Outer Billiards on Kites

by

Richard Evan Schwartz

Preface

Outer billiards is a basic dynamical system defined relative to a convex shape in the plane. B.H. Neumann introduced outer billiards in the 1950s, and J. Moser popularized outer billiards in the 1970s as a toy model for celestial mechanics. Outer billiards is an appealing dynamical system because of its simplicity and also because of its connection to such topics as interval exchange maps, piecewise isometric actions, and area-preserving actions. There is a lot left to learn about these kinds of dynamical systems, and a good understanding of outer billiards might shed light on the more general situation.

The *Moser-Neumann question*, one of the central problems in this subject, asks *Does there exist an outer billiards system with an unbounded orbit*? Until recently, all the work on this subject has been devoted to proving that all the orbits are bounded for various classes of shapes. We will detail these results in the introduction.

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 tiling. Our proof involves special properties of the Penrose kite, and naturally raises questions about generalizations.

In this monograph we will give a more general and robust answer to the Moser-Neumann question. We will prove that outer billiards has unbounded orbits when defined relative to any irrational kite. A *kite* is probably best defined as a "kite-shaped" quadrilateral. (See the top of §1.2 for a non-circular definition.) The kite is irrational if it is not affinely equivalent to a quadrilateral with rational vertices. Our proof uncovers some of the deep structure underlying outer billiards on kites, and relates the subject to such topics as self-similar tilings, polytope exchange maps, and the modular group.

I discovered every result in this monograph by experimenting with my computer program, Billiard King, a Java-based graphical user interface. For the most part, the material here is logically independent from Billiard King, but I encourage the serious reader of this monograph to download Billiard King from my website [1] and play with it. My website also has an interactive guide to this monograph, in which many of the basic ideas and constructions are illustrated with interactive Java applets.

[1] www.math.brown.edu/∼res

There are a number of people I would like to thank. I especially thank Sergei Tabachnikov, whose great book *Geometry and Billiards* first taught me about outer billiards. Sergei has constantly encouraged me as I have investigated this topic, and he has provided much mathematical insight along the way.

I thank Yair Minsky for his work on the punctured-torus case of the Ending Lamination Conjecture. It might seem strange to relate outer billiards to punctured-torus bundles, but there seems to me to be a common theme. In both cases, one studies the limit of geometric objects indexed by rational numbers and controlled in some sense by the Farey triangulation.

I thank Eugene Gutkin for the explanations he has given me about his work on outer billiards. The work of Gutkin-Simanyi and others on the boundedness of the orbits for rational polygons provided the theoretical underpinnings for some of my initial computer investigations.

I thank Jeff Brock, Peter Doyle, David Dumas, Richard Kent, Howie Masur, Curt McMullen, John Smillie, and Ben Wieland, for their mathematical insights and general interest in this project.

I thank the National Science Foundation for their continued support, currently in the form of the grant DMS-0604426.

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

Table of Contents

1 Introduction

1.1 History of the Problem

B.H. Neumann introduced *outer billiards* in the late 1950s. In the 1970s, J. Moser popularized outer billiards as a toy model for celestial mechanics. One appealing feature of polygonal outer billiards is that it gives rise to a piecewise isometric mapping of the plane. Such maps have close connections to interval exchange transformations and more generally to polygon exchange maps. See [**T1**] and [**DT**] for an exposition of outer billiards and many references.

To define an outer billiards system, one starts with a bounded convex set $S \subset \boldsymbol{R}^2$ and considers a point $x_0 \in \boldsymbol{R}^2 - S$. One defines x_1 to be the point such that the segment $\overline{x_0 x_1}$ is tangent to S at its midpoint and S lies to the right of the ray $\overrightarrow{x_0 x_1}$. (See Figure 1.1 below.) The iteration $x_0 \to x_1 \to x_2$... is called the *forwards outer billiards orbit* of x_0. It is defined for almost every point of $\boldsymbol{R}^2 - S$. The backwards orbit is defined similarly.

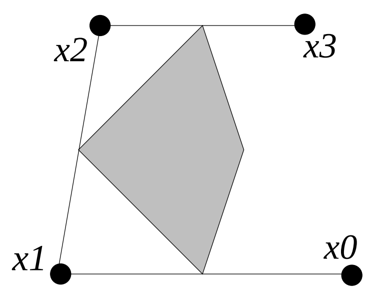

Figure 1.1: Outer Billiards

Moser [**M**, p. 11] attributes the following question to Neumann *circa* 1960, though it is sometimes called *Moser's Question*.

Question: *Is there an outer billiards system with an unbounded orbit?*

This question is an idealized version of the question about the stability of the solar system. The Moser-Neumann question has been considered by various authors. Here is a list of the main results on the question.

- J. Moser [**M**] sketches a proof, inspired by K.A.M. theory, that outer billiards on S has all bounded orbits provided that ∂S is at least C^6 smooth and positively curved. R. Douady gives a complete proof in his thesis, [**D**].

- 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.

- In [**VS**], [**Ko**], and (later, but with different methods) [**GS**], it is proved that outer billiards on a *quasirational polygon* has all orbits bounded. This class of polygons includes rational polygons and also regular polygons. In the rational case, all defined orbits are periodic.

- S. Tabachnikov analyzes the outer billiards system for the regular pentagon and shows that there are some non-periodic (but bounded) orbits. See [**T1**, p 158] and the references there.

- D. Genin [**G**] shows that all orbits are bounded for the outer billiards systems associated to trapezoids. He 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.

- Recently, in [**S**] we proved 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. See §1.6 for a discussion.

The work in [**S**] naturally raises questions about the generality in which the discovered phenomena holds true. The purpose of this monograph is to give a robust affirmative answer to the Moser-Neumann problem, and to somewhat develop the theory of outer billiards on kites. We expect that some of the theory we develop here will work for polygonal outer billiards in general, though right now a general theory is beyond us.

We discovered all the results and phenomena discussed in this monograph using our computer program, Billiard King. One can download this Java-based program from my website and run it on most platforms. Even though we have tried to make this monograph logically independent of Billiard King, we strongly encourage the reader to use the program. We also mention that our website [2] has an interactive guide to this monograph, in which many of the main ideas are explained with simple Java applets.

[2]http://www.math.brown.edu/∼res/BilliardKing

1.2 Main Results

For us, a *kite* is a quadrilateral of the form $K(A)$, with vertices

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

Figure 1.1 shows an example. We call $K(A)$ *(ir)rational* iff A is (ir)rational. Outer billiards is an affinely invariant system, and any quadrilateral Q that is traditionally called a kite is affinely equivalent to some $K(A)$.

Let $\boldsymbol{Z}_{\text{odd}}$ denote the set of odd integers. Reflection in each vertex of $K(A)$ preserves $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$. Hence, outer billiards on $K(A)$ preserves $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$. We say that a *special orbit* on $K(A)$ is an orbit contained in $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$. We call a point in $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$ *special erratic* (relative to the kite) if both the forwards and backwards orbits of this point are unbounded, and return infinitely often to every neighborhood of the vertex set of the kite. Define the *special erratic set* to be the set of special erratic points.

We say that a set S *essentially contains* a set C is there is a countable set S' such that $C - S' \subset S$.

Theorem 1.1 (Erratic Orbit) *Relative to any irrational kite, the special erratic set essentially contains a Cantor set.*

The Erratic Orbit Theorem says, in particular, that outer billiards on any irrational kite has uncountably many unbounded special orbits.

Theorem 1.2 *Relative to any irrational kite, any special orbit is periodic or else unbounded in both directions.*

Theorem 1.3 *Relative to any irrational kite, the set of periodic special orbits is open dense in the set of all special orbits.*

The rational cases of Theorems 1.2 and 1.3 follow from the work of [**VS**], [**K**], and [**GS**], In this case, all orbits are periodic. In the irrational case, Theorem 1.1 makes Theorems 1.2 and 1.3 especially interesting. Computer evidence suggests that Theorems 1.2 and 1.3 hold without the restriction to special orbits. See §9.4 for a discussion.

The monograph has two other main results as well, namely the Hexagrid Theorem (§4) and the Master Picture Theorem (§10), but these results are not easily stated without a buildup of terminology. We will discuss these two results later on in the introduction.

1.3 Rational Kites

We say that p/q is *odd* or *even* according as to whether pq is odd or even. In §6 we will construct, for each irrational $A \in (0,1)$, a canonical sequence $\{p_n/q_n\}$ of odd rationals that converges to A. The sequence $\{p_n/q_n\}$ is similar to the sequence of continued fraction approximants to A. Our approach to the main results is to get a good understanding of the special orbits relative to $K(p_n/q_n)$ and then to take suitable limits.

We find it convenient to work with the square of the outer billiards map rather than the map itself, and our results implicitly refer to the square map. (This has no effect on the results above.)

Let $O_2(x)$ denote the square outer billiards orbit of x. Let

$$\Xi = \boldsymbol{R}_+ \times \{-1,1\}. \tag{2}$$

Our Pinwheel Lemma, from §11, shows that $O_2 \cap \Xi$ pretty well controls O_2 for any special orbit O_2. The intuitive idea is that $O_2(x)$ generally circulates around the kite, and hence returns to Ξ in a fairly regular fashion. We will state the results about rational kites in terms of $O_2 \cap \Xi$.

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 an orbit that starts out "right next to" the kite vertex $(0,1)$. This orbit plays a crucial role in our proof of Theorem 1.1.

Let $\lambda(p/q) = 1$ if p/q is odd and $\lambda(p/q) = 2$ if p/q is even. Our results refer to outer billiards on $K(p/q)$.

Theorem 1.4 *$O_2(1/q, 1) \cap \Xi$ has diameter at least $\lambda(p/q)(p+q)/2$.*

When the time comes, we will assume that $p > 1$ in our proof. The case $p = 1$ is still true, but the proof is more tedious.

