_n. From Equation 2.10 we see that v_n is N_0-low, where $N_0 = (N/2) + 1$. Let β_n be the component of $\widehat{\Gamma}_n$ that contains v_n. By the Continuity Principle in §2.7,

$$\mathrm{diam}(\beta_n) \to \infty. \tag{23.3}$$

By Equation 23.3, the diameter of β_n exceeds N_1 for n large.

The Low Vertex Theorem says that there is some arc β_n' of β_n, having length at most N_2, that connects v_n to a 1-low vertex.

By the Continuity Principle, the first N_2 iterates of Ψ_n are defined on ξ for n large. Interpreting β_n' dynamically we see that there is a sequence $\{k_n\}$ such that

$$\Psi_n^{k_n}(\xi) \in J, \qquad |k_n| < N_2. \tag{23.4}$$

By the Pidgeonhole Principle, some k appears infinitely often in the sequence $\{k_n\}$. Applying the Continuity Principle to this subsequence, we see that $\Psi^k(\xi) \in J$. $\square$

Remark: Referring to the proof we just gave, one might worry that some of the points involved actually lie in the boundary of J. However, the boundary points of J do not have well defined orbits and all the points we considered do have well defined orbits. Hence this problem does not occur.

Corollary 23.3 *If $\xi \in \Xi \cap U_A$ and $\|\xi\| < N$, then $\psi^k(\xi) \in J$ for some $|k| < N_5$.*

Proof: By Corollary 23.2, there is some $m \in (-N_2, N_2)$ such that $\Psi^m(\xi) \in J$. We will consider the case when $m \geq 0$. The proof in the other case is essentially the same. Let $\xi_0 = \xi$ and inductively define

$$\xi_j = \Psi(\xi_{j-1}), \qquad j = 1, ..., m. \tag{23.5}$$

Examining the proof of the Pinwheel Lemma, we see that there is some constant N_3 such that

$$\|\xi_j\| < N_3, \qquad j = 0, ..., m. \tag{23.6}$$

Again examining the proof of the Pinwheel Lemma, we see that there are constants $n_1, ..., n_m$ such that

$$\xi_j = \psi^{n_j}(\xi_{j-1}), \qquad n_j \in (0, N_4). \tag{23.7}$$

Setting $N_5 = N_2 N_4$ we see that $\psi^k(\xi) \in J$ for some $|k| < N_5$. $\square$

Corollary 23.4 *If $\zeta \in U_A$ and $\|\zeta\| < N$, then $\psi^k(\zeta) \in J$ for some $|k| < N_8$.*

Proof: Examining the proof of the Pinwheel Lemma, we see that there is some constant N_6, some $|m| < N_6$, and some $\xi \in \Xi$ such that

$$\xi = \psi^m(\zeta), \qquad \|\xi\| < N_6. \tag{23.8}$$

Applying Corollary 23.3 with N_6 in place of N, we have $\psi^n(\xi) \in J$ for some $|n| < N_7$. Therefore $\psi^k(\zeta) \in J$ for some $|k| < N_8$. Here we have set $N_8 = N_6 + N_7$. $\square$

Corollary 23.4 is identical to statement 1 of the Comet Theorem, except that it uses ψ in place of ψ' and J in place of I. This completes the proof.

23.2 THE CANTOR SET

Before we prove the remaining statements of the Comet Theorem we first need to resolve the technical point that the set C_A is actually well defined. For convenience, we repeat the definition.

$$C_A = \bigcup_{\kappa \in \Pi} \Big(X(\kappa), -1\Big), \qquad X(\kappa) = \sum_{i=0}^{\infty} 2k_i |Aq_i - p_i|. \tag{23.9}$$

Lemma 23.5 *The infinite sums in Equation 23.9 converge. Hence C_A is well defined.*

Proof: Combining Equation 21.5 with the bound $0 \le k_n < d_n$, we see that the nth term in the sum defining $X(\kappa)$ is at most $2q_n^{-1}$. Given that $2q_k < q_{k+1}$ for all k, we get

$$2q_n^{-1} < 2^{-n+1}. \tag{23.10}$$

The sequence defining $X(\kappa)$ decays exponentially and hence converges. □

For the purposes of this section, we equip the product space Π with the lexicographic ordering and the product topology. For instance, if $d_n = 1$ for all n, then Π is just the space of binary sequences. The lexicographic order treats these sequences as binary expansions of real numbers and then orders them as usual. The general case is similar.

Lemma 23.6 *The map $X \colon \Pi \to C_A$ is a homeomorphism that maps the lexicographic order to the linear order. Hence C_A is a Cantor set.*

Proof: We first show that the map X is injective. In fact, we will show that X is order preserving. If $\kappa = \{k_i\} \prec \kappa' = \{k_i'\}$ in the lexicographic ordering, then there is some smallest index m such that $k_i = k_i'$ for all indices $i = 0, ..., (m-1)$ and $k_m < k_m'$. Let $\lambda_m = |Aq_m - p_m|$, as in Equation 21.5. Then

$$X(\kappa') - X(\kappa) \ge 2\lambda_m - \sum_{k=m+1}^{\infty} 2d_k\lambda_k = \lambda_m - \lambda_{m+1}' > 0 \tag{23.11}$$

by Equation 21.7.

The map $X \colon \Pi \to [0, 2]$ is continuous with respect to the topology on Π because the nth term in the sum defining X is always less than 2^{-n+1}. We also know that X is injective. Hence X is bijective onto its image. Any continuous bijection from a compact space to a Hausdorff topological space is a homeomorphism. □

Remark: In Chapter 25 we will have much more to say about the geometry of C_A. For instance, C_A always has length 0.

23.3 A PRECURSOR OF THE COMET THEOREM

In this section we present two auxilliary results that combine to prove almost all the remaining statements of the Comet Theorem.

Let C_A be the Cantor set considered in the previous section. Define

$$C'_A = C_A - (2\mathbf{Z}[A] \times \{-1\}). \tag{23.12}$$

One can view our next result as a precursor of the Comet Theorem.

Theorem 23.7 (Comet Precursor) *Let U_A denote the set of unbounded special orbits relative to an irrational $A \in (0, 1)$.*

1. $C'_A \subset U_A$.
2. *The first return map $\rho_A \colon C'_A \to C'_A$ is defined precisely on $C'_A - \phi(-1)$. The map ϕ^{-1} conjugates ρ_A to the restriction of the odometer on $\mathcal{Z}_A$.*
3. *For $\zeta \in C'_A - \phi(-1)$, the orbit portion between ζ and $\rho_A(\zeta)$ has excursion distance in*

$$\left[\frac{d^{-1}}{2} - 4, 2d^{-1} + 20\right]$$

and length in

$$\left[\frac{d^{-2}}{32} - \frac{d^{-1}}{4}, 100d^{-3} + 100d^{-2}\right].$$

Here $d = d(-1, \phi^{-1}(\zeta))$.

Remarks:

(i) The constants in item 3 are not optimal; some tedious elementary arguments would improve them.

(ii) Since $d^{-1} \geq 1$, the estimates in item 3 imply the less precise estimates in the Comet Theorem – once we establish that $C^{\#}_A = C'_A$.

(iii) As we remarked following the Comet Theorem, the only nonsharp bound in item 3 is the length upper bound. For instance, our proof in [**S1**], which establishes a kind of coarse self-similarity structure, would give a better bound for $A = \sqrt{5} - 2$ if carefully examined. We conjecture that -3 is the best bound that works for all parameters at once.

To relate Theorem 23.7 to the Comet Theorem, we prove the following double identity.

Lemma 23.8 $U_A \cap I = C^{\#}_A = C_A - (2\mathbf{Z}[A] \times \{-1\})$.

Statements 2 and 3 of the Comet Theorem follow from this result and Lemma 23.7. Lemma 23.8 also contains the first claim in statement 4 of the Comet Theorem. At the end of the chapter, we will prove the second claim made in statement 4 of the Comet Theorem.

23.4 CONVERGENCE OF THE FUNDAMENTAL ORBIT

Let $\{p_n/q_n\}$ denote the superior sequence associated to A. We use the notation from the previous chapter. Here Γ_n denotes the corresponding arithmetic graph and

$$C_n = \bigcup_{\kappa\in\Pi_n} \Big(X_n(\kappa), -1\Big), \qquad X_n(\kappa) = \frac{1}{q_n} + \sum_{i=0}^{n-1} 2k_i|A_n q_i - p_i|. \tag{23.13}$$

We have already proved that

$$C_n \subset O_2(1/q_n, -1). \tag{23.14}$$

Let $\kappa \in \Pi$ be some infinite sequence. Let $\kappa_n \in \Pi_n$ be the truncated sequence. Let

$$\sigma_n = (X_n(\kappa_n), -1), \qquad \sigma = (X(\kappa), -1) \tag{23.15}$$

Here is our basic convergence result.

Lemma 23.9 *$\sigma_n \to \sigma$ as $n \to \infty$.*

Proof: For $i < n$, let $\tau_{i,n}$ denote the ith term in the sum for $X_n(\kappa_n)$. Let τ_n be the corresponding term in the sum for $X(\kappa)$. By Lemma 17.1, the sign of $A - A_i$ is the same as the sign of $A_n - A_i$. Therefore

$$\begin{aligned}|\tau_n - \tau_{i,n}| &=\\ 2k|A - A_n|q_n < 2q_n^{-1} &<\\ 2^{-n+1}&.\end{aligned} \tag{23.16}$$

Therefore

$$\begin{aligned}|\sigma - \sigma_n| &=\\ |X(\kappa) - X(\kappa_n)| &=\\ \sum_{i=0}^{n-1}|\tau_n - \tau_{i,n}| + \sum_{i=n}^{\infty}\tau_i &<\\ 2\sum_{i=0}^{n-1}2^{-n} + 2\sum_{i=n}^{\infty}2^{-i} &<\\ (2n+4)2^{-n}&.\end{aligned} \tag{23.17}$$

This completes the proof. $\square$

The uniformity of convergence gives us the following immediate corollary.

Corollary 23.10 *C_A is the Hausdorff limit of $\{C_n\}$.*

23.5 AN ESTIMATE FOR THE RETURN MAP

Let $\{k_i\}$ be a point in the sequence space Π. We call $\{k_i\}$ *first* if $\tilde{k}_i = 0$ for all i, and *last* if $\tilde{k}_i = d_i$ for all i. The map $\phi_2 \colon \Pi_A \to C_A$ is a homeomorphism. Using ϕ_2, we transfer the notions of *first* and *last* to points of C_A.

It turns out that the first and last points of C_A are the special points mentioned in connection with the Comet Theorem. If these orbits are well defined, then it turns out that the last point leaves C_A under the forward dynamics and never returns. Likewise, the first point of C_A leaves under the backward dynamics and never returns. (We prove these statements later on.) Here we will estimate the nature of how the nonlast points of C_A return to C_A under the forward direction of the dynamics. The idea is to essentially take the geometric limit of the result from Lemma 22.2.

Let $\zeta \in C'_A$ denote a point that is not last. Let κ denote the corresponding sequence in Π_A. Say that two sequences in Π are *equivalent* if they have the same infinite tail end. We can define the reverse lexicographic order on any equivalence class. Likewise, we can extend the twirl order to any equivalence class. In particular, we extend the twirl order to the equivalence class of κ, the sequence currently of interest to us.

Remark: These orders on equivalence classes cannot be defined on the entire space; points in different equivalence classes are often not comparable.

Since κ is not last, we can find some smallest index $m = m(\zeta)$ where $\tilde{k}_m < d_i$. In other words, m is the smallest index such that κ differs from the last sequence in the mth spot.

The successor κ_+ of κ is obtained by incrementing $\tilde{k}_m$ by 1 and setting $\tilde{k}_i = 0$ for all $i < m$. This notion of successor is compatible with the twirl ordering on the finite truncations Π_n. Define

$$\zeta_+ = (X(\kappa_+), -1), \qquad (\zeta_n)_+ = (X(\kappa_n)_+, -1). \tag{23.18}$$

Lemma 23.11 *Let $\zeta \in C'_A$ be a point that is not last. Let $m = m(\zeta)$. The forward Ψ orbit of ζ returns to C_A as ζ_+ in at most $5q_m^2$ steps. This portion of the orbit wanders between $q_m/2 - 2$ units and $2q_m + 2$ units away from* $(0, -1)$.

Proof: By Lemma 2.2, the orbit of ζ is well defined. Referring to the notation in Lemma 22.2, we get $\mathrm{ind}(\kappa_n) = m$ for n large enough. Hence the forward Ψ_n orbit of ζ_n returns to $(\zeta_n)_+$ after at most $5q_m^2$ steps, moving away from $(0, -1)$ by at least $q_m/2 - 2$ units and at most $2q_m + 2$ steps. Here m is independent of n. Since X is continuous, we have $(\zeta_n)_+ \to \zeta_+$ as $n \to \infty$. The Contintuity Principle implies that the forward Ψ orbit of ζ returns as ζ_+ after at most $5q_m^2$ steps, moving away from $(0, -1)$ at least $q_m/2 - 2$ units and at most $2q_m + 2$ steps. □

There is an entirely analogous result for the backward return map. This result holds for all but the first point.

23.6 PROOF OF THE COMET PRECURSOR THEOREM

23.6.1 Statement 1

Here we prove statement 1 of Theorem 23.7. We say that a sequence of Π_A is *equivalent-to-first* if it differs from the first sequence in only a finite number of positions. We call a sequence *equivalent-to-last* if it differs from the last sequence in a finite number of positions. As in the previous section, we transfer these notions to C_A. It is immediate from the definitions that no sequence in Π_A is both equivalent-to-first and equivalent-to-last.

Let ζ be a point in C'_A that is not equivalent-to-last. We will show that the forward orbit of ζ is unbounded. Let $m = m(\kappa)$ be as in the proof of Lemma 23.11. Lemma 22.2 says that the portion of the orbit between ζ and ζ_+ wanders at least $q_m/5$ from the origin. Since we can achieve any initial sequence we like with iterated successors of κ, we can find iterated successors κ' of κ such that $m(\kappa')$ is as large as we like. But this shows that the forward orbit of ζ is unbounded. Here we are using the fact that $\lim_{m\to\infty} q_m = \infty$. This shows that ζ has an unbounded forward orbit.

Essentially the same argument works for the backward orbit of points that are not equivalent-to-first. This establishes statement 1.

23.6.2 Statement 2

The successor map on Π_A is defined except on the last sequence κ of Π_A. Referring to the homeomorphism ϕ_1 given in Equation 1.8, we have

$$\phi_1(-1) = \kappa.$$

Thus the point $\phi_2(\kappa) \in C_A$ corresponding to κ is precisely $\phi(-1)$. By Lemma 23.11, the return map $\rho_A\colon C'_A \to C'_A$ is defined on $C'_A - \phi(-1)$.

The map ϕ_1 conjugates the odometer map on $\mathcal{Z}_A$ to the successor map on Π_A. Combining this fact with Lemma 23.11, we see that ϕ^{-1} conjugates ρ_A to the restriction of the odometer map on $\mathcal{Z}_A$.

It remains to understand what happens to the forward orbit of $x = \phi(-1)$ in the case when $x \in C'_A$. The following result completes the proof of statement 2.

Lemma 23.12 *If $x \in C'_A$, then the forward orbit of x does not return to C'_A.*

Proof: Suppose that the forward orbit of x returns to C'_A after N steps. Since outer billiards is a piecewise isometry, there is some open neighborhood U of x such that every point of $C'_A \cap U$ returns to C'_A in at most N steps. But there is some uniformly small m such that every point $\zeta \in C'_A - U$ differs from the last sequence κ at or before the mth spot. Lemma 23.11 says that such points return to C'_A in a uniformly bounded number of steps. In short, all points of C'_A return to C'_A in a uniformly bounded number of steps. But then all orbits in C'_A are bounded. This is a contradiction. □

23.6.3 Statement 3

Let $\zeta \in C'_A$. Let O_ζ denote the portion of the forward outer billiards orbit of ζ between ζ and $\rho_A(\zeta)$. We mean to use the original outer billiards map ψ' here. Let m be such that

$$d(\phi^{-1}(\zeta), -1) = q_m^{-1}. \tag{23.19}$$

By definition $\phi^{-1}(\zeta)$ and -1 disagree by $\mathbf{Z}/D_{m+1}$, but then they agree in $\mathbf{Z}/D_k$ for $k = 1, ..., m$. In the case when $m = 0$, the points $\phi^{-1}(\zeta)$ and -1 already disagree in $\mathbf{Z}/D_1$. Let $\kappa \in \Pi_A$ denote the sequence corresponding to ζ.

Finding the Index: Let $\text{ind}(\kappa)$ be as in §22.1. Let λ be the sequence corresponding to $\phi(-1)$. Then λ is the last sequence in the twirl order. The sequences κ and λ agree in positions $k = 0, ..., m-1$ but then disagree in position m. Hence m is the first index where κ disagrees with the last sequence in the twirl order. But then κ and κ_+ disagree in positions $0, ..., m$ and agree in position k for $k > m$. Hence $\text{ind}(\kappa) = m$.

Excursion Distance Bounds: Lemma 23.11 tells us that the Ψ-orbit of ζ between ζ and $\rho_A(\zeta)$ wanders between $q_m/2 - 4$ and $2q_m + 4$ units from the origin. Here we are interested in the full outer billiards O_ζ. Since the Ψ-orbit of ζ between ζ and ρ_ζ is a subset of O_ζ, the lower bound follows from Lemma 23.11.

The upper bound follows from a simple geometric analysis of the Pinwheel Lemma. Starting at a point on Ξ that is R units from the origin, the ψ-orbit remains within $2R + 8$ units of the origin before returning to Ξ. Essentially, the ψ-orbit follows an octagon once around the kite before returning, as shown in Figure 7.3. The constant of 10 amply takes care of the small deviations from this path that arise in the proof of the Pinwheel Lemma. Since ψ' always acts as the reflection in a vertex that is within 1 unit of the origin, we see that the entire ψ'-orbit of interest to us is at most $2R + 12$ units from the origin. Hence the portion of the orbit of interest wanders at most $2(q_m + 4) + 12 = 2q_m + 20$ units from the origin.

Orbit Length Bounds: The Ψ-orbit of ζ between ζ and $\rho_A(\zeta)$ has length at most $5q_m^2$. Examining the proof of the Pinwheel Lemma, we see that a point on Ξ that is R units from the origin returns to Ξ in less than $10R$ iterates. Given that $R = 2q_m + 2$, the orbit O_ζ is at most $20q_m + 20$ times as long as the corresponding Ψ-orbit. This gives the upper bound.

Now we prove the lower bound. Some point in the Ψ-orbit of ζ between ζ and $\rho_A(\zeta)$ lies at least $q_m/2 - 4$ vertical units from the origin. Consecutive iterates in the Ψ-orbit have vertical distance at most 4 units apart. Hence there are at least $q_m/8 - 1$ points in the Ψ-orbit that are at least $q_m/4$ horizontal units from the origin. Inspecting the Pinwheel Lemma, we see that the length of the ψ'-orbit between two such points is at least $q_m/4$. Hence O_ζ has length at least $q_m^2/32 - q_m/4$.

This completes the proof of statement 3.

23.7 THE DOUBLE IDENTITY

In this section we will prove Lemma 23.8. Our proof of this result relies on the following technical theorem.

Theorem 23.13 (Period) *For any $\epsilon > 0$, there is an $N > 0$ with the following property. If $\zeta \in I$ is more than ϵ units from C_n, then the period of ζ is at most N. The constant N depends only on ϵ and A.*

We will prove the Period Theorem in Part 6. Here is a corollary of this result.

Corollary 23.14 $U_A \cap I \subset C_A$.

Proof: The constants $N_1, N_2, \ldots$ depend only on ϵ and A.

We will suppose that U_A contains a point $\zeta \notin C_A$ and derive a contradiction. By compactness, there is some $\epsilon > 0$ such that ζ is at least 3ϵ from any point of C_A. Since C_A is the geometric limit of C_n, we see that there is some N_1 such that $n > N_1$ implies that ζ is at least 2ϵ from C_n.

Let $\{\zeta_n\} \in I$ be a sequence of points converging to ζ. We can choose these points so that the orbit of ζ_n relative to A_n is well defined. There is a constant N_2 such that $n > N_2$ implies that ζ_n is at least ϵ from C_n. But then, by the Period Theorem, there is some N_3 such that the period of ζ_n is at most N_3.

On the other hand, by the Continuity Principle in §2.7, the arithmetic graph $\Gamma(\zeta_n, A_n)$ converges to the arithmetic graph $\Gamma(\zeta, A)$. In particular, the period of $\Gamma(\zeta_n, A_n)$ tends to ∞. This is a contradiction. Hence ζ cannot exist. $\square$

Now we state a useful principle that will help with the remainder of the proof of Lemma 23.8.

Odometer Principle: Let Π_A be the sequence space from §1.7. Say that two sequences in Π_A are equivalent if they have the same infinite tail ends. Given the nature of the odometer map, we have the following useful principle. Any two equivalent sequences are in the same orbit of the odometer map. Call this the *Odometer Principle*. We will use this principle several times in our proofs.

Lemma 23.15 *No point of $C_A - C_A^{\#}$ has a well defined orbit.*

Proof: Let $\{d_n\}$ be the renormalization sequence, as above. Call a sequence in Π_A *equivalent-to-trivial* if it either differs from the 0 sequence by a finite number of terms or it differs from the sequence $\{d_i\}$ by a finite number of terms. The homeomorphism ϕ_2 bijects the equivalent-to-trivial points in Π_A to $C_A - C_A^{\#}$.

Suppose first that the superior sequence for A is not eventually monotone. Referring to §23.6 for definitions, in this case an equivalent-to-trivial sequence is neither equivalent-to-first nor equivalent-to-last.

Suppose $\sigma \in C_A - C_A^{\#}$ has a well defined orbit. Let κ be the equivalent-to-trivial sequence corresponding to σ. By Lemma 23.11 and the analog for the backward orbit, both directions of the orbit of σ return infinitely often to $C_A - C_A^{\#}$. If κ is eventually 0, then by the Odometer Principle, κ is in the same sequence orbit as the 0 sequence κ_0. But the point in C_A corresponding to κ_0 is exactly the vertex $(0, -1)$. This vertex does not have a well defined orbit. This is a contradiction. If κ is such that $k_i = d_i$ for large i, then by the Odometer Principle, κ is in the same orbit as the sequence $\{d_i\}$. By Equation 21.5, the corresponding point in C_A is $(2, -1)$. One can easily check that the orbit of $(2, -1)$ is not defined after the second iterate. Again we have a contradiction.

Now suppose the superior sequence is eventually monotone. We will treat the case when $A - A_n$ is eventually positive. In this case, $\{A_n\}$ is eventually monotone increasing. Suppose that κ is equivalent to the 0-sequence. We can iterate backward a finite number of times until σ returns as the first point of C_A. Hence, without loss of generality, we can assume that κ is the first sequence in Π_A. But now we can iterate forward indefinitely, and we will reach every equivalent-to-zero sequence by the Odometer Principle. Eventually, we reach the 0 sequence and get the same contradiction as above. If κ is such that $k_i = d_i$ for large i, we run the same argument backward. □

Corollary 23.16 $U_A \cap I \subset C_A^{\#}$.

Proof: Corollary 23.14 says that $U_A \cap I \subset C_A$. Since all orbits of U_A are well defined, Lemma 23.15 implies that

$$U_A \cap (C_A - C_A^{\#}) = \emptyset.$$

Our result follows immediately. □

Lemma 23.17 *No point of $C_A^{\#}$ has a first coordinate in $2\mathbf{Z}[A]$.*

Proof: Let $\{A_n\}$ be the superior sequence approximating A. We assume that $A_n < A$ infinitely often. The other case has the same treatment. Suppose that

$$\alpha = (2MA + 2N, -1) \in C_A^{\#}. \tag{23.20}$$

By Equation 21.7, the set $C_A^{\#}$ is invariant under the map

$$(x, -1) \to (2 - x, -1).$$

Indeed, the twist automorphism of Π induces this map on C_A. From this symmetry, we can assume that $M > 0$.

Let $P\Gamma_k$ denote the pivot arc. Suppose, for the sake of contradiction, that (M, N) is a vertex of $P\Gamma_k$ for some k. Then $2AM + 2N$ is a finite sum of terms

$$\lambda_j = |2Aq_j - p_j|, \tag{23.21}$$

by Theorem 1.8. But such points all lie in $C_A - C_A^{\#}$. This contradiction shows that (M, N) is not a vertex of $P\Gamma_k$ for any k.

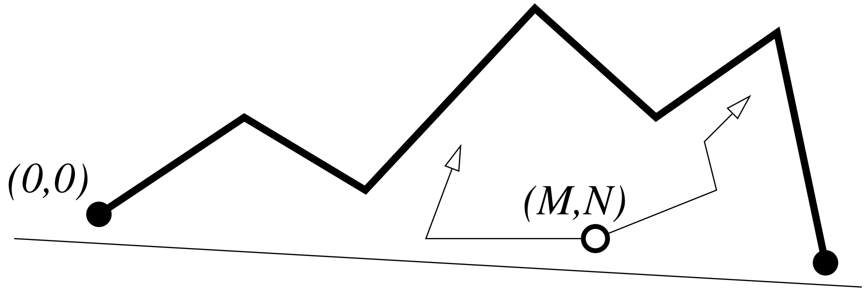

Figure 23.1: One arc traps another.

Let $P\Gamma_k^+$ denote the portion of $P\Gamma_k$ that moves rightward from $(0,0)$. We define Γ_k similarly. From the definition of the pivot points, the length of $P\Gamma_k^+$ tends to ∞ with k. Hence $\{\Gamma_k^+\}$ and $\{P\Gamma_k^+\}$ have the same Hausdorff limit. We can choose k large enough so that $P\Gamma_k^+$ contains a low vertex (M', N') to the right of (M, N). So, $P\Gamma_k^+$ connects $(0,0)$ to (M', N') and skips right over (M, N). See Figure 23.1.

Since $\alpha \in C_A^\#$, we can find a sequence of points $\{\alpha_n\} \in C_A^\# - \mathbf{Z}[A]$ such that the first coordinate of $\alpha_n - \alpha$ is positive. Let $\zeta_n = \alpha_n - \alpha$. Note that

$$\zeta_n \notin 2\mathbf{Z}[A]. \tag{23.22}$$

Let $\widehat{\Gamma}(\zeta_n, A)$ be the whole arithmetic graph corresponding to ζ_n. Let

$$\gamma_n = \Gamma(\zeta_n, A) \tag{23.23}$$

be the component containing $(0,0)$. By the Rigidity Lemma, the sequences

$$\{\Gamma(\zeta_n, A_n)\}, \qquad \{\Gamma_n\}$$

have the same Hausdorff limit. Hence $P\Gamma_k^+ \subset \gamma_n$ once n is large. In particular, some arc of γ_n connects $(0,0)$ to (M', N') and skips over (M, N). Call this the *barrier arc*.

Since $\alpha_n - \zeta_n = \alpha \in 2\mathbf{Z}[A]$, there is another component $\beta_n \subset \widehat{\Gamma}(\zeta_n)$ that tracks the orbit of α_n. One of the vertices of β_n is exactly (M, N). The component β_n is unbounded in both directions because all defined orbits in $C_A^\#$ are unbounded. On the other hand, β_n is trapped beneath the barrier arc. It cannot escape out either end, and it cannot intersect the barrier arc, by the Embedding Theorem. But then β_n cannot be unbounded in either direction. This is a contradiction. $\square$

Now we observe 3 facts.

- Corollary 23.16 says that $U_A \cap I \subset C_A^\#$.
- Lemma 23.17 shows that $C_A^\# \subset C_A'$.
- Theorem 23.7 shows that $C_A' \subset U_A \cap I$.

Putting these facts together gives Equation 23.8.

Remark: Lemma 23.17 is a purely number-theoretic statement and ought to have a number-theoretic proof. We do not know one, however.

23.8 STATEMENT 4

We have already established the first part of statement 4 of the Comet Theorem. Now we prove the second part.

By statements 1 and 2 of the Comet Theorem, it suffices to consider pairs of points in $C_A^\#$. It follows immediately from Equation 2.1 that two points of $C_A^\#$ lie on the same orbit only if their first coordinates differ by an element of $2\mathbf{Z}[A]$. Our goal is to prove the converse.

Lemma 23.18 *All but at most* 2 *orbits in* $C_A^\#$ *are erratic.*

Proof: By Theorem 23.7, Lemma 23.11, and the backward analog of Lemma 23.11, all orbits in $C_A^\#$ are erratic except those corresponding to the eqivalent-to-first sequences and the equivalent-to-last sequences. By the Odometer Principle, all the points in $C_A^\#$ corresponding to equivalent-to-first sequences lie on the same orbit. Likewise, all the points in $C_A^\#$ corresponding to equivalent-to-last sequences lie on the same orbit. These two orbits are the only ones that can fail to be erratic. □

Lemma 23.19 *Suppose that two points in* $C_A^\#$ *have first coordinates that differ by* $2\mathbf{Z}[A]$. *Suppose also that at least one of the points has an erratic orbit. Then the two points lie on the same orbit.*

Proof: One direction follows immediately from Equation 2.1. For the converse, suppose that the two points have first coordinates that differ by $2\mathbf{Z}[A]$. The first coordinates of the points do not lie in $2\mathbf{Z}[A]$, by Lemma 23.17. Hence one and the same arithmetic graph $\widehat{\Gamma}$ contains components γ_1 and γ_2 that, respectively, track the two orbits.

Since both orbits are dense in $C_A^\#$, we know that both are erratic in at least one direction. Suppose first that γ_1 is erratic in both directions. Since γ_2 is erratic in one direction, we can find a low vertex v of γ_1 that is not a vertex of γ_2. Since γ_2 is erratic in both directions, we can find vertices w_1 and w_2 of γ_1 lying to the left and to the right of v, respectively. But then the arc of γ_1 starting at v is trapped beneath the arc of γ_2 connecting w_1 to w_2. This contradicts the Embedding Theorem. In short, $\widehat{\Gamma}$ is not big enough to contain both components. □

It remains only to deal with the case when both points lie on orbits that are only erratic in only one direction..

Lemma 23.20 *Suppose that two points in* $C_A^\#$ *have first coordinates that differ by* $2\mathbf{Z}[A]$. *Suppose also that neither point lies on an erratic orbit. Then the two points lie on the same orbit.*

Proof: Let $\alpha \in C_A^\#$ (respectively, β) be the unique point such that the forward (respectively, backward) first return map to $C_A^\#$ at α (respectively, β) does not exist. There are exactly 2 one-sided erratic orbits. α is one orbit, and β is on the other.

It suffices to prove that $\alpha - \beta \notin 2\mathbf{Z}[A] \times \{0\}$. We will suppose the contrary and derive a contradiction. Suppose that $\alpha - \beta = (2Am + 2n, 0)$ for some $(m, n) \in \mathbf{Z}^2$.

α is the last point in the twirl order, and β is the first point. In terms of sequences, α corresponds to the sequence $\{\tilde{d}_i\}$ and β corresponds to the sequence $\{\tilde{0}_i\}$. Let $\{\alpha_j\}$ be a sequence of points in $C_A^\#$ converging to α, chosen so that the corresponding orbit is erratic. Define

$$\beta_j = \alpha_j + (\beta - \alpha). \tag{23.24}$$

Then

$$\alpha_j - \beta_j = (2Am + 2n, 0). \tag{23.25}$$

By the case we have already considered, β_j lies in the same orbit as α_j.