We think of Theorem 1.4 as a "rational precursor" of Theorem 1.1. The next result shows that Theorem 1.4 is sharp to within a factor of 2.

Theorem 1.5 *Let $\lambda = \lambda(p/q)$. Each special orbit intersects Ξ in exactly one set of the form $I_k \times \{-1,1\}$, where*

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

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

Theorem 1.5 is similar in spirit to a result in [**K**]. See §4.4 for a discussion.

An outer billiards orbit on $K(A)$ is called *stable* if there are nearby and combinatorially identical orbits on $K(A')$ for all A' sufficiently close to A. Otherwise, the orbit is called *unstable*. In the odd case, $O_2(1/q,1)$ is unstable, and this fact is crucial to our proof of Theorem 1.1. It also turns out that $O(3/q,1)$ is stable, and this fact is crucial to our proof of Theorem 1.3. Here is the complete stability classification of special orbits.

Theorem 1.6 *In the even rational case, all special orbits are stable. In the odd case, the set $I_k \times \{-1,1\}$ contains exactly two unstable orbits, U_k^+ and U_k^-, and these are conjugate by reflection in the x-axis. In particular, we have $U_0^\pm = O_2(1/q,\pm 1)$.*

1.4 The Arithmetic Graph

All our results about special orbits derive from our analysis of a fundamental object, which we call the *arithmetic graph*. The principle guiding our construction is that sometimes it is better to understand the abelian group $\boldsymbol{Z}[A] := \boldsymbol{Z} \oplus \boldsymbol{Z}A$ as a module over $\boldsymbol{Z}$ rather than as a subset of $\boldsymbol{R}$.

In this section we will explain the idea behind the arithmetic graph. In §2.3 we will give a precise construction.

Let ψ denote the square of the outer billiards map. For all $p \in \boldsymbol{R}^2$ on which ψ is defined, we have

$$\psi(p) = p + v_p; \qquad v_p \in 2\boldsymbol{Z}[A] \times 2\boldsymbol{Z} \tag{3}$$

This equation derives from the fact that $v_p = 2(v_1 - v_2)$, where v_1 and v_2 are two vertices of $K(A)$. See Figure 1.2.

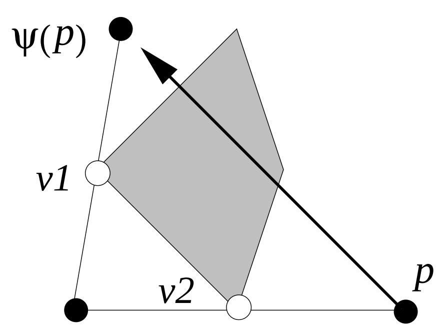

Figure 1.2: The square of the outer billiards map

Let $\Xi = \boldsymbol{R}_+ \times \{-1, 1\}$, as above. The arithmetic graph encodes the arithmetic behavior of the *first return map* $\Psi : \Xi \to \Xi$. Assuming that the orbit of $(2\alpha, 1)$ is defined, it turns out that

$$\Psi^k(2\alpha, 1) = (2\alpha + 2m_k A + 2n_k, \pm 1); \qquad m_k, n_k \in \boldsymbol{Z}. \tag{4}$$

The *arithmetic graph* $\Gamma_\alpha(A)$ is the path in $\boldsymbol{Z}^2$ with vertices (m_k, n_k), connected in the obvious way. We will prove that

$$(m_{k+1}, n_{k+1}) - (m_k, n_k) \in \{-1, 0, 1\}^2 \tag{5}$$

Thus $\Gamma_\alpha(A)$ is a lattice path that connects nearest neighbors in $\boldsymbol{Z}^2$. In the next section we show some pictures. The reader can see pictures for any smallish parameter using either Billiard King or the interactive guide to the monograph.

When $A = p/q$ is rational, we set $\alpha = 1/(2q)$, and simplify our notation to $\Gamma(p/q)$. In the odd case, $\Gamma(p/q)$ is an open polygonal curve, invariant under translation by $(q, -p)$. In the even case, $\Gamma(p/q)$ is an embedded polygon.

We will prove Theorem 1.4 by showing that $\Gamma(p/q)$ rises on the order of $p + q$ units away from the line of slope $-p/q$ through the origin. This line is invariant under translation by $(q, -p)$. Theorems 1.5 and 1.6 are based on a more global version of $\Gamma(p/q)$, where we consider simultaneously the arithmetic graphs of all orbits of the form

$$(1/q + 2\boldsymbol{Z}[A]) \times \{-1, 1\} \tag{6}$$

The resulting object, which we call $\widehat{\Gamma}(p/q)$, turns out to be a disjoint union of embedded polygons and embedded infinite polygons. This is the content of our Embedding Theorem. See §2.4. Each component corresponds to a combinatorial equivalence class of special orbit. The component is a polygon iff the orbit is stable.

In §4 we will explain a general structural result for the arithmetic graph, the Hexagrid Theorem. Part III of the monograph is devoted to the proof of the Hexagrid Theorem. It turns out that the large scale structure of $\widehat{\Gamma}$ is controlled by a grid made from 6 infinite families of parallel lines. We call this grid the hexagrid. The lines of the hexagrid confine the stable components and force the unstable components to oscillate in predictable ways. Figure 4.3 shows a picture of how the arithmetic graph interacts with the hexagrid. We view the Hexagrid Theorem as a being similar in spirit to De Bruijn's pentagrid construction of the Penrose tilings. See [**DeB**].

1.5 Period Copying

What allows us to pass from our results about rational kites, especially Theorem 1.4, to the results about irrational kites is a period copying phemomenon. We will illustrate the idea with some pictures. Each picture shows $\Gamma(p/q)$ in reference to the line of slope $-p/q$ through the origin. Figure 1.3 shows $\Gamma(7/13)$.

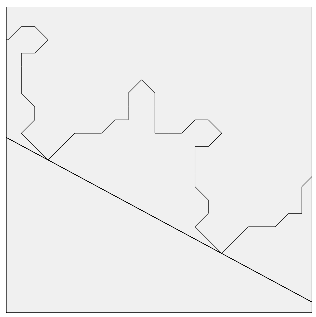

Figure 1.3: The graph $\Gamma(7/13)$.

Figure 1.4 shows a picture of $\Gamma(19/35)$. Notice that $\Gamma(19/35)$ has a much wider oscillation, but also manages to copy a bit more than one period of $\Gamma(7/13)$. What makes this work is that $7/13$ is a very good approximation to $19/35$. The results in §7, which we call Copy Theorems, deal with this phenomenon. Our interactive guide to the monograph has several applets that let the user see this phenomenon for arbitrary pairs of smallish parameters.

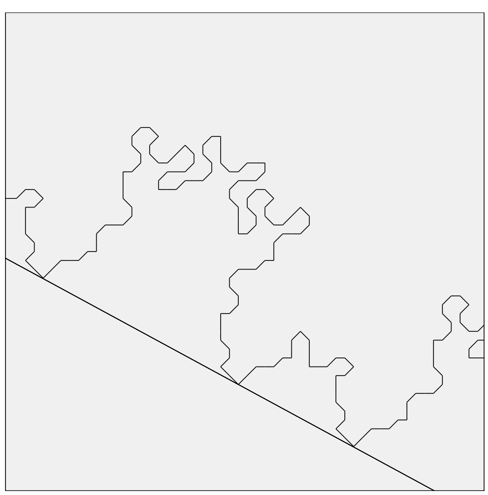

Figure 1.4: The graph $\Gamma(19/35)$.

Figure 1.5 shows the same phenomenon for $\Gamma(45/83)$. This graph oscillates on a large scale but still manages to copy a bit more than one period of $\Gamma(19/35)$. Hence $\Gamma(45/83)$ copies a period of $\Gamma(7/13)$ and a period of $\Gamma(19/35)$. That is, $\Gamma(45/83)$ oscillates on 3 scales.

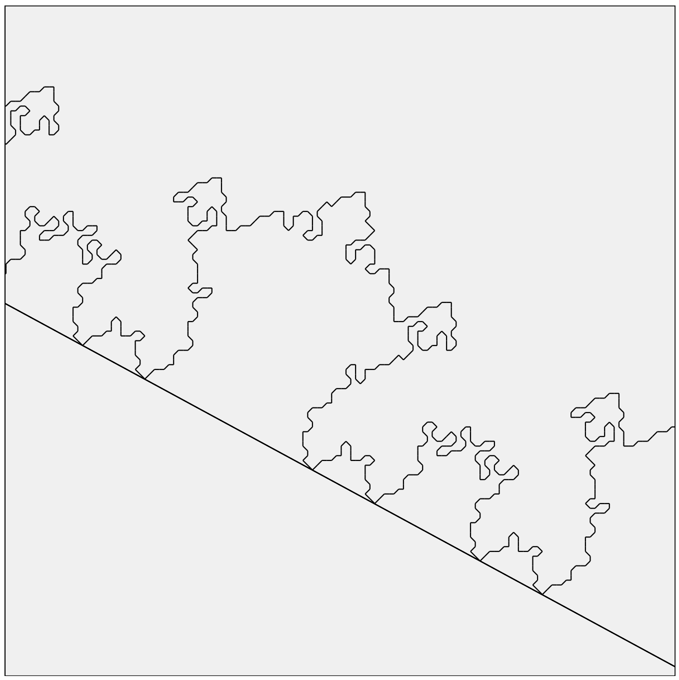

Figure 1.5: The graph $\Gamma(45/83)$.

Continuing in this way, and taking a limit, we produce a graph $\Gamma(A)$ that oscillates on all scales. By taking the limit carefully, we arrange that the limit is the arithmetic graph of a well-defined, erratic orbit.

We first prove Theorem 1.1 in §3, under the hypothesis that A has a fairly good approximation by odd rationals. Our argument takes care of almost every A. To handle the general parameter, we will build up a little theory of odd rational approximation, similar to the theory of continued fractions. What distinguishes the general case from the special case considered in §3 is that the general case requires a more delicate period copying theorem. We will deduce Theorem 1.2 from Theorem 1.1. Theorem 1.3 requires a few other ideas.

1.6 The Case of the Penrose Kite

The Penrose kite is affinely equivalent to $K(\phi^{-3})$, where ϕ is the golden ratio. In [**S**], we proved that outer billiards on $K(\phi^{-3})$ has erratic orbits by considering the arithmetic graph $\Gamma(\phi^{-3})$ corresponding to the orbit of $(\phi^{-2}/2, 1)$. We showed that $\Gamma(\phi^{-3})$ oscillates on infinitely many scales by showing that $\Gamma(\phi^{-3})$ lies in a small tubular neighborhood of the dilated graph $\phi^3\Gamma(\phi^{-3})$. In other words, $\Gamma(\phi^{-3})$ has a kind of self-similar structure. It is a large-scale fractal. Compare [**Ke**].

The self-similarity of $\Gamma(\phi^{-3})$ derives from the fact that ϕ^{-3} has a periodic continued fraction expansion, namely $[0; 4, 4, 4, 4...]$. Put another way, there is a kind of recurrent pattern to the different scales on which $\Gamma(\phi^{-3})$ oscillates. Something like this should work for all quadratic irrationals, but we have not worked it out. For the general irrational parameter A, we do not get such a symmetric picture, and the scales of oscillation of $\Gamma(A)$ depend on the Diophantine properties of A.

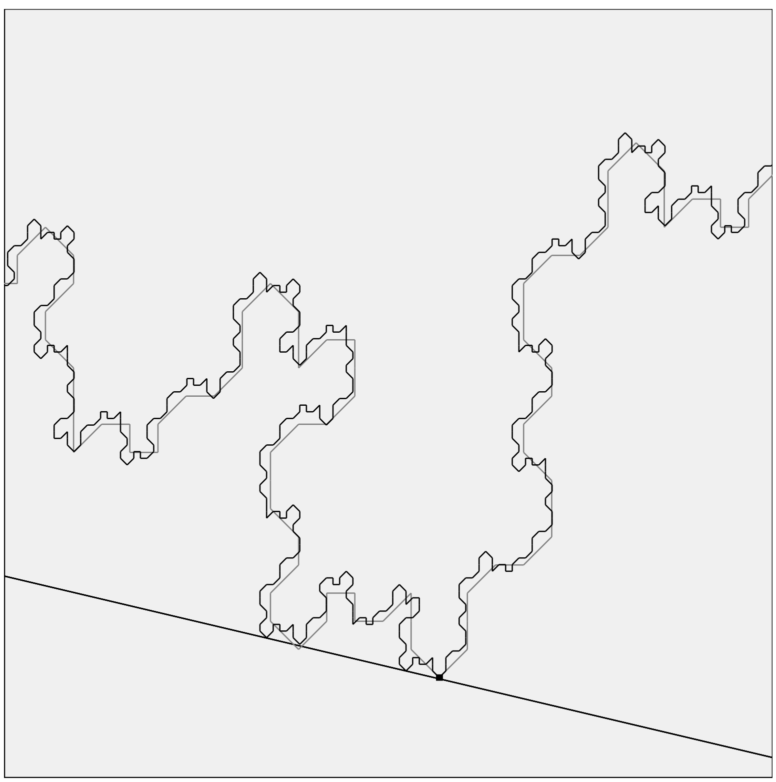

Figure 2.5: The graph $\Gamma(\phi^{-3})$ superimposed over $\phi^3\Gamma(\phi^{-3})$.

1.7 The Master Picture Theorem

All our main theorems follow from a combination of the Embedding Theorem, the Hexagrid Theorem, and the Copy Theorems. These three results have a common source, a result that we call the Master Picture Theorem. We formulate and prove this result in Part II of the monograph. Here we will give the reader a feel for the result.

Recall that $\Xi = \boldsymbol{R}_+ \times \{-1, 1\}$. The arithmetic graph encodes the dynamics of the first return map $\Psi : \Xi \to \Xi$. It turns out that Ψ is an infinite interval exchange map. The Master Picture Theorem reveals the following structure.

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 $\widehat{\Psi} : \widehat{\Xi} \to \widehat{\Xi}$, defined relative to a partition of $\widehat{\Xi}$ into 28 polyhedra.

3. The map μ is a semi-conjugacy 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 our Copy Theorems. 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.

I believe that a version of the Master Picture Theorem should hold for general convex n-gons. For instance, one sees a more general picture extending the one we have explained here when one considers all the orbits on a kite. John Smillie and I have some ideas on how to work out the general case, and we hope to pursue this at a later time.

1.8 Computational Issues

As I mentioned above, I discovered all the structure of outer billiards by experimenting with Billiard King. Ultimately, I am trying to verify the structure I noticed on the computer, and so one might expect there to be some computation in the proof. The proof here uses considerably less computation than the proof in [**S**], but I still use a computer-aided proof in several places. For example, I use the computer to check that various 4 dimensional convex integral polytopes have disjoint interiors.

To the reader who does not like computer-aided proofs (however mild) I would like to remark that the experimental method here has some advantages over a traditional proof. I checked all the main steps in the proof with massive and visually-based computation. These checks make sure that I am not led astray by logical or conceptual errors arising from steps taken *in vacuo*. I came to the Moser-Neumann problem as a kind of blank state, and only got the ideas for general structural statements by looking at concrete evidence.

Again, I mention that my website has an interactive java-based guide to this monograph. The interested reader can play with simple java applets that illustrate and explain many of the ideas. If nothing else, this interactive guide provides extensive color versions of the pictures we have shown above.

1.9 Organization of the Monograph

This monograph comes in 4 parts. In Part I, we prove all the main results modulo the Hexagrid Theorem, the Embedding Theorem, the Copy Theorems, and a few smaller auxilliary results. In Part II, we prove the Master Picture Theorem. In Parts III and IV we deduce all the auxilliary theorems from the Master Picture Theorem. Before each part of the monograph, we include an overview of the contents of that part.

Part I

Here is an overview of this part of the monograph.

- In §2 we give background material on polygonal outer billiards, and then define the arithmetic graph. The arithmetic graph is our main object of study. In the last section we state two results, Lemma 2.5 and 2.6, that suffice to prove Theorem 1.1 for a full measure set of parameters.
- In §3 we prove Theorem 1.1, modulo the results from §2, for almost every irrational number in $(0, 1)$. Some readers might be interested mainly in the resolution of the Moser-Neumann problem, and our result in §3 does the job. These readers can skip §5,6,8,9.
- In §4 we state the Hexagrid Theorem, the main geometric result about arithmetic graphs, and then deduce Lemma 2.5, Theorems 1.4, 1.5, and 1.6. We prove the Hexagrid Theorem in Part III of the monograph.
- In §5, we explain some further structural properties of the arithmetic graph. The basic themes in this chapter are symmetry. stability, and topological orbit confinement.
- In §6 we develop a little theory about approximating irrational numbers by odd rational numbers. This theory parallels the ordinary theory of continued fraction approximants.
- In §7 we state our main period copying theorem, the Copy Theorems I and II.
- In §8 we prove Theorems 1.1, 1.2, and 1.3 in full generality. Our proofs put together all the material from previous chapters.
- In §9 we discuss some interesting experimental phenomena we discovered but have not yet proved. In particular, we discuss the behavior of general outer billiards orbits on kites.

2 The Arithmetic Graph

2.1 Polygonal Outer Billiards

Let P be a polygon. We denote the outer billiards map 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 $\boldsymbol{R}^2 - P$. This countable set of line segments is sometimes called the *limit set*.

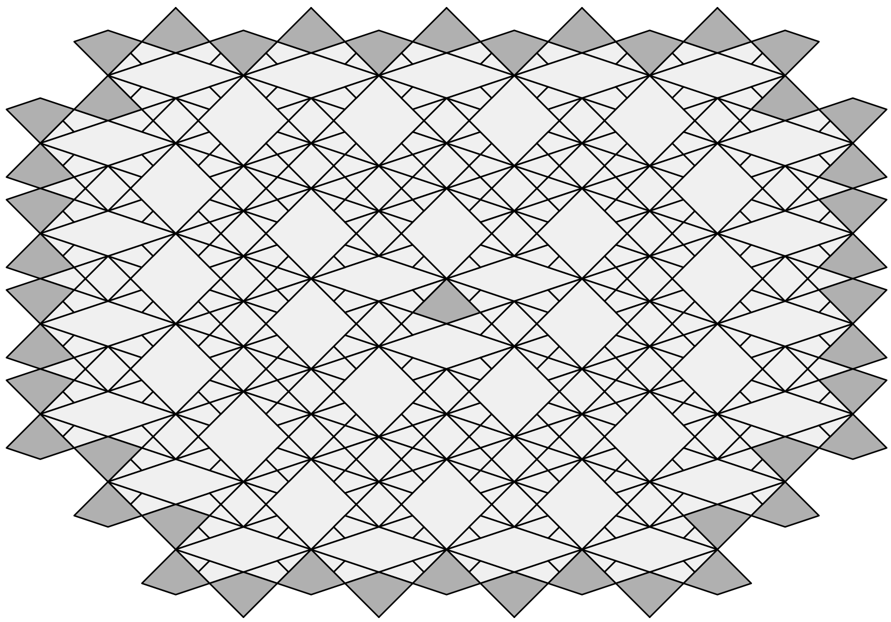

Figure 2.1: Part of the Tiling for $K(1/3)$.