For j large, the sequence corresponding to α_j matches the terms of the sequence for α for many terms. Likewise, the sequence corresponding to β_j matches the terms of the sequence for β for many terms. Hence these two sequences disagree for many terms. Given that the return dynamics to $C_A^\#$ is conjugate to the odometer map on the sequence space, we have

$$2Am + 2n = \pi_1(\alpha_j - \beta_j) = \sum_{i=0}^{N_j} a_{ji}\lambda_i, \qquad |a_{ji}| \le d_i. \tag{23.26}$$

Here $N_j \to \infty$ as $j \to \infty$, and π_1 denotes projection onto the first coordinate.

Let M be the map from Equation 2.10. We have

$$M(m, n) = \sum_{i=0}^{N_j} b_{ji} M(V_i), \qquad |b_{ji}| \le d_i. \tag{23.27}$$

Here $b_{ji} = \pm a_{ji}$, depending on the sign of $A_i - A$. Since A is irrational, M is injective. Therefore, setting $N = N_j$ for ease of notation, we have

$$(m, n) = \sum_{i=0}^{N} b_{ji} V_i = b_{Ni} V_N + \sum_{i=0}^{N-1} b_{ji} V_i. \tag{23.28}$$

Looking at the second coordinates, we see that

$$q_N - \sum_{i=0}^{N-1} d_i q_i \le \left| b_{Ni} q_N - \sum_{i=0}^{N-1} b_{ji} q_i \right| = |n|. \tag{23.29}$$

However, it follows fairly easily from Equation 21.6 that the left hand side tends to ∞ as $N_j \to \infty$. This contradiction finishes the proof. $\square$

This page intentionally left blank

Chapter Twenty-Four

Dynamical Consequences

24.1 MINIMALITY

Here we prove Theorem 1.3. Statement 3 of this Theorem is contained in the Comet Theorem. We just have to prove statements 1 and 2.

Recall from the introduction that a set $S \subset \mathbf{R}^2$ is *locally homogeneous* if every two points of S have arbitrarily small neighborhoods that are translation equivalent. Note that the points themselves need not be in the same positions within these sets.

Statements 1 and 2 of Theorem 1.3 say, respectively, that U_A is dynamically minimal and locally homogeneous. Statement 3 of Theorem 1.3 is an immediate consequence of the Comet Theorem.

Proof of Statement 1: Since every orbit in U_A intersects $C_A^\#$, it suffices to prove that every point of $C_A^\#$ lies on an orbit that is forward dense in U_A, backward dense in U_A, or both.

Let $\zeta \in C_A^\#$ be the point. By the Comet Theorem, the orbit of ζ is forward dense in $C_A^\#$, backward dense in $C_A^\#$, or both. Assume that ζ lies on an orbit that is forward dense in $C_A^\#$. The case of backward-dense orbits requires a similar treatment.

Let $\beta \in U_A$ be some other point. Some point $\alpha \in C_A^\#$ lies on the orbit of β. Hence $(\psi')^k(\alpha) = \beta$ for some k. Here ψ' is the outer billiards map. But $(\psi')^k$ is a piecewise isometry. Hence $(\psi')^k$ maps small intervals centered at α isometrically to small intervals centered at β. The forward orbit of ζ enters any interval about α infinitely often. Hence the forward orbit of ζ enters every interval about β infinitely often. □

Proof of Statement 2: For any $p \in U_A$, there is some integer k such that

$$(\psi')^k(p) \in C_A^\#.$$

Here ψ' is the outer billiards map. But ψ^k is a local isometry. Hence there are arbitrarily small neighborhoods of p that are isometric to neighborhoods of points in $C_A^\#$. For this reason, it suffices to prove that $C_A^\#$ is locally homogeneous. This is a purely geometric problem.

Let $\{d_k\}$ denote the renormalization sequence. The set C_A breaks into $d_0 + 1$ isometric copies of a smaller Cantor set. Each of these breaks into $d_1 + 1$ isometric copies of still smaller Cantor sets. And so on. From this we see that both C_A and $C_A^\#$ are locally homogeneous. □

24.2 TREE INTERPRETATION OF THE DYNAMICS

Let A be an irrational kite parameter. We can illustrate the return dynamics to $C_A^\#$ using infinite trees. The main point here is that the dynamics is conjugate to an odometer. The conjugacy is given by the map $\phi\colon \mathcal{Z}_A \to C_A$ from the Comet Theorem. Our figures encode the structure of ϕ graphically.

We think of C_A as the ends of a tree T_A. We label T_A according to the sequence of signs $\{A - A_n\}$. Since $A - A_0$ is negative, we label the level 1 vertices 0, ..., d_0 from right to left. Each level 1 vertex has d_1 downward vertices. We label all these vertices from left to right if $A - A_1 > 0$ and from right to left if $A - A_1$ is negative. And so on. This idea of switching left and right according to the sign of $A - A_k$ corresponds precisely to our method of identification in Equations 1.8 and 1.9. Figure 24.1 shows the example for the renormalization sequence $\{1, 3, 2\}$ and the sign sequence $-, +, -$.

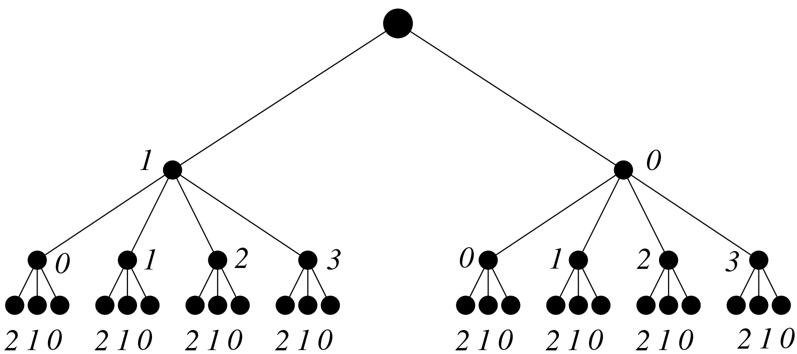

Figure 24.1: Tree labelling.

We have the return map

$$\rho_A\colon C_A^\# - \phi(-1) \to C_A^\# - \phi(-1),$$

and this map is conjugate to the restriction of the odometer on $\mathcal{Z}_A$. Accordingly, we can extend ρ_A to all of C_A even though the extension no longer describes outer billiards dynamics on the extra points. Nonetheless, it is convenient to have this extension.

To see what ρ_A does, we write this code for a given end. We write the code "backward" so that the topmost level of the tree corresponds to the rightmost digit, and so on. So, the sequences trail off to the left. For the odometer, we add 1, carrying to the right. For instance, we have

$$(\ldots 000) \to (\ldots 001) \to (\ldots 010), \qquad (\ldots 031) \to (\ldots 100)$$

Every time many of the initial digits in the odometer turn over, the corresponding orbit makes a large excursion before it returns. We will formalize this below.

24.3 PROPER RETURN MODELS AND CUSPED SOLENOIDS

24.3.1 Proper Models

Here we will describe the sense in which the Comet Theorem allows us to combinatorially model the dynamics on U_A, the set of unbounded special orbits. The results in this section are really just a repackaging of some of the statements of the Comet Theorem.

Let X be an unbounded metric space and let $f: X \to X$ be a bijection. We assume that f^2 moves points by a small amount. That is, there is a universal constant C such that

$$d(x, f^2(x)) < C, \qquad \forall x \in X. \tag{24.1}$$

The example we have in mind, of course, is the outer billiards map

$$\psi': U_A \to U_A. \tag{24.2}$$

The square map ψ moves points by at most 4 units.

Say that a compact subset $X_0 \subset X$ is a *proper section* for f if for every N there is some N' such that $d(x, X_0) < N$ implies that $f^k(x) \in X_0$ for some $|k| < N'$. In particular, every orbit of f intersects X_0. This condition is just the abstract version of statement 1 of the Comet Theorem. Informally, all the orbits head either directly toward X_0 or directly away from X_0.

Let $f_0: X_0 \to X_0$ be the first return map. This is a slight abuse of notation because f_0 might not be defined on all points of X_0. Some points might exit X_0 and never return. We define two functions

$$e_1, e_2: X_0 \to \mathbf{R}_+ \cup \infty. \tag{24.3}$$

The function $e_1(x)$ is the maximum distance the forward orbit of x gets away from X_0 before returning as $f_0(x)$. The function $e_2(x)$ is the length of this same portion of the orbit. If f_0 is not defined on x, then obviously $e_2(x) = \infty$. The proper section condition guarantees that $e_1(x) = \infty$ as well.

The condition that X_0 is a proper section guarantees that e_1 and e_2 are proper functions of each other. That is, if $\{x_n\}$ is a sequence of points in X_0, then $e_1(x) \to \infty$ if and only if $e_2(x) \to \infty$. This observation includes the statement that $e_1(x) = \infty$ iff $e_2(x) = \infty$ iff f_0 is not defined on x_0. For the purpose of getting a rough qualitative picture of the orbits, we consider just the function e_1. We set $e = e_1$ and call e the *excursion function*.

Suppose now that $f': X' \to X'$ is another bijection and X_0' is a proper section. Let $e': X_0' \to \mathbf{R} + \cup\infty$ denote the excursion function for this system. We say that (X, X_0, f) is *properly equivalent* to (X', X_0', f') if there is a homeomorphism $\phi: X \to X'$ such that

- ϕ conjugates f_0 to f_0'.
- $e' \circ \phi$ and e are proper functions of each other on X_0.

These conditions guarantee that ϕ carries the points where f_0 is not defined to the points where f_0' is not defined.

The notion of proper equivalence turns out to be a little too strong for our purposes. We say that (X, X_0, f) and (X', X_0', f') are *essentially properly equivalent* if ϕ has all the above properties but is defined only on the complement of a finite number of orbits of X_0. In this case, the inverse map has the same property: It will be well defined on all but a finite number of orbits of X_0'. In other words, an essential proper equivalence is a proper equivalence, provided that we first delete a finite number of orbits from the spaces. We call (X, X_0, f) an *essentially proper model* for (X', X_0', f').

24.3.2 The Cusped Solenoid

Statement 1 of the Comet Theorem says that $C_A^\#$ is a proper section for the map in Equation 24.2. Now we can describe the proper models for the triple $(U_A, C_A^\#, \psi')$. Statements 2 and 3 in particular describe the excursion function up to a bi-Lipschitz constant. Here we convert this information into a concrete essentially proper model for this dymamics.

Let $\mathcal{Z}_A$ denote the metric Abelian group from the Comet Theorem. For convenience, we recall the definition of the metric d here. $d(x, y) = q_{n-1}^{-1}$, where n is the smallest index such that $[x]$ and $[y]$ disagree in $\mathbf{Z}/D_n$. Here $\{p_n/q_n\}$ is the superior sequence approximating A.

We denote the odometer map on $\mathcal{Z}_A$ by f_0. That is, $f_0(x) = x+1$. Topologically, the *solenoid* based on $\mathcal{Z}_A$ is defined as the mapping cylinder

$$\mathcal{S}_A = \mathcal{Z}_A \times [0, 1]/\sim, \qquad (x, 1) \sim (x + 1, 0). \tag{24.4}$$

This is a compact metric space.

We now modify this space a bit. First of all, we remove the point

$$(-1, 1/2)$$

from $\mathcal{S}_A$. This deleted point, the cusp, lies halfway between $(-1, 0)$ and $(0, 0)$. We now change the metric on the space by declaring the length of the segment between $(x, 0)$ and $(x, 1)$ to be

$$\frac{1}{d(x, -1)}$$

Metrically, we simply rescale the length element on each interval by the appropriate amounts. We call the resulting space $\mathcal{C}_A$, the *cusped solenoid* based on A.

24.3.3 The Main Results

We define $f : \mathcal{C}_A \to \mathcal{C}_A$ to be the map such that

$$f(x, t) = \left(x, \frac{t}{d(x, -1)}\right). \tag{24.5}$$

From the way we have scaled the distances, f maps each point by 1 unit. Indeed, some readers will recognize f as the time-one map of the geodesic flow on $\mathcal{C}_A$. The original set $\mathcal{Z}_A$ is a proper section for the map, and the return map is precisely f_0. Put another way, f is a suspension flow over f. Note that f also depends on A, but we suppress this from our notation.

Theorem 24.1 *$(\mathcal{C}_A, \mathcal{Z}_A, f)$ is an essentially proper model for $(U_A, C_A^\#, \psi')$.*

Proof: This is just a repackaging (and weakening) of statements 2 and 3 of the Comet Theorem. □

Remarks:
(i) The model forgets the linear ordering on $C_A^\#$ that comes from its inclusion in I, but one can recover this from the discussion in §24.2.
(ii) In a certain sense, the triple $(\mathcal{C}_A, \mathcal{Z}_A, f)$ provides a *bi-Lipschitz model* for the nature of the unboundedness of the orbits in U_A. However, it would be misleading to call the model an actual bi-Lipschitz model for the dynamics on U_A because we are not saying much about what happens to the orbits in the two systems after they leave their proper section. For instance, the excursion times could be wildly different from each other even though they are proper functions of each other.

The following result contains statements 1 and 2 of Theorem 1.4.

Theorem 24.2 *The time-one map of the geodesic flow on any cusped solenoid serves as an essentially proper model for the dynamics of the special unbounded orbits relative to uncountably many different parameters.*

Proof: Up to a proper change of the excursion function, the model depends on only the renormalization sequence, and there are uncountably many parameters realizing any renormalization sequence. □

24.3.4 Equivalence and Universality

To each parameter A, we associate the renormalization sequence $\{d_n\}$. We then associate the sequence $\{D_n\}$, where

$$D_n = \prod_{i=0}^{n-1}(d_i + 1). \tag{24.6}$$

We call A and A' *broadly equivalent* iff for each m there is some n such that D_m divides D'_n and D'_m divides D_n. Each broad equivalence class has uncountably many members.

Lemma 24.3 *If A and A' are broadly equivalent, then there is a homeomorphism from $\mathcal{Z}_A$ to $\mathcal{Z}_{A'}$ that conjugates one odometer to the other.*

Proof: Each element of $\mathcal{Z}_A$ is a compatible sequence $\{a_m\}$ with $a_m \in \mathbf{Z}/D_m$. Using the divisibility relation, this element determines a corresponding sequence $\{a'_m\}$. Here a'_m is the image of a_n under the factor map $\mathbf{Z}/D_n \to \mathbf{Z}/D'_m$, where n is such that D'_m divides D_n. One can easily check that this map is well defined and determines the desired homeomorphism. □

Theorem 24.4 *If A and B are broadly equivalent, then there is an essentially proper equivalence between $(U_A, C_A^\#, \psi'_A)$ and $(U_B, C_B^\#, \psi'_B)$. In particular, the return maps to C_A and C_B are essentially conjugate.*

Proof: The homeomorphism from $\mathcal{Z}_A$ to $\mathcal{Z}_B$ maps -1 to -1. By construction, this homeomorphism sets up a proper equivalence between $(\mathcal{C}_A, \mathcal{Z}_A, f_A)$ and $(\mathcal{C}_B, \mathcal{Z}_B, f_B)$. This result now follows from Theorem 24.1. □

One might wonder about the nature of the topological equivalence between the return maps to $C_A^\#$ and $C_B^\#$. One can reconstruct the conjugacy from the tree labellings given in §24.2. The conjugacy is well defined for all points of C_A and C_B, but we typically have to ignore the countable sets of points on which the relevant return maps are not defined. This accounts for the precise statement of the theorem above.

Let $\mathcal{Z}$ denote the inverse limit over all finite cyclic groups. The map $x \to x+1$ is defined on $\mathcal{Z}$. This dynamical system is called the *universal odometer*. Sometimes $\mathcal{Z}$ is called the *profinite completion* of $\mathbf{Z}$.

We call A *universal* if every $k \in \mathbf{N}$ divides some D_n in the sequence. If A is universal, then there is a group isomorphism from $\mathcal{Z}$ to $\mathcal{Z}_A$ that respects the odometer maps. In short, when A is universal, $\mathcal{Z}_A$ is the universal odometer. See [**H**, §5] for a proof of this fact – stated in slightly different terms – and for a detailed discussion of the universal odometer.

Lemma 24.5 *Almost every parameter is universal.*

Proof: A sufficient condition for a parameter to be universal is that every integer appears in the renormalization sequence. We can express the fact that a certain number appears in the renormalization sequence as a statement that a certain combination appears in the continued fraction expansion of A. Geometrically, as one drops a geodesic down from ∞ to A, the appearance of a certain pattern of geodesics in the Farey graph forces a certain number in the renormalization sequence. As is well known, the continued fraction expansion for almost every number in $(0, 1)$ contains every finite string of digits. □

Statement 3 of Theorem 1.4 is contained in the following result.

Theorem 24.6 *For almost every $A \in (0, 1)$, the triple $(U_A, C_A^\#, \psi')$ is properly modelled by the time-one map of the geodesic flow on the universal cusped solenoid.*

Proof: This result is an immediate consequence of the previous result and Theorem 24.1. □

Remark: One might wonder if there is a concrete parameter that exhibits this universal behavior. It seems that the parameter $A = e - 2$ has the following inferior sequence.

$$\frac{1}{1} \leftarrow \frac{5}{7} \leftarrow \frac{51}{71} \leftarrow \frac{719}{1001} \cdots, \qquad r_{n+2} = (4n+10)r_{n+1} + r_n, \qquad n \geq 0.$$

One can easily check that this sequence leads to the universal odometer. Thus the fractional part of e has universal behavior.

24.4 SOME OTHER EQUIVALENCE RELATIONS

Call A and B *narrowly equivalent* if they have the same renormalization sequence and if the sign of $A - A_j$ is the same as the sign of $B - B_j$ for all j. Here $\{A_j\}$ and $\{B_j\}$ are the superior sequences approximating A and B, respectively. Referring to Equation 1.9, the definition of $\tilde{k}_j$ relative to the narrowly equivalent parameters is the same for every index. Each narrow equivalence class is uncountable.

Theorem 24.7 *If A and B are narrowly equivalent, then there is an order-preserving homeomorphism from I to itself that conjugates the return map on $C_A^\#$ to the return map on $C_B^\#$. This map is a proper equivalence from $(U_A, C_A^\#, \psi'_A)$ to $(U_B, C_B^\#, \psi'_B)$.*

Proof: The two spaces Π_A and Π_B are exactly the same, and the extended twirl orders on the (equivalence classes) of these spaces are the same. Thus the successor maps on the two spaces are identical. The map $h = \phi'_2 \circ \phi_2^{-1}$ is a homeomorphism from C_A to C_B that carries $C_A^\#$ and $C_B^\#$ and conjugates one return dynamics to the other. By construction, h preserves the linear ordering on I, and we can extend h to the gaps of $I - C_A$ in the obvious way. By construction, this map carries $\phi_A(-1)$ to $\phi_B(-1)$ and is continuous. Hence it is a proper equivalence. $\square$

The *first renormalization* of the odometer map $x \to x+1$ on the inverse system

$$\ldots \to \mathbf{Z}/D_3 \to \mathbf{Z}/D_2 \to \mathbf{Z}/D_1 \tag{24.7}$$

is the D_1th power of the map. This corresponds to the map $x \to x+1$ on the inverse system

$$\ldots \to \mathbf{Z}/D'_3 \to \mathbf{Z}/D'_2 \to \mathbf{Z}/D'_1, \qquad D'_n = D_{n+1}/D_1. \tag{24.8}$$

As in the Comet Theorem, each D_n divides D_{n+1} for all n, so the construction makes sense. In terms of the symbolic dynamics on the sequence space Π, the renormalization consists of the first return map to the subspace

$$\Pi' = \{\kappa \in \Pi |\ k_0 = 0\}. \tag{24.9}$$

In terms of the dynamics on C_A, the first renormalization is the first return map to the Cantor subset corresponding to Π'. The *second renormalization* is the first renormalization of the first renormalization. And so on.

Let Γ denote the $(2, \infty, \infty)$-triangle group from the Comet Theorem. Given the construction of the inferior sequence, the enhanced renormalization sequences for two Γ-equivalent parameters have the same tail ends. Thus the tail ends of the renormalization sequences are the same, and the tail ends of the sign sequences are the same. This gives us the following result.

Corollary 24.8 *Suppose that A and B are equivalent under Γ_2. Then some renormalization of the return map to $C_A^\#$ is conjugate to some renormalization of the return map to $C_B^\#$. The conjugacy is given by an order-preserving homeomorphism.*

Remark: The homeomorphism mentioned in the last corollary is a similary. Compare Statement 2 of Theorem 1.5 and see §25.3 for more details.

Chapter Twenty-Five

Geometric Consequences

25.1 PERIODIC ORBITS

Here we prove statement 1 of Theorem 1.5.

Lemma 25.1 *U_A has length* 0.

Proof: Since U_A is locally homogeneous, it suffices to prove that C_A has length 0. Let $\lambda_n = |Aq_n - p_n|$, as in Equation 21.5. We define

$$G_n = \sum_{k=n+1}^{\infty} 2\lambda_k d_k. \tag{25.1}$$

Then

$$C_A \subset \sum_{\kappa \in \Pi_n} \Big(I_n + X(\kappa)\Big). \tag{25.2}$$

Here I_n is the interval with endpoints $(0, 1)$ and $(G_n, 1)$. In other words, C_A is contained in D_n translates of an interval of length G_n. We just need to prove that $D_n G_n \to 0$. It suffices to prove this when n is even. By Equation 21.6,

$$D_n < \epsilon^{-n} q_n, \qquad \epsilon = \sqrt{5/4}. \tag{25.3}$$

By Equation 21.5 we have

$$G_n < 2 \sum_{k=n+1}^{\infty} q_k^{-1} < 2q_n^{-1} \sum_{k=1}^{\infty} 2^{-k} < 2q_n^{-1}. \tag{25.4}$$

Here we have used the trivial bound that $q_m/q_n < 2^{n-m}$ when $m > n$. Therefore

$$D_n G_n < 2\epsilon^{-n}. \tag{25.5}$$

This completes the proof. □

Theorem 25.2 *Any defined orbit in $I - C_A$ is periodic. There is a uniform bound on the period depending only on the distance from the point to C_A.*

Proof: The Comet Theorem combines with statement 2 of Theorem 1.1 to prove that any defined orbit in $I - C_A$ is periodic. The period bound comes from taking a limit of the Period Theorem as $n \to \infty$ in the rational approximating sequence. In other words, if this result were false, then we could contradict the Period Theorem using the Continuity Principle. □

Combining these results, we have statement 1 of Theorem 1.5: Almost every point of $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ lies on a periodic orbit.

25.2 A TRIANGLE GROUP

Let $\mathbf{H}^2$ denote the upper half-plane model of the hyperbolic plane. Let $\Gamma \subset \mathrm{Isom}(\mathbf{H}^2)$ denote the $(2, \infty, \infty)$-reflection triangle group generated by reflections in the sides of the geodesic triangle with vertices $(0, 1, i)$. Figure 25.1 shows this triangle. See [**Be**] for details.

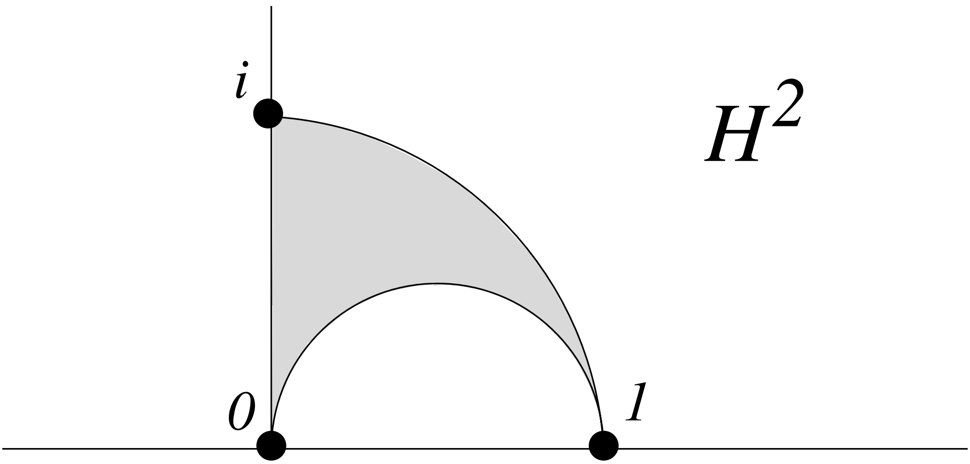

Figure 25.1: The geodesic triangle with vertices $(0, 1, i)$.

Γ is the largest subgroup of $\mathrm{Isom}(\mathbf{H}^2)$ with the following 3 properties.

1. Γ preserves the Farey graph.

2. Γ permutes the odd rationals and also the even rationals.

3. Every element $T \in \Gamma$ acts on $\mathbf{R} \cup \infty$ via an equation of the form

$$T(x) = \frac{ax+b}{cx+d}, \qquad |ad - bc| = 1. \tag{25.6}$$

   On the upper half-plane, T acts either as a linear fractional transformation or as the composition of a linear fractional transformation with complex conjugation. This depends on the sign of $ad - bc$.

These properties guarantee that elements of Γ are well adapted to the construction of the inferior and superior sequences. See §17.1.

Remark: It seems worthwhile to mention the connection between Γ and other familiar groups. Γ contains the ideal triangle group generated by reflections in the sides of the ideal geodesic triangle with vertices $(0, 1, \infty)$. The ideal triangle group in turn contains $P\Gamma_2$, where $\Gamma_2 \subset SL_2(\mathbf{Z})$ is the *level 2 congruence subgroup* consisting of matrices congruent to the identity mod 2. Here P means that we take these matrices mod $\pm I$. Finally, Γ is *commensurable* with the modular group $PSL_2(\mathbf{Z})$.

25.3 MODULARITY

Here we will prove statement 2 of Theorem 1.5. Let λ_k and λ'_k be the quantities associated to parameters A and A', as in Equation 1.10

Lemma 25.3 *Let $T \in \Gamma$ be such that $T(A) = A'$. Then there is some integer m such that*

$$d'_k = d_{k+m}, \qquad \frac{\lambda'_k}{\lambda_{k+m}} = |T'(A)|^{1/2}$$

provided that k is sufficiently large.

Proof: Here $T'(A) = (cA+d)^{-2}$ when T is as in Equation 25.6. Given our construction of the inferior sequence, and the first two properties of Γ, we have $T(A_{k+m}) = A'_k$ for some m and all sufficiently large k. Hence $d'_k = d_{k+m}$ for these choices of k and m. We compute

$$\lambda'_k = |A'q'_k - p'_k|, \qquad A' = \frac{aA+b}{cA+d}, \qquad T\left(\frac{p}{q}\right) = \frac{ap+bq}{cp+dq}. \quad (25.7)$$

An exercise in modular arithmetic shows that the fraction on the right is in lowest terms. Hence

$$p'_k = aq_{k+m} + bp_{k+m}, \qquad q'_k = cq_{k+m} + dp_{k+m}. \quad (25.8)$$

Combining the last two equations, we have

$$\lambda'_k = \left|\frac{(aA+b)(cq_{k+m}+dp_{k+m}) - (cA+d)(aq_{k+m}+bp_{k+m})}{cA+d}\right|$$

$$= \left|\frac{(ad-bc)(Aq_{k+m}-p_{k+m})}{cA+d}\right| = \left|\frac{Aq_{k+m}-p_{k+m}}{cA+d}\right| = \frac{\lambda_{k+m}}{cA+d} = \lambda_{k+m}|T'(A)|^{1/2}.$$

This completes the proof. $\square$

Recall that C_A is defined by the formula

$$C_A = \bigcup_{\kappa\in\Pi} (X(\kappa), 1), \qquad X(\kappa) = \sum_{i=0}^{\infty} 2k_i\lambda_i, \qquad \lambda_i = |Aq_i - p_i|. \quad (25.9)$$

If A and A' are Γ-equivalent, as above, then we have the obvious map

$$\sum_{i=k_0}^{\infty} k_{m+i}\lambda_{m+i} \to \sum_{i=k_0}^{\infty} k_i\lambda'_i. \quad (25.10)$$

By the previous result, this map is well defined if k_0 is large enough. Also by the previous result, this map is a similarity. Hence $C_A^\#$ and $C_{A'}^\#$ are locally similar. Hence U_A and $U_{A'}$ are locally similar.

25.4 HAUSDORFF DIMENSION

In this section, we review some basic properties of the Hausdorff dimension. See [**F**] for more details.

We will work with sets in $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$ and especially sets in our favorite interval $I = [0, 2] \times \{-1\}$. Given an interval J, let $|J|$ denote its length. Given $S \subset \mathbf{R} \times \mathbf{Z}_{\text{odd}}$ and some $\delta > 0$, we define

$$\mu(S, s, \delta) = \inf \sum |J_n|^s. \tag{25.11}$$

The infimum is taken over all countable covers of S by intervals $\{J_n\}$ such that $\text{diam}(J_n) < \delta$. Next, we define

$$\mu(S, s) = \lim_{\delta \to 0} \mu(S, s, \delta) \in [0, \infty]\,. \tag{25.12}$$

This limit exists because $\mu(S, s, \delta)$ is a monotone function of δ. Note that $\mu(S, 1) < \infty$ because I has finite total length. Finally,

$$\dim(S) = \inf\{s|\ \mu(S, s) < \infty\}. \tag{25.13}$$

The number $\dim(S)$ is called the *Hausdorff dimension* of S.

Given an explicit family of covers, as we have constructed in the proof of Lemma 25.1, it is easy for us to compute upper bounds on the Hausdorff dimension. Here we recall a method for getting lower bounds on the Hausdorff dimension. Let $S \subset I$ be a compact set. We say that $f\colon I \to \mathbf{R}$ is a ρ*-density* relative to S if f is monotone nondecreasing and constant on the complementary intervals of S and

$$C|a - b|^{\rho} \geq f(b) - f(a) \tag{25.14}$$

for some $C > 0$ and all intervals $[a, b] \subset I$ such that $|a - b|$ is sufficiently small.

Lemma 25.4 *If S admits a ρ-density, then* $\dim(S) \geq \rho$.

Proof: This is essentially the Mass Distribution Principle 4.2 in [**F**, p. 55]. The function f is the integral of the mass distribution described in connection with this principle. $\square$

In computing the function $u(A) = \dim(U_A)$, we would prefer to work with the sets C_A. The following lemma justifies this.

Lemma 25.5 *U_A and C_A and $C_A^{\#}$ all have the same dimension.*

Proof: Since $C_A - C_A^{\#}$ is countable, we have $\dim(C_A) = \dim(C_A^{\#})$. Since U_A is locally homogeneous $\dim(U_A) = \dim(U_A \cap J)$ for any interval J about a point in U_A. In particular, $\dim(U_A) = \dim(C_A^{\#})$. $\square$

25.5 QUADRATIC IRRATIONAL PARAMETERS

25.5.1 Self Similarity

First, we prove statement 3 of Theorem 1.5.

A Cantor set is commonly called *self-similar* if it is a finite union of similar copies of itself.

Lemma 25.6 *Suppose that $A \in (0, 1)$ is a quadratic irrational. Then C_A is a finite disjoint union of self-similar Cantor sets.*

Proof: A has an eventually periodic continued fraction expansion. Hence A is the fixed point of some infinite-order element $T \in SL_2(\mathbf{R})$, acting by linear fractional transformations. But some power of T lies in the group Γ. Hence, without loss of generality, we can take $T \in \Gamma$. But then the map from Equation 2.9 carries one clopen subset V_2 of C_A to a larger clopen subset V_1. (Here *clopen* means simultaneously closed and open.) Looking at Equation 2.9 and recalling the definition of C_A from Equation 1.11, we see that C_A is a finite disjoint union of translates of V_1, and V_1 is a finite disjoint union of translates of V_2. Hence V_1 is a finite disjoint union of similar copies of itself. Hence C_A is a finite union of translates of V_1, each of which is a self-similar Cantor set. □

A self-similar Cantor set has the property that every point in it has arbitrarily small neighborhoods that are also self-similar Cantor sets. Statement 2 of Theorem 1.3 says that any point of U_A has a neighborhood that is isometric to a neighborhood in $C_A^{\#}$. Shrinking this neighborhood appropriately, we get a self-similar trimmed Cantor set surrounding the point in U_A. This proves statement 3 of Theorem 1.5.

25.5.2 Dimension Formula

Now we present a dimension formula in the quadratic irrational case. Actually, the formula is slightly more general. Let $A \in (0, 1)$ be irrational. Let $\{p_n/q_n\}$ be the associated superior sequence and let $\{d_n\}$ be the renormalization sequence. We call A tame if

1. $q_{n+1} < Cq_n$ for some constant C that is independent of n.
2. The following limits exist.
$$D(A) = \lim_{n\to\infty} \frac{\log(D_n)}{n}, \qquad Q(A) = \lim_{n\to\infty} \frac{\log(q_n)}{n}.$$

There are uncountably many tame parameters. In particular, we have the following result.

Lemma 25.7 *Quadratic irrational parameters are tame.*

Proof: Let A be a quadratic irrational parameter. From the work in §25.3, we see that the renormalization sequence $\{d_k\}$ is eventually periodic. But this implies that the limit $D(A)$ exists. At the same time, we have integers c, d, n such that $q_{k+n} = cq_k + d$ for all sufficiently large k. This easily implies that $Q(A)$ exists and that q_{k+1}/q_k is uniformly bounded. □

Lemma 25.8 *Suppose A is a tame parameter. Let $\{p_n/q_n\}$ be the associated superior sequence. Then $\lambda_n \in [C_1, C_2]\, q_n^{-1}$ for positive constants C_1, C_2.*

Proof: For tame parameters, the renormalization sequence $\{d_n\}$ is bounded. We have

$$\lambda_n = q_n|A - A_n| < 2d_n^{-1}q_n^{-1} < C_2 q_n^{-1},$$

by Lemma 17.4. For the lower bound, note first that $\lambda_{n+1} < \lambda'_{n+1} < \lambda_n$, by Equation 21.7. By the triangle inequality,

$$|A - A_n| + |A - A_{n+1}| \geq |A_n - A_{n+1}| \geq \frac{2}{q_n q_{n+1}}.$$

Hence

$$2\lambda_n > \lambda_n + \lambda_{n+1} = q_n|A - A_n| + q_{n+1}|A - A_{n+1}|$$

$$> q_n\Big(|A - A_n| + |A - A_{n+1}|\Big) \geq 2q_{n+1}^{-1} \geq 2C_1 q_n^{-1}.$$

This gives the lower bound. □

Here is the main result for this section.

Theorem 25.9 *If A is a tame parameter then $u(A) = D(A)/Q(A)$.*

Proof: Let $\mathcal{C}_n$ be the covering we constructed in the proof of Lemma 25.1. The intervals in $\mathcal{C}_n$ are pairwise disjoint and have the same length. Each interval of $\mathcal{C}_n$ contains $(d_n + 1)$ evenly and maximally spaced intervals of $\mathcal{C}_{n+1}$.

We first use these covers to get an upper bound on $\dim(U_A)$. There are D_n intervals in $\mathcal{C}_n$, all having length G_n. Choose any $\epsilon > 0$. For n large,

$$D_n \in \bigg(\exp\Big(n(D - \epsilon)\Big), \exp\Big(n(D + \epsilon)\Big)\bigg). \tag{25.15}$$

We have

$$G_n = 2\lambda_{n+1}^* \in \big[2\lambda_{n+1}, \lambda_n\big] \in \big[C_1 q_{n+1}^{-1}, C_2 q_n^{-1}\big] \in [C_3, C_2]\, q_n^{-1}, \tag{25.16}$$

by the preceding lemma. Hence

$$G_n \in \bigg(\exp\Big(-n(Q + \epsilon)\Big), \exp\Big(-n(Q - \epsilon)\Big)\bigg). \tag{25.17}$$

Setting $s = (D + \epsilon)/(Q - \epsilon)$ and letting $n \to \infty$, we have $\mu(U_A, s) \leq 1$. Hence $\dim(U_A) \leq s$. But ϵ is arbitrary. Hence $\dim(U_A) \leq D/Q$.

For the lower bound, we set $\rho = (D - \epsilon)/(Q + \epsilon)$ and construct a ρ-density. Let $\mathcal{X}_n$ denote the partition of [0, 1] into D_n equally sized intervals. Going from left to right, we map the jth interval of $\mathcal{C}_n$ into the jth interval of $\mathcal{X}_n$. We map the gaps between consecutive intervals of $\mathcal{C}_n$ to the obvious points common to consecutive intervals of $\mathcal{X}_n$. The maps $\{f_n\}$ form a uniformly continuous family, and the limit $f: I \to [0,1]$ exists. By construction, f is monotone nondecreasing and constant on the components of $I - U_A$.

Consider $[a,b] \subset I$. By Equation 25.16, the sequence $\{G_n/G_{n+1}\}$ is uniformly bounded. Hence we can assume without loss of generality that $|a-b| = G_n$ for some n. By construction $[a,b]$ intersects at most 2 consecutive intervals of $\mathcal{C}_n$. Hence $f(b) - f(a) \leq 2D_n^{-1}$. Hence

$$\begin{aligned} 2|a-b|^\rho = 2G_n^\rho &\geq \\ 2\exp\Big(-\rho n\big(Q+\epsilon\big)\Big) &= \\ 2\exp\Big(-n\big(D-\epsilon\big)\Big) &> \\ 2D_n^{-1} &\geq \\ f(b) - f(a). \end{aligned} \tag{25.18}$$

This shows that f is a ρ-density relative to U_A. Hence $\dim(U_A) \geq \rho$. Again, ϵ is arbitrary, so $\dim(U_A) \geq D/Q$. $\square$

Example 1: Let $A = \sqrt{5} - 2 = \phi^{-3}$, the Penrose kite parameter. Here ϕ is the golden ratio. The inferior sequence for A is

$$\mathbf{\frac{1}{1}} \leftarrow \frac{1}{3} \leftarrow \mathbf{\frac{1}{5}} \leftarrow \frac{3}{13} \leftarrow \mathbf{\frac{5}{21}} \leftarrow \frac{13}{55} \leftarrow \mathbf{\frac{21}{89}} \leftarrow \frac{55}{233} \leftarrow \mathbf{\frac{89}{377}} \cdots .$$

The superior sequence obeys the recurrence relation $r_{n+2} = 4r_{n+1} + r_n$, where r stands for either p or q. The inferior renormalization sequence is 1, 0, 1, 0, ... The renormalization sequence is 1, 1, 1, ... Hence $D = \log(2)$. From the recurrence relation, we compute $Q = \log(\sqrt{5}+2)$. Hence

$$u(A) = \frac{\log(2)}{\log(\sqrt{5}+2)} = \frac{\log(2)}{\log(\phi^3)}.$$

Example 2: Let $A = \sqrt{2} - 1$. The inferior sequence for A is

$$\frac{1}{1} \leftarrow \frac{1}{3} \leftarrow \frac{3}{7} \leftarrow \frac{7}{17} \leftarrow \cdots, \qquad r_{n+2} = 2r_{n+1} + r_n.$$

All terms are superior. From the recurrence relation, we get $D(A) = \log(2)$ and $Q(A) = \log(\sqrt{2}+1)$. The inferior sequence for $1 - A$ is

$$\frac{1}{1} \leftarrow \frac{3}{5} \leftarrow \frac{17}{29} \leftarrow \frac{99}{169} \leftarrow \cdots, \qquad r_{n+2} = 6r_{n+1} - r_n.$$

All terms are superior. From the recurrence relation, we have $D(1-A) = \log(3)$ and $Q(1-A) = 2\log(\sqrt{2}+1)$. Hence

$$u(A) = \frac{\log(2)}{\log(\sqrt{2}+1)}, \qquad u(1-A) = \frac{\log(3)}{2\log(\sqrt{2}+1)}.$$

In particular, $u(A) \neq u(1-A)$. The hyperbolic isometry $z \to 1 - \overline{z}$ is a symmetry of the Farey graph that does not belong to the group Γ. The calculation shows that the dimension function does not in general have this additional symmetry.

25.6 THE DIMENSION FUNCTION

Here we prove statement 4 of Theorem 1.5.

The *Borel σ-algebra* of subsets of $\mathbf{R}^n$ is the smallest collection that contains the open sets and is closed under complementation and countable unions. A *Borel set* is a member of this σ-algebra. A function $f: \mathbf{R}^n \to \mathbf{R}$ is *Borel-measurable* if the set $\{x \mid f(x) \geq a\}$ is a Borel set for all $a \in \mathbf{R}$.

Lemma 25.10 *Let $S \subset [0, 1]^2$ be a Borel subset. Let S_A denote the intersection of S with the line $\{y = A\}$. Suppose S_A is compact for all A. Let $f(A) = \dim(S_A)$. Then f is a Borel-measurable function of $[0, 1]$.*

Proof: This is a special case of [**MM**, Theorem 6.1]. □

Recall that $u(A) = \dim(U_A)$ the Hausdorff dimension of the set of unbounded special orbits.

Lemma 25.11 *The function u is Borel-measurable.*

Proof: When $A = p/q$, we let $C_A = O_2(J) \cap I$. Here J is the interval of length $2/q$ in I whose left endpoint is $(0, 1)$. Thus C_A is just a thickened version of part of the fundamental orbit. Having stated this definition, we define C as in Equation 1.13. By construction, C_A is compact for all $A \in [0, 1]$. In order to apply Lemma 25.10, we just have to show that C is a Borel set.

In the proof of Lemma 25.1 we produced a covering $\mathcal{C}_n$ of C_A by intervals all having the same length. One can extend this definition to the rational case in a fairly obvious way. Let $C_A^{(n)}$ denote the union of these intervals. Let $C^{(n)}$ be the corresponding union, with $C_A^{(n)}$ replacing C_A in Equation 1.13. The sizes and positions of the intervals in $C_A^{(n)}$ vary with A in a piecewise continuous way. Hence $C^{(n)}$ is a Borel set. Hence $C = \cap C^{(n)}$ is a Borel set. □

Lemma 25.12 *The function u is almost everywhere constant.*

Proof: The function u is a Γ-invariant Borel-measurable function on $[0, 1]$. We can extend u by the action of Γ so that the extended function $\widehat{u}$ has the same properties on all of $\mathbf{R} \cup \infty$. As is well known, Γ acts *ergodically* on $\mathbf{R} \cup \infty$. See [**BKS**]. But then any invariant Borel-measurable function is almost everywhere constant. This applies to $\widehat{u}$. Hence u is almost everywhere equal to some constant u_0. □

Let

$$S = [0, 1] - \mathbf{Q}. \tag{25.19}$$

Now we want to see that u maps every open subset of S onto $[0, 1]$. Since u is Γ-invariant and the Γ-orbits of S are dense in $[0, 1]$, it suffices to prove that $u(S) = [0, 1]$. We will prove this below.

Say that $A \in (0, 1)$ is *superior* if all the terms in the inferior sequence are superior. Let $D = D(A)$ be as in the dimension formula above.

Lemma 25.13 *If A is tame and superior, then $u(A) \geq D/(D + \log 2)$.*

Proof: Referring to the inferior sequence $\{p_n/q_n\}$ and the inferior renormalization sequence $\{d_n\}$, we always have

$$q_{n+1} < 2(d_n + 1)q_n.$$

This bound directly applies to the superior sequence when A is superior. By induction,

$$q_n \leq 2D_n.$$

Hence $Q \leq D + \log 2$. The bound follows immediately. □

Lemma 25.14 *Let A be a superior parameter whose renormalization sequence $\{d_n\}$ diverges to ∞. If d_{n+1}/d_n grows subexponentially, $u(A) = 1$.*

Proof: The same argument as in Lemma 25.8 shows that

$$\lambda_n > (h_n q_n)^{-1}, \tag{25.20}$$

where $\{h_n\}$ grows subexponentially. From Equation 21.7, we get

$$G_n = 2\lambda_n' > 2\lambda_n > 2(h_n q_n)^{-1}. \tag{25.21}$$

Therefore

$$\begin{aligned} &\lim_{n\to\infty} \frac{\log(D_n)}{\log(G_n^{-1})} \\ &\geq \lim_{n\to\infty} \frac{\log(D_n)}{\log(h_n q_n)} \\ &=^* \lim_{n\to\infty} \frac{\log(D_n)}{\log(q_n)} \\ &\geq \lim_{n\to\infty} \frac{\log(D_n)}{\log(D_n) + \log(2)} = 1. \end{aligned} \tag{25.22}$$

The starred equality comes from the subexponential growth of h_n. The same construction as in Theorem 25.9 shows that $u(A) \geq 1$. Hence $u(A) = 1$. □

Lemma 25.15 *There exists $A \in S$ such that $u(A) = 0$.*

Proof: We take A so that the inferior renormalization sequence consists entirely of 0s and 1s. Our argument for the upper bound in Theorem 25.9 gives $u(A) = 0$ if the number of 0s between each pair of 1s grows at a fast enough rate. □

Now we come to the main result.

Lemma 25.16 $u(S) = [0, 1]$.

Proof: In light of the results above, it suffices to prove $(0, 1) \subset u(S)$. Let $x \in (0, 1)$. We will consider only parameters having an odd enhanced inferior renormalization sequence. Such parameters are determined by their inferior renormalization sequences.

Let $A(M, N)$ denote the parameter with inferior renormalization sequence $N, 0_M$ repeating. Here 0_M denotes M consecutive 0s. These parameters are all quadratic irrational and hence tame. By Lemma 25.13, we have $u(A(0, N)) > x$ for N large. Moreover, for fixed N, we have $u(A(M, N)) \to 0$ as $M \to \infty$. Hence we can choose M and N such that

$$u(A(M + 1, N)) < x < u(A(M, N)). \tag{25.23}$$

(If we have equality on either side, we are finished, so we can assume strict inequality.) We fix this pair (M, N) for the rest of the proof.

For any binary sequence, $\epsilon = \{\epsilon_k\}$ we let $A(\epsilon)$ be the parameter with inferior renormalization sequence

$$N, 0_{M+\epsilon_1}, N, 0_{M+\epsilon_2}, N, 0_{M+\epsilon_3}, \ldots.$$

By construction,

$$D(A(\epsilon)) = D = \log(N) \tag{25.24}$$

independent of ϵ and M. Consider the sequence $\{q_k\}$ of denominators of superior terms corresponding to $A(\epsilon)$. By Equation 25.23, we have

$$\frac{\log q_n}{n} < xD, \qquad \frac{\log q_n}{n} > xD,$$

respectively, for the 0-sequence and for the 1-sequence, once n is large. Inserting an additional 0 into the inferior renormalization sequence has the effect of at most doubling the terms in the denominator sequence. (Compare the proof of Lemma 21.1.) Therefore we can choose the first n terms of ϵ such that

$$\left|\frac{\log(q_n)}{n} - xD\right| \leq \frac{\log 2}{n},$$

provided n is large. Passing to a subsequence and taking a limit, we can choose ϵ so that $Q(A(\epsilon)) = xD$. But then $A(\epsilon)$ is tame and $u(A(\epsilon)) = x$. □

We have already shown that the function u is almost everywhere constant. Let $r \in [0, 1]$ be arbitrary. We have just shown that $u^{-1}(r)$ is nonempty. But u is invariant under the $(2, \infty, \infty)$-triangle group. Hence $u^{-1}(r)$ is dense in S. This finishes the proof of statement 4 of Theorem 1.5.

Part 6. More Structure Theorems

In this part of the book, we will prove all the results left over from Part 5. The material in this part is probably the most difficult, so it seems worthwhile to point out that one can stop reading early and still take away partial results.

- In Chapter 26, we prove the Copy Theorem from §22.2. Knowing just the Copy Theorem, we can conclude that $C_A^{\#} \subset U_A$. That is, the trimmed Cantor set from the Comet Theorem is *contained* in the union of special unbounded orbits. All the dynamical results on $C_A^{\#}$ – e.g., the essential conjugacy to the odometer – follow just from the Copy Theorem. This might be a nice result for the reader interested mainly in the existence and nature of unbounded orbits.

- In Chapter 27, we define what we mean by the pivot arc relative to an even rational kite parameter. Along the way we will prove another version of the Diophantine Lemma from §18.2. This lemma works for pairs of odd rationals, and the result here works for pairs of Farey-related rationals, either even or odd. This whole chapter is a prelude to the last 4 chapters.

- In Chapter 28, we prove the Pivot Theorem from §22.2. The Pivot Theorem works in both the even and odd cases, and is proved in an inductive way that requires both cases. From the Pivot Theorem and the Copy Theorem combined, we have Theorem 1.8.

- In Chapter 29, we prove the Period Theorem. Combining the Copy Theorem, the Pivot Theorem, and the Period Theorem, we prove that $U_A \cap I = C_A^{\#}$. In other words, we completely characterize the set of unbounded orbits inside the special interval I from the Comet Theorem.

- In Chapter 30, we prove a technical result, the Hovering Lemma, which rules out the existence of certain pathological components of the arithmetic graph. We use the Hovering Lemma as a step in the proof of the Low Vertex Theorem.

- In Chapter 31, we prove the Low Vertex Theorem. This is the technical result we needed for statement 1 of the Comet Theorem. Statement 1 of the Comet Theorem is the result that gives us the minimality and homogeneity of U_A. So, one needs to read all the way to the end to obtain the global structural results for U_A.

Chapter Twenty-Six

Proof of the Copy Theorem

26.1 A FORMULA FOR THE PIVOT POINTS

Let A be an odd rational. Let A_- be as in Equation 4.1. Let $V_- = (q_-, -p_-)$. Here we give a formula for the pivot points $E^\pm$ associated to A. Recall that these points are the endpoints of the pivot arc, the subject of the Copy Theorem.

Lemma 26.1 *The following are true.*

- *If $q_- < q_+$, then $E^+ + E^- = -V_- + (0, 1)$.*
- *If $q_+ < q_-$, then $E^+ + E^- = V_+ + (0, 1)$.*

Proof: We will establish this result inductively. Suppose first that $1/1 \leftarrow A$. Then

$$A = \frac{2k-1}{2k+1}, \qquad E^- = (-k, k), \qquad E^+ = (0, 0), \qquad V_- = (k, -k+1).$$

$$A_- = \frac{k-1}{k}, \qquad q_- = k-1 < k = q_+.$$

The result works in this case.

In general, we have

$$A = A_2, \qquad A_0 \leftarrow A_1 \leftarrow A_2.$$

There are 4 cases, depending on Lemma 17.2. Here the index is $m = 1$. We will consider case 1. The other cases are similar. By case 1, we have $(q_1)_+ < (q_1)_-$. Hence, by induction,

$$E_1^+ + E_1^- = (V_1)_+ + (0, 1).$$

Since $A_1 < A_2$, we have

$$E_2^- = E_1^-, \qquad E_2^+ = E_1^+ + d_1 V_1.$$

Therefore

$$E_2^+ + E_2^- = (V_1)_+ + d_1 V_1 + (0, 1) = (V_2)_+ + (0, 1).$$

The last equality comes from case 1 of Lemma 17.2. As we remarked after stating Lemma 17.2, this result works for both numerators and denominators.) In case 1, we have $(q_2)_+ < (q_2)_-$, so the result holds. $\square$

Recall that $R_1 = R_1(A)$ and $R_2 = R_2(A)$ are the two parallelograms from the Decomposition Theorem. See Chapter 19.

Lemma 26.2 *E^- lies to the left of R_1, and E^+ lies to the right of R_1.*

Proof: Let π_1 denote the projection to the first coordinate. One of the bottom vertices of R_1 is $(0, 0)$. We will consider the case when the left bottom vertex is $(0, 0)$. In all cases one can easily check from the definitions that $\pi_1(E^-) \leq -1$. Hence E^- lies to the left of R_1.

Consider the right side. We have $q_+ < q_-$ in our case. By case 2 of Lemma 26.1 and the result for the left hand side, we have

$$\pi_1(E^+) \geq \pi_1(V_+) + 1.$$

But V_+ lies on the line extending the bottom right edge of R_1, exactly $1/q$ vertical units beneath the bottom edge of R_1. This right edge has a slope greater than 1. Finally, the line connecting V_+ to $\pi_1(E^+)$ has a nonpositive slope because E^+ is a low vertex lying to the right of V_+. From all this geometry, we see that E_1^+ lies to the right of R_1. □

Figure 26.1 illustrates this result for the parameter $13/57$. The smaller of the two parallelograms is R_1 in this case. The pivot arc starts out on the far left and extends about to the bottom middle of the figure.

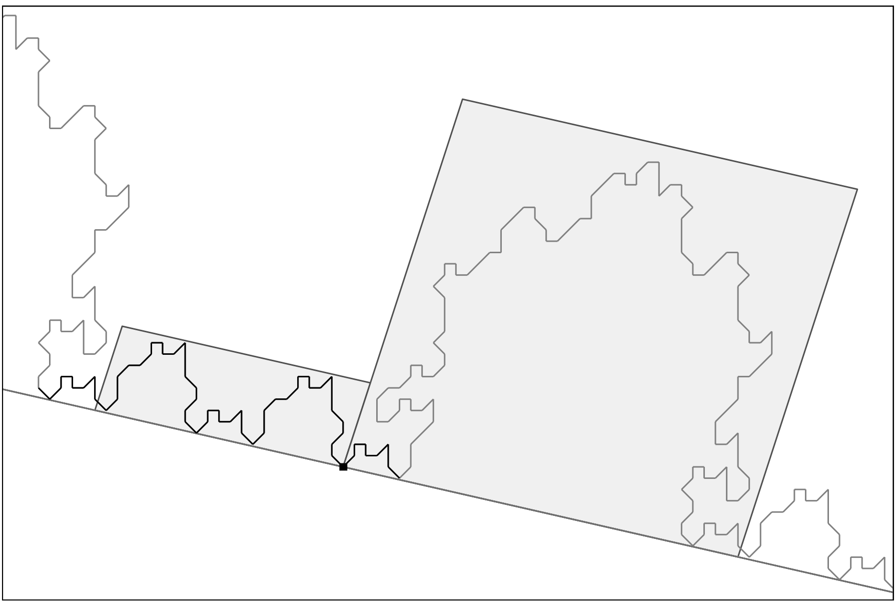

Figure 26.1: $P\Gamma(13/57)$ and $R_1(13/57)$ and $R_2(13/57)$.

Note that the pivot arc is symmetrically situated with respect to R_1. This always happens, as we shall see later on.

26.2 A DETAIL FROM PART 5

While we are in the neighborhood, we will clear up a detail from Part 5, namely, the proof of Lemma 22.5. For convenience, we repeat the statement here. In this statement, E^+ and E^- are the pivot points relative to the odd rational parameter $A = p/q$.

Lemma 26.3

$$-\frac{q}{2} < \pi_1(E^-) < \pi_1(E^+) < \frac{q}{2}.$$

Proof: We will prove this result inductively. Suppose that

$$A_1 \leftarrow A_2,$$

and the result is true for A_1. We consider the case when $A_1 < A_2$. The case when $A_1 > A_2$ requires the same treatment. When $A_1 < A_2$, we have

$$E_1^- = E_2^-,$$

so certainly the bound holds for E_1^-.

For the $(+)$ case, we have

$$\pi_1(E_2^+) = \pi_1(E_1^+) + d_1 q_1 \qquad d_1 = \text{floor}\left(\frac{q_2}{2q_1}\right). \tag{26.1}$$

There are two cases to consider.

Case 1: Suppose that $\delta_1 = \text{floor}(q_2/q_1)$ is odd. In this case

$$(2d_1 + 1)q_1 < q_2, \qquad \Longrightarrow \qquad d_1 q_1 < \frac{q_2}{2} - \frac{q_1}{2}.$$

The first equation implies the second. Hence, by induction,

$$\pi_1(E_2^+) < \frac{q_1}{2} + \frac{q_2 - q_2}{2} < \frac{q_2}{2}.$$

Case 2: Suppose that δ_1 is even. Then we have case 2 of Lemma 17.2 applied to the index $m = 1$. This is to say that

$$(q_1)_- < (q_1)_+. \tag{26.2}$$

From the formula above, the first coordinate of $E_2^- + E_2^+$ is negative. Hence

$$|\pi_1(E^-)| > |\pi_1(E^+)|.$$

This fact finishes the proof. $\square$

26.3 PRELIMINARIES

As preparation for the main argument of our proof, we prove several easy results in this section and also set up some notation.

The pivot points are well defined vertices, but so far, we do not know that the pivot arc is well defined. That is, we do not know that E^- and E^+ are actually vertices of Γ. These points might be vertices of some other component of $\widehat{\Gamma}$. We will prove the well definedness result along with the proof of the Copy Theorem.

To start things off, we prove that the pivot arcs are well defined in the simplest cases.

Lemma 26.4 *If* $1/1 \leftarrow A$*, then the pivot arc is well defined relative to* A*.*

Proof: Here

$$A = \frac{2k-1}{2k+1}. \tag{26.3}$$

In §19.5, we showed that the line segment connecting $(0,0)$ to $(-k,k)$ is contained in the arithmetic graph. So, the pivot arc is well defined. □

Notation: Here we introduce some notation that we will use repeatedly below. Let A_1 be an odd rational. For each integer $\delta_1 \geq 1$, there is a unique odd rational $A_2 = A_2(\delta_1)$ such that $A_1 \leftarrow A_2$ and

$$\delta_1 = \text{floor}\left(\frac{q_2}{q_1}\right).$$

Thus the numbers $A_2(1)$, $A_2(2)$, ... give the complete list of odd rationals having A_1 as an inferior predecessor. Lemma 17.2 describes how to construct $A_2(\delta_1)$. For instance, if $A_1 = 1/3$ then

$$A_2(1) = \frac{1}{5}, \quad A_2(3) = \frac{3}{11}, \quad A_2(5) = \frac{5}{17}, \qquad A_2(2) = \frac{3}{7}, \quad A_2(4) = \frac{5}{13}.$$

We have listed the numbers this way to show the pattern better.

Lemma 26.5 *Let* $E_j^{\pm}$ *be the pivot points associated to the parameter* A_1*. There is an arc* $P\Gamma_1(\delta_1) \subset \Gamma_1$ *whose endpoints are* E_2^- *and* E_2^+*.*

Proof: Suppose that $A_1 < A_2$. When $A_2 < A_1$ the proof is similar. Then, by Equation 22.3, we have

$$E_2^- = E_1^-, \qquad E_2^+ = E_1^+ + d_1 V_1, \qquad V_1 = (q_1, -p_1).$$

Here d_1 is as in Equation 4.5 and Lemma 17.2. But Γ_1 is invariant under translation by V_1. Hence $E_2^{\pm}$ is a vertex of Γ_1. □

26.4 THE GOOD PARAMETER LEMMA

Call A_1 a *good parameter* if

$$P\Gamma_1 \subset \Delta_1(I). \tag{26.4}$$

Here $\Delta_1(I)$ is the region from the Diophantine Lemma defined relative to the pair $(A_1, A_2(1))$. We call I the *base interval*. We will give a formula below. Here is the main result in this section.

Lemma 26.6 (Good Parameter) *If A_1 is good and $A_1 \leftarrow A_2$, then the Copy Theorem holds for the pair (A_1, A_2).*

We will prove this result in a case-by-case way. That is, we will treat the pair $\big(A_1, A_2(k)\big)$ for $k = 1, 3, 5, \ldots$ and $k = 2, 4, 6, \ldots$. The case when $k = 1$ is pretty easy. The cases when $k = 2$ and $k = 3$ are the critical cases, and they have essentially the same proof. The remaining cases are easy. For the rest of this section, we assume that A is good.

Our proof will use the Mismatch Principle established in Chapter 19. For convenience, we repeat it here.

Mismatch Principle: Let Γ and Γ' be two arithmetic graphs. If Γ' and Γ fail to agree in R_1, then there are two adjacent vertices of $\Gamma' \cap R_1$ where the two arithmetic graphs $\widehat{\Gamma}$ and $\widehat{\Gamma}'$ do not agree.

26.4.1 An Easy Case

Here we show that the Copy Theorem holds for A_1 and $A_2(1)$. Note that

$$P\Gamma_1(1) = P\Gamma_1. \tag{26.5}$$

The pivot points do not change in this case: $E_1^{\pm} = E_2^{\pm}$. So, if A_1 is good, then the Diophantine Lemma immediately implies that $P\Gamma_1(1) = P\Gamma_1 \subset \Gamma_2$. But then there is an arc of Γ_2 that connects E_2^- to E_2^+, the two endpoints of $P\Gamma_1(1)$. This shows that this pivot arc for A_2 is well defined as a subarc of $\Gamma(A_2)$ and moreover that his pivot arc is a subarc of Γ_1.

Before we leave this section, we establish some notation to be used below. Let A_0 be such that the sequence

$$A_0 \leftarrow A_1 \leftarrow A_2(1)$$

is part of the inferior sequence. Let I be the base interval, as above. We will consider the case when $A_0 < A_1$. The other case is similar. In the case at hand, the base interval is given by

$$I = \big[-q_1 + 2, q_1 + (q_2)_+ - 2\big] = \big[-q_1 + 2, q_1 + (q_1)_+ - 2\big]. \tag{26.6}$$

The first equality is Lemma 17.8. The second equality is case 1 of Lemma 17.2, with $d_1 = 0$.

For later purposes, we write

$$I = \big[I_{\text{left}}, I_{\text{right}}\big]. \tag{26.7}$$

26.4.2 The Critical Odd Case

Now we show that the Copy Theorem holds for A_1 and $A_2(3)$. The basic idea is to build from the easy case we have already handled. We again consider the case when $A_0 < A_1$ for ease of exposition. The other case is essentially the same. From Lemma 17.2, we know that $A_1 < A_2(3)$, just as we knew above that $A_1 < A_2(1)$.

We define

$$A_2 = A_2(1), \qquad A_2^* = A_2(3). \tag{26.8}$$

We attach a $(*)$ to objects associated to A_2^*. Let I be the base interval. Let I^* denote the interval corresponding to the pair (A_1, A_2^*). By Lemma 17.2, we have $(q_2^*)_+ = q_1 + (q_1)_+$. Hence, by Lemma 17.8 and by definition,

$$I^* = \left[I_{\text{left}}, I_{\text{right}} + q_1\right]. \tag{26.9}$$

Lemma 26.7

$$P\Gamma_1(3) = P\Gamma_1 \cup \gamma \cup \Big(P\Gamma_1 + V_1\Big), \qquad \gamma \in (R_2 + V_1). \tag{26.10}$$

That is, $P\Gamma_1(3)$ is obtained from $P\Gamma_1(1)$ by concatenating one period of Γ_1 to the right.