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 dy dynamically invariant convex polygons. Figure 2.1 shows the picture for the kite $K(1/3)$.

This is the simplest tiling [3] we see amongst all the kites. We have only drawn part of the tiling. The reader can draw more of these pictures, and in color, using Billiard King. The existence of these tilings was what motivated me to study outer billiards. I wanted to understand how the tiling changed with the rational parameter and saw that the kites gave rise to highly non-trivial pictures.

[3]Note that the picture is rotated by 90 degrees from our 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: We note first that the order 2 rotations about the vertices of $K(A)$ send the point (x, y) to the point:

$$(-2-x, -y); \qquad (-x, 2-y); \qquad (-x, -2-y); \qquad (2A-x, -y). \tag{7}$$

Let ψ' denote the outer billiards map on $K(A)$. The map ψ' is built out of the 4 transformations from Equation 7. The set $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$ is a countable collection of lines. Let $\Lambda \subset \boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$ denote the set of points of the form $(2a+2bA, 2c+1)$, with $a, b, c \in \boldsymbol{Z}$. The complementary set $\Lambda^c = \boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}} - \Lambda$ is the union of the special intervals.

Looking at Equation 7, we see that $\psi'(x) \in \Lambda^c$ provided that $x \in \Lambda^c$ and ψ' is defined on x. To prove this lemma, it suffices to show that ψ' is defined on any point of Λ^c.

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

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

Here $n \in \boldsymbol{Z}$. Notice that all these points lie in Λ. ♠

Let $\boldsymbol{Z}[A] = \boldsymbol{Z} \oplus \boldsymbol{Z}A$. More generally, the same proof gives:

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

When A is irrational, the set $2\boldsymbol{Z}[A]$ is dense in $\boldsymbol{R}$. However, it is always a countable set.

2.3 The Return Map

Let $\Xi = \boldsymbol{R}_+ \times \{-1, 1\}$, as in the introduction. Let $A \in (0,1)$ be any parameter. As in §1 we mean to discuss outer billiards relative to the kite $K(A)$. The following result is a consequence of the Pinwheel Lemma, proved in §11.

Lemma 2.3 (Return) *If the outer billiards orbit of $p \in \Xi$ is defined, then both the forwards and backwards direction return to Ξ.*

We let $\Psi(p)$ denote the first point in the forwards ψ-orbit of $p \in \Xi$. In short, $\Psi : \Xi \to \Xi$ is the first return map for ψ. Given the nature of the maps in Equation 7 comprising ψ, we see that

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

Combining the Pinwheel Lemma from §11 and the Master Picture Theorem, we have a much stronger result.

$$\Psi(p) - (p) = 2(A\epsilon_1 + \epsilon_2, \epsilon_3); \quad \epsilon_j \in \{-1, 0, 1\}; \quad \sum_{j=1}^{3} \epsilon_j \equiv 0 \bmod 2. \tag{8}$$

On a nuts-and-bolts level, this monograph concerns how to determine the *return pair* $(\epsilon_1(p), \epsilon_2(p))$ as a function of $p \in \Xi$. The pair (ϵ_1, ϵ_2) and the parity condition determine ϵ_3.

Remarks:

(i) Some version of the Return Lemma also appears in [**K**] and [**GS**]. For instance, in [**GS**], the authors are concerned with the return of an orbit to a certain region called *the runway*. The runway is a generalization of Ξ.

(ii) Once we develop the picture a bit, the reader will see that the Return Lemma is obvious for points of Ξ that are far from the kite: Geometrically, the ψ orbit circulates around the kite, skipping at most 1 line of $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$ at a time, and so it must keep returning to Ξ. The tricky part of the analysis is for points near the kite.

(iii) 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 $p = (x, \pm 1)$ only depends on x.

2.4 The Arithmetic Graph

Recall that $\Xi = \boldsymbol{R}_+ \times \{-1, 1\}$. Define $M = M_{A,\alpha} : \boldsymbol{R} \times \{-1, 1\}$ by

$$M_{A,\alpha}(m, n) = \Big(2Am + 2m + 2\alpha, (-1)^{m+n+1}\Big) \tag{9}$$

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 8. We call M a *fundamental map*. Each choice of α gives a different map.

When A is irrational, M is injective. In the rational case, M is injective on any disk of radius q. Given $p_1, p_2 \in \boldsymbol{Z}^2$, we write $p_1 \to p_2$ iff the following holds.

- $\zeta_j = M(p_j) \in \Xi$.
- $\Psi(\zeta_1) = \zeta_2$.
- $\|p_1 - p_1\| < q$.

(The third condition is redundant in the irrational case.) This construction gives a directed graph with vertices in $\boldsymbol{Z}^2$. We call this graph the *arithmetic graph* and denote it by $\widehat{\Gamma}_\alpha(A)$. When $A = p/q$ we have the canonical choice $\alpha = 1/(2q)$ and then we set

$$\widehat{\Gamma}(p/q) = \widehat{\Gamma}_{1/(2q)}(p/q). \tag{10}$$

We say that the *baseline* of $\widehat{\Gamma}(A)$ is the line $M^{-1}(0)$. The whole arithmetic graph lies above the baseline. In Part III of we will prove the following result.

Theorem 2.4 (Embedding) *For any $A \in (0, 1)$ and $\alpha \notin \boldsymbol{Z}[A]$, the graph $\widehat{\Gamma}_\alpha(A)$ is a disjoint union of embedded polygons and embedded infinite polygonal curves.*

Remark: In the arithmetic graph, there are some lattice points having no edges emanating from them. These isolated points correspond to points where the return map is the identity and hence the orbit is periodic in the simplest possible way. We usually ignore these trivial components.

We are mainly interested in the component of $\widehat{\Gamma}$ that contains $(0, 0)$. We denote this component by Γ. In the rational case, $\Gamma(p/q)$ encodes the structure of the orbit $O_2(1/q, -1)$. (The orbit $O_2(1/q, 1)$, the subject of Theorem 1.4, is conjugate to $O_2(1/q, -1)$ *via* reflection in the x-axis.)

In §5.5 we will see that $\Gamma(p/q)$ is a closed polygon when p/q is an even rational. In the next section, we will discuss the odd rational case in detail.

2.5 The Odd Rational Case

Let p/q be an odd rational. We define

$$V = (q, -p); \qquad W = \left(\frac{pq}{p+q}, \frac{pq}{p+q} + \frac{q-p}{2}\right). \tag{11}$$

Note that W typically does not belong to $\boldsymbol{Z}^2$. We let $R(p/q)$ denote the parallelogram whose vertices are

$$(0,0); \qquad V; \qquad W; \qquad V+W. \tag{12}$$

Let

$$d_0 = (x, \text{floor}(y)); \qquad x = \frac{p+q}{2}; \qquad y = \frac{q^2 - p^2}{4q} \tag{13}$$

Lemma 2.5 *$\Gamma(p/q)$ 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 $R(p/q)$.*

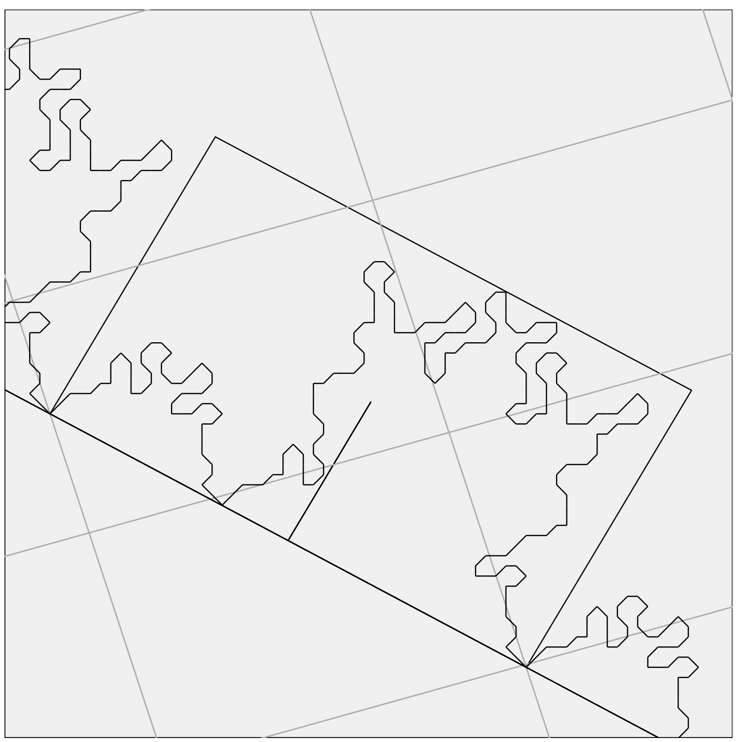

Figure 2.3: $\Gamma(27/45)$ and $R(27/45)$

Figure 2.3 illustrates Lemma 2.5 when $p/q = 25/47$. In all cases, the point $(x, y$ lies on the line segment connecting the midpoints of the top and bottom edges of $R(p/q)$, closer to the top edge. In Figure 2.3, the point (x, y) lies at the intersection of the two grey grid lines. These grid lines are part what we call the hexagrid, a system of 6 infinite families of parallel lines. The sides of $R(p/q)$ are also part of the hexagrid.

Lemma 2.5 is a corollary of the of our main structural result, the Hexagrid Theorem. We will deduce Lemma 2.5 in §4.2. The odd case of Theorem 1.4 is a quick corollary of Lemma 2.5. See §20.5 for a proof of Theorem 1.4 in the even case.