Proof: Let $R_j = R_j(A_1)$, as in the Decomposition Theorem for A_1. As in Lemma 26.2, we know that R_1 lies to the right of the origin and R_2 to the left. This is because $(q_1)_+ < (q_1)_-$ in the case we are considering. By Lemma 26.2, the arc $P\Gamma_1$ completely crosses R_1. The left endpoint lies in R_2, and the right endpoint lies in $R_2 + V_1$, the translate of R_2 that lies on the other side of R_1. By symmetry, one endpoint of $P\Gamma_1(1)$ enters $R_2 + V_1$ from the left, and one endpoint of $P\Gamma_1(1) + V_1$ enters $R_2 + V_1$ from the right. The arc γ joins two points already in $R_2 + V_1$. This arc cannot cross out of $R_2 + V_1$, by Lemma 19.2. □

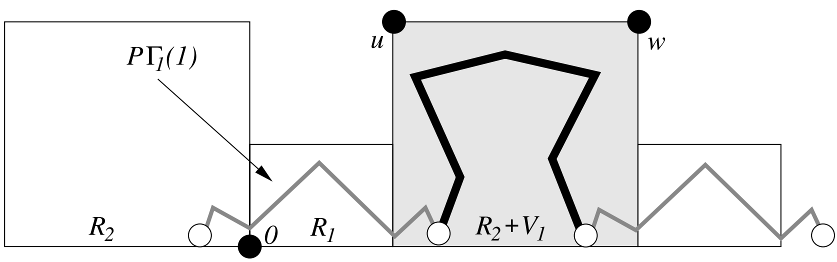

Figure 26.2: Decomposition of $P\Gamma_1(3)$.

Now that we have broken $P\Gamma_1(3)$ into three pieces, as shown in Figure 26.2, we have three pieces to consider. The left piece is is easy.

$$P\Gamma_1 \subset \Delta_1(I) \subset \Delta_1(I^*) \qquad \Longrightarrow \qquad P\Gamma_1 \subset \Gamma(A_2^*). \tag{26.11}$$

The first containment is the definition of goodness. By the Diophantine Lemma, the first equation implies the second.

The right piece is also easy. Any $v^* \in P\Gamma_1 + V_1$ has the form $v + V_1$, where $v \in P\Gamma_1$. By Lemma 18.1, we have

$$G_1(v^*) = G_1(v) + q_1, \qquad H_1(v^*) = H_1(v) + q_1.$$

Hence $v \in \Delta_1(I)$ implies $v^* \in \Delta(I^*)$. Therefore

$$P\Gamma_1 + V_1 \subset \Delta_1(I^*) \quad \Longrightarrow \quad P\Gamma_1 + V_1 \subset \Gamma(A_2^*). \tag{26.12}$$

The middle piece is harder.

Lemma 26.8 $\gamma \subset \Gamma(A_2^*)$.

Proof: We will use the same argument that we used in §19.4. Since $\gamma \subset R_2$, we just have to show that the vertices of R_2 belong to the set $\Delta_1(I^*) \cup \Delta_2(I^*)$. This is a calculation just like the one in §19.4.

Now for the calculation. Let u and w, respectively, be the upper left and upper right vertices of $R_2 + V_1$. We have

$$u \approx W_1 + \frac{(q_1)_+}{q_1} V_1, \qquad w = W_1 + V_1. \tag{26.13}$$

Here the vectors are as in Equation 3.2, as usual. The approximation is good to within $1/q_1$. To avoid approximations, we consider the very slightly altered parallelogram $\tilde{R}_2 + V_1$. The vertices are

$$\begin{aligned} &(V_1)^+, \qquad && \tilde{u} = W_1 + \frac{(q_1)_+}{q_1} V_1, \\ &V_1, && w = V_1 + W_1. \end{aligned} \tag{26.14}$$

Each vertex of the new parallelogram is within $1/q_1$ of the corresponding old parallelogram. Using the Mismatch Principle, it suffices to do the calculation in $\tilde{R}_2 + V_1$. Here is the calculation.

$$\begin{aligned} &G_1(\tilde{u}) - (-q_1) \\ =&(2q_1 + q_+) - H_1(w) \\ =&q_1 + (q_1)_+ - \frac{q_1^2}{p_1 + q_1} \geq 2. \end{aligned} \tag{26.15}$$

These bounds hold for all but a few exceptional parameters, as in Lemma 19.4. The remaining few cases can be treated using exactly the same tricks as in §19.5. □

Now we have shown that $P\Gamma_1(3) \subset \Gamma_2$, as desired.

26.4.3 The Rest of the Odd Cases

We will consider the case when $\delta_1 = 5$. The cases $\delta = 7, 9, 11, \ldots$ have the same treatment.

In the case at hand, $P\Gamma_1(5)$ is obtained by concatenating 2 periods of Γ_1 to the right of $P\Gamma_1(1)$. We have decomposition of the form

$$P\Gamma_1(5) = P\Gamma_1(1) \cup \gamma \cup \Big(P\Gamma_1(1) + 2V_1\Big), \qquad \gamma \subset (R_2 + V_1) \cup (R_2 + 2V_1). \tag{26.16}$$

Here γ is contained in a parallelogram that is twice as long as in the case $\delta = 3$. The calculations are exactly the same in this case. The key point is that $I^* = [a, b + 2q_1]$.

26.4.4 The Even Cases

Once we take care of the critical even case, we will treat the remaining even cases just as we treated the remaining odd cases. We will show that the Copy Theorem holds for A_1 and $A_2(2)$. As in the odd case, we assume that $A_0 < A_1$. The other case is entirely similar.

Our proof here is about the same as in §26.4.2. We will just indicate the highlights. The same argument as in §26.4.2 gives the decomposition

$$P\Gamma_1(2) = \Big(P\Gamma_1(1) - V_1\Big) \cup \gamma \cup P\Gamma_1(1), \qquad \gamma \subset R_2. \tag{26.17}$$

The same arguments as in §26.4.2 take care of the left and right pieces of $P\Gamma_1(2)$, as shown in Figure 26.3. Now we repeat the analysis for the middle arc γ. Combining case 4 of Lemma 17.2 with Lemma 17.8, we have

$$I^* = \big[-q_1 - (q_1)_- + 2, q_1 - 2\big]. \tag{26.18}$$

We have

$$u \approx \frac{-(q_1)_-}{q_1} V_1 + W_1, \qquad w = W_1. \tag{26.19}$$

Again, the approximation holds up to $1/q_1$.

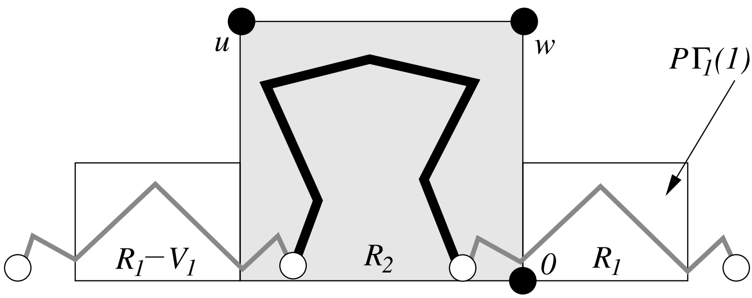

Figure 26.3: Decomposition of $P\Gamma_1(2)$.

To avoid approximations, we use the modified parallelogram $\tilde{R}_2$ with vertices

$$\frac{-(q_1)_-}{q_1} V_1, \qquad \tilde{u} = \frac{-(q_1)_-}{q_1} V_1 + W_1, \qquad (0,0), \qquad w = W_1. \tag{26.20}$$

Again, this is justified by the Mismatch Principle. The following estimate combines with the Diophantine Lemma to show that $\gamma \subset \Gamma_2(A_2^*)$.

$$\begin{aligned} &G_1(\tilde{u}) - (-q_1 - (q_1)_-) \\ =&q_1 - H(w) \\ =&q_1 - H(w) \\ =&q_1 - \frac{q_1}{p_1 + q_1} \geq 2. \end{aligned} \tag{26.21}$$

This calculation takes care of the same parameters as in Lemma 19.4, and then the same tricks as in §19.5 take care of the exceptional cases.

26.5 THE END OF THE PROOF

The Good Parameter Lemma reduces our job to showing that any odd rational parameter is good. We will give an inductive argument.

Lemma 26.9 *If* $1/1 \leftarrow A$*, then* A *is good.*

Proof: We write $1/1 \leftarrow A \leftarrow \widehat{A}$. In this case, Lemma 17.2 tells us that

$$1/1 > A > \widehat{A} \tag{26.22}$$

(The first inequality is obvious.) We have

$$A = \frac{2k-1}{2k+1}, \qquad \widehat{A} = \frac{4k-3}{4k+1}, \qquad \widehat{q}_- = k.$$

By Lemma 17.8, we have

$$I = \big[-q - q_- + 2, q - 2\big] = [-3k+1, 2k-1]$$

The left vertex of $P\Gamma_1$ is $u = (-k, k)$, and the right vertex is $v = (0, 0)$. We compute

$$G(u) = -k - 1 \geq -3k + 1, \qquad H(w) = 0 \leq 2k - 1.$$

The extreme case occurs when $k = 1$. □

Lemma 26.10 $A = p/q$ *is good if* $q < 20$ *or if* $p = 1$.

Proof: We check the case $q < 20$ by hand. If $p = 1$, the pivot arc is just the edge connecting $(-1, 1)$ to $(0, 0)$, whereas the interval I contains $[-q, q]$, a huge interval. This case is obvious. □

Now we establish the inductive step. Suppose that $A_1 \leftarrow A_2$ and that A_1 is good. Having eliminated the few exceptional cases by the result above, the argument in the previous section shows that $P\Gamma_2 \subset \Delta_1(I_1)$. Here I_1 is the interval based on the constant $\Omega(A_1, A_2)$. This is the Diophantine constant defined in §17.4 relative to the pair (A_1, A_2). To finish the proof of the Copy Theorem, we just have to establish the following equation.

$$P\Gamma_2 \subset \Delta_2(I_2), \tag{26.23}$$

where I_2 is the different interval based on the pair $A_2 \leftarrow A_3$, with $\delta(A_2, A_3) = 1$. Here we establish two basic facts.

Lemma 26.11 $I_1 \subset I_2$*, and either endpoint of* I_1 *is more than* 1 *unit from the corresponding endpoint of* I_2*.*

Proof: By Lemma 17.8, applied to both parameters, we have

$$I_1 \subset [-q_2+3, q_2-3] \subset [-q_2-2, q_2-2] \subset I_2.$$

This completes the proof. $\square$

Let G_j and H_j be the linear functionals associated to A_j in the Diophantine Lemma. See §18.1.

Lemma 26.12 $|G_1(v) - G_2(v)| < 1$ *and* $|H_1(v) - H_2(v)| < 1$ *for* $v \in \Delta_1(I_1)$.

Proof: From Lemma 17.8 and a bit of geometry, we get the bound

$$(m,n) \in \Delta_1(I_1) \qquad \Longrightarrow \qquad \max(|m|,|n|) \leq q_2. \tag{26.24}$$

Looking at Equation 18.2, we see that

$$G(m,n) = \left(\frac{1-A}{1+A}, \frac{-2}{1+A}\right) \cdot (m,n) = (G_1, G_2) \cdot (m,n).$$

$$H(m,n) = \left(\frac{1+4A-A^2}{(1+A)^2}, \frac{2-2A}{(1+A)^2}\right) \cdot (m,n) = (H_1, H_2) \cdot (m,n). \tag{26.25}$$

A bit of calculus shows that

$$|\partial_A G_j| \leq 2, \qquad |\partial_A H_1| \leq 6, \qquad |\partial_A H_2| \leq 2. \tag{26.26}$$

Since $A_1 \leftarrow A_2$, we have

$$|A_1 - A_2| = \frac{2}{q_1 q_2}. \tag{26.27}$$

Putting everything together, and using basic calculus, we arrive at the bound

$$|G_1(v) - G_2(v)|, |H_1(v) - H_2(v)| < 16/q_1 < 1, \tag{26.28}$$

at least for $q_1 > 16$. $\square$

We have already remarked, during the proof of the Decomposition Theorem, that no lattice point lies between the bottom of $\Delta_2(I_2)$ and the bottom of $\Delta_1(I_2)$. Hence $F_1(v) > 0$ iff $F_2(v) > 0$. The two lemmas now show that $\Delta_1(I_1) \subset \Delta_2(I_2)$. This was our final goal, from Equation 26.23.

This completes the proof of the Copy Theorem.

Chapter Twenty-Seven

Pivot Arcs in the Even Case

27.1 MAIN RESULTS

Given two rationals $A_1 = p_1/q_1$ and $A_2 = p_2/q_2$, we introduce the notation

$$A_1 \vdash A_2 \qquad \Longleftrightarrow \qquad |p_1 q_2 - q_1 p_2| = 1, \qquad q_1 < q_2. \tag{27.1}$$

In this case, we say that A_1 and A_2 are *Farey-related*. We sometimes call (A_1, A_2) a *Farey pair*.

We have the notions of *Farey addition* and *Farey subtraction*, respectively.

$$A_1 \oplus A_2 = \frac{p_1 + p_2}{q_1 + q_2}, \qquad A_2 \ominus A_1 = \frac{p_2 - p_1}{q_2 - q_1}. \tag{27.2}$$

Note that $A_1 \vdash A_2$ implies that $A_1 \vdash (A_1 \oplus A_2)$ and that A_1 is Farey-related to $A_2 \ominus A_1$.

Lemma 27.1 *Let A_1 be an even rational. Then there is a unique odd rational A_2 such that $A_1 \vdash A_2$ and $2q_1 > q_2$. In this case, we write $A_1 \models A_2$.*

Proof: Equation 4.1 works for both even and odd rationals. When A_1 is even, exactly one of the rationals $(A_1)_\pm$ is also even. Call this rational A_1'. Then $A_1' \vdash A_1$. We define $A_2 = A_1 \oplus A_1'$. If B_2 were another candidate, then $B_2 \ominus A'$ would be the relevant choice of $(A_1)_\pm$. Hence $B_2 = A_2$. □

Let A be an odd rational. Then either $A_- \models A$ or $A_+ \models A$ when A is an odd rational. If $A_- \models A$, then we write $A_+ \Leftarrow A$. The relationship implies that $2q_+ < q$. Likewise we write $A_- \Leftarrow A$ when $2q_- < q$. Here is an example: Let $A = 3/7$. Then

$$A_+ = 1/2 \Leftarrow 3/7, \qquad A_- = 2/5 \models 3/7.$$

So far we have defined pivot points and arcs for odd parameters. Now we define them for even parameters. We have

$$E^\pm(A_1) = E^\pm(A_2), \qquad A_1 \models A_2. \tag{27.3}$$

This makes sense because we have already defined the pivot points in the odd case. We still need to prove that these vertices lie on Γ_1. We will do this below.

Assuming that the pivot points $E_1^\pm$ are vertices of Γ_1, we define $P\Gamma_1$ to be the lower arc of Γ_1 that connects E_1^- to E_1^+. Since Γ_1 is a polygon in the even case, it makes sense to speak of the lower arc. Figure 27.1 shows an example. Here $P\Gamma_1 = P\Gamma_2$. We will show that this always happens.

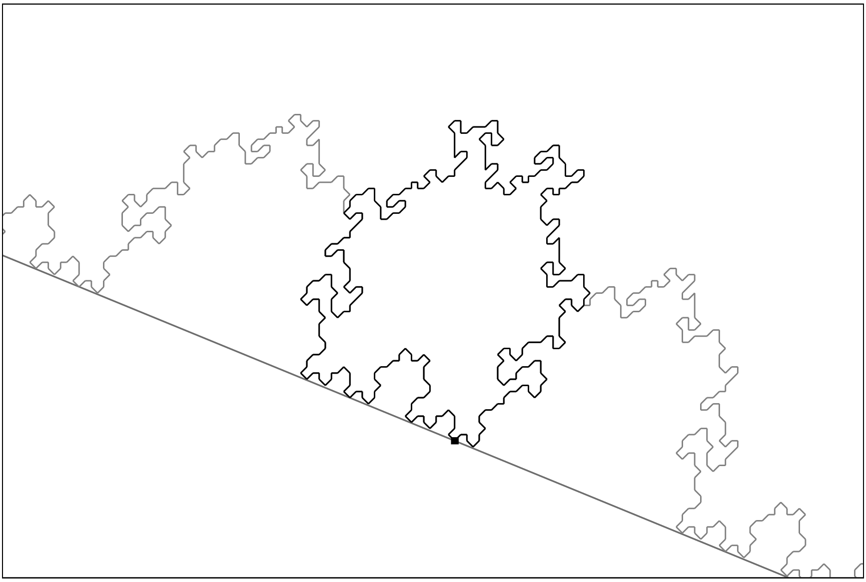

Figure 27.1: $\Gamma(16/39)$, in black, overlays $\Gamma(25/61)$, in gray.

In this chapter we prove the following results.

Lemma 27.2 *Let* $A_1 \models A_2$. *Then* $P\Gamma_1$ *is well defined and* $P\Gamma_1 = P\Gamma_2$.

Lemma 27.3 (Structure) *The following are true.*

1. *If* $A_- \Leftarrow A$, *then* $E^+(A) = E^+(A_-)$.
2. *If* $A_+ \Leftarrow A$, *then* $E^-(A) = E^-(A_+)$.
3. *If* $A_- \Leftarrow A$, *then* $E^-(A) + V = E^-(A_-) + kV_-$ *for some* $k \in \mathbf{Z}$.
4. *If* $A_+ \Leftarrow A$, *then* $E^+(A) - V = E^+(A_+) + kV_+$ *for some* $k \in \mathbf{Z}$.

The Structure Lemma is of crucial importance in our proofs of the Pivot Theorem and the Period Theorem. Here we illustrate its meaning and describe a bit of the connection to the Pivot Theorem. Figure 27.2 shows slightly more than one period of $\Gamma(25/61)$, in black. This black arc overlays $\Gamma(9/22)$ on the left and

$$\Gamma(9/22) + 2V(9/22)$$

on the right. Call these two gray components the eggs. Here

$$9/22 \Leftarrow 25/61.$$

The points

$$E^+(25/61), \qquad E^-(25/61) + V(25/61)$$

are the left and right endpoints, respectively, of the big central bump of $\Gamma(25/61)$. Call this black arc the bump. The content of the Structure Lemma (in this case) is that the endpoints of the bump are simultaneously pivot points on the eggs. The reader can draw many figures like this using Billiard King.

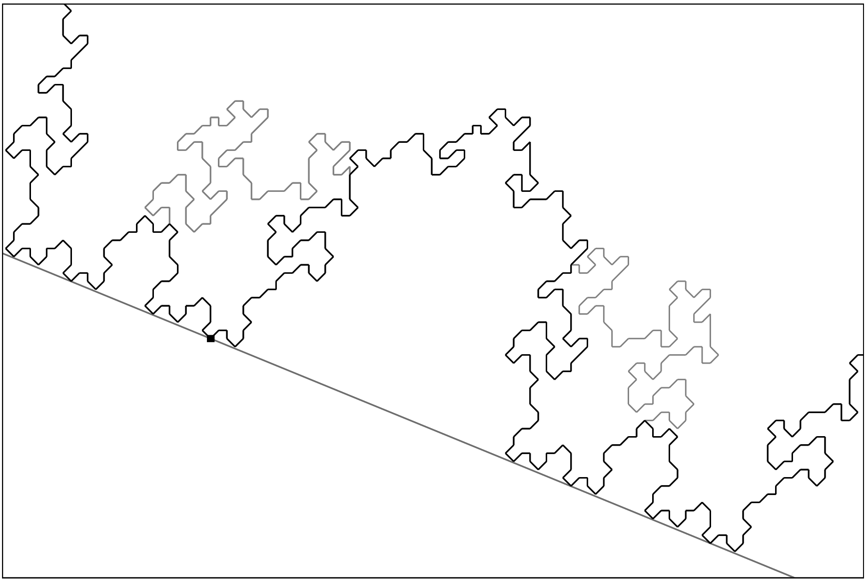

Figure 27.2: $\Gamma(25/61)$ overlays several components of $\widehat{\Gamma}(9/22)$.

The content of the Pivot Theorem for $25/61$ is that the bump has no low vertices except its endpoints. Note that the ends of the bump copy pieces of the eggs. If we understand the behavior of the eggs – meaning how they rise away from the baseline – then we understand the behavior of the ends of the bump. Knowing that the bump behaves nicely near its endpoints gets our proof off the ground so to speak.

The eggs are based on a simpler rational, and this suggests an inductive approach to the Pivot Theorem: In this way, the behavior of the arithmetic graph for a simpler rational gives us information about what happens for a more complicated rational. This is (some of) the strategy for our proof of the Pivot Theorem. In the first section of the next chapter, we will present a long and somewhat informal discussion about the remainder of the strategy.

Remarks:
(i) In §27.5 below we will describe the precise relationship between the two pivot arcs in the cases of interest to us. After reading the description, the reader will perhaps be able to see this connection as illustrated in Figure 1.5.
(ii) Notice in Figure 27.2 that the gray curves lie completely above the black one except for the edges where they coincide. There is nothing in our theory that explains such a clean kind of relationship, but it always seems to hold. There is a similar phenomenon for pairs of even rationals. See Figure 1.5.
(iii) The Structure Lemma has a crisp result that is easily checked computationally for individual cases. However, as the reader will see, our proof is rather tedious and we wish we had a better one.

27.2 ANOTHER DIOPHANTINE LEMMA

Here we prove a copying lemma that helps with Lemma 27.2. Our result works for Farey pairs. Let $\Delta_1(I)$ and $\Delta_2(I)$ be the sets defined exactly as in the Diophantine Lemma. See §18.2. The result we prove here is actually more natural than our original result. However, the original result better suited our more elementary purposes.

Lemma 27.4 *Suppose that* $A_1 \vdash A_2$.

1. *If* $A_1 < A_2$, *let* $I = [-q_1 + 2, q_2 - 2]$.

2. *If* $A_1 > A_2$, *let* $I = [-q_2 + 2, q_1 - 2]$.

Then $\widehat{\Gamma}_1$ *and* $\widehat{\Gamma}_2$ *agree on* $\Delta_1(I) \cup \Delta_2(I)$.

Proof: We will consider the case when $A_1 < A_2$. The other case requires a very similar treatment. In the proof of the Diophantine Lemma we used only the oddness of the rationals in Lemma 17.5. Once we prove the analog of this result in the even setting, the rest of the proof works verbatim.

Recall that an integer μ is *good* if

$$[A_1\mu] = [A_2\mu]. \tag{27.4}$$

Here [ ] denotes the floor function. The analog of Lemma 17.5 is the statement that an integer μ is good provided that $\mu \in (-q_1, q_2)$. We will give a geometric proof. Let L_1 (respectively, L_2) denote the line segment of slope $-A_1$ (respectively, $-A_2$) joining the two points whose first coordinates are $-q_1$ and q_2. If we have a counterexample to our claim, then there is a lattice point (m, n) lying between L_1 and L_2.

If $m < 0$, we consider the triangle T with vertices $(0, 0)$ and $-V_1$ and (m, n). Here $V_1 = (q_1, -p_1)$. The vertical distance between the left endpoints of L_1 and L_2 is $1/q_2$. By the base-times-height formula for triangles,

$$\text{area}(T) < q_1/(2q_2) < 1/2. \tag{27.5}$$

But this contradicts the fact that $1/2$ is a lower bound for the area of a lattice triangle. If $m > 0$, we consider the triangle T with vertices $(0, 0)$ and V_1 and (m, n). The lattice point (m, n) is closer to the line containing L_1 than is the right endpoint of L_2, namely, $(q_2, -p_2)$. Hence

$$\text{area}(T) < \text{area}(T'), \tag{27.6}$$

where T' is the triangle with vertices $(0, 0)$ and V_1 and V_2. But

$$\text{area}(T') = 1/2 \tag{27.7}$$

because A_1 and A_2 are Farey-related. We have the same contradiction as in the first case. □

27.3 COPYING THE PIVOT ARC

Here we prove Lemma 27.2. As we did for the Decomposition Theorem, we will first establish the result for most parameters. Then we will treat the exceptional cases.

Suppose that $A_1 \models A_2$. To show that $P\Gamma_1$ is well defined, we just have to show $P\Gamma_2 \subset \Gamma_1$. This result simultaneously shows that $P\Gamma_1 = P\Gamma_2$ because the endpoints of these two arcs are the same by definition.

In the case at hand, we have $A_1 = (A_2)_-$. To simplify the notation, we write $A = A_2$. Then $A_1 = A_-$. By Lemma 27.4, it suffices to prove that

$$P\Gamma \subset \Delta(J), \qquad J = \big[-q_- + 2, q - 2\big]. \tag{27.8}$$

We have actually already proved this, but it takes some effort to recognize the fact.

Let $A' \leftarrow A$ denote the inferior predecessor of A. Since $q_- > q_+$, we have

$$A' = A_- \ominus A_+. \tag{27.9}$$

In the previous chapter, when we proved the Copy Theorem, we established

$$P\Gamma \subset \Delta'(J'), \qquad J' = \big[-q' + 2, q' + q_+ - 2\big], \tag{27.10}$$

as long as $p' \geq 3$ and $q' \geq 7$. Here Δ' is defined relative to the linear functionals G' and H', which are defined relative to A'. The right endpoint in Equation 27.10 comes from Lemma 17.8. The point is that the calculation in Lemma 26.8 gives the same bounds as the calculation for Lemma 19.4.

Now we observe that

$$q' = q_- - q_+ < q_- \tag{27.11}$$

and

$$q' + q_+ < (q_- - q_+) + q_+ = q_- < q. \tag{27.12}$$

These calculations show that $J' \subset J$. Usually J is much larger.

The region $\Delta(J)$ is computed relative to the parameter A, whereas the region $\Delta'(J')$ is computed relative to the parameter A'. The same argument as in Lemma 26.12 shows that

$$\Delta(J) \subset \Delta'(J') \tag{27.13}$$

except when $q < 20$. We check the cases when $q < 20$ by hand, using Billiard King.

The distance between the left endpoint of J' and the left endpoint of J is $q_- - q'$. The same is true for the right endpoints. As long as $q_- - q' \geq 2$, the argument in the proof of Lemma 19.4 shows that $P\Gamma \subset \Delta(J)$. The point is that Equation 19.9 is replaced by

$$\frac{-q'}{1 + A'} \geq -q', \tag{27.14}$$

which is always true. When $q_- = q' + 1$, we must have $p = 1$. In this case, the pivot points are $E_- = (-1, 1)$ and $E_+ = (0, 0)$. This case is trivially true.

Remark: The reader might wonder why we have so much slack in the (supposedly) tightest possible situation. The slack comes from the fact that, in Lemma 26.8, the arc γ is well inside the parallelogram R_2. For the sake of robustness, we mention that any small size of $q_- - q'$ leads to a similar proof.

27.4 PROOF OF THE STRUCTURE LEMMA

We will consider the case when $A_- \Leftarrow A$. The other case is similar. Let B be the odd rational such that $A_- \models B$. Then $P\Gamma(A_-) = P\Gamma(B)$, by definition.

Lemma 27.5 *The Structure Lemma holds when* $1/1 \leftarrow A$.

Proof: In this case

$$A = \frac{2k-1}{2k+1}, \qquad A_- = \frac{k-1}{k}, \qquad B = \frac{2k-3}{2k-1}. \tag{27.15}$$

Then $P\Gamma(A)$ is the line segment connecting $(0,0)$ to $(-k,k)$, and $P\Gamma(B)$ is the line segment connecting $(0,0)$ to $(-k+1, k-1)$. □

In all other cases, we have $A' \leftarrow A$, where $A' \neq 1/1$. As in Lemma 17.2, let

$$\delta = \delta(A', A) = \mathrm{floor}(q'/q).$$

Lemma 27.6 *If* $\delta = 1$*, then the Structure Theorem holds.*

Proof: If $\delta(A', A) = 1$, then $d(A', A) = 0$. If $d(A', A) = 0$, then $P\Gamma = P\Gamma'$ by the Copy Theorem and the definition of pivot arcs. At the same time, we can apply Lemma 17.2 to the pair $A_m = A'$ and $A_{m+1} = A$. Since $\delta(A', A) = 1$, we must have Case 1 or Case 3. But we also have $A_- < A_+$. Hence we have Case 3. But then $A'_- = A_-$. Hence we can replace the pair (A_-, A) by the pair (A'_-, A'), and the result follows by induction on the size of the denominator of A. □

Lemma 27.7 *Suppose that* $\delta = 2$*. Then* $A' = B$.

Proof: B is characterized by the property that A_- and B are Farey-related, and

$$2q_- > \mathrm{denominator}(B) > q_-.$$

We will show that A' has this same property. Note that A' and A_- are Farey-related. The equations

$$2q' < q, \qquad q = q_+ + q_-, \qquad q' = q_+ - q_-$$

lead to

$$3q_- > q_+ \qquad \Longrightarrow \qquad 2q_- > (q_+ - q_-) = q'.$$

This establishes the first property for A'. The fact that $\delta = 2$ gives $3q' > q$. This leads to

$$q_+ > 2q_-, \qquad \Longrightarrow \qquad q' = q_+ - q_- > q_-.$$

This is the second property for A'. □

Lemma 27.8 *Suppose $\delta \geq 3$. Then $A' \leftarrow B$.*

Proof: There is some even rational C such that

$$B = A_- \oplus C. \tag{27.16}$$

The denominator of C is smaller than the denominator of A_- because of the fact that $A_- \models B$. The inferior predecessor of B is $A_- \ominus C$. At the same time,

$$A' = A_+ \ominus A_-. \tag{27.17}$$

So, we are trying to show that $A_+ \ominus A_- = A_- \ominus C$. This is the same as showing that

$$C = D = A_- \oplus A_- \ominus A_+. \tag{27.18}$$

Since A_+ and A_- are Farey-related, D and A_- are Farey-related. We claim that

$$2q_- - q_+ = \text{denominator} \in (0, q_-). \tag{27.19}$$

The upper bound comes from the fact that $q_+ > q_-$. The lower bound comes from the fact that $q_+ < 2q_-$. To see this last equation, note that

$$q = q_+ + q_-, \qquad q' = q_+ - q_-, \qquad 3q' < q.$$

But C is the only even rational that is Farey related to A_- and satisfies equation 27.19. Hence $C = D$. □

As we have already proved, the case $\delta = 1$ is handled by induction on the denominator of A. The case $\delta = 2$ gives

$$P\Gamma_- = P\Gamma'.$$

In this case, the Structure Lemma follows from the definition of the pivot points.

When $\delta \geq 3$, the rational A' is a common inferior predecessor of A and B. Since $A_+ = A' \oplus A_-$ and $A_- < A_+$, we have $A' > A_+$. Hence $A' > A$.

Lemma 27.9 $A' > B$.

Proof: Lemma 27.8 gives

$$A' = A_- \ominus C, \quad A_+ = A' \oplus A_-, \quad A = A_+ \oplus A_-, \quad B = A_- \oplus C. \tag{27.20}$$

This gives

$$A \ominus B = A_+ \ominus C = A' \oplus A_- \ominus C = A' \oplus A'.$$

Hence

$$A = B \oplus A' \oplus A'. \tag{27.21}$$

Since $A_+ = A' \oplus A_-$ and $A_- < A_+$, we have $A' > A_+$. Hence $A' > A$. By Equation 27.21, A lies between A' and B. Hence $B < A < A'$. Hence $A' > B$. In short, $A' > A$ and $A' > B$. □

Finally, from the definition of pivot points, we have $E^+(A) = E^+(B)$. This establishes statement 1. Statement 2 has a similar proof.

Now for statement 3. By Lemma 26.1,

$$E^+(A) + E^-(A) = -A_- + (0, 1), \qquad E^+(B) + E^-(B) = -B_+ + (0, 1).$$

Since $E^+(A) = E^+(B)$, we have

$$E^-(B) - E^-(A) = A_- - B_+ = V(C) \tag{27.22}$$

Here $V(C)$ is, as in Equation 3.2, defined relative to C. We now have

$$\begin{aligned} & E^-(A) + V - E^-(A_-) \\ =& E^-(A) - E^-(B) + V \\ =& -V(C) + V(A) \\ =& V(A_+ \oplus A_-) - V(A_+ \ominus A_-) \\ =& 2V(A_-) \in \mathbf{Z}(V_-). \end{aligned} \tag{27.23}$$

This completes the proof of statement 3. The proof of statement 4 is similar.

An Even Version: Now that we have established the Structure Lemma, we prove a variant. For each even rational $A_2 \in (0, 1)$ that is not of the form $1/q_2$, there is another even rational $A_1 = p_1/q_1 \in (0, 1)$ such that $q_1 < q_2$ and $A_1 \vdash A_2$.

Lemma 27.10 *The Structure Lemma holds for the pair* (A_1, A_2).

Proof: We will deduce this new version of the Structure Lemma from the original version we have just finished proving.

Consider statement 1. Let

$$A_3 = A_1 \oplus A_2. \tag{27.24}$$

Then

$$A_1 \Leftarrow A_3, \qquad A_2 \models A_3 \tag{27.25}$$

Note that $E_2^+ = E_3^+$ by definition. Also, $E_1^+ = E_3^+$, by the Structure Lemma. Hence $E_1^+ = E_2^+$. This proves statement 1 for the pair (A_1, A_2). Statement 2 has the same kind of proof.

Consider statement 3. We have $E_2^- = E_3^-$ and

$$E_3^- - E_1^- + V_3 \in \mathbf{Z}V_1. \tag{27.26}$$

On the other hand

$$V_3 = V_2 + V_1, \qquad \Longrightarrow \qquad E_3^- - E_1^- + V_2 \in \mathbf{Z}V_1. \tag{27.27}$$

The first equation implies the second. But $E_3^- = E_2^-$. This completes the proof of statement 3. Statement 4 has the same kind of proof. □

27.5 THE DECREMENT OF A PIVOT ARC

Here we work out the precise relationship between the pivot arcs in the Structure Lemma. One can see the structure we describe here in Figure 1.5.

Let A be an odd rational and let A' be the superior predecessor of A. By the Copy Theorem, $P\Gamma$ contains at least one period of Γ' starting from either end. Let γ' be one period of Γ' starting from the right endpoint of $P\Gamma$. We define $DP\Gamma$ by the following formula.

$$P\Gamma = DP\Gamma * \gamma'. \tag{27.28}$$

The operation on the right hand side of the equation is the concatenation of arcs. We call $DP\Gamma$ the *decrement* of $P\Gamma$.

The arc $DP\Gamma$ is a pivot arc relative to a different parameter. (See the next lemma.) $DP\Gamma$ is obtained from $P\Gamma$ by deleting one period of Γ'. Now we present an *addendum* to the Structure Lemma.

Lemma 27.11 *If $B \Leftarrow A$, then $P\Gamma(B) = DP\Gamma(A)$, up to translation.*

Proof: We will consider the case when $A_- \Leftarrow A$. The other case, when $A_+ \Leftarrow A$, has essentially the same proof. We reexamine Lemmas 27.7 and 27.8. In Lemma 27.7, we have

$$P\Gamma_- = P\Gamma'.$$

However, in this case, $\delta(A, A') = 2$, and from the definition of pivot points, we see that $P\Gamma$ is obtained from $P\Gamma'$ by concatenating a single period of Γ'. This gives us what we want.

In Lemma 27.8, we have Equation 27.21, which implies

$$\text{denominator}(A) = \text{denominator}(B) + 2q'. \tag{27.29}$$

But this implies that $d(A', A) = d(A', B) + 1$. Applying the Copy Theorem to both pairs, we see that $P\Gamma$ is obtained from $P\Gamma'$ by concatenating $d(A, B) + 1$ periods of Γ' where $P\Gamma_-$ is obtained from $P\Gamma'$ by contatenating $d(A', B)$ periods of Γ'. This gives us the desired relationship. $\square$

27.6 AN EVEN VERSION OF THE COPY THEOREM

Let A_2 be an even rational. We write $A_2 = A_0 \oplus A_1$, where A_0 is odd and A_1 is even.

Lemma 27.12 $P\Gamma_2 \subset \Gamma_0$.

Proof: We have $P\Gamma_2 = P\Gamma(A_3)$, where A_3 is the odd rational such that $A_2 \models A_3$. Since $A_1 \vdash A_2$ and both A_1 and A_2 are even, we have $A_3 = A_1 \oplus A_2$. At the same time, we have $A_0 = A_2 \ominus A_1$. Hence $A_0 \leftarrow A_3$. But now we can apply the Copy Theorem to the pair (A_0, A_3) to conclude that $P\Gamma_3 \subset \Gamma_0$. But $P\Gamma_3 = P\Gamma_2$. $\square$

This page intentionally left blank

Chapter Twenty-Eight

Proof of the Pivot Theorem

28.1 AN EXCEPTIONAL CASE

We first prove the Pivot Theorem for $A = 1/q$. This case does not fit the pattern we discuss below.

Let Γ be the arithmetic graph associated to $A = 1/q$ and let $P\Gamma$ denote the pivot arc. In all cases, $P\Gamma$ contains the vertices $(0, 0)$ and $(-1, 1)$. These vertices correspond to the two points

$$\left(\frac{1}{q}, -1\right), \qquad \left(2 - \frac{1}{q}, -1\right). \tag{28.1}$$

These two points are the midpoints of the special intervals

$$I_1 = \left(0, \frac{2}{q}\right) \times \{-1\}, \qquad I_2 = \left(2 - \frac{2}{q}, 2\right) \times \{-1\}. \tag{28.2}$$

By *special interval*, we mean intervals in the sense of §2.2. Recall from that section that these special intervals are permuted by the outer billiards map.

The special intervals in Equation 28.2 appear at either end of

$$I = [0, 2] \times \{-1\}. \tag{28.3}$$

For any $A < 1/2$, the phase portrait in Figure 2.5 (repeated here for convenience) shows that the interval

$$I_3 = (2A, 2 - 2A) \times \{-1\} \tag{28.4}$$

returns to itself under one iterate of Ψ. When $A = 1/q$, we have

$$I - I_3 = I_1 \cup I_2. \tag{28.5}$$

But then the orbit of I_1 intersects I only in $I_1 \cup I_2$. Hence the only low vertices on Γ are equivalent to $(0, 0)$ and $(-1, 1)$ modulo translation by $V = (-q, 1)$. This establishes the Pivot Theorem for $A = 1/q$.

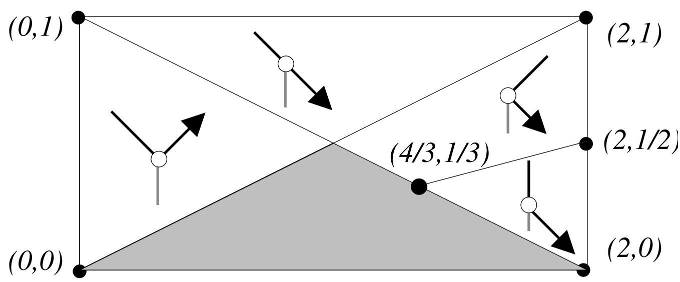

Figure 28.1: Low-vertex phase portrait. (Repeat of Figure 2.5.)

28.2 DISCUSSION OF THE PROOF

Now we consider the general case of the Pivot Theorem. We will not consider the odd case until the last section of the chapter. At the end, we will explain the minor differences in the even case.

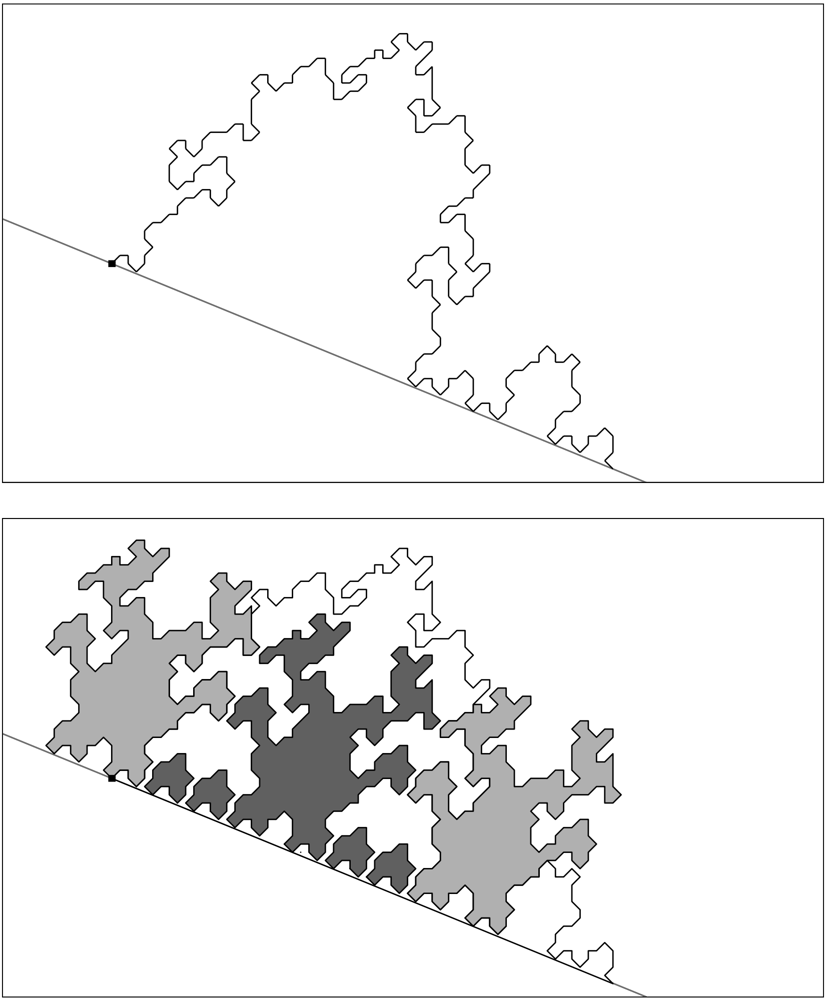

Figure 28.2: Components of $\widehat{\Gamma}(25/61)$ and $\widehat{\Gamma}(9/22)$.

The top of Figure 28.2 shows one period of $\Gamma(25/61)$ and the bottom shows an enhanced version of Figure 27.2. The light-gray regions are the eggs we discussed in connection with Figure 27.2. These are components of $\Gamma(9/22)$. The dark-gray components lie underneath the bump. (See below for a formal definition.) There is one large dark-gray component and 4 small ones. These dark-gray components, it turns out, belong to both $\widehat{\Gamma}(25/61)$ and $\widehat{\Gamma}(9/22)$.

For any odd rational $A_2 \neq 1/q_2$, we have $A_1 \Leftarrow A_2$, where $A_1 \in (0,1)$ is an even rational. What we mean is that A_1 and A_2 are Farey-related and $2q_1 < q_2$. See §27.1 for details. We will argue by induction, assuming that the Pivot Theorem is true for A_1.

Now we introduce some notation.

- The *bump* is the arc γ of Γ connecting $P\Gamma$ to either $P\Gamma + V$ or $P\Gamma - V$. We write $H(\Gamma)$. Whether we take γ to lie on the left or the right depends on the rationals involved. In any case, $P\Gamma \cup \gamma$ is one period of Γ.
- A *low component* of $\widehat{\Gamma}_1$ is a component that contains a low vertex.
- A *major low component* of $\widehat{\Gamma}_1$ is a low component that is a translate of Γ_1.
- We call the other low components of $\widehat{\Gamma}_1$ *minor components*.
- The *eggs* are the two major components of $\widehat{\Gamma}_1$ that contain the endpoints of the bump. The Structure Lemma guarantees that these components are major.

Figure 28.3 shows an abstract and slightly generalized version of Figure 28.2. We will base the discussion on Figure 28.3, but we will use Figure 28.2 as a reality check. The numbered regions are major components of $\widehat{\Gamma}_1$. The small dark-gray regions are minor components of $\widehat{\Gamma}_1$. The regions labelled 0 and 4 are the eggs, as discussed above. The black arc is the bump. Lemma 27.4 gives a large region Δ where $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ agree. Δ is white.

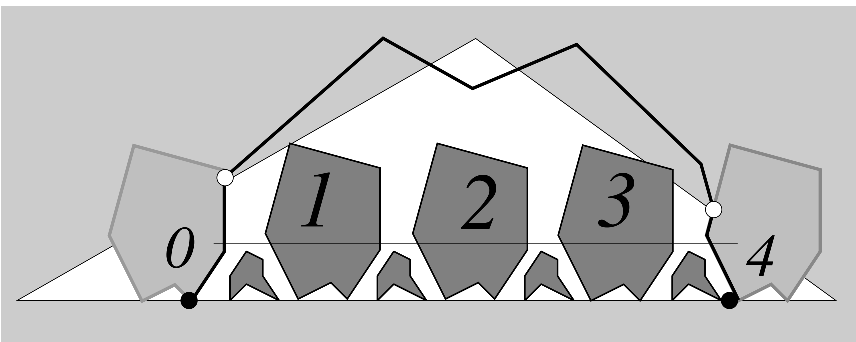

Figure 28.3: Cartoon view of the proof.

We want to determine that the bump has no low vertices except for its endpoints. By the Structure Lemma, the endpoints of the bump are also endpoints of the pivot arcs of C_0 and C_4. By induction, the only low vertices of C_0 and C_4 are contained in the pivot arcs. These pivot arcs are on the other sides of the endpoints we are considering. Hence there are no low vertices on the black arc as long as it coincides with either C_0 or C_4.

There is one subtle point to our argument. When we refer to *low vertices* of the black arc, the vertices are low with respect to the parameter A_2. However, when we refer to low vertices of C_0 and C_4, the vertices are low with respect to A_1. We will discuss this subtle point in the next section. What saves us is that the two notions of *low* coincide, because of the way in which A_1 approximates A_2.

So, either end of the black arc starts out well: It rises away from the baseline. This is exactly the situation we discussed in the last chapter in connection with Figure 27.2. Once the bump gets off the ground, what could go wrong? Answer: One of the ends could dip back down into Δ and (at the boundary) merge with a component of $\widehat{\Gamma}_1$. In other words, some component of $\widehat{\Gamma}_1$ would have to stick out of Δ.

We will analyze the various possibilities in turn. We distinguish 3 basic cases.

The End Major Components: These are the components labelled C_1 and C_3 in Figure 28.3. In Figure 28.2, the single large component is the only end major component. These components seem to give us the most trouble because they come closest to sticking out of Δ. In fact, we cannot show that these components are contained in Δ even though experimentally we can see that this is true. However, Lemma 2.6 comes to the rescue. The low vertices on these components have odd parity, and the low vertices on the bump have even parity. We will see that this implies that the bump cannot merge with C_1 and C_3. The parity argument steps in where geometry fails.

The Middle Major Components: This is the component labelled C_2 in Figure 28.3. In Figure 28.2 there are no middle major components even though the large dark-gray component there sits in the middle in some obvious sense. In general, there are $n+1$ major components and $n-1$ middle major components. The middle major components are much farther inside Δ. We will show that the other major components are contained entirely inside Δ.

Minor Components: These are the remaining small dark-gray components in Figure 28.3. The Barrier Theorem from Chapter 14 handles these. The black horizontal line in Figure 28.3 represents the barrier which no minor component can cross. Equipped with the Barrier Theorem, we will be able to show that all minor components lie in Δ. The barrier line keeps them from sticking out.

This takes care of all the potential problems. Since the bump cannot merge with any of the small dark-gray components, it just skips over everything and has no low vertices except for its endpoints. As with the proof of the Decomposition Theorem, the estimates we make are true by a wide margin when A_1 is large. However, when A_1 is small, the estimates are close and we need to consider the situation in a case-by-case way. We hope that this dealing with small cases does not obscure the basic ideas in the proof.

Remark: As we remarked above, it seems that $\widehat{\Gamma}_2$ copies all the low components of $\widehat{\Gamma}_2$ that lie between the two endpoints of the bump. In light of what we said in the case-by-case analysis, we will show that this is true except perhaps for the end major components. Our methods are not quite good enough to get these as well. This deficiency in our methods causes our proofs to be more complicated in a few places.

28.3 CONFINING THE BUMP

We continue with the notation from the previous section. For ease of exposition, we assume that $A_1 < A_2$. The other case is similar. For ease of notation, we set $A = A_2$. Until the end of this section, we consider only A. We write one period of Γ as $P\Gamma \cup \gamma$. Here $P\Gamma$ is the pivot arc, and γ is the bump considered in the previous section.

Let W be the vector from Equation 3.2. Let S be the infinite strip whose left edge is the line through $(0, 0)$ parallel to W and whose right edge is the line through V_+ and parallel to W. Here $V_+ = (q_+, -p_+)$, and p_+/q_+ is as in Equation 4.1. Figure 28.4 is a schematic picture.

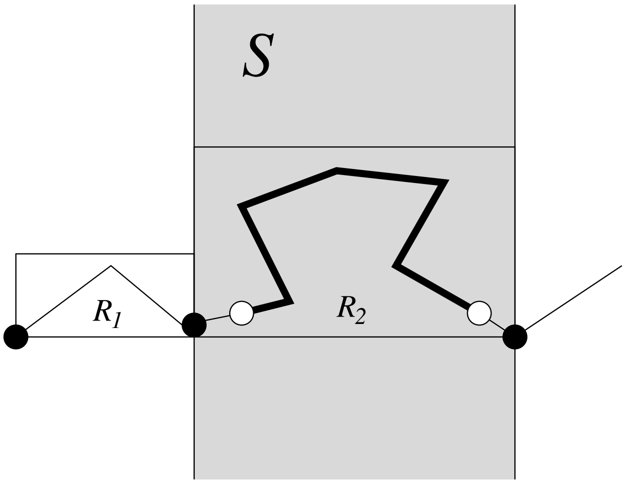

Figure 28.4: The strip S contains the bump.

Lemma 28.1 *γ does not cross the lines bounding S.*

Proof: The lines of S are precisely the extensions of the sides of R_2, the larger of the two parallelograms from the Decomposition Theorem. We know that Γ crosses these lines only once. These are the black dots shown in Figure 28.4. The thick arc represents γ. By Lemma 26.2 and symmetry, both endpoints of γ belong to R_2. These are the white dots in Figure 28.4. The endpoints of γ occur between the crossing points. Since there are no other crossings, $\gamma \subset R_2$. Hence $\gamma \subset S$.

Now we can clear up the subtlety mentioned in the previous section. We set $S_2 = S$, the strip defined relative to the odd rational A_2.

Lemma 28.2 *A vertex in S is low with respect to A_1 iff it is low with respect to A_2. Hence a vertex of γ is low with respect to A_1 iff it is low with respect to A_2.*

Proof: Let L_j denote the baseline with respect to A_j. The conclusion of this lemma is equivalent to the statement that there does not exist a lattice point between $L_1 \cap S$ and $L_2 \cap S$. This is a consequence of our proof of Lemma 27.4. $\square$

28.4 A TOPOLOGICAL PROPERTY OF PIVOT ARCS

Let A be a rational kite parameter, either even or odd. Let $P\Gamma$ denote the pivot arc of $\Gamma = \Gamma(A)$. The two endpoints of $P\Gamma$ are low vertices. Here we prove a basic structural result about $P\Gamma$.

Lemma 28.3 *$P\Gamma$ contains no low vertex to the right of its right endpoint. Likewise, $P\Gamma$ contains no low vertex to the left of its left endpoint.*

Proof: We will prove the first statement. The second statement has the same proof. We argus as in the proof of Lemma 2.6. Note that Γ right-travels at $(0, 0)$. Hence $P\Gamma$ right-travels at its right endpoint ρ. Suppose that $P\Gamma$ contains a low vertex σ to the right of ρ. Then some arc β of $P\Gamma$ connects ρ to σ. Since Γ right-travels at ρ, some arc γ of $\Gamma - P\Gamma$ enters into the region between ρ and σ and beneath β. But γ cannot escape from this region, by the Embedding Theorem. The point here is that γ cannot squeeze beneath a low vertex because the only vertices below a low vertex are also below the baseline. Figure 28.5 shows the situation.

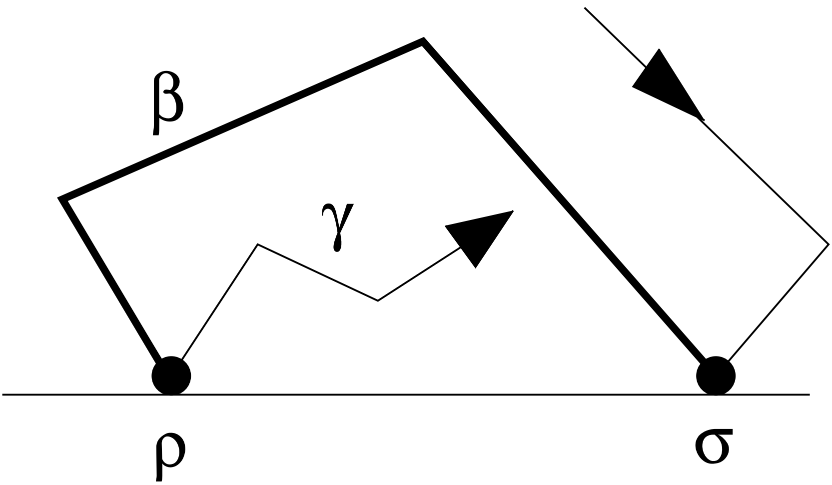

Figure 28.5: $P\Gamma$ creates a pocket.

In the odd case we have an immediate contradiction. In the even case, we see that there must be a loop containing both ρ and σ. This loop must be a closed polygon and a subset of $P\Gamma$. Since $P\Gamma$ is also a closed (and embedded) polygon, the loop must equal $P\Gamma$. But by definition, $P\Gamma$ lies below $\Gamma - P\Gamma$. From Figure 28.4, we see that $P\Gamma$ (which contains β) in fact lies above $\Gamma - P\Gamma$ (which contains γ). This is a contradiction. $\square$

28.5 COROLLARIES OF THE BARRIER THEOREM

Here we derive a few corollaries of the Barrier Theorem. See Chapter 14 for the statement. Let L_0 be the line through $(0,0)$ and parallel to the vector W, from Equation 3.2. Referring to our proof of statement 2 of the Hexagrid Theorem, L_0 is the wall line we considered in detail.

In this section we will suppose that A is an even rational parameter. Let $\widehat{\Gamma} = \widehat{\Gamma}(A)$ be the corresponding arithmetic graph.

Corollary 28.4 *A minor component of* $\widehat{\Gamma}$ *cannot cross* L_0.

Proof: Our line is one of the lines in the Hexagrid Theorem. By the Hexagrid Theorem, only Γ crosses this line beneath the barrier, and the crossing takes place at $(0,0)$. By definition, Γ is a major component. □

We are trying to construct a parallelogram that bounds the minor components. The baseline contains the bottom edge. The barrier contains the top edge. The line in Corollary 28.4 contains the left edge. Now we supply the right edge. Actually, there are many choices for this right edge.

Lemma 28.5 *Let* $V_+ = (q_+, -p_+)$. *Let* L *be the line parallel to* L_0 *and containing the point* $V_+ + kV$ *for some* $k \in \mathbf{Z}$. *A minor component cannot cross* L.

Proof: Since $\widehat{\Gamma}$ is invariant under translation by V, it suffices to prove this result for $k = 0$. Let L be the line through V_+ parallel to L_0. Our result follows from Corollary 28.4 and the rotational symmetry we established in §12.3.

Let Λ be the barrier. Consider the symmetry ι defined in §12.3. The two lines Λ and $\iota(\Lambda)$ are equally spaced above and below the baseline up to an error of at most $1/q$. Suppose that some minor component β crosses the line L. Then the component $\iota(\beta)$ crosses the line $\iota(L)$. But $\iota(L)$ is the line from Lemma 28.4. Inspecting the hexagrid, we see that $\iota(L)$ contains the door $(0,0)$, but no other door between the baseline and $\iota(\Lambda)$. Indeed, the doors above and below the baseline are just about evenly spaced from $(0,0)$ going in either direction. See Figure 3.2, a representative figure. (In this figure, we are talking about the long axis of the kite, and $(0,0)$ is the bottom tip of the kite.)

The component γ' of $\widehat{\Gamma}$ that crosses $\iota(L)$ near $(0,0)$ is the same size as Γ. Hence this component crosses through $\iota(\Lambda)$. Hence $\iota(\gamma')$ is a major component. Hence $\beta \neq \iota(\gamma')$. Hence $\iota(\beta) \neq \gamma$. Hence $\iota(\beta)$ does not cross $\iota(L)$. Hence β does not cross L. □

Now that we have found some parallelograms that completely confine the minor components, we will embed this picture, so to speak, in our proof of the Pivot Theorem. This requires us to juggle two parameters at once.

28.6 THE MINOR COMPONENTS

28.6.1 The Minor Box

In our proof of the Pivot Theorem, we have two parameters $A_1 \Leftarrow A_2$. As above, we focus our attention on the case when $A_1 < A_2$. The other case involves a completely parallel discussion. See §30.3 for a brief discussion of the other case.

Lemma 28.5 applies to vectors defined in terms of A_1, but ultimately we would like to make a statement about the parameter A_2. So, we would like to translate the information in Lemma 28.5 into a statement about some lines that are defined (partly) in terms of A_2. Let $(V_j)_+$ be as in §28.3. Then Lemma 28.5 applies to the vectors of the form

$$(V_1)_+ + kV_1. \tag{28.6}$$

However, we are also interested in the vector $(V_2)_+$.

Lemma 28.6 *Suppose that $A_1 < A_2$. Then, there is some integer k such that $(V_2)_+ = (V_1)_+ + kV_1$.*

Proof: We set $A = A_2$. Then $A_- = A_1$. Let A_{-+} denote the parameter that relates to A_- in the same way that A_+ relates to A. That is, $A_{-+} > A_-$ are Farey-related and A_{-+} has a smaller denominator than A_-. We want to prove that $V_+ = V_{-+} + kV_-$ for some k. The rationals A_{-+} and A_- are Farey-related. Therefore so are the parameters

$$A_-, \qquad\qquad A_{-+} \oplus A_- \oplus \cdots \oplus A_-. \tag{28.7}$$

Here we are doing Farey addition. Conversely, if any rational A' is Farey-related to A_- and has a larger denominator, then $A' \ominus A_-$ is also Farey-related to A_-. Thus the rationals in Equation 28.7 account for all the rationals A' with the properties just mentioned. But A_+ is one such rational. Hence A_+ has the form given in Equation 28.7. This completes the proof. □

Let R denote the parallelogram defined by the following lines.

- The baseline relative to A_1.
- The barrier for A_1.
- The line parallel to W_1 through $(0, 0)$.
- The line parallel to W_1 through $(V_2)_+$.

Then any minor component with one vertex in R stays completely in R. This is a consequence of the Barrier Theorem, its corollaries, and the lemma in this section. Modulo a tiny adjustment in the slopes, the left and right edges of R are contained in the left and right edges of the strip S considered in §28.3. We call R the *minor box*.

28.6.2 Trapping the Minor Components

We continue with $A_1 \Leftarrow A_2$, as above, and $A_1 < A_2$. Define

$$\Delta = \Delta_1(I) \cup \Delta_2(I), \qquad I = [-q_1 + 2, q_2 - 2]. \tag{28.8}$$

Here Δ is as in Lemma 27.4, the second Diophantine lemma. Let R be the minor box.

Lemma 28.7 *Let $\beta \subset \widehat{\Gamma}_1$ be any component that is contained in R. Then $\beta \subset \widehat{\Gamma}_2$.*

Proof: Our proof follows the same strategy as in the Decomposition Theorem. We will work with the functionals G_1 and H_1 defined relative to A_1. We want to show $R \subset \Delta$ and apply Lemma 27.4. To avoid a messy calculation, we use the Mismatch Principle from Chapter 19. We replace R by the nearby parallelogram $\tilde{R}$ with vertices

$$(0,0), \qquad \lambda W_1, \qquad (V_2)_+, \qquad (V_2)_+ + \lambda W_1. \tag{28.9}$$

The constant λ has the following definition. The top left vertex of R lies on the line through $(0,0)$ and parallel to W_1, as discussed above. Hence this vertex has the form λW_1. We compute

$$M_1(\lambda W_1) = p_1' + q_1' < p_1 + q_1 = M_1(W). \tag{28.10}$$

Hence $\lambda < 1$. Here A_1' is the rational that appears in the Barrier Theorem. The point here is that the barrier contains the point $(0, (p_1' + q_1')/2)$.

Let u and w be the top left and top right vertices of R, respectively. As usual, it suffices to show that the quantities

$$G_1(u) - (-q_1 + 2) > 0, \qquad (q_2 - 2) - H_1(w) > 0. \tag{28.11}$$

By affine symmetry (or a calculation, as we do), these quantities are equal. We compute

$$G_1(u) - (-q_1 + 2) = q_1 - \lambda \frac{q_1^2}{p_1 + q_1} - 2 \tag{28.12}$$

By Lemma 28.6, we have

$$(V_2)_+ + V_1 = (V_2)_+ + (V_2)_- = V_2 \qquad \Longrightarrow \qquad V_2 - w = V_1 - \lambda W_1.$$

The first equation implies the second. Hence

$$\begin{aligned} &(q_2 - 2) - H_1(w) \\ &= -2 + H_1(V_2 - w) \\ &= -2 + H_1(V_1 - \lambda W_1) \\ &= q_1 - \lambda \frac{q_1^2}{p_1 + q_1} - 2. \end{aligned} \tag{28.13}$$

Since $\lambda \le 1$, the quantities in Equation 28.11 are nonnegative as long as $p_1 \ge 3$ and $q_1 \ge 7$. This is exactly the same estimate as in Lemma 19.3. When $p_1 = 2$, we see that

$$p_1' = 1, \qquad q_1' = \frac{q_1 - 1}{2}.$$

Thus $\lambda \approx 1/2$, and we get massive savings. When $p_1 \ge 2$ and $q_1 \le 7$, we check the cases by hand using the same trick as in §19.5.

When $p_1 = 1$, the graph $\widehat{\Gamma}_1$ has no minor components, as we saw in §28.1. □

28.7 THE MIDDLE MAJOR COMPONENTS

We keep the parameters $A_1 \Leftarrow A_2$ as above, with $A_1 < A_2$. We have already defined the pivot points of Γ_1. We define the pivot points of the translates $C_k = \Gamma_1 + kV_1$ in the obvious way, by translation.

By the Structure Lemma, there is some component C_k whose left pivot point is $E_2^- + V_2$, the right endpoint of the bump. The components $C_0, ..., C_k$ are exactly as in §28.2. By Lemma 2.6, the index k is even. More generally, C_j contains low vertices of even parity if and only if j is even.

As in §28.2, we are interested in bounding the components $C_2, ..., C_{k-2}$. Actually, we care only about the even components, but the bound works equally well for the odd components between C_2 and C_{k-2}. If $k = 2$, as in Figure 28.2, this section is vacuous.

By the Hexagrid Theorem, C_0 is contained in the parallelogram R_0 with vertices

$$-V_1, \qquad -V_1 + 2W_1, \qquad V_1 + 2W_1, \qquad V_1. \tag{28.14}$$

This means that C_j is contained in the translated parallelogram

$$R_j = R_0 + jV_1 \tag{28.15}$$

We choose $j \in \{2, ..., k-2\}$.

Here we describe some features of R_j, as well as a method for symmetrizing it.