Proof of Theorem 1.4, Odd Case: Let M_1 be the first coordinate for the fundamental map of p/q. We compute that $M_1(d_0) > (p+q)/2$, at least when $p > 1$. Technically, $\Gamma(p/q)$ describes $O_2(1/q, -1)$, but the two orbits $O_2(1/q, 1)$ and $O_2(1/q, -1)$ are conjugate by reflection in the x-axis. ♠

Let $\Gamma^1(p/q)$ denote the period of $\Gamma(p/q)$ that is contained in $R(p/q)$. For any $\epsilon > 0$, let

$$\Gamma^{1+\epsilon} = \Gamma \cap (R \cup B_{\epsilon q}(V)). \tag{14}$$

Here $B_{\epsilon q}(V)$ is the ball of radius ϵq about V, the bottom right vertex of $R(p/q)$. We think of $\Gamma^{1+\epsilon}$ as being slightly more than one period of Γ.

Lemma 2.6 *Let $\epsilon = 1/16$. Let p_1/q_1 and p_2/q_2 be two odd rationals such that*

$$0 < \frac{p_2}{q_2} - \frac{p_1}{q_1} < \frac{1}{2q_1^2}.$$

If p_1 is sufficiently large and ϵ is sufficiently small, then $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ agree on $R_1 \cup B_{\epsilon q_1}(V_1)$. In particular, $\Gamma_1^{1+\epsilon} \subset \Gamma_2^1$.

Remarks:
(i) In §7 we will deduce Lemma 2.6 from the Copy Theorem I.
(ii) The constant $\epsilon = 1/16$ is not optimal, but we don't care about finding the optimal constant.
(iii) There is another version of Lemma 2.6, designed for the case when we have $p_2/q_2 < p_1/q_1$. This alternate version has a very similar formulation, involving the parallelogram $R(p_1/q_1) - (q_1, -p_1)$ instead of $R(p_1/q_1)$.
(iv) Lemma 2.5 and Lemma 2.6 together suffice to prove Theorem 1.1 for a full measure set of parameters. We give the proof in the next chapter.

3 Erratic Orbits Almost Always

3.1 Good Approximation Properties

Let $\Delta(k)$ denote the set of numbers $A \in (0,1)$ such that

$$\left|A - \frac{p}{q}\right| < \frac{1}{kq^2}; \qquad p, q \in \boldsymbol{Z}_{\text{odd}} \tag{15}$$

holds infinite often. It is a fairly easy exercise to show that $\Delta(k)$ has full measure in $(0,1)$ for any $k > 0$. As Curt McMullen pointed out to me, every irrational in $(0,1)$ belongs to $\Delta(1)$. In this chapter, we will prove Theorem 1.1 for all $A \in \Delta(2)$.

If $A \in \Delta(2)$ then A is the limit of a monotone sequence $\{p_n/q_n\}$ of odd rationals satisfying

$$0 < \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| < \frac{1}{2q_n^2}; \qquad \forall n. \tag{16}$$

For ease of exposition, we will treat the case when this sequence is monotone increasing. This case is designed to work with Lemma 2.6. The other case would work with the unstated version of Lemma 2.6 that we mentioned in Remark (iii) at the end of the last chapter.

As we mentioned in the introduction, the basic idea in our proof is to take a limit of the graphs $\Gamma_n = \Gamma(p_n/q_n)$. In order to guarantee that the sequence of graphs converges, we arrange that Γ_{n+1} copies $1 + \epsilon$ periods of Γ_n for all n, in the sense of Lemma 2.6. Having Γ_{n+1} copy slightly more than one period of Γ_n allows us to see a kind of large-scale Cantor set structure embedded in the arithmetic graphs.

The straightforward limit $\lim_{n\to\infty} \Gamma_n$ does exist, but it is not the arithmetic graph of a well-defined orbit. Γ_n describes the orbit of $(1/q_n, -1)$ and our limit would describe the orbit of $(0, -1)$, a kite vertex.

In order to get around this pitfall, we take a more sophisticated kind of limit. We set $N = 2n+1$ and consider translations of the form $\Gamma_n' = \Gamma_N - \omega_n$. Here ω_n is a certain vertex of Γ_N. Our construction works for any binary sequence with infinitely many 0s and infinitely many 1s. The great flexibility of the construction produces arithmetic graphs that track an uncountable set of different orbits.

3.2 A Large Scale Cantor Set

Let $v_n = (q_n, -p_n)$. Lemma 2.6 gives us $\Gamma_n^{1+\epsilon} \subset \Gamma_{n+1}^1$. Define

$$\Gamma_n^2 = \Gamma_n^1 + v_{n+1}. \tag{17}$$

Note that $\Gamma_n^2 \subset \Gamma_{n+1}$, because Γ_{n+1} is invariant under translation by v_{n+1}.

If necessary, we pass to a further subsequence so that

$$\epsilon q_{n+1} > 100 q_n. \tag{18}$$

Then

$$\Gamma_n^{1+\epsilon} \subset B_{\epsilon q_{n+1}}(0,0) \cap \Gamma_{n+1}. \tag{19}$$

Combining translational symmetry, Lemma 2.6, and Equation 19, we have

$$\Gamma_n^2 \subset B_{\epsilon q_{n+1}}(v_{n+1}) \subset \Gamma_{n+1}^{1+\epsilon} \subset \Gamma_{n+2}^1. \tag{20}$$

In summary,

- $\Gamma_n^1 \subset \Gamma_{n+1}^1$ for all n.
- $\Gamma_n^2 = \Gamma_n^1 + v_{n+1} \subset \Gamma_m^1$ for all (m,n) such that $m \geq n+2$.

Figure 3.3 shows the sort of structure we have arranged. In the figure, the notation nk stands for Γ_n^k.

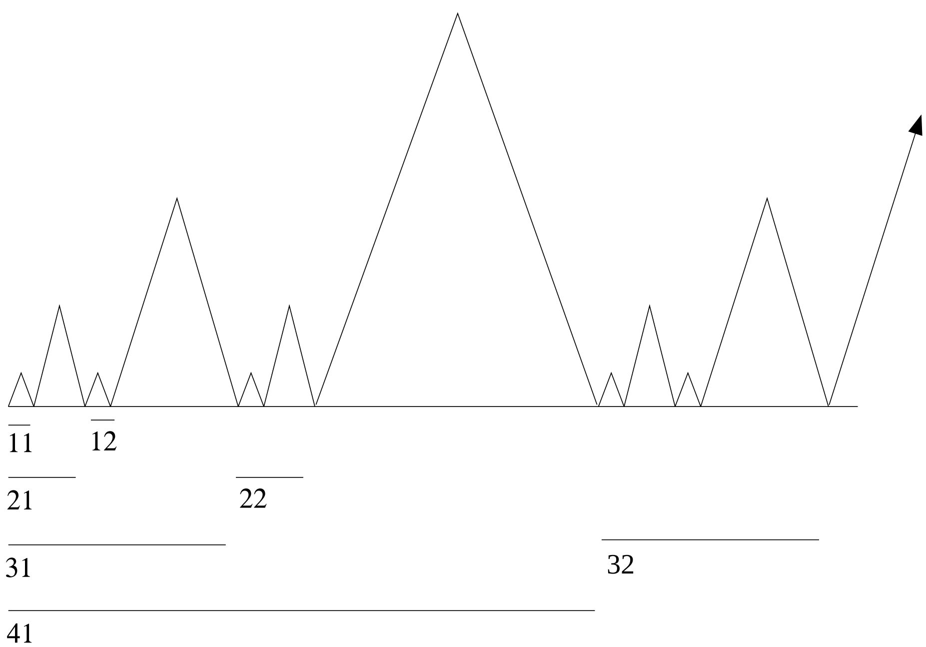

Figure 3.3: large scale Cantor set structure

Let $\epsilon \in \{0,1\}$. First, we have $\epsilon v_2 \in \Gamma_2^1 \subset \Gamma_3^1$. Second, for any $n \geq 1$ we have $\Gamma_{2n-1}^1 + \epsilon v_{2n} \subset \Gamma_{2n+1}^1$. From these facts and induction, we have

$$\omega(\sigma) := \sum_{k=1}^{n} \epsilon_k v_{2k} \subset \Gamma_{2n+1}^1 \tag{21}$$

for any binary sequence $\sigma = \epsilon_1, ..., \epsilon_n$. For $k \leq n$, we let σ_k denote the sequence obtained by just switching the kth digit of σ. Let σ_k' denote the sequence obtained by setting the first $k-1$ digits of σ to 0 and the kth digit to 1.

Lemma 3.1 (Translation) *Let γ be the arc of Γ_{2k+1}^1 connecting $\omega(\sigma)$ to $\omega(\sigma_k)$. Then γ is a translation equivalent to a single period of Γ_{2k}. Up to translations, γ only depends on the first k digits of σ. Finally, $\omega(\sigma_k') \subset \gamma$.*

Proof: We first consider the case when $n = k$. Our claim is independent of the value of the last digit of σ. So, we consider the case when σ ends in a 0 and σ_k ends in a 1. Then γ connects the two points

$$\omega(\sigma) \subset \Gamma_{2k-1}^1 \subset \Gamma_{2k}^1; \qquad \omega(\sigma_k) \in \Gamma_{2k-1} + v_{2k} \subset \Gamma_{2k} \cap \Gamma_{2k+1}^1.$$

The arc $\Gamma_{2k-1}^1 + v_{2k}$ starts out the second period of Γ_{2k}. From this structure we see that γ is exactly one period of Γ_{2k}, and $v_{2k} = \omega(\sigma_k') \in \gamma$.

In general, let σ^* denote the sequence obtained by setting all digits after the kth one to 0. Let γ^* be the arc connecting $\omega(\sigma^*)$ to $\omega(\sigma_k^*)$. Then, by the special case we have already considered, γ^* is one period of Γ_{2k} and $\gamma^* \subset \Gamma_{2k+1}^1$. By an inductive argument, we establish that

$$\gamma^* + \sum_{j=k+1}^{n} \epsilon_j v_{2j} \subset \Gamma_{2n+1}^1.$$

But the arc on the left hand side of this equation connects $\omega(\sigma)$ to $\omega(\sigma_k)$, and hence equals γ, the arc of interest to us. In short

$$\gamma = \gamma^* + \sum_{j=k+1}^{n} \epsilon_j v_{2j}. \tag{22}$$

Equation 22 combines with the special case we have already considered to establish the lemma. ♠

3.3 Taking a Limit

Our limit construction is based on any infinite binary sequence $\sigma = \{\epsilon_n\}$ with infinitely many 0s and infinitely many 1s. Let $\sigma(n)$ denote the first n digits of σ. For $k < n$, let $\sigma_k(n)$ denote the sequence obtained from $\sigma(n)$ by switching the kth digit. Let $\omega(n) = \omega(\sigma(n))$ and $\omega_k(n) = \omega(\sigma_k(n))$.

Let $N = 2n + 1$. Define

$$\Gamma'_n = \Gamma_N - \omega(n). \tag{23}$$

Note that Γ'_n is the arithmetic graph for the point $\zeta_n = M_N(\omega(n))$. Here M_N is the fundamental map associated to p_N/q_N. (See Equation 28 below.)

Lemma 3.2 *There are divergent sequences $\{E_{0n}\}$ and $\{E_{1n}\}$ such that the first E_{0n} edges of Γ'_m in the forwards direction are independent of $m \geq n$ and the first E_{1n} edges of Γ'_m in the backwards direction are independent of $m \geq n$.*

Proof: Let γ_k be the arc connecting $\omega(n)$ to $\omega_k(n)$. By the Translation Lemma, the arc $\gamma'_k = \gamma_k - \omega_n$ connecting 0 to $\omega_k(n) - \omega(n)$ belongs to Γ'_n and is independent of $n > k$. For each n, we let $n_i < n$ be the largest place where the n_ith digit of σ is an i. Let E_{in} denote the number of edges in γ'_{n_i}. These sequences do the job. ♠

Lemma 3.2 implies that the graphs $\{\Gamma'_n\}$ converge to a limiting graph Γ_∞. We want to control the rate of this convergence. Here are 4 basic definitions.

- Let L_N denote the baseline of Γ_N. Let

$$L'_n = L_N - \omega(n) \tag{24}$$

- Let B_n denote the ball of radius $2\max(E_{0n}, E_{1n})$ about 0. Note that the first E_{0n} forwards edges of Γ_∞ lie in B_n and the first E_{1n} backwards edges of Γ_∞ lie in B_n.

- Let d_n denote the distance from $\omega(n)$ to L_N. Note that d_n is also the distance from 0 to L'_n.

- Define $M_0(m, n) = 2Am + n$. Here M_0 is the linear part of the first coordinate of any fundamental map associated to the parameter A.

Passing to a subsequence we can arrange the following.

- $M_0(v_n) < 10^{-n}$.
- $|d_m - d_n| < 10^{-n}$ for all $m > n$.
- $|A_n - A_m| < 10^{-n}$ for all $m > n$.
- The limit $L_\infty = \lim L'_n$ exists, and $L'_n \cap B_n$ lies within 10^{-n} of L_∞.

Only the last item requires some explanation. The size of B_n is at most $O(q_{2n})$ whereas the slope of L'_n differs from the slope of L_∞ by $O(q_{2n}^{-2})$.

Lemma 3.3 Γ_∞ *rises unboundedly far away from* L_∞ *in either direction.*

Proof: The point d_0 in Lemma 2.5 lies at least $q/4$ units above the baseline. Hence, by Lemma 2.5, the arc γ_{0n} rises at least $q_{2n_0}/4$ units above L_N. Hence, the first E_{0n} forward edges of Γ'_n rise at least $q_{2n_0}/4$ above L'_n. Hence, the first E_{0n} forward edges of Γ_∞ rise at least $q_{2n_0}/4 - 1$ above L_∞. The backwards direction is similar. ♠

Lemma 3.4 *Both directions of* $\Gamma_\infty(\sigma)$ *come arbitrarily close to* L_∞.

Proof: M_0 maps L_∞ to a single point, namely

$$M_0(L_\infty) = -\sum_{n=1}^{\infty} \epsilon_j M_0(v_{2n}). \tag{25}$$

By the last statement of the Translation Lemma, the arc $\gamma'_{n_0} \subset \Gamma_\infty$ contains a vertex of the form

$$\mu_n = \omega(\sigma'_k(n)) - \omega(\sigma(n)) = -\sum_{i=1}^{n_0} \epsilon_n v_{2n} + \sum_{n=n_0+1}^{\infty} \epsilon'_n v_{2n}. \tag{26}$$

Here $\{\epsilon'_n\}$ is a binary sequence whose composition is irrelevant. We therefore have

$$|M_0(L_\infty) - M_0(v_n)| < \sum_{n=n_0+1}^{\infty} M(v_{2n}) < 2 \times 10^{-n_0}.$$

This last equation does the job for us. ♠

3.4 Recognizing the Limit

Referring to our sequence $\sigma = \{\epsilon_i\}$, define

$$\xi = \bigg(\sum_{n=1}^{\infty} \epsilon_j M_0(v_{2n}), -1\bigg). \tag{27}$$

Our rapid decay rates imply that this limit exists and that the map $\sigma \to \xi(\sigma)$ is injective. Hence, our construction produces uncountably many choices of ξ. Throwing out a countable set, we choose so that the first coordinate does not lie in $2\boldsymbol{Z}[A]$. In this case, the outer billiards orbit of ξ is well defined relative to $K(A)$.

Lemma 3.5 *Γ_∞ is the arithmetic graph of ξ and L_∞ is the baseline of Γ_∞.*

Proof: Recall that $N = 2n+1$. Let M_N be the fundamental map associated to Γ_N. That is,

$$M_N(x,y) = 2A_N x + 2y + \frac{1}{q_N}. \tag{28}$$

For any lattice point (x,y), the first coordinate of $M_N(x,y)$ converges to $M_0(x,y)$ as $n \to \infty$. Let $\xi_n = M_N(\omega_n)$. By construction, $\xi_n \to \xi$. As we already remarked, Γ'_n is the arithmetic graph for ξ_n, relative to A_N.

Let $O_2(\xi_n; N)$ denote the outer billiards orbit of ξ_n relative to $K(A_N)$. Since $\xi_n \to \xi$ and $A_N \to A$ and $O_2(\xi)$ is defined, the graph Γ'_n describes $O_2(\xi)$ on increasingly large balls. Taking the limit, we see that Γ_∞ is the arithmetic graph corresponding to $O_2(\xi)$. We have $M_0(L_\infty) = -\xi$ by Equation 25. Let M be the fundamental map such that $M(0,0) = \xi$. The first coordinate of M differs from M_0 by the addition of ξ. Hence, the first coordinate of M maps L_∞ to 0. Hence L_∞ is the baseline for Γ_∞. ♠

The last two results in the previous section show that $(\xi, 1)$ lies in the special erratic set. Given that our construction works for any binary sequence having infinitely many 0s and 1s, we see that the special erratic set essentially contains a Cantor set. This proves Theorem 1.1 for $A \in \Delta(2)$.

Terminology: Since $\xi \notin 2\boldsymbol{Z}[A]$, the entire arithmetic graph $\widehat{\Gamma}_\infty$ is defined. We call this graph a *limit graph*. The component of $\widehat{\Gamma}_\infty$ through $(0,0)$, namely Γ_∞, is the one we analyzed in this chapter.

4 The Hexagrid Theorem

We already mentioned that Lemma 2.5 is a consequence of a general structural result, the Hexagrid Theorem. In this chapter, we explain the Hexagrid Theorem and its consequences.

4.1 The Arithmetic Kite

In this section we describe a certain quadrilateral, which we call the *arithmetic kite*. 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.

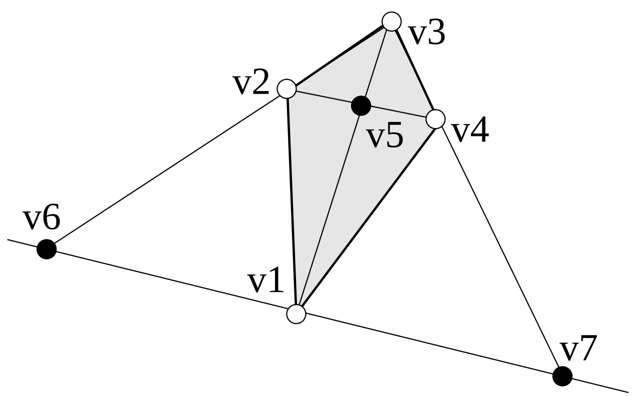

Figure 4.1: The arithmetic kite

Let $A = p/q$. Figure 4.1 shows a schematic picture of $Q(A)$. The vertices are given by the equations.

1. $v_1 = (0, 0)$.
2. $v_2 = \frac{1}{2}(0, p+q)$.
3. $v_3 = \frac{1}{2q}(2pq, (p+q)^2 - 2p^2)$.
4. $v_4 = \frac{1}{2(p+q)}(4pq, (p+q)^2 - 4p^2)$.
5. $v_5 = \frac{1}{2(p+q)}(2pq, (p+q)^2 - 2p^2)$.
6. $-v_6 = v_7 = (q, -p)$.

Referring to the parallelogram $R(p/q)$ from Equation 12, we have

$$V = v_7; \qquad W = v_5. \tag{29}$$

A short calculation, which we omit, shows that $K(A)$ and $Q(A)$ are actually affinely equivalent. $Q(A)$ does not have Euclidean bilateral symmetry, but it does have affine bilateral symmetry.