1. The bottom edge of R_j is contained in the line through $(0, 0)$ and is parallel to V_1–i.e., the baseline– as usual.

2. The top edge of R_j is contained in the line through $2W_1$ and is parallel to V_1. These lines are independent of j.

3. The left edge of R_j is parallel to, and to the right of, the line Λ parallel to W_1 and containing V_1. When $j = 2$, the left edge of R_j is contained in Λ.

4. The same argument as in Lemma 28.5 shows that C_2 lies to the left of the line through $(V_2)_+ - V_1$ and parallel to W_1. Referring to the symmetry ι in Lemma 28.5, this is the line $\iota(\Lambda)$. In brief, if C_j crosses $\iota(\Lambda)$, then $\iota(C_j)$ crosses Λ, and this contradicts the Hexagrid Theorem, applied below the baseline. So, $\iota(\Lambda)$ is the fourth line bounding the symmetrized parallelogram R.

Let R be the parallelogram defined by the 4 lines above. By construction, $C_j \subset R$ for $j \in \{2, ..., k-2\}$. We call R the *major box*.

Lemma 28.8 *Let $\beta \subset \Gamma_1$ be any component of $\widehat{\Gamma}_1$ that is contained in R. Then $\beta \subset \widehat{\Gamma}_2$.*

Proof: The proof is exactly the same. Let u and w denote the top left and top right vertices of R. we have the same symmetry as in the previous bound, and so we just have to compute $G_1(u) \geq -q_1 + 2$. We compute

$$G_1(u) - (-q_2 + 2) = 2q_1 - \frac{2q_1^2}{p_1 + q_1} - 2. \tag{28.16}$$

This time we always get a positive number, though in small cases it is pretty close. □

28.8 EVEN IMPLIES ODD

Having assembled all the necessary technical ingredients, we now formalize the discussion we gave in §28.2. We will present an inductive proof of the Pivot Theorem. This section contains half the proof, and the next section contains the other half. Again, we assume that $A_1 < A_2$. Let $P(A)$ be the statement that the Pivot Theorem is true for A.

Lemma 28.9 *Let $A_1 \Leftarrow A_2$. Then $P(A_1)$ implies $P(A_2)$.*

Our proof follows the format of the discussion in §28.2. As in §28.3, we define the *complementary arc* $\gamma_2 \subset \Gamma$ to be the arc to the right of $P\Gamma_2$ such that $P\Gamma_2 \cup \gamma_2$ is one period of Γ_2. The endpoints of γ_2 are

$$E_2^+, \qquad\qquad E_2^- + V_2. \tag{28.17}$$

Here γ_2 is the bump in §28.2.

We say that a *spoiler* is a low vertex of γ_2 that is not an endpoint of γ_2. The Pivot Theorem is equivalent to the statement that there are no spoilers.

Let $L(\gamma_2)$ denote the left endpoint of γ_2. Likewise, let $R(\gamma_2)$ denote the right endpoint of γ_2.

Lemma 28.10 *Any spoiler lies between $L(\gamma_2)$ and $R(\gamma_2)$.*

Proof: We will show that any spoiler lies to the right of $L(\gamma_2)$. The statement that any spoiler lies to the left of $R(\gamma_2)$ is similar. By Lemma 28.1, all spoilers lie in the strip S_2. But $P\Gamma_2$ crosses the left boundary of S_2. Any low vertices in S_2 to the left of $L(\gamma_2)$ lie either on $P\Gamma_2$ or beneath it. By the Embedding Theorem, γ cannot contain these vertices. $\square$

Recall that Δ is the region from Lemma 27.4. This is the white triangle in Figure 28.3.

Lemma 28.11 *Δ contains all the spoilers.*

Proof: We will work with the linear functionals G_2 and H_2 defined relative to A_2. Thus we are really showing that the smaller set $\Delta_2(I)$ contains all the spoilers.

Let $v = (m, n)$ be a spoiler. It suffices to prove that $G_2(v) \geq -q_0 + 2$ and $H_2(v) \leq q_2 - 2$. We have $m \geq 1$. Since v is a low vertex, we have $n \leq 0$. We compute that $\partial_y G_2 < 0$. Hence

$$G_2(v) \geq G_2(m, 0) = m\frac{1 - A_2}{1 + A_2} > 0 \geq -q_1 + 2.$$

This takes care of G_2.

Let $w = v - V_2 = (r, s)$. By Lemma 18.1, it suffices to show that $H_2(w) \leq -2$. We compute $\partial_y H > 0$. Since w lies at most one vertical unit above the line of slope $-A_2$ through the origin, we have

$$H_2(w) \leq H_2(w'), \qquad\qquad w' = (r, -A_2 r + 1). \tag{28.18}$$

We compute

$$H_2(w') = r + \frac{2(1 - A_2)}{(1 + A_2)^2} < r + 2. \tag{28.19}$$

This shows that $H(w) < -2$ as long as $r \leq -4$. By Lemma 2.6, we have $r + s$ even. We just have to rule out $(-2, 2)$ and $(-3, 1)$ as spoilers.

Case 1: If $A_2 < 1/2$, then $(-2, 2)$ is not a low vertex. If $A_2 > 1/2$, then

$$\frac{2k-1}{2k+1} \leftarrow \cdots \leftarrow A_2$$

for some $k \geq 2$. In this case, E_2^- has first coordinate less than or equal to -2. But then $r \leq -3$. This rules out $(-2, 2)$.

Case 2: We compute that

$$A \geq \frac{1}{9} \quad \Longrightarrow \quad H_2(-3, 1) < -2.$$

When $A < 1/9$, we use the phase portrait in §2.6 to check that $\widehat{\Gamma}_2$ is trivial at $(-3, 1)$. This rules out $(-3, 1)$. □

Let v be a spoiler. By the previous result, there is some component β of $\widehat{\Gamma}_1$ that has v as a vertex.

Lemma 28.12 *β is not a subset of $\widehat{\Gamma}_2$.*

Proof: Suppose that $\beta \subset \widehat{\Gamma}_2$. Note that β is a closed polygon. Recall that γ_2 is the bump. Supposedly, γ_2 and β share the vertex v. Let us start at v and trace γ_2 in some direction. If the conclusion of this lemma is false, we remain simultanously on γ_2 and β until we loop around and return to v_2. This is because β is a closed polygon. This contradicts the fact that γ_2 never visits the same vertex twice. □

Here is the end of the argument. β cannot be a minor component, given the bound in §28.6.2. Next, $\beta \not\in \{C_2, ..., C_{k-2}\}$, given the bounds in §28.7. Next,

$$\beta \not\in \{C_1, C_{k-1}\}, \tag{28.20}$$

by Lemma 2.6. Next, $\beta \neq C_0$: By induction, all the low vertices of C_0 lie on PC_0. By Lemma 28.3, these low vertices all lie to the left of the spoiler. Likewise, $\beta \neq C_k$. We have exhausted all the possibilities. β cannot exist. Hence there is no spoiler. Therefore $P(A_2)$ holds.

We have shown that $P(A_1)$ implies $P(A_2)$ when $A_1 \Leftarrow A_2$ and A_1 is even and A_2 is even. We have given the proof under the assumption that $A_1 < A_2$, but the other case is essentially the same. See §30.3. It remains to consider the case when both A_1 and A_2 are even.

28.9 EVEN IMPLIES EVEN

28.9.1 A Decomposition Result

As a prelude to tackling the even case in the induction argument, we revisit the construction in §28.3, but for even parameters. Now A_1 and A_2 are both even parameters, with $A_1 \vdash A_2$. We set $A = A_2$ and consider just objects relative to A. We define the strip S exactly as in §28.3. For any set β, let $\beta^{\#}$ denote the translate $\beta + V$. We define

$$\gamma = (\beta \cup \beta^{\#}) \cap S, \qquad \beta = \Gamma - P\Gamma. \tag{28.21}$$

In Figure 28.6, the arc γ is the union of 2 thick arcs In Figure 28.6.

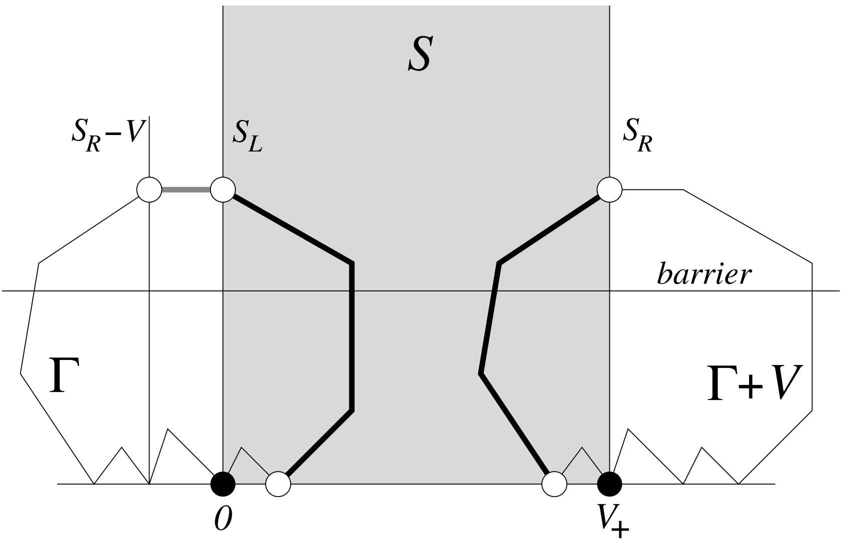

Figure 28.6: The even version of γ.

Lemma 28.13 *γ consists of two connected arcs. Any low vertex of $\Gamma - P\Gamma$ is translation-equivalent to a low vertex of γ.*

Proof: By the Hexagrid Theorem Γ crosses S_L only once. The door on S_L lies above the barrier line. Hence the crossing occurs above the barrier line. Likewise, $\iota(\Gamma + V)$ crosses S_L only once. The relevant door lies below the image of the barrier line under ι. Here ι is as in the proof of Lemma 28.1. But then $\Gamma + V$ crosses S_R only once, and the crossing occurs above the barrier line. Hence γ consists of 2 connected arcs.

The line $S_R - V$ is parallel to S_L and lies to the left of S_L. By symmetry, Γ crosses $S_R - V$ only once, and the crossing takes place above the barrier line. By the Barrier Theorem, the gray arc of Γ between S_L and $S_R - V$ lies above the barrier line and hence has no low vertices. Finally, any vertex of $\Gamma - P\Gamma$ not translation-equivalent to a vertex of γ lies on the gray arc of Γ between S_L and $S_R - V$. $\square$

28.9.2 The Induction Argument

Let $A_1 \vdash A_2$ be a pair of even rationals as in §27.4. This pair exists as long as $A_2 \neq 1/q_2$. Referring to the terminology in Lemma 28.9, we prove the following result in this section.

Lemma 28.14 *Let $A_1 \Leftarrow A_2$. Then $P(A_1)$ implies $P(A_2)$.*

We have already taken care of the base case in the induction, the case when $A = 1/q$. Lemmas 28.14 and 28.9 then imply the Pivot Theorem by induction. The proof is essentially the same as in the odd case, once we see that the basic structural results hold. The result in §27.4 gives us the even/even version of the Structure Lemma.

We consider the case when $A_1 < A_2$. The other case is similar. We define spoilers just as in the odd case. We just need to show that the arc γ_2 defined in the previous section has no spoilers. The same argument as in the odd case shows that a spoiler must lie between $L(\gamma_2)$ and $R(\gamma_2)$, the left and right endpoints, respectively.

Let Δ be the region of agreement between $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$, as above. The formulas are exactly the same. Here is the even version of Lemma 28.11.

Lemma 28.15 *Δ contains all the spoilers.*

Proof: The general argument in Lemma 28.11 works exactly the same here. It is only at the end, when we consider the vertices $(2, -2)$ and $(3, -1)$, that we use the fact that A_2 is odd. Here we consider these special cases again. The argument for $(-3, 1)$ does not use the parity of A_2. We have to consider just $(-2, 2)$.

If $A_2 < 1/2$, then $(-2, 2)$ is not a low vertex. We do not need to treat the extremely trivial case when $A_2 = 1/2$. When $A_2 > 1/2$, we have $A_1 > 1/2$ as well. The point is that no edge of the Farey graph crosses from $(0, 1/2)$ to $(1/2, 1)$. Hence $A_3 = A_1 \oplus A_2 > 1/2$ as well. But, by definition, the pivot points relative to A_2 are the same as for A_3. This is as in §27.4. Hence the same argument as in Lemma 28.11 now rules out $(2, -2)$. $\square$

Essentially the same argument as in the odd case now shows that γ_2 contains no spoilers.

The Pivot Theorem now follows from induction. This completes the proof.

Chapter Twenty-Nine

Proof of the Period Theorem

29.1 INHERITANCE OF PIVOT ARCS

Let A be some rational parameter. For each polygonal low component β of $\Gamma(A)$, we define the pivot arc $P\beta$ to be the lower arc of β that joins the two low vertices that are farthest apart. We say *lower arc* because all the components are closed polygons, and hence two arcs join the pivot points in all cases. When A is an even rational and $\beta = \Gamma$, this definition coincides with the definition of $P\Gamma$, by the Pivot Theorem. In general, we say that a pivot arc of Γ is a pivot arc of some low component of $\widehat{\Gamma}$. We call a pivot arc of $\widehat{\Gamma}$ *minor* if it is not a translate of $P\Gamma$.

For each rational in (0, 1), we are going to define an *odd predecessor* and an *even predecessor*. Aside from a few trivial cases, the predecessors exist and are rationals in (0, 1). The odd predecessor of A will be denoted by A', and we will use a single arrow, as in $A' \leftarrow A$. The even predecessor of A will be denoted by A'', and we will use the notation $A'' \Leftarrow A$. This notation should be compatible with our previous similar notation.

1. When A is odd, A' is as in the inferior sequence.
2. When A is odd, A'' is as in the Structure Lemma and Lemma 28.9.
3. When A is even, A' is as in the Barrier Theorem.
4. When A is even, A'' is as in Lemma 28.14.

It is worthwhile to mention another characterization of these numbers.

$$A \text{ even} \qquad \Longrightarrow \qquad A = A' \oplus A''. \tag{29.1}$$

$$A \text{ odd} \qquad \Longrightarrow \qquad A = A' \oplus A'' \oplus A''. \tag{29.2}$$

Just to cement the idea, we give an example.

$$\frac{3}{7} \leftarrow \frac{7}{17}, \qquad \frac{2}{5} \Leftarrow \frac{7}{17}, \qquad \frac{3}{7} \leftarrow \frac{5}{12}, \qquad \frac{2}{5} \Leftarrow \frac{5}{12}.$$

Here is our main technical tool for the Period Theorem.

Lemma 29.1 (Inheritance) *Let A be any rational. Suppose that*

$$A' \leftarrow A, \qquad A'' \Leftarrow A.$$

Then, every minor pivot arc β of $\widehat{\Gamma}$ is either a minor pivot arc of $\widehat{\Gamma}'$ or a pivot arc of $\widehat{\Gamma}''$. The set of low vertices of β is the same when considered either in A or in the relevant predecessor.

We first prove the odd case, and then we prove the even case. The proof is almost the same in both cases.

Proof in the Odd Case: Recall that $P\Gamma \cup \gamma$ is one period of Γ. There are 2 kinds of minor components of $\widehat{\Gamma}$.

1. Pivot arcs that lie underneath $P\Gamma$.

2. Pivot arcs that lie underneath γ.

We can push harder on Lemma 27.2. Since $P\Gamma$ lies in the set Δ, from Lemma 27.4, so does every low component of $\widehat{\Gamma}$ underneath $P\Gamma$. To see this, recall that our proof involved showing that $P\Gamma \subset \Delta$. But if a point of $P\Gamma$ lies in Δ, then so does the entire line segment connecting this point to the baseline. Hence all components of $\widehat{\Gamma}$ beneath $P\Gamma$ also belong to Δ. Hence the low components of $\widehat{\Gamma}$ lying underneath $P\Gamma$ coincide with the low components of $\widehat{\Gamma}'$ lying underneath $P\Gamma'$. This takes care of the first case.

Consider the second case. Our proof of Lemma 28.9 shows that every minor component of $\widehat{\Gamma}''$ lying inside $\Delta(A'', A)$ is contained in $\widehat{\Gamma}$. We showed the same result for every major component except the ones we labelled C_1 and C_{k-1}. Note that the pivot arcs are subject to the Barrier Theorem. That is, the two crossings from the Barrier theorem occur on the upper arcs rather than on the pivot arcs. Hence the pivot arcs behave exactly like the minor components. Hence the pivot arcs of C_1 and C_{k-1} are copied by $\widehat{\Gamma}$ even though the upper arcs might not be. By Lemma 28.11, every low vertex of $\widehat{\Gamma}$ lying underneath γ lies on the pivot arcs of the components we have just considered. This takes care of the second case.

There is only one detail we need to take care of. A vertex of the kind we are considering is low relative to A' or A'' if and only if it is low with respect to A. This follows from the basic property of Δ. See the geometric proof of Lemma 27.4. Thus every low component of $\widehat{\Gamma}$ of the kind we have considered is also low relative to $\widehat{\Gamma}'$ or $\widehat{\Gamma}''$, whichever is relevant. Likewise, the converse holds. $\square$

Proof in the Even Case: The minor pivot arcs of $\widehat{\Gamma}$ are of two kinds, those that lie underneath $P\Gamma$ and those that do not. By the same argument as in the odd case, the pivot arcs of the first kind are all minor pivot arcs of $\Gamma(A^*)$, where A^* is such that $A \bowtie A^*$. But then $A^* = A \oplus A''$. Hence $A'' \Leftarrow A^*$. At the same time, $A' = A \ominus A''$. Hence $A' \leftarrow A^*$. Applying the odd case of the Inheritance Lemma to the triple (A^*, A', A''), we see that every pivot arc of $\widehat{\Gamma}$ beneath $P\Gamma$ is a pivot arc of either $\widehat{\Gamma}'$ or $\widehat{\Gamma}''$. This takes care of the first case. The second case is just like the odd case. $\square$

Remark: Implicit in the definitions of *predecessor* is the idea of a *tree* of rationals. Each rational has 2 ancestors who are simpler in some sense. The Inheritance Lemma esplains how the traits – here meaning the pivot arcs – of the arithmetic graph for a complicated parameter are inherited from the ancestors.

29.2 FREEZING NUMBERS

Every rational parameter has an odd and an even predecessor. Starting with (say) an odd rational A, we can iterate the construction and produce a tree of simpler rationals. If B lies on this tree, we write $B \prec A$. Here is an immediate corollary of the Inheritance Lemma.

Corollary 29.2 *Every minor pivot arc of* $\widehat{\Gamma}(A)$ *is a pivot arc of* $\widehat{\Gamma}(B)$ *for some even* B *such that* $B \prec A$.

Let A be an odd rational. Let β be a minor component of $\widehat{\Gamma}(A)$. We define $F(\beta, A)$ to be the smallest denominator of a rational $B \prec A$ such that $P\beta$ is a pivot arc of $\widehat{\Gamma}(B)$. We call $F(\beta, A)$ the *freezing number* of β. Our terminology has the following meaning. As we move through the tree of rationals, from simple to complicated, various features of the corresponding graphs change, but at various states certain features freeze. The freezing number of a component marks the point when the component becomes a permanent feature.

Lemma 29.3 *The* Ψ*-period of a minor component* β *is at most*

$$20s^2, \qquad s = F(\beta, A).$$

Proof: This is an immediate consequence of the Hexagrid Theorem applied to the rational $B = r/s$ such that β is a component of $\widehat{\Gamma}(B)$. The Hexagrid Theorem confines β to a parallelogram of area less than $20s^2$. □

Let $x \in I$ correspond to a point not on $C(A_n)$. We let

$$F(x, n) = F(\beta_x, A_n),$$

where β_x is the component of $\widehat{\Gamma}_n$ corresponding to x. We say that a *growing sequence* is a sequence $\{x_n\}$ such that

$$F(x_n, n) \to \infty. \tag{29.3}$$

Recall that C_A is the Cantor set from the Comet Theorem.

Lemma 29.4 *Suppose every growing sequence accumulates on* C_A*. Then the Period Theorem is true for* A.

Proof: If the Period Theorem is false, then we can find a sequence of points $\{x_n\}$ in G_n such that the distance from x_n to C_n is uniformly bounded away from 0 and yet the period of x tends to ∞. But then Lemma 29.3 shows that $\{x_n\}$ is a growing sequence. By construction, $\{x_n\}$ does not have a limit point on C_A. □

29.3 THE END OF THE PROOF

Let $\{A_n\}$ be the odd sequence of rationals above. For each n, we can form the tree of predecessors, as above. Suppose we choose some proper function $m(n)$ such that $B_m \prec A_n$ is some even rational in the tree for A_n.

Lemma 29.5 $\lim_{n\to\infty} B_m = A$.

Proof: We consider the situation in the hyperbolic plane relative to the Farey triangulation. See §17.1 for definitions. We consider the portion G of the Farey graph consisting of edges having both endpoints in $[0, 1]$. We direct each edge in G so that it points from the endpoint of smaller denominator to the endpoint of larger denominator. The two endpoints never have the same denominator, so the definition makes sense. Say that the *displacement* of a directed path in G is the maximum distance between a vertex of the path and its initial vertex.

Given an $\epsilon > 0$, there are only finitely many vertices in G that are the initial points of directed paths having displacement greater than ϵ. This follows from the nesting properties of the half-disks bounded by the edges in G, and from the fact that there are only finitely many edges in G having a diameter greater than ϵ.

Given the nature of the tree of predecessors, there is a directed path in G connecting B_m to A_n. The displacement of this path tends to 0 as $n \to \infty$ because $\{B_m\}$ is an infinite list of rationals with only finitely many repeaters. Also, the distance from A_n to A tends to 0. Hence the distance from B_m to A tends to 0 by the triangle inequality. $\square$

Now we bring in an idea from the Rigidity Lemma. See §2.7. Let $\{B_m\}$ be any sequence of even rationals converging to the irrational parameter A. Then the Rigidity Lemma implies that the limits

$$\lim_{m\to\infty} \Gamma(A_m), \qquad \lim_{m\to\infty} \Gamma(B_m) \tag{29.4}$$

agree. In other words, longer and longer portions of $\Gamma(A_m)$ look like longer and longer pictures of $\Gamma(B_m)$. This is all we need to know from the Rigidity Lemma.

Now let $M_{m,A}$ be the fundamental map associated to A_m. This map is defined in Equation 2.10. In the proof of Theorem 1.6, we showed that

$$C_A = \lim_{m\to\infty} M_{m,A}(\Sigma(A_m)). \tag{29.5}$$

The limit takes place in the Hausdorff topology. Here $\Sigma(A_m)$ is the set of low vertices on Γ_m. Given Equation 29.4, we get the analogous result

$$C_A = \lim_{n\to\infty} M_{m,B}(\Sigma(B_m)). \tag{29.6}$$

Let us generalize this result. For each m, suppose there is some $n \geq m$. We also have

$$C_A = \lim_{m\to\infty} M_{n,A}(\Sigma(B_m)). \tag{29.7}$$

The reason is that the maps $M_{m,A}$ and $M_{n,B}$ converge to each other on any compact subset of $\mathbf{R}^2$, and compact pieces of the limit in Equation 29.4 determine increasingly dense subsets of C_A.

Lemma 29.6 *Suppose that $\Sigma_n \subset \widehat{\Gamma}(A_n)$ is a translate of Σ_m consisting entirely of low vertices. Then*

$$C_A = \lim_{m\to\infty} M_{n,A}(\Sigma_n).$$

Proof: We have some vector U_m such that

$$\Sigma_n = \Sigma(A_m) + U_m. \tag{29.8}$$

Since $M_{n,A}$ is affine, we have

$$M_{n,A}(\Sigma_n) = M_{n,A}\Sigma(A_m) + \lambda_m. \tag{29.9}$$

Now we get to the moment of truth. Since $\Sigma(B_m)$ consists entirely of low vertices, we have

$$M_{A,n}(x) \in [0,2]$$

for all $x \in \Sigma(B_m)$. Since Σ_n consists entirely of low vertices, we have $M_{A,n}(x) + \lambda_n \in [0,2]$ as well. Putting $t = M_{A,n}(x)$, we have

$$t, \qquad t + \lambda_m \in [0,2]. \tag{29.10}$$

This last equation puts constraints on λ_m.

By the case when $n = 0$ of Equation 21.7, the set C_A contains both 0 and 2. Therefore, once m is large, we can choose $x \in \Sigma(B_m)$ such that $t = M_{A,n}(x)$ is very close to 0. But this forces

$$\liminf \lambda_m \geq 0.$$

At the same time, we can choose x such that $M_{A,m}(x)$ is very close to 2. This shows that

$$\limsup \lambda_m \leq 0.$$

In short, $\lambda_m \to 0$. □

We just have to tie the discussion above together with the notion of a growing sequence. Suppose that $\{x_n\}$ is a growing sequence. Let β_n denote the component of $\widehat{\Gamma}_n$ corresponding to x_n. There is a proper function $m = m_n$ such that the pivot arc $P\beta_n$ is a translate of the major pivot arc $P\Gamma(B_m)$. Here $\{B_m\}$ is a sequence of even rationals that satisfies the hypotheses of Lemma 29.5. Hence $\{B_m\} \to A$. Hence the application of the Rigidity Lemma above applies.

Every low vertex on $P\beta_n$ is a translate of a low vertex on $P\Gamma(B_m)$. By the Inheritance Lemma, every low vertex on $P\beta_n$ relative to B_m is also low with respect to A_n. Thus we have exactly the situation described in Lemma 29.6.

Let Σ_n denote the set of low vertices of $P\beta_n$. Then Σ_n is a translate of the set $\Sigma(B_m)$ of low vertices on $P\Gamma(B_m)$, as in the lemma above. Since

$$x_n \in M_{A,n}(\Sigma_n), \tag{29.11}$$

we see that the Hausdorff distance from $\{x\}$ to C_A tends to 0 as n (and m) tend to ∞.

This completes the proof of the Period Theorem.

29.4 A USEFUL RESULT

While we are in the neighborhood, we establish a technical result related by Lemma 29.5 that we will use in the next chapter.

Let $\{B_n\}$ be any sequence of rationals that converges to A. Recall from §29.2 that any rational parameter B has a tree $T(B)$ of predecessors. We can consider $T(B_n)$ for each parameter B_n in the sequence.

Lemma 29.7 *Let N be any integer. Then there are only finitely many rationals in the union*

$$\bigcup_{n=1}^{\infty} T(B_n)$$

having complexity less than N.

Proof: We will argue as in the proof of Lemma 29.5. Suppose $C = r/s$ is a rational in the tree $T(B_n)$ such that r is small and s and n are large. Then the directed Farey path connecting C to B_n has tiny displacement and $|B_n - A|$ is small. Hence $|C - A|$ is small. Also, C is near 0. Hence A is near 0. This is a contradiction once s and n are large enough. Hence there is some function f, depending on the sequence, such that $s < f(r)$. Hence the union contains only finitely many rationals having a numerator less than N. Our result follows from this fact. □

Chapter Thirty

Hovering Components

30.1 THE MAIN RESULT

Let $A \in (0,1)$ be a rational parameter. We say that $v \in \mathbf{Z}^2$ is *D-low* if the baseline of $\Gamma(A)$ separates v from $v - (0, D)$. Here $D \in \mathbf{Z}$. We have the usual convention that the baseline is the line of slope $-A$ through the point $(0, -\epsilon)$, where ϵ is an infinitesimally small positive number. Thus $(0,0)$ is 1-low. Previously, we were interested in 1-low vertices, which we called *low*.

Let β be a component of $\widehat{\Gamma}(A)$. We call β a *hovering component* if it has no 1-low vertices. More specifically, we call β a *D-hovering component* of $\widehat{\Gamma}(A)$ if β has no 1-low vertices and if β contains a D-low vertex. The goal of this chapter is to prove the following result.

Lemma 30.1 (Hovering) *Let $\{A_n\}$ be the superior sequence approximating A. Fix D. Then there is a constant D' with the following property. If n is sufficiently large, then $\widehat{\Gamma}_n$ has no D-hovering components having diameter greater than D'. Here D' is independent of n.*

Now we start the proof of the Hovering Lemma. For each rational B, we form a tree of depth 2 by considering the 2 predecessors of B and their 2 predecessors. We define the complexity of B to be the minimum value of all the numerators of the rationals involved in this list of 7 rationals. In the case when some of these predecessors are not defined, we set the complexity to 0.

Lemma 30.2 *Fix D. Let A_2 be any rational with predecessors A_0 and A_1. Let β be a D-hovering component of $\widehat{\Gamma}(A)$. Assuming that A_2 has sufficiently high complexity, β is either a translate of a D-hovering component of $\widehat{\Gamma}_0$ or a translate of a D-hovering component of $\widehat{\Gamma}_1$.*

Proof of the Hovering Lemma: Applying the Hovering Lemma recursively, we see that β is the translate of a D-hovering component of $\widehat{\Gamma}(B_n)$, where B_n belongs to the tree of predecessors of A_n and has uniformly bounded complexity. But then, by Lemma 29.7, the sequence $\{B_n\}$ has only finitely many different terms. Hence β is the translate of one of finitely many different polygons. $\square$

The rest of the chapter is devoted to proving Lemma 30.2.

30.2 TRAPS

Let A be a rational parameter. As usual, $\widehat{\Gamma}(A)$ is invariant under translation by $\mathbf{Z}[V]$. Here $V = (q, -p)$. We say that a *major component* of $\widehat{\Gamma}(A)$ is one that is translation-equivalent to $\Gamma(A)$.

Let $X \subset \mathbf{R}^2$ be a solid parallelogram. We call X a *cap* if the the following hold.

- The only components of $\widehat{\Gamma}$ that cross ∂X are major components.
- If γ is a major component that crosses ∂X, then $\gamma \cap X$ is a finite union of connected arcs, each of which contains a 1-low vertex.

Remark: The second item requires a bit of interpretation. When we take $\gamma \cap X$, we might cut an edge off right in the middle. We always add the full edge to this intersection. Thus $\gamma \cap X$ could stick out a tiny bit from X, and the low vertex in question could be just outside of X. This small annoyance causes no trouble.

Let A_0 and A_1 be the predecessors of A_2. We take

$$A_0 \leftarrow A_2, \qquad A_1 \Leftarrow A_2 \tag{30.1}$$

so that A_0 is odd and A_1 is even. For $j = 0, 1$, let Δ_j denote the region of agreement between $\widehat{\Gamma}_j$ and $\widehat{\Gamma}_2$, as in the Diophantine lemma. Between the Diophantine Lemma and Lemma 27.4, we cover all cases.

We say that a pair (X_0, X_1) of parallelograms is a *D-trap* for A_2 if the following axioms hold.

1. $X_j \subset \Delta_j$.
2. X_j is a cap relative to A_j.
3. Any vertex in X_j is 1-low with respect to A_j iff this vertex is 1-low with respect to A_2.
4. Any D-low vertex relative to A_2 is translation-equivalent, mod $\mathbf{Z}[V_2]$, to a point in $X_0 \cup X_1$.

Lemma 30.3 *Fix D. If A_2 has sufficiently high complexity, then there is a D-trap for A_2.*

Before we prove this result, we use it to prove Lemma 30.2.

Proof of Lemma 30.2: Let β_2 be a D-hovering component of $\widehat{\Gamma}_2$. Let $v \in \beta_2$ be a D-low vertex. By axiom 4, we can translate so that v lies in either X_0 or X_1. Suppose without loss of generality that $v \in X_0$. Since translation by multiples of V_2 preserves the baseline for Γ_2, we see that v is D-low with respect to A_2.