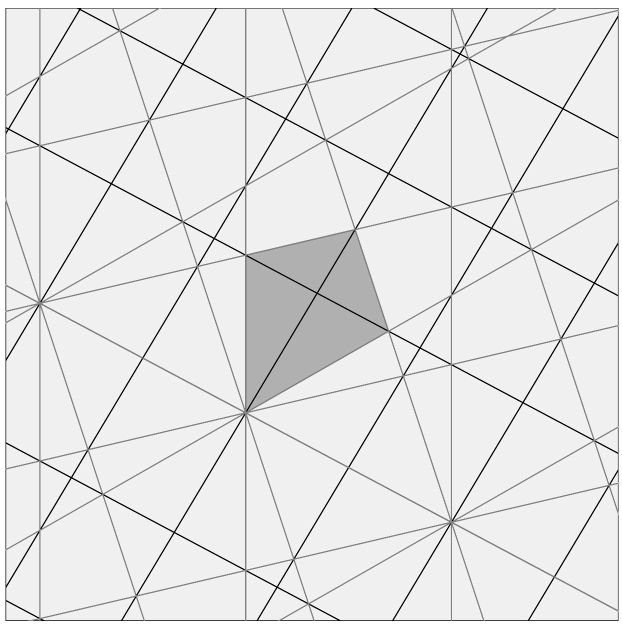

Figure 4.2: $G(25/47)$. and $Q(25/47)$.

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 $Q(A)$ and then taking the orbit under the lattice $\boldsymbol{Z}[V/2, W]$. These are the black lines in Figure 4.2. In case A is an even rational, we would make the same definition, but use the lattice $\boldsymbol{Z}[V, 2W]$ 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 $Q(A)$ and then taking their orbit under the one dimensional lattice $\boldsymbol{Z}[V]$. These are the grey lines in Figure 4.2.

4.2 The Hexagrid Theorem

The Hexagrid Theorem relates two kinds of objects, *wall crossings* and *doors*. Informally, the Hexagrid Theorem says that the arithmetic graph only crosses a wall at a door. Here are formal definitions.

Rooms and Walls: $RG(A)$ divides $\boldsymbol{R}^2$ 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 make the same definition, except that we exclude crossing 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 17.6.) 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 of Γ incident 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.

In Part III of the monograph we will prove the following result. Let $\underline{y}$ denote the greatest integer less than y.

Theorem 4.1 (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 $(m, \underline{y}) \in \boldsymbol{Z}^2$. 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 $(x, y - 1)$.*

Remark: We really only care about the Hexagrid Theorem when A is an odd rational. We will not discuss the even case until §20.

Figure 4.3 illustrates the Hexagrid Theorem for $p/q = 25/47$. The shaded parallelogram is $R(25/47)$, the parallelogram from Lemma 2.5. We have only drawn the unstable components in Figure 4.3. The reader can see much better pictures of the Hexagrid Theorem using either Billiard King or our interactive guide to the monograph. (The interactive guide only shows the odd case, but Billiard King also shows the even case.)

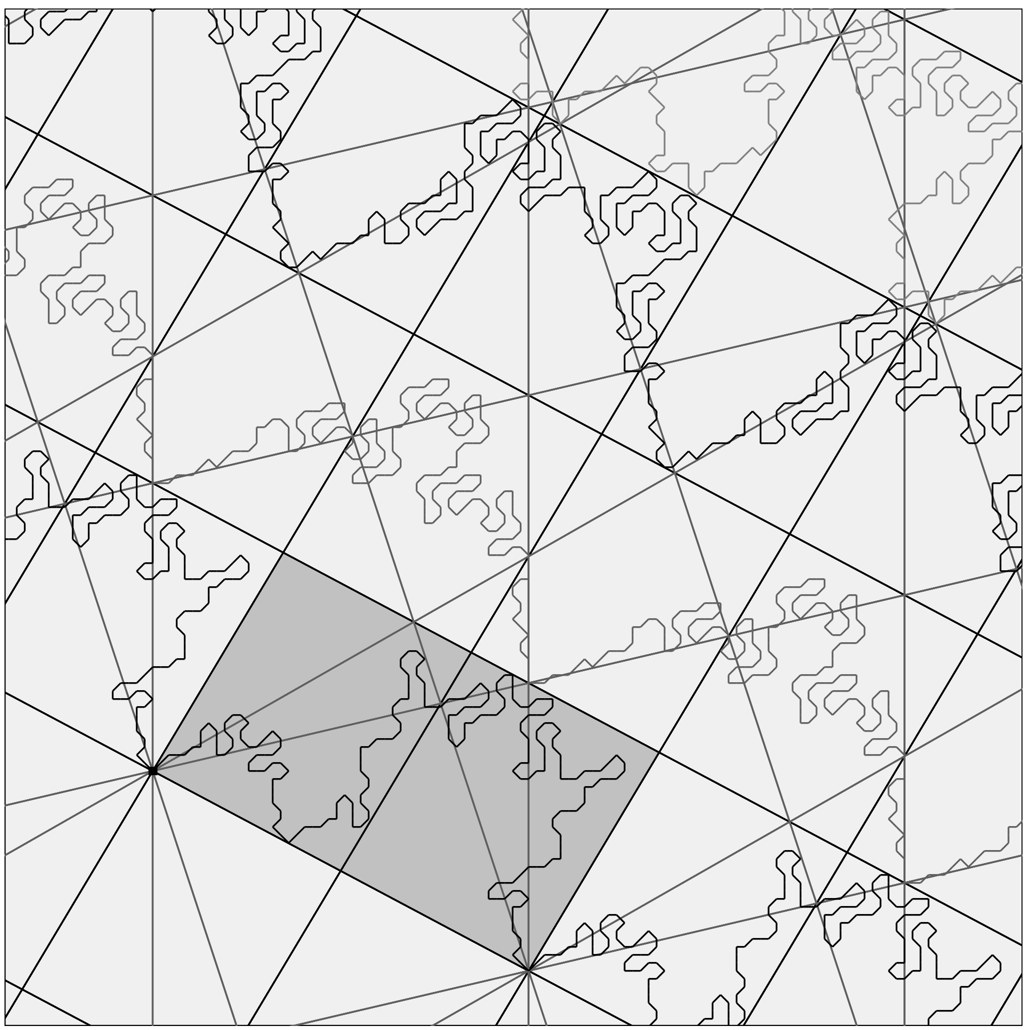

Figure 4.3: $G(25/47)$, $R(25/47)$, and some of $\widehat{\Gamma}(25/47)$.

4.3 Proof of Lemma 2.5

First of all, for any value of A, it is easy to check that $\Gamma(A)$ contains the arc $(-1,1)\to(0,0)\to(1,1)$. 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 13. From the Hexagrid 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 picture repeats endlessly to the left and the right of $R(p/q)$. Hence $\Gamma(p/q)$ is an open polygonal curve.

4.4 Proof of Theorem 1.5

First suppose that p/q is odd. Let M_1 be the first coordinate of the fundamental map associated to p/q. 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 $\boldsymbol{Z}^2$, 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

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

Modulo translation by V, the point ζ_k is the only lattice point on L_k. Statement 1 of the Hexagrid Theorem contains that statement 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

$$M_1(\zeta_k)=k(p+q)+\frac{1}{q}.$$

For all lattice points (m,n) between L_k and L_{k+1} we therefore have

$$M_1(m,n)\in I_k, \tag{30}$$

the interval from Theorem 1.5. Theorem 1.5 now follows from Equation 30, Statement 1 of the Hexagrid Theorem, and our remarks about ζ_k.

The proof of Theorem 1.5 in the even case is exactly the same, except that we get a factor of 2 due to the different definition of the room grid.

Remark: We compare Theorem 1.5 to a result in [**K**]. The result in [**K**] is quite general, and so we will specialize it to kites. In this case, a kite is quasi-rational 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 $J_0, J_1, J_2, \ldots$, where

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

The special floors corresponding to 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. Our result is a refinement in a special case.

4.5 Proof of Theorem 1.6, Odd Case

Now we assume that $A = p/q$ is 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 lies in the kth suite. With this notation $\Gamma_0(p/q) = \Gamma(p/q)$ is the component that contains $(0,0)$.

We have just described the infinite family of unstable components listed in Theorem 1.6. All the other components of $\widehat{\Gamma}(p/q)$ are closed polygons and must be confined to single rooms. The point here is that the infinite polygonal arcs we have already described have used up all the doors. Nothing else can cross any of the walls.

Each vertex (m,n) in the arithmetic graph corresponds to the two points $(M_1(m,n), \pm 1)$. Thus, each component of $\widehat{\Gamma}$ tracks either 1 or 2 orbits. By the parity result in Equation 8, these two points lie on different ψ-orbits. Therefore, each component of $\widehat{\Gamma}$ tracks two special orbits. In particular, there are exactly two unstable orbits U_k^+ and U_k^- contained in the interval I_k, and these correspond to $\Gamma_k(p/q)$.

See §5.5 for the proof in the even case.

5 Additional Structure of the Graph

5.1 Translational Symmetry

See §17.1 for a proof of the claims in this section. Let p/q be an odd rational. We have already seen that the arithmetic graph $\widehat{\Gamma}(p/q)$ is invariant under translation by $V = (q, -p)$.

In Part II we will prove our general structural result, the Master Picture Theorem. We will see, as a consequence of the Master Picture theorem, that $\widehat{\Gamma}(p/q)$ has a canonical extension to all of $\boldsymbol{Z}^2$, and that this extension does not cross the baseline. Moreover, the extension turns out to be invariant under the lattice

$$\Theta = \boldsymbol{Z}V + \boldsymbol{Z}V'. \tag{31}$$

Here

$$V' = \left(0, \frac{(p+q)^2}{4}\right) \tag{32}$$

A direct calculation, which we make in §17, shows that the hexagrid $G(p/q)$ is also invariant under Θ. The vector V' is a kind of hidden symmetry of the picture. Some version of this fact is implicit in the work of [**K**], though he doesn't consider anything like the arithmetic graph.