Axiom 3 says that a vertex in X_0 is 1-low with respect to A_0 iff it is 1-low with respect to A_2. But clearly this implies that a vertex in X_0 is k-low with respect to A_0 iff it is k-low with respect to A_2. So, when we use the term *k-low*, it applies equally well relative to A_0 and A_2.

Let β_0 be the component of $\widehat{\Gamma}_0$ that contains v. Suppose first that β_0 crosses ∂X_0. Then β_0 is a major component. Since X_0 is a cap relative to A_0, the component of $\beta_0 \cap R$ that contains v also contains a low vertex. So, tracing β_0 from v, we take a path

$$\gamma \subset X_0 \subset \Delta_0 \tag{30.2}$$

whose endpoint is a low vertex in X_0. The second containment is axiom 1 above. But then $\gamma \subset \widehat{\Gamma}_2$. Since β_2 and γ agree at v, they must agree (by the Embedding Theorem) on the whole path. But then β_2 contains a 1-low vertex. This is a contradiction.

Now we know that β_0 does not cross ∂X_0. But then $\beta_0 \subset \Delta_0$. Hence β_0 is a component of $\widehat{\Gamma}_2$. Since β_0 and β_2 agree at v, we have $\beta_0 = \beta_2$. By construction, $\beta_0 = \beta_2$ contains a D-low vertex and no 1-low vertex. Therefore $\beta_0 = \beta_2$ is a D-hovering component of $\widehat{\Gamma}_0$. □

The rest of the chapter is devoted to the proof of Lemma 30.3. We have 4 cases to consider, and we will consider these cases in turn.

1. A_2 is odd and $A_1 < A_2$.
2. A_2 is odd and $A_1 > A_2$.
3. A_2 is even and $A_1 < A_2$.
4. A_2 is even and $A_1 > A_2$.

Now we reconcile the notation here with the notation in §4.1.

In case 1, we have

$$A_0 = (A_2)_+ - (A_2)_-, \qquad A_1 = (A_2)_-. \tag{30.3}$$

In case 2, we have

$$A_0 = (A_2)_- - (A_2)_+, \qquad A_1 = (A_2)_+. \tag{30.4}$$

In case 3 we have

$$A_0 = (A_2)_+, \qquad A_1 = (A_2)_-. \tag{30.5}$$

In case 4 we have

$$A_0 = (A_2)_-, \qquad A_1 = (A_2)_+. \tag{30.6}$$

We will concentrate on cases 1 and 3. case 2 is essentially the same as case 1, and case 4 is essentially the same as case 3. When it comes time to deal with cases 2 and 4, we will briefly indicate the modifications needed and then show some illustrations from Billiard King.

The parallelograms come from two sources:

- The Decomposition Theorem in Chapter 19.
- The minor box in §28.6.1.

We will explain this precisely below. Mainly, we are repackaging constructions we have already made. When it comes to verifying the axioms, we have essentially already done all the hard work. The proof is mainly a matter of locating the relevant results in previous chapters.

30.3 CASES 1 AND 2

Case 1: We state the following definitions.

- $X_0 = R_1(A_2)$, the small parallelogram from the Decomposition Theorem for the parameter A_2. Here X_0 lies to the left of the origin.
- X_1 is the minor box, defined relative to the parameter A_1, in §28.6.1. Here X_1 lies to the right of the origin.

Remark: The top/bottom of X_0 has a slightly different slope from the top/bottom of X_1, but the difference is tiny when A_2 has high complexity. X_0 and X_1 may or may not have about the same height. The figures below show one case where this happens and one case where it does not.

Lemma 30.4 (X_0, X_1) *satisfies axiom 1.*

Proof: In §28.6.1, we showed that $X_1 \subset \Delta_1$. We just have to consider X_0. The argument for X_0 is really the same as that for the Decomposition Theorem. However, since we considered a different case there, we will work out the details here.

We will apply the Diophantine Lemma. Since we do not care about small cases, we write $I_1 \approx I_2$ to denote the relation where two intervals are with 2 units of each other. We work with the linear functionals G_2 and H_2 associated to the parameter A_2. Let u and w denote the top left and right vertices of X_0, respectively. The interval in the Diophantine Lemma is

$$I_2 \approx \left[-(q_2)_- - q_0, q_0\right]. \tag{30.7}$$

The lower bound comes from case 2 of Lemma 17.8.

Hence it suffices to show that

$$G_2(u) \gg -(q_2)_- - q_0, \qquad H_2(w) \ll q_0. \tag{30.8}$$

The symbol $(\gg)$ indicates an inequality in which the difference between the two sides tends to ∞ with the complexity of A_2.

We have the estimates

$$u \approx -(V_2)_- + \lambda W_2, \qquad w \approx \lambda W_2, \qquad \lambda = \frac{q_2^*}{q_2} \leq \frac{q_0}{q_2}. \tag{30.9}$$

Here $A_2^* = p_2^*/q_2^*$ is the superior predecessor of A_2. The approximation becomes arbitrarily good as the complexity of A_2 tends to ∞. Hence the approximation is good to within 1 unit once A_2 has sufficiently high complexity.

We compute

$$G_2(u) \approx -(q_2)_- - \lambda \frac{q_2^2}{p_2 + q_2} \gg -(q_2)_- - \lambda(q_2) \geq -(q_2)_- - q_0.$$

This takes care of the vertex u. Now we compute

$$H_2(w) \approx \lambda \frac{q_2^2}{p_2 + q_2} \ll \lambda q_2 = q_0.$$

This takes care of the vertex w. □

Lemma 30.5 *X_0 is a cap.*

Proof: Consider X_0 first. We are interested in how $\widehat{\Gamma}_0$ sits with respect to X_0, but the Decomposition Theorem gives us information about $\widehat{\Gamma}_2$. By the Decomposition Theorem, the only component of $\widehat{\Gamma}_2$ that crosses ∂X_0 is Γ_2, a major component. The intersection $\Gamma_2 \cap X_0$ is a single arc that crosses ∂X_0 at its endpoints. These endpoints are the low vertices. However, $\widehat{\Gamma}_0$ and $\widehat{\Gamma}_2$ agree in X_0. Moreover, X_0 contains $(0,0)$. From this we see that Γ_0 is the only component to cross ∂X_0, and the description of the intersections is exactly the same. □

Lemma 30.6 *X_1 is a cap.*

Proof: This argument is really a repeat of the argument given in the proof of the Pivot Theorem. Consider first the infinite strip S obtained by extending the top and bottom sides of X_1. By the Barrier Theorem, each major component of $\widehat{\Gamma}_1$ intersects S in a connected arc that contains 1-low vertices. Now we analyze what happens near the side walls of X_1. The bottom left vertex $(0,0)$ is a low vertex of a major component of $\widehat{\Gamma}_1$. The same is true for the bottom right vertex of X_1. Indeed, the bottom right vertex of X_1 is the right endpoint of the bump associated to A_2, as discussed in §28.2. This was a key part of the proof of the Pivot Theorem. By the Hexagrid Theorem, the major components of $\widehat{\Gamma}_1$ intersect X_1 in arcs connecting a low vertex to the top of X_1. □

Combining these results, we see that (X_0, X_1) satisfies axiom 2.

Lemma 30.7 *(X_0, X_1) satisfies axiom 3.*

Proof: This follows from the geometric interpretation of the Diophantine constant given in the Goodness Lemma in §17.4.2. See also §22.4. □

Lemma 30.8 *(X_0, X_1) satisfies axiom 4.*

Proof: The left bottom vertex of X_0 is $-(V_2)_-$, whereas the bottom right vertex of X_1 is $(V_2)_+$. These two vertices differ by V_2. The bottom right vertex of X_0 is $(0,0)$, the same as the bottom left vertex of X_1, as shown in Figure 30.1. We have emphasized the gap between the two parallelograms, which is usually tiny, for the sake of highlighting the important issues.

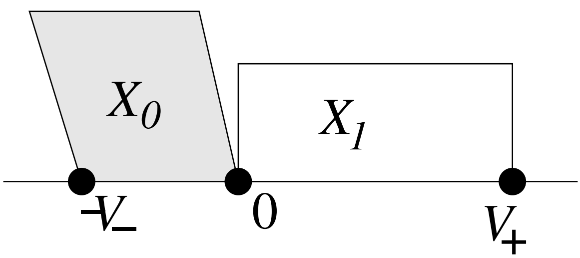

Figure 30.1: The trap.

Suppose for the moment that the sides of X_0 have the same slope as the sides of X_1. Then, once A_2 has high complexity, the tops of both parallelograms are more than D units from the baseline. But then the union of translations

$$\bigcup_{k\in\mathbf{Z}} \Big(X_0 + X_1 + kV_2) \tag{30.10}$$

contains all D-low vertices, as desired.

The slight complication is that the sides of X_0 are parallel to W_2, whereas the sides of X_1 are parallel to W_1. These are the vectors from Equation 3.2 relative to A_2 and A_1. As the complexity of A_2 tends to ∞, the slopes converge, and no D-low lattice point lies between the two lines emanating from the same point. Thus the union in Equation 30.10 still contains all D-low vertices once A_2 has high complexity. $\square$

Case 2: We use the same definitions as for case 1 except that $-(V_2)_-$ replaces $(V_2)_+$ in the definition of the minor box for X_1. Aside from switching the roles played by left and right, and $(+)$ and $(-)$, the proofs for case 2 are exactly the same as the proofs for case 1.

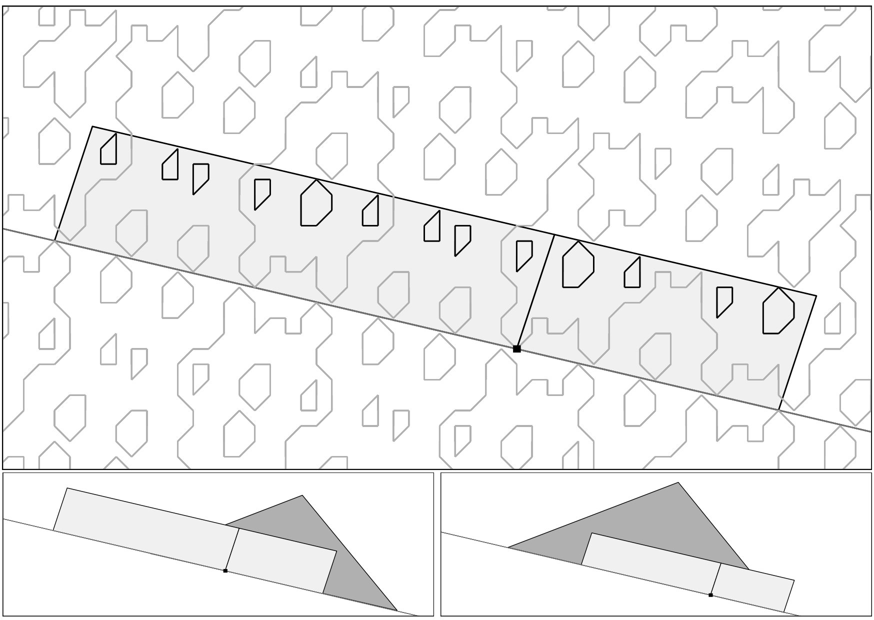

Figure 30.2: The traps and hovering components for $11/47$.

Figure 30.2 shows an example in this case. We have

$$A_0 = \frac{3}{13}, \qquad A_1 = \frac{4}{17}, \qquad A_2 = \frac{11}{47}.$$

The top frame shows some of the components of $\widehat{\Gamma}(11/47)$. Note that the low hovering components, outlined in black, are trapped. Other components, however, are allowed to cross out of the traps. Figure 30.2 also shows $\Delta(A_0, A_2)$ and $\Delta(A_1, A_2)$. We have $X_j \subset \Delta(A_j, A_2)$.

30.4 CASES 3 AND 4

Case 3: We define X_0 to be the parallelogram bounded by the following lines.

1. The baseline relative to A_0.
2. The line parallel to V_0 and containing W_0. Compare the Room Lemma.
3. The line parallel to W_0 and containing $(0, 0)$.
4. The line parallel to W_0 and containing $-(V_2)_-$.

We define X_1 to be the minor box, as in §28.6.1. (This definition does not use the parity of A_2.)

Lemma 30.9 (X_0, X_1) *satisfies axiom 1.* $X_0 \subset \Delta_0$.

Proof: As in case 1, the work in §28.6.1 takes care of X_1. We just have to show that $X_0 \subset \Delta_0$. We will apply Lemma 27.4. This time we work with the linear functionals G_0 and H_0 associated to the parameter A_0. Let u and w denote the top left and right vertices of X_0, respectively. The interval in the Diophantine Lemma is

$$I \approx [-q_2, q_0]. \tag{30.11}$$

Hence it suffices to show that

$$G_2(u) \gg -q_2, \qquad H_2(w) \ll q_0. \tag{30.12}$$

We have

$$u = -(V_2)_- + W_0, \qquad w = W_0. \tag{30.13}$$

We compute

$$G_0(u) \approx -(q_2)_- - \frac{q_0^2}{p_0 + q_0} \gg -(q_2)_- - q_0 = -(q_2)_- - (q_2)_+ = -q_2.$$

This takes care of the vertex u. Now we compute

$$H_2(w) = \frac{q_0^2}{p_0 + q_0} \ll q_0.$$

This takes care of the vertex w. □

Lemma 30.10 X_0 *is a cap*

Proof: We use an argument similar to Lemma 30.6. Consider first the infinite strip S obtained by extending the top and bottom sides of X_0. By Statement 1 of the Hexagrid Theorem, no edge of $\widehat{\Gamma}_0$ crosses the top of S. By this theorem, the only component to cross the right side of X_0, namely, the wall line through $(0, 0)$, is Γ_0. By rotational symmetry, the same is true for the left side of X_0. The argument is essentially the same as that given in §19.3. The point is that some rotational symmetry of $\widehat{\Gamma}_0$ carries the left side of X_0 to the right side. To be sure, compare Lemma 28.6. □

Lemma 30.11 (X_0, X_1) *satisfies axiom 2.*

Proof: The argument for X_1 is essentially the same as in case 1. The only difference is that we use the setup from §28.9.2 because A_1 and A_2 are both even rationals. □

Combining these results, we see that (X_0, X_1) satisfies axiom 2. The verification of axioms 3 and 4 is the same as in case 1.

Case 4: We use the same definitions as in case 3 except that we interchange the roles played by $-(V_2)_-$ and $(V_2)_+$. The proof in this case is essentially the same as in case 3, modulo the same switching of left and right. Figure 30.3 shows an example for

$$A_0 = \frac{9}{31}, \qquad A_1 = \frac{7}{24}, \qquad A_2 = \frac{16}{55}$$

Figure 30.3 also shows the hovering components that are trapped in the parallelograms.

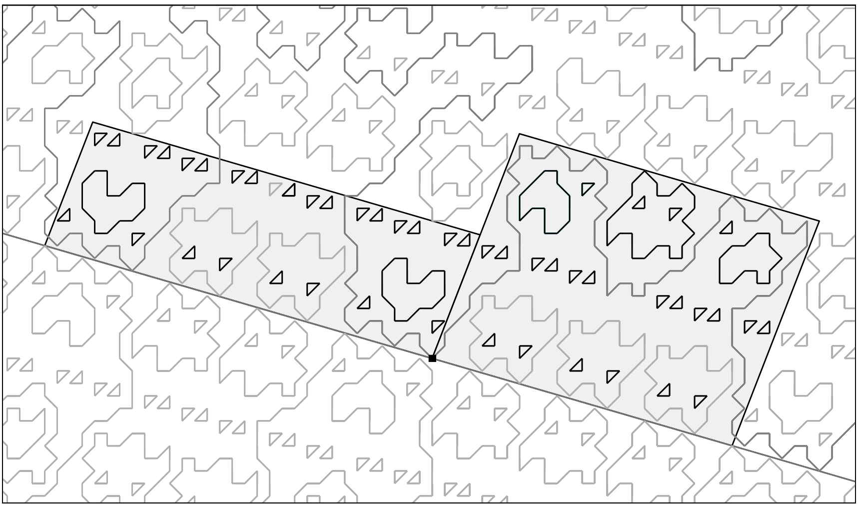

Figure 30.3: The traps and hovering components for 16/55.

Chapter Thirty-One

Proof of the Low Vertex Theorem

31.1 OVERVIEW

The Low Vertex Theorem in Chapter 23 is a consequence of the following result.

Lemma 31.1 (Descent) *Let $A \in (0, 1)$ be irrational. Let $\{B_n\}$ be any sequence of rationals in $(0, 1)$ that converges to A. Let β be a low component of $\widehat{\Gamma}(B_n)$. There is some constant D' such that every D-low vertex of β can be connected to a low vertex of β in less than D' steps. Here D' depends on D and on A but not on n.*

Proof of the Low Vertex Theorem: Let N_0 and $\{v_n\}$ be as in the Low Vertex Theorem. Let β_n be the component of $\widehat{\Gamma}_n$ that contains v_n. Here is the imput from the Hovering Lemma. If the constant N_1 is chosen sufficiently large, then the inequality

$$\mathrm{diam}(\beta_n) > N_1$$

implies that β_n is a low component. We choose N_1 in this way. Applying the Descent Lemma to the sequence $\{A_n\}$, the component $\beta = \beta_n$, and the constant $D = N_0$, we immediately obtain the conclusion of Low Vertex Theorem with $N_2 = D'$. $\square$

The rest of the chapter is devoted to proving the Descent Lemma. Our proof of the Descent Lemma is somewhat complicated by the fact that we cannot quite prove a very useful conjecture. Experimentally, we observe the following improvement for the Inheritance Lemma.

Conjecture 31.2 *Let A_2 be any rational having the predecessors $A_0 \leftarrow A_2$ and $A_1 \Leftarrow A_2$. Then every minor low component of $\widehat{\Gamma}_2$ is either the translate of a low component of $\widehat{\Gamma}_0$ or the translate of a low component of $\widehat{\Gamma}_1$.*

Referring to the proof of the Pivot Theorem, the end major components give us trouble. See the discussion at the end of §28.2.

As we will explain below, Conjecture 31.2 would be very useful in proving the Descent Lemma. See the remark in §31.3. Our strategy for proving the Descent Lemma is to prove a somewhat weaker version of Conjecture 31.2 that captures all the necessary features. We state this weaker result, Lemma 31.3, in the next section. One strategy for understanding this chapter is to first assume the truth of Conjecture 31.2. Then, once the overall logic of the argument makes sense, one can learn the complications that arise from the fact that we must use Lemma 31.3 in place of Conjecture 31.2.

31.2 A MAKESHIFT RESULT

Let A be an even rational. Previously, we divided the polygon $\Gamma(A)$ into two arcs, the pivot arc $P\Gamma(A)$ and the upper arc. These two arcs join together at the pivot points.

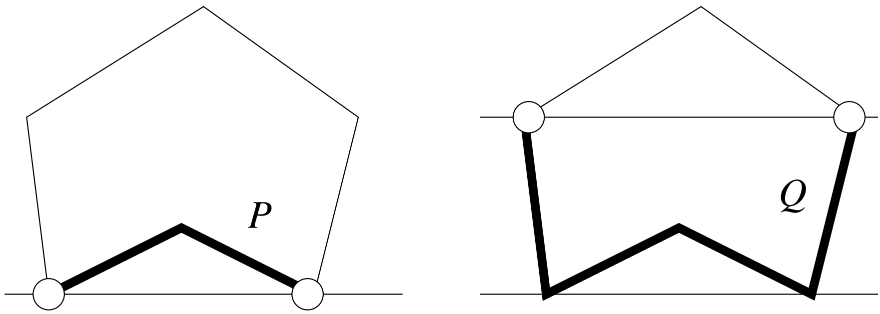

Figure 31.1 $P\Gamma$ and $Q\Gamma$.

Referring to the Barrier Theorem, recall that $\Gamma(A)$ passes through the barrier at 2 points. One arc of Γ lies below the barrier and one above. Let $Q\Gamma$ denote the component that lies below. Then $P\Gamma \subset Q\Gamma$. We call $Q\Gamma$ an *extended pivot arc*. We think of $Q\Gamma$ as a kind of compromise between the whole component Γ and the pivot arc $P\Gamma$. If A has sufficiently high complexity, then $Q\Gamma$ contains all the vertices within D of the baseline. This is a consequence of the Barrier Theorem.

So far we have defined $Q\beta$ only when $\beta = \Gamma(A)$ and A is an even rational. The result next serves both as a lemma and a definition. It will allow us to apply the definition of *extended pivot arc* to all polygonal low components of $\widehat{\Gamma}(A)$ when A is any rational parameter. The result we prove here is both a lemma and a definition.

Lemma 31.3 *Let A_2 be a rational having predecessors $A_0 \leftarrow A_2$ and $A_1 \Leftarrow A_2$. If A_2 has high enough complexity, then every low component of $\widehat{\Gamma}_2$ has a well defined extended pivot arc, and this pivot arc is the translate of an extended pivot arc of $\widehat{\Gamma}_j$ for one of $j = 0, 1$.*

Proof: We will suppose that A_2 is odd. The even case is similar. In the proof of the Inheritance Lemma, the same constructions and arguments work for the whole components and not just their pivot arcs – except perhaps in the case of the end major components. Again compare the discussion at the end of §28.2. To deal with the end major components, we consider the trap (X_0, X_1) constructed in the previous chapter. The important point here is that the top of X_1 is the barrier line for the parameter A_1. The two end major components β_1 and β_2 intersect X_1 precisely in the arcs $Q\beta_1$ and $Q\beta_2$. Hence $Q\beta_1$ and $Q\beta_2$ are copied whole by $\widehat{\Gamma}_2$. Let $\tilde{\beta}$ denote the component of $\widehat{\Gamma}_2$ that contains $\beta \cap X_1$. We define $Q\tilde{\beta} = Q\beta$. Then $Q\tilde{\beta}$ is copied from $\widehat{\Gamma}_1$ by construction. $\square$

Remark: Lemma 31.3 is not stated in a way that makes it obviously parallel to Conjecture 31.2. Below we will explain why Lemma 31.3 plays a role in the proof of the Descent Lemma that is similar to the role that Conjecture 31.2 would play.

The following result is an addendum to the proof of Lemma 31.3.

Lemma 31.4 *Let N be fixed. If A_2 has sufficiently high complexity and β is an end major component of $\widehat{\Gamma}_1$, then $\tilde{\beta}_1 - Q\beta_1$ does not contain any vertices within N units of the baseline.*

Proof: As in our proof of the Pivot Theorem, we consider the case when $A_1 < A_2$. The other case is entirely similar.

Let

$$\gamma = \tilde{\beta} - Q\tilde{\beta}. \tag{31.1}$$

Here γ is an arc of $\widehat{\Gamma}_2$. Let X_1 be as above. $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ agree in X_1. The component $\tilde{\beta}$ has a low vertex in X_1. The arc γ has both its endpoints on the top edge of X_1.

Let S denote the infinite strip obtained by extending the left and right sides of X_1. We claim that $\tilde{\beta}$ does not cross either side of S. To prove this claim, let S_L and S_R denote the left and right boundaries of S. Then $\tilde{\beta}$ does not cross S_L, by the Hexagrid Theorem applied to A_2. Likewise, $\iota(\tilde{\beta})$ does not cross S_L, by the Hexagrid Theorem. Here ι is the same symmetry as in Lemma 28.1. By construction, ι swaps S_L and S_R. Hence $\tilde{\beta}$ does not cross S_R. This establishes our claim.

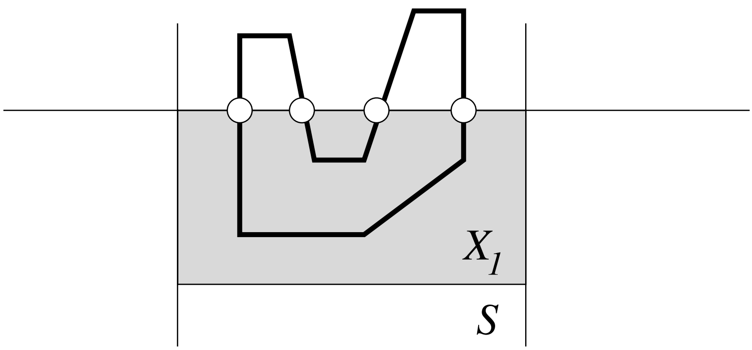

Figure 31.2: γ crosses X_1 four times.

Now we know that γ does not cross the sides of S. Hence, if γ contains a vertex within N units of the baseline, this vertex must lie in X_1. But then $\tilde{\beta}$ crosses the top edge of X_1 at least 4 times, as shown in Figure 31.2. But these 4 crossing points are then copied from $\widehat{\Gamma}_1$. This contradicts the Barrier Theorem because the top edge of X_1 is contained in the barrier line for $\widehat{\Gamma}_1$. □

31.3 ELIMINATING MINOR ARCS

Suppose that the Descent Lemma is false. This means that we can find a sequence $\{v_n\}$ of vertices, all uniformly close to the baseline, such that the n-neighborhood of β_n contains no low vertices. Here β_n is the component of $\widehat{\Gamma}_n$ that contains v_n. In this section we reduce the several possible situations to one situation that is easier to manage.

Passing to a subsequence and using translation symmetry, we can arrange one of two cases.

- β_n is a minor component of $\widehat{\Gamma}_n$ for all n.
- $\beta_n = \Gamma_n$ for all n.

Here we will show that a counterexample of the first kind forces a counterexample of the second kind.

Remark: Assuming Conjecture 31.2, we can argue as follows. By Conjecture 31.2, the component β_n is the translate of $\Gamma(B_n')$ for some $B_n' \in T(B_n)$. Since β_n is a low component, and yet the n-ball about v_n contains no low vertices, we see that the diameter of β_n tends to ∞ with n. But then the complexity of B_n' tends to ∞ with n. Hence, by Lemma 29.5, $B_n' \to A$. Thus a counterexample to Lemma 31.1 involving minor components leads to a counterexample involving major components. The new counterexample uses the parameters $\{B_n'\}$.

Since we cannot prove Conjecture 31.2, we have to make do with Lemma 31.3. We need one last result before we can make Lemma 31.3 work for us.

Lemma 31.5 *Let β_n be a low component of $\widehat{\Gamma}(B_n)$. Suppose that the diameter of β_n tends to ∞. Then the distance from any point on $\beta_n - Q\beta_n$ to the baseline of $\widehat{\Gamma}(B_n)$ tends to ∞ as well.*

Proof: This is a consequence of Lemma 31.4. Each β_n is a translate of a component of the form

$$\tilde{C}, \qquad C = \Gamma(B_n'). \tag{31.2}$$

Here B_n' is on the tree of predecessors of B_n. Since the diameter of $\tilde{C}$ tends to ∞ with n, we see than the complexity of B_n' tends to ∞ with n by Lemma 29.7. Hence the distance from $\tilde{C} - Q\tilde{C}$ to the relevant baseline tends to ∞ with n. $\square$

Now let us revisit the argument above. By Corollary 31.5, the points v_n lie on $Q\beta_n$ once n is sufficiently large. Indeed, by Lemma 31.5, the distance from v_n to a point on $\beta_n - Q\beta_n$ tends to ∞ with n. By Lemma 31.3, we know that $Q\beta_n$ is the translate of $Q\Gamma(B_n')$ for some B_n'. The sequence $\{B_n'\}$ converges to A. Then $Q\Gamma(B_n')$ has a vertex v_n' that is uniformly close to the baseline but has an n-neighborhood with no low vertices. This is a counterexample of the second kind.

To finish the proof, we just have to rule out counterexamples of the second kind. We will first present a topological lemma and then complete the proof.

31.4 A TOPOLOGICAL LEMMA

The result concerns the trap (X_0, X_1) constructed in the previous chapter. Let γ_2 be the bump associated to the parameter A_2, as in §28.2.

Lemma 31.6 *When A_2 has sufficiently high complexity, the set $\gamma_2 \cap X_1$ consists of 2 connected arcs, each joining an endpoint of γ_2 to the top of X_1.*

Proof: In the even case, this is a restatement of Lemma 28.13. Consider the odd case. We take $A_1 < A_2$. The other case is entirely similiar.

The two endpoints of γ_2 are E_2^+ and $E_2^- + V_2$. Both these points belong to X_1. The line parallel to W_2 through $V_2/2$ divides X_1 into two pieces. (See Figure 31.3.) By the Hexagrid Theorem, γ_2 crosses a door on this line. This door lies above the top of X_1. At the same time, γ_2 can cross the top of X_1 only twice. This follows from the Barrier Theorem, as applied to A_1, and from the fact that $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ agree in a neighborhood of X_1. So, starting from the left endpoint of γ_2, some initial arc of γ_2 rises up to the top of X_1. The next arc of γ_2 crosses through a door and returns to the top of X_1. The final arc of γ_2 connects the top of X_1 to the right endpoint of γ_2. $\square$

Figure 31.3 illustrates our argument for $A = 21/55$. The dark-gray parallelogram is X_1. The line parallel to W_2 through $V_2/2$ is the line of high positive slope on the right side of the figure. (The vectors V and W are as in the definition of the Hexagrid given in Chapter 3.) The relevant door is the triple point on this line at the far right. We have shown part of the hexagrid so as to point out the door.

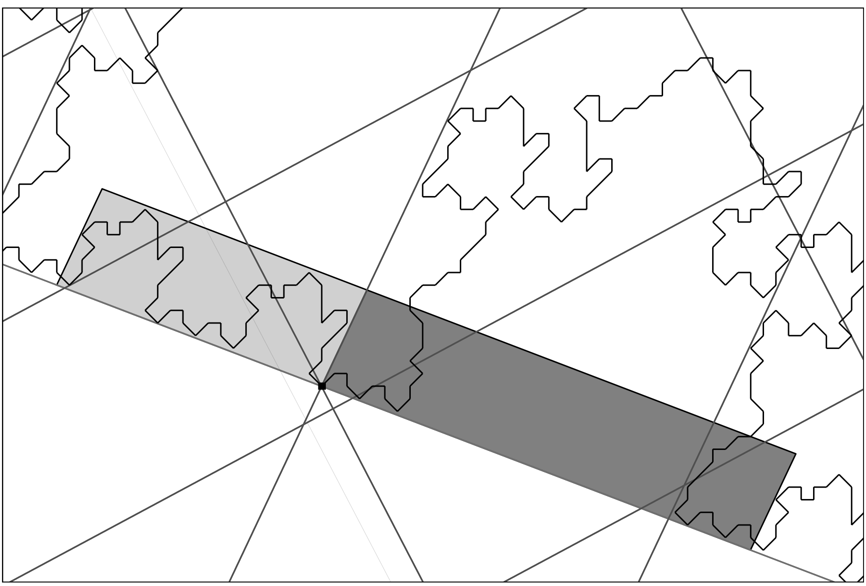

Figure 31.3: Lemma 31.6 for $A_2 = 21/55$.

31.5 THE END OF THE PROOF

Let $\Gamma_2 = \Gamma(A_2)$, as in the previous section. We say that a D-arc of Γ_2 is a connected arc α that joins a low vertex to a D-low vertex. Let $|\alpha|$ denote the smallest integer N such that α contains no vertices that are more than N vertical units above the baseline. Given a D-low vertex $v \in \alpha$, let

$$F(A_2) = \max f(v; A_2), \qquad f(v; A_2) = \min |\alpha|. \tag{31.3}$$

In the first equation, the maximum is taken over all D-low vertices. In the second equation, the minimum is taken over all D-arcs having v as an endpoint. (Actually, this minimum it taken over the two shortest D-arcs, each going out in a different direction from v.) These functions depend implicitly on D, which is fixed throughout the discussion.

Before we prove any results, we give some intuition about the function F. If $F(A_2)$ is large, it means that there exists a D-low vertex v such that the only arcs connecting v to an actual low vertex rise up very high away from the baseline. At least in a large neighborhood of v, the component containing v would imitate a hovering component. This is the sort of thing we want to rule out.

Lemma 31.7 *If A_2 has sufficiently high complexity, then*

$$F(A_2) \leq \max\Big(F(A_0), F(A_1)\Big).$$

Proof: We treat the odd case. The even case has the same proof except that we use Lemma 27.12 in place of the Copy Lemma.

Let (X_0, X_1) be the trap for A_2. Choose a D-low vertex $v \in \Gamma_2$ such that $F(A_2) = f(v)$. Recall that γ_2 is the bump corresponding to A_2. The union $\Gamma_2 \cup \gamma_2$ is one period of Γ modulo translations by V_2. We have two cases.

Case 1: Suppose that $v \in P\Gamma_2$. By the Copy Theorem, $P\Gamma_2 \subset \Gamma_0$. By the argument in §22.4, a vertex on $P\Gamma_2$ is k-low with respect to A_0 iff it is k-low with respect to A_2. Since both endpoints of $P\Gamma_2$ are 1-low with respect to both parameters, the D-arcs of Γ_2 realizing $f(v, A_2)$ coincide with the D-arcs of Γ_0 realizing $f(v, A_0)$. Hence

$$F(A_0) \geq f(v, A_0) = f(v, A_2) = F(A_2).$$

Case 2: Suppose that $v \in \gamma_2$. Then $v \subset X_1$, and v is in one of the two arcs from Lemma 31.6. Let us say that v is on the left arc λ. Then $\lambda \subset \Gamma_1 \cap \Gamma_2 \cap X_1$, by axiom 1 for traps combined with Lemma 27.4. By axiom 3 for traps, a vertex of λ is k-low with respect to A_1 iff it is k-low with respect to A_2. Let α be a D-arc of Γ_1 such that $f(v; A_1) = |\alpha|$. The left endpoint of λ is 1-low, and the right endpoint lies on the top of X_1. When A_2 has high complexity, $\alpha \subset \lambda$. The idea here is that the D-arc connecting v to the left endpoint of λ remains in X_1, whereas any D-arc exiting λ must pass through the top of X_1. Since $\alpha \subset \lambda$, we have $F(A_1) \geq F(A_2)$ as in case 1. $\square$

Let $\{B_n\}$ be the sequence in the Descent Lemma.

Corollary 31.8 *$F(B_n)$ is uniformly bounded independent of n.*

Proof: Applying the previous result recursively, we see that there is some parameter $C_n \in T(B_n)$, of uniformly bounded complexity, such that

$$F(B_n) \le F(C_n).$$

But the sequence $\{C_n\}$ has only finitely many distinct members, by Lemma 29.7. □

In light of the work in §31.3, the following corollary finishes the proof of the Descent Lemma.

Corollary 31.9 *A D-low vertex of $\Gamma(B_n)$ can be connected to a low vertex of $\Gamma(B_n)$ by an arc that has length less than D'. Here D' is independent of n.*

Proof: Let v_n be the D-low vertex in question. By Corollary 31.8 we can find a D-arc α_n connecting v_n to a low vertex of $\Gamma(B_n)$ such that $|\alpha_n| < N$ and N is independent of n. But the same argument as in the proof of Lemma 5.7 shows that the diameter of α_n is uniformly bounded. The idea here is that α_n cannot grow a long way in a thin neighborhood of the baseline. □

This completes the proof of the Low Vertex Theorem. This was the last remaining piece of business. Our work is done.

This page intentionally left blank

Appendix

In this appendix, we describe some additional experimental observations we have made about outer billiards on kites and quadrilaterals.

A.1 STRUCTURE OF PERIODIC POINTS

A.1.1 Irrational Case

Suppose A is an irrational parameter. Let C_A and I be as in the Comet Theorem. It follows from the Comet Theorem that all defined orbits in $I - C_A$ are periodic. Here we discuss a conjectural picture of the dynamics of these points. We use the notation from the Comet Theorem.

As in §24.2, we can naturally identify C_A with the ends of an infinite directed tree T_A. Using the homeomorphism

$$\phi: \mathcal{Z}_A \to C_A,$$

we can formally extend the return map on $C_A^{\#} - \phi(-1)$ to all of C_A, even though the extended return map does not correspond to the outer billiards dynamics on the extra points. The extended return map is induced by an automorphism

$$\Theta_A: T_A \to T_A$$

as discussed in §24.2. The complementary open intervals in $I - C_A$ – the *gaps*– are naturally in bijection with the forward cones of T_A.

Conjecture A.1 *The outer billiards map is entirely defined on a gap. The first return map to $I - C_A$ permutes the gaps according to the action of Θ_A on the forward cones of T_A.*

Some reflection should convince the reader that this is the simplest possible description of the periodic dynamics that is compatible with the Comet Theorem.

With a lot of effort, we can prove the weaker result that Conjecture A.1 correctly describes the first return map for every *defined* orbit in $I - C_A$. The part we cannot prove is that all the orbits of $I - C_A$ are actually defined. This is a big difference. If all points in the same gap have well defined orbits, then the whole gap moves as a single orbit. That is, all points in the same gap have the same combinatorial type of orbit. Without knowing that all points in the gap have well defined orbits, all we can say is that two points in the same gap return to I in the correct way. The orbits might have different itineraries outside of I.

We might have included the proof of the weak version of Conjecture A.1 in this book, but we would prefer to hold out for the definitive result.

A.1.2 Rational Case

Now we describe a rational version of Conjecture A.1 which, combined with the results we have proved, implies Conjecture A.1. Let $A = p_n/q_n$, as in Theorem 1.8. Let $C(A)$ be the set from Theorem 1.8. Each $\xi \in C(A)$ is the midpoint of a special interval in the sense of §2.2. Call this interval $J(\xi)$. Define

$$\widehat{C}(A) = \bigcup_{\xi \in C(A)} J(\xi). \tag{A.1}$$

Figure A.1 shows three examples. Here we have thickened the intervals to get a better picture. We have also added white bars to clarify the spacing.

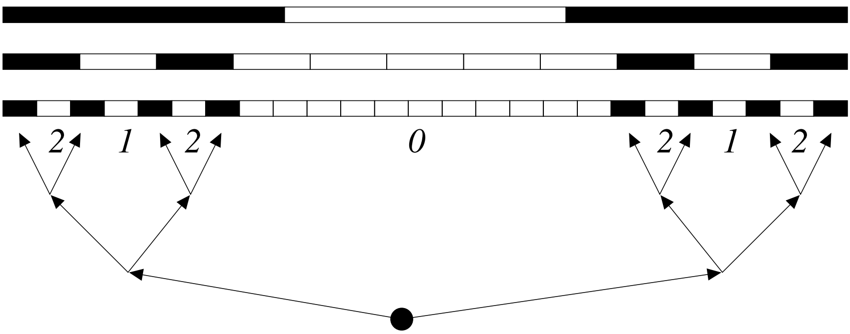

Figure A.1: $\widehat{C}(A)$ for $A = 1/3$ and $3/11$ and $7/25$.

The three rationals in Figure A.1 are part of a superior sequence, and one can see that each level sort of refines the one above it. It is a consequence of Lemma 2.6 that, in the odd case, there is a gap between every pair of intervals in $\widehat{C}(A)$. In the even case, this need not be true. One can compute the positions of the intervals using the formula in Theorem 1.8.

Say that a *gap* is an maximal interval of $I - \widehat{C}$. For $\widehat{C}(7/25)$ there are 7 gaps. Each gap has a *level*, as indicated in the figure. The levels go from 0 to $n-1$ in $\widehat{C}(A)$. (Here A is the nth term in the superior sequence that leads up to A.) Informally, the gaps of level $k \leq n-2$ are inherited from previous terms in the superior sequence, and the gaps of level $n-1$ are newly created with the last parameter.

Given this notion of levels, there is a natural identification of $C(A)$ with the ends of a directed finite tree. The return map $\Theta_A \colon C(A) \to C(A)$ comes from an automorphism of this tree. The union of all the gaps is bijective with the forward cones of the tree. The automorphism of the tree induces an automorphism on the set of forward cones. With all this notation in place, the conjecture for rational parameters is exactly like Conjecture A.1.

The Inheritance Lemma in Chapter 29 makes some progress toward proving the rational version of Conjecture A.1, but this lemma is not powerul enough. (Neither is Lemma 31.3.) We know how to deduce the rational version of Conjecture A.1 from Conjecture 31.2, but we do not know how to prove Conjecture 31.2.

A.2 SELF-SIMILARITY

Figure A.2 shows the arithmetic graphs for the parameters 169/408 and 72/305. These rationals are close approximations to $\sqrt{2}-1$ and $\sqrt{5}-2$, respectively. The second parameter is the Penrose kite parameter. It seems that the arithmetic graphs associated to quadratic irrational parameters are self-similar on a large scale.

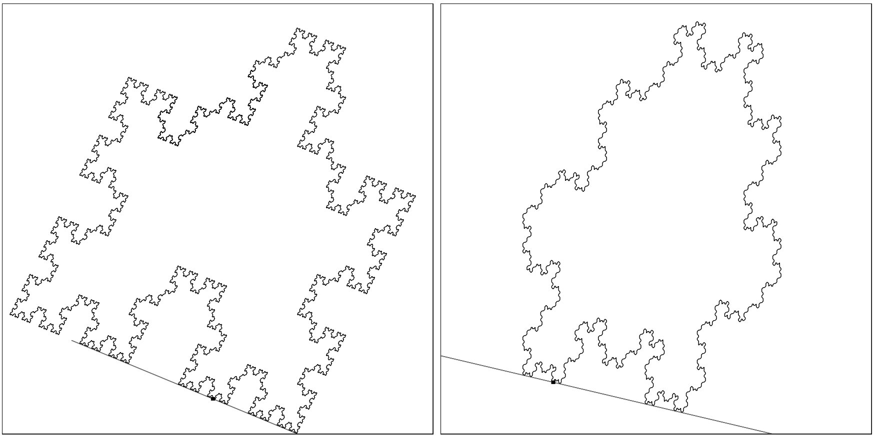

Figure A.2: The arithmetic graph for rationals close to $\sqrt{2}-1$ and $\sqrt{5}-2$.

Let Γ denote the $(2,\infty,\infty)$-triangle group, from Theorem 1.5. Let I and ϕ be the interval from the Comet Theorem.

Conjecture A.2 *Let $g \in \Gamma$ and let $A \in (0,1)$ be a fixed point of g. Suppose that $\alpha = \phi^{-1}(-1)$ has a well defined orbit relative to the parameter A. Then the arithmetic graph $\widehat{\Gamma}_\alpha(A)$ is quasi-invariant under dilation by $|g'(A)|^{1/2}$.*

By *quasi-invariant* we mean that there is a dilation T such that $\widehat{\Gamma}$ and $T(\widehat{\Gamma})$ are contained in bounded tubular neighborhoods of each other. Sometimes $\phi^{-1}(-1)$ does not have a well defined orbit. In these cases, there is a replacement for Conjecture A.2, but it is more difficult to state.

Conjecture A.2 for $A = \sqrt{5}-2$ is a consequence of the results in [**S1**]. This kind of self-similarity is stronger than the kind in item 3 of Theorem 1.5. Indeed, item 3 of Theorem 1.5 is really just a reflection of the fact that the set of low vertices of the component Γ behaves like a large-scale fractal. Conjecture A.2 deals with the whole arithmetic graph and not just the bottom layer of one component.

One consequence of Conjecture A.2 is that suitably rescaled limits of arithmetic graphs, at quadratic irrational parameters, are self-similar curves – or perhaps closely akin to self-similar tilings in the sense of [**Ke**] if all components are rescaled at once. We think that the following conjecture would be another consequence.

Conjecture A.3 *For each quadratic irrational $A \in (0,1)$, there is some exponent $a = a(A) \in (2,3)$ such that the bound $\left[c_2^{-1}d^{-2}, c_2 d^{-3}\right]$ in item 3 of the Comet Theorem can be replaced by $\left[c_2^{-1}d^{-a}, c_2 d^{-a}\right]$.*

A.3 GENERAL ORBITS ON KITES

This entire book is concerned with the special orbits on kites, those that lie on $\mathbf{R} \times \mathbf{Z}_{\text{odd}}$. For any $y \in \mathbf{R}$, let

$$S_y = \{y + 2k|\ k \in \mathbf{Z}\}, \qquad \boldsymbol{\Omega}_y = \{(x, y')|\ y' \in S_y\}. \tag{A.2}$$

$\boldsymbol{\Omega}_y$ consists of an infinite family of parallel lines, each spaced 2 apart from its nearest neighbors. The special orbits all lie on $\boldsymbol{\Omega}_1$. The square of the outer billiards map on a kite preserves $\boldsymbol{\Omega}_y$ for any choice of y.

Once we choose an offset $\alpha \in \mathbf{R}$, we can define the arithmetic graph $\widehat{\Gamma}_\alpha(A; y)$. When A is rational, there is a canonical choice for α and we omit it from our notation. As $y \to 0$, the nature of $\widehat{\Gamma}(A; y)$ changes in a fascinating way. In Figure A.3, we show $\widehat{\Gamma}(17/37; y)$ for the y-values

$$1, \quad \frac{1}{2}, \quad \frac{1}{4}, \quad \frac{1}{8}.$$

As $y \to 0$, the graph starts to concentrate along straight lines. These lines are asymptotically parallel to the lines of the door grid from the Hexagrid Theorem.

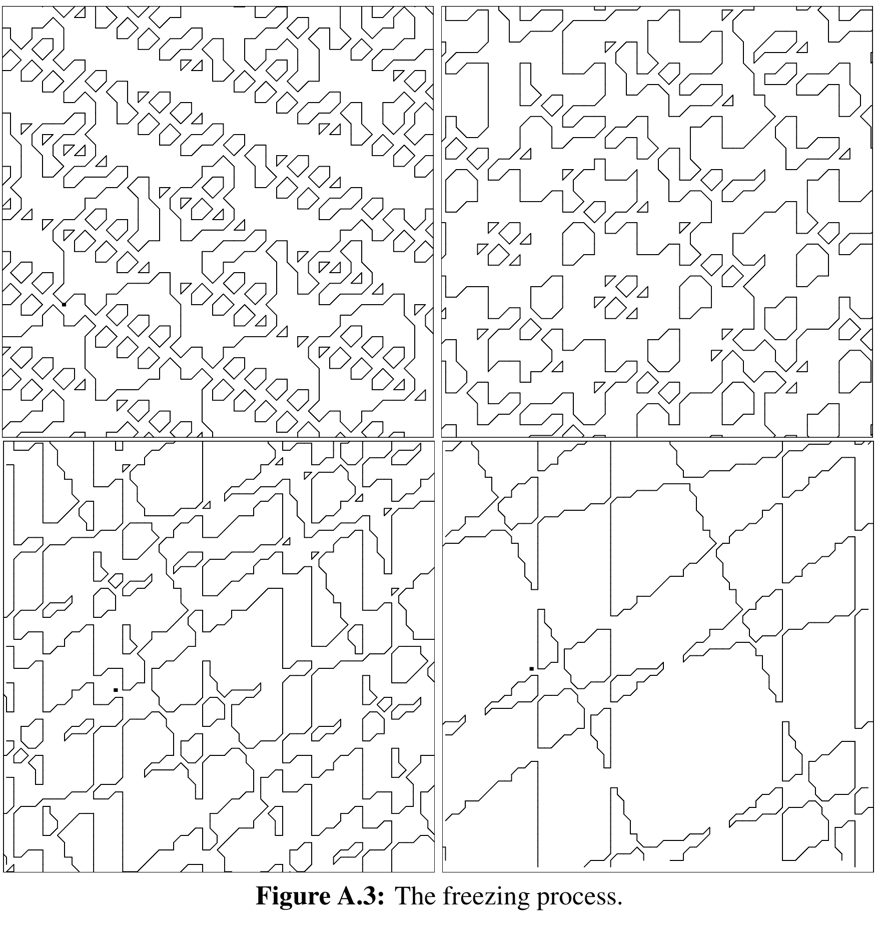

Figure A.3: The freezing process.

Informally, we think of $y \in [0, 1]$ as being a kind of temperature, with 0 corresponding to freezing and 1 corresponding to boiling. Note that the figure for $y \in [-1, 0]$ is symmetric. Thus one sees a similar freezing process as $y \to 0$ from below.

We do have an explanation of sorts for the freezing phenomenon, though we have not worked through all the details. The Master Picture Theorem seems to hold for the general orbits. That is, there is one 5-dimensional picture that works for all orbits and all parameters at once. The Master Picture Theorem we proved here is a boundary case.

As $y \to 0$, the regions in this master partition that assign nontrivial edges to the arithmetic graph seem to concentrate along a finite union of hyperplanes. The preimages of these hyperplanes are the asymptotic lines we see in the freezing process.

Here are some other observations about these generalized arithmetic graphs.

- The Embedding Theorem seems true in general.
- The Hexagrid Theorem is false in general.
- The Diophantine Lemma is false in general.
- All the results in §1.5 are false in general.

We think that most of our theorems ought to have (probably weaker) analogs for the general orbit. We do not know which way to bet on the answer, however. Here are some obvious questions one might ask:

Question 1: Is every orbit in a kite either periodic or unbounded?

Question 2: Is almost every orbit in a kite periodic?

Question 3: Are there any unbounded orbits that are not special orbits?

Question 4: Is every unbounded orbit oscillatory in at least one direction?

In the last question, an orbit is *oscillatory* if its ω-limit set is nonempty. Erratic orbits are oscillatory in both directions. Note that the Comet Theorem completely answers all these questions for orbits in Ω_1.

What makes these questions difficult for us to answer (aside from a general lack of understanding of the situation) is the fact that the Hexagrid Theorem no longer holds. This precise result played a huge role in our overall proof. It is interesting that one sees remnants of the hexagrid, as the asymptotic lines, as the temperature y tends to 0. One might wonder if there is a united Hexagrid Theorem that somehow governs the whole picture. Another difficulty is that the Copy Theorem no longer seems to hold in such a precise way as they did for special orbits.

A.4 GENERAL QUADRILATERALS

First we discuss the situation for trapezoids. As mentioned in the introduction, Dan Genin worked out the complete picture for trapezoids. See [**Ge**]. His work is similar in spirit to the work discussed in this book, though ultimately the situation for trapezoids is simpler. Genin finds that all orbits are bounded, and most are aperiodic. Thus the orbit dichotomy, periodic or unbounded, does not work for trapezoids.

One appealing feature about studying the general quadrilateral is that one can perhaps interpolate between the work in this book and Genin's results. The final picture ought to be compatible with both kites and trapezoids. We have no idea how to carry this out at present. However, in this section, we will present some interesting figures. Our latest version of Billiard King contains a separate program that generalizes some of the features of Billiard King to general quadrilaterals. Indeed, Figure A.3 is taken from this other program.

The space $\mathcal{Q}$ of convex quadrilaterals modulo the affine group is 2-dimensional. For (a, b, c) in the positive orthant of $\mathbf{R}^3$, we let $Q(a, b, c)$ denote the quadrilateral with vertices

$$(0,0), \quad (1,0), \quad (0,1), \quad v = \left(\frac{a+b}{a+b+c}, \frac{b+c}{a+b+c}\right).$$

Any convex quadrilateral is affinely equivalent to some $Q(a, b, c)$. Our coordinatization is adapted to a certain action of the positive matrices in $SL_3(\mathbf{Z})$ on $\mathcal{Q}$, which we will not discuss. The trapezoids correspond to points of the form $(0, b, c)$ and (symmetrically) $(a, b, 0)$.

For the first return map, we take Ξ to be the strip $\mathbf{R}_+ \times [-1, 1]$. This time we consider the solid strip and not just its boundary. Picking a point $(\alpha_1, \alpha_2) \in \Xi$ and watching the first return map, we see a sequence of points

$$(\alpha_1, \alpha_2) + (2m_k, 2n_k) + 2o_k v, \qquad m_k, n_k, o_k \in \mathbf{Z}. \tag{A.3}$$

The lattice path corresponding to the orbit, namely, $\{(m_k, n_k, o_k\}$, lies very close to a plane in $\mathbf{R}^3$. The fact that the y-coordinate lies in $[-1, 1]$ places a relationship on n_k and o_k. We can project into this plane and draw a 2 dimensional figure.

When we do this carefully, taking into account the parity as in Equation 2.10, we get a notion of the arithmetic graph that extends what we have for kites. We show some illustrations below. In all the figures, we start with the offset value $(\alpha_1, \alpha_2) = (0, -1)$. As for the case with kites, we mean to add an infinitesimally small vector to the offset, so as to track well defined orbits. Compare the discussion in §2.5.

Figure A.4.1 shows the figure for the trapezoid with coordinates $(0, 233, 377)$. One of the main diagonals of our bounding box is approximately the baseline. Here 233 and 377 are fairly large Fibonacci numbers. This figure is typical of what one sees for trapezoids.

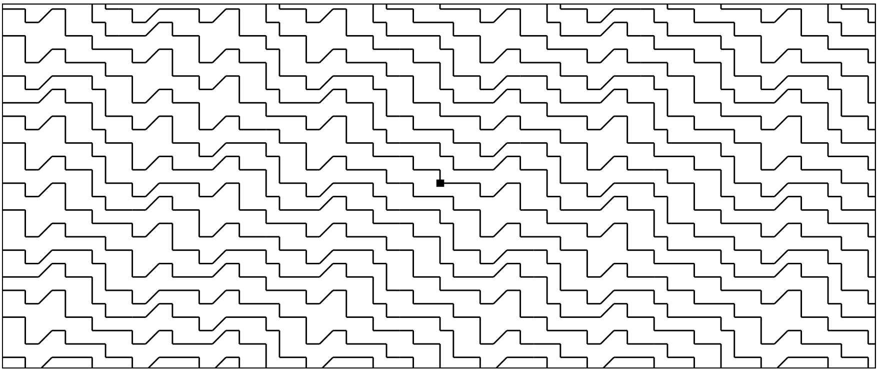

Figure A.4.1: The arithmetic graph for (0, 233, 377).

When we perturb away from the trapezoids, the orbits become much more complicated. Figure A.4.2 shows part of what we would call the fundamental component $\Gamma(1, 233, 377)$. This component tracks essentially the same orbit we considered extensively in the book. The path is part of a single immersed polygonal arc!

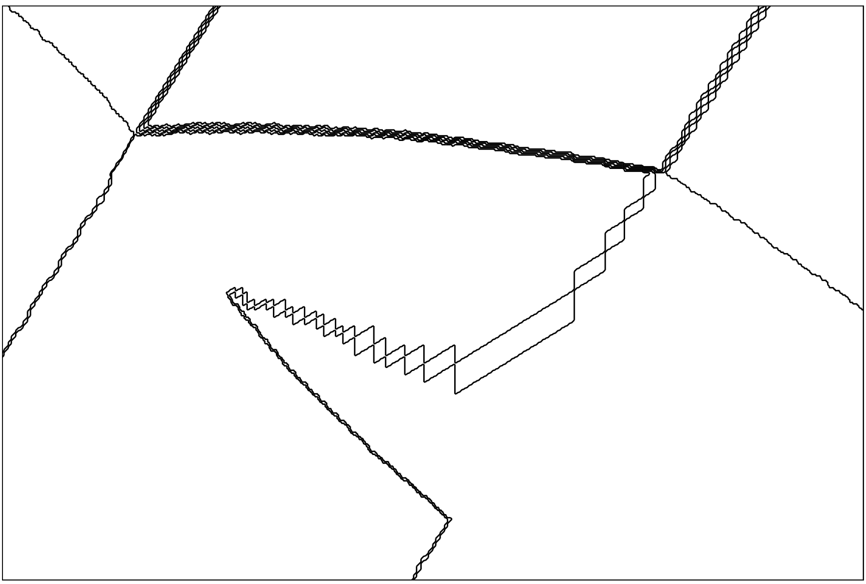

Figure A.4.2: Part of $\Gamma(1, 233, 377)$.

Looking closely at the figure, it seems as if several of the strands approximate curved arcs. It seems that one can get genuinely curved arcs by taking rescaled limits. For instance, a suitable limit of the graphs corresponding to the family $\{(1, F_n, F_{n+1})\}$ seems to have this property. Here F_n is the nth Fibonacci number.

Figure A.4.3 shows a similar phenomenon for a messier fundamental component.

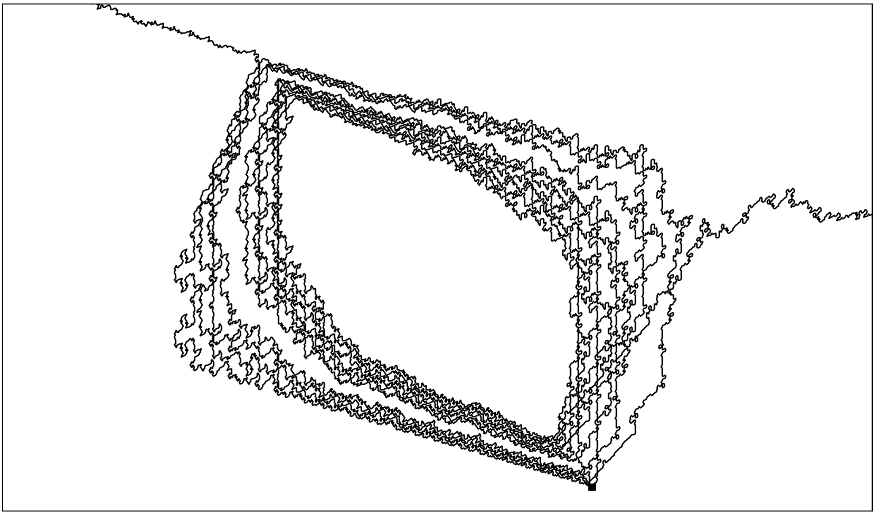

Figure A.4.3: Part of $\Gamma(336, 237, 238)$.

Sometimes the figure for the fundamental orbit dissolves into an incomprehensible cloud, as in Figure A.4.4. We are sure that one can state something interesting about the structure of a polygonal path like this, but we do not know what that statement is. Perhaps the reader can see why we confined our attention to special orbits on kites.

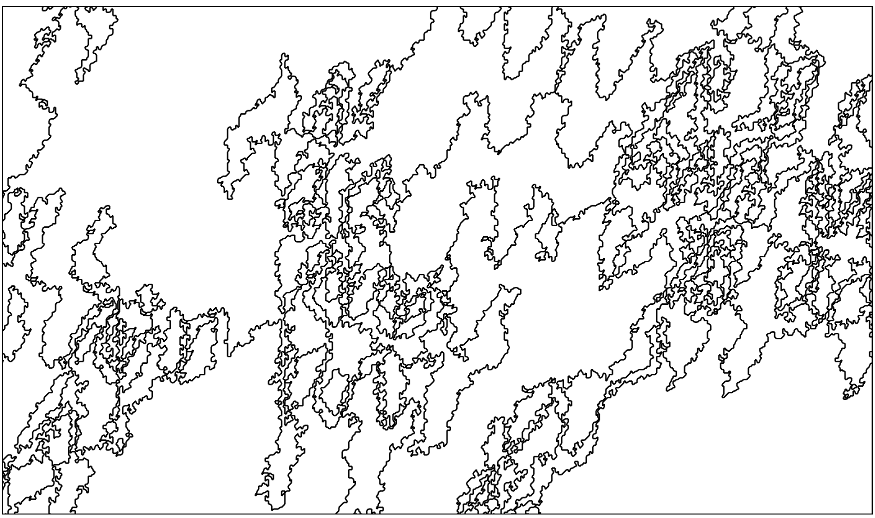

Figure A.4.4: Part of $\Gamma(336, 239, 611)$.

Bibliography

[**B**] P. Boyland, *Dual billiards, twist maps, and impact oscillators*, Nonlinearity **9**:1411–1438 (1996).

[**Be**] A. Beardon, *The Geometry of Discrete Groups*, Graduate Texts in Mathematics 91, Springer, New York (1983).

[**BKS**] T. Bedford, M. Keane, and C. Series, eds., *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces*, Oxford University Press, Oxford (1991).

[**DeB**] N. E. J. De Bruijn, *Algebraic theory of Penrose's nonperiodic tilings*, Nederl. Akad. Wentensch. Proc. **84**:39–66 (1981).

[**Da**] Davenport, *The Higher Arithmetic: An Introduction to the Theory of Numbers*, Hutchinson and Company, London (1952).

[**D**] R. Douady, *These de 3-eme cycle*, Université de Paris 7, 1982.

[**DF**] D. Dolyopyat and B. Fayad, *Unbounded orbits for semicircular outer billiards*, Annales Henri Poincaré, to appear.

[**DT1**] F. Dogru and S. Tabachnikov, *Dual billiards*, Math. Intelligencer **26**(4):18–25 (2005).

[**DT2**] F. Dogru and S. Tabachnikov, *Dual billiards in the hyperbolic plane*, Nonlinearity **15**:1051–1072 (2003).

[**F**] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley and Sons, New York (1990).

[**G**] D. Genin, *Regular and Chaotic Dynamics of Outer Billiards*, Pennsylvania State University Ph.D. thesis, State College (2005).

[**GS**] E. Gutkin and N. Simanyi, *Dual polygonal billiard and necklace dynamics*, Comm. Math. Phys. **143**:431–450 (1991).

[**H**] M. Hochman, *Genericity in topological dynamics*, Ergodic Theory Dynam. Systems **28**:125–165 (2008).

[**Ke**] R. Kenyon, *Inflationary tilings with a similarity structure*, Comment. Math. Helv. **69**:169–198 (1994).

[**Ko**] Kolodziej, *The antibilliard outside a polygon*, Bull. Pol. Acad Sci. Math. **37**:163–168 (1994).

[**M1**] J. Moser, *Is the solar system stable?*, Math. Intelligencer **1**:65–71 (1978).

[**M2**] J. Moser, *Stable and random motions in dynamical systems, with special emphasis on celestial mechanics*, Ann. of Math. Stud. 77, Princeton University Press, Princeton, NJ (1973).

[**MM**] P. Mattila and D. Mauldin, *Measure and dimension functions: measurability and densities*, Math. Proc. Cambridge Philos. Soc. **121**(1):163–168 (1997).

[**N**] B. H. Neumann, *Sharing ham and eggs*, Summary of a Manchester Mathematics Colloquium, 25 Jan 1959, published in Iota, the Manchester University Mathematics Students' Journal.

[**S**] R. E. Schwartz, *Unbounded Orbits for Outer Billiards*, J. Mod. Dyn. **3**:371–424 (2007).

[**T1**] S. Tabachnikov, *Geometry and billiards*, Student Mathematical Library 30, Amer. Math. Soc. (2005).

[**T2**] S. Tabachnikov, *A proof of Culter's theorem on the existence of periodic orbits in polygonal outer billiards*, Geometriae Dedicata **129**(1):83–87 (2007).

[**T3**] S. Tabachnikov, *Billiards*, Société Mathématique de France, "Panoramas et Syntheses" 1, 1995

[**VL**] F. Vivaldi and J. H. Lowenstein, lit Arithmetical properties of a family of irrational piecewise rotations, *Nonlinearity* **19**:1069–1097 (2007).

[**VS**] F. Vivaldi and A. Shaidenko, *Global stability of a class of discontinuous dual billiards*, Comm. Math. Phys. **110**:625–640 (1987).

[**W**] S. Wolfram, *The Mathematica Book*, 4th ed., Wolfram Media/Cambridge University Press, Champaign/Cambridge (1999).

Index