Periodicity Proof: Statement I of the Hexagrid Theorem implies that all special orbits in the odd rational case are bounded, and hence periodic. Here we will give another proof in the rational case that all special orbits are bounded and hence periodic. The proof we give here only depends on the Embedding Theorem and the translational symmetry of the arithmetic graph. These are softer results than the Hexagrid Theorem. A similar proof works in the even case, but we omit the details.

The quotient $\boldsymbol{Z}^2/\Theta$ is a flat torus containing finitely many canonical integer points. By the Embedding Theorem and symmetry, the quotient $\widehat{\Gamma}/\Theta$ is an embedded graph whose vertices lie at the integer points. Hence, every component of $\widehat{\Gamma}/\Theta$ is a closed embedded polygon. Let P be one of these components. Let $\widetilde{P}$ be the corresponding periodic component of $\widehat{\Gamma}$.

If $\widetilde{P}$ is a closed polygon, then we are done. If $\widetilde{P}$ is invariant under translation by $aV + bW$, with $b \neq 0$, then $\widetilde{P}$ crosses the baseline and we have a contradiction. Hence $b = 0$. But then $\widetilde{P}$ lies within a uniformly bounded distance of the baseline, since translation by V preserves the baseline.

5.2 Rotational Symmetry

From now on we work with the extended version of $\widetilde{\Gamma}(p/q)$. There is a unique rational p_+/q_+ such that $p/q < p_+/q_+$ and $q_+ \in (0,q)$, and $qp_+ - pq_+ = 1$. We discuss this rational in detail in the next chapter. In §17.2, we will prove that rotation by 180 degrees about any of the points

$$\beta + \theta; \qquad \beta = \frac{1}{2}(q_+, -p_+); \qquad \theta \in \Theta \tag{33}$$

is a symmetry of $\widehat{\Gamma}$. These points are half-integers, and rotation about them preserves $\boldsymbol{Z}^2$. The point β is very close to the baseline of $\widehat{\Gamma}$.

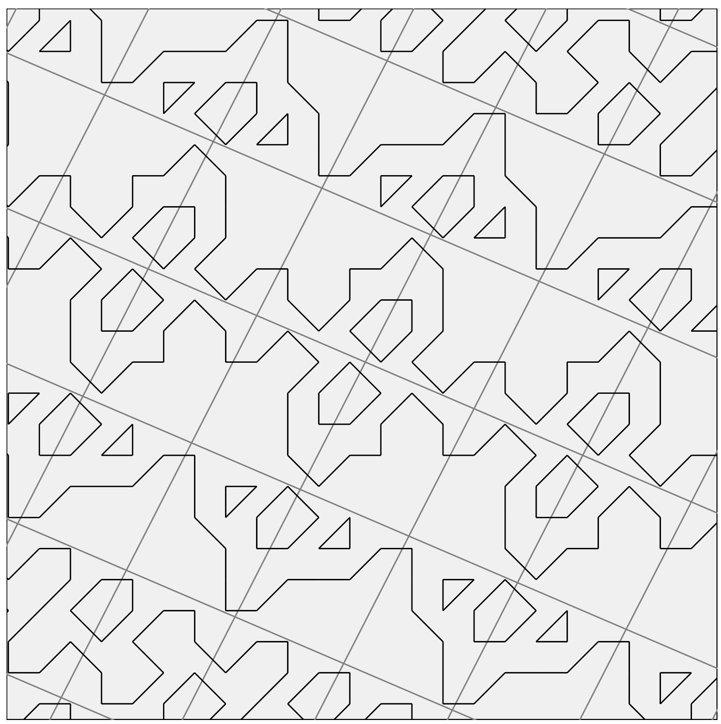

Figure 5.1: $\widehat{\Gamma}(3/7)$ centered on a point of symmetry.

Figure 5.1 shows the picture for the parameter 3/7. The grid in the picture is a translated version of the room grid $RG(3/7)$, arranged so that some of the points where the grid lines cross are points of rotational symmetry of the arithmetic graph. In the picture, we can also see a near-bilateral symmetry. We will explain this in §5.4.

5.3 The Decomposition Theorem

Let $G(p/q)$ be the hexagrid associated to p/q. We will see when we prove the Hexagrid Theorem that the result holds for the extended version of $\widehat{\Gamma}(p/q)$. Here we distingish two lines of $G(p/q)$. Let λ_0 and λ_1 denote the lines extending the left and right edges of $R(p/q)$.

Let ι denote rotation by 180 degrees about the point β mentioned in the last section. We have already mentioned that ι preserves $\widehat{\Gamma}(p/q)$. However, ι does not preserve $G(p/q)$. One problem is that β does not quite lie on a grid line, but this is really a minor problem. Actually, $G(p/q)$ and $\iota(G(p/q))$ are not even close. We have

$$\lambda_0, \lambda_1 \subset G(p/q); \qquad \iota(\lambda_0) \subset \iota(G(p/q)) - G(p/q). \tag{34}$$

The line $\iota(\lambda_0)$ lands somewhere in the middle of $R(p/q)$ and divides it into pieces. We take the smaller of the two pieces and shrink it further, so that its ceiling is half as high as formerly. The 3 lines of positive slope in Figure 5.2 are λ_0, $\iota(\lambda_0)$, and λ_1, for $p/q = 31/59$. Here $p_+/q_+ = 10/19$.

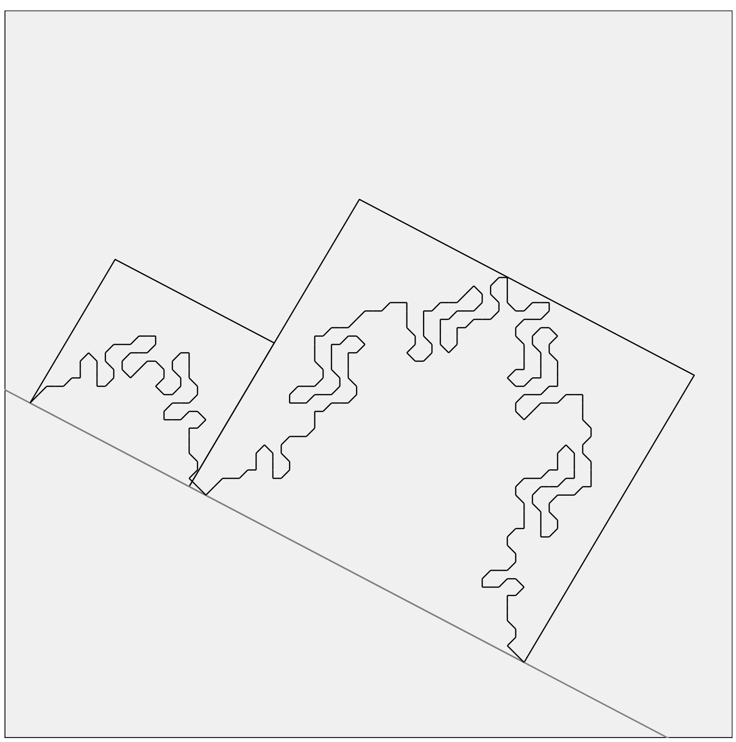

Figure 5.2: Father and Son decomposition of $\Gamma(31/59)$.

The large parallelogram to the right of $\iota(\lambda_0)$ shares the top edge with the top edge of $R(31/59)$. The small rectangle on the left of $\iota(\lambda_0)$ is exactly half as tall. Our decomposition theorem, stated below, says that the picture in Figure 5.2 is the general one. We call the small parallelogram the *son's room* and we call the large parallelogram the *father's room*.

Looking at the picture, we can see that the portion of $\Gamma(p/q)$ contained in each of the two parallelograms has near-bilateral symmetry. We will explain this phenomenon in the next section.

How we come to the formal definitions of the Father and Son Decomposition. We first define the regions and then we state the result.

Dividing Line: The *dividing line* is the line $\iota(\lambda_0)$, parallel to the vector V and containing the point $(q_+, -p_+)$. We change notation and denote the dividing line by $DR(p/q)$. Let $q_- = q - q_+$.

Father's Room: If $q_+ > q_-$ then the father's room is the closure of the left component of $R(p/q) - DR(p/q)$. If $q_+ < q_-$ (as in the figure above), then the father's room is the closure of the right component of $R(p/q) - DR(p/q)$. In either case, we denote this parallelogram by $FR(p/q)$.

Son's Room: If $q_+ > q_-$ then $SR(p/q)$ is the bottom half of the right component of $R(p/q) - DR(p/q)$. If $q_+ < q_-$ (as in the figure above) then $SR(p/q)$ is the bottom half of the left component of $R(p/q) - DR(p/q)$. In both case, the line extending the top of $SR(p/q)$ is exactly halfway between the floor and ceiling of $R(p/q)$.

Let Γ^1 denote the connected arc of Γ that has endpoints $(0,0)$ and $(q,-p)$. In Part IV we will prove the following result.

Theorem 5.1 (Decomposition) $\Gamma^1(p/q) \subset SR(p/q) \cup FR(p/q)$.

The Decomposition Theorem is a vital tool in the proof Theorems 1.1 and 1.3 in general. The containment in the decomposition is more efficient than the containment $\Gamma^1(p/q) \subset R(p/q)$ established in Lemma 2.5. The better control on $\Gamma^1(p/q)$ allows us to prove a stronger version of Lemma 2.6. See Corollary 7.5.

5.4 Near Bilateral Symmetry

A glance at all our pictures of arithmetic graphs suggests that these graphs have an approximate bilateral symmetry. Figure 5.3 shows a different view of this symmetry.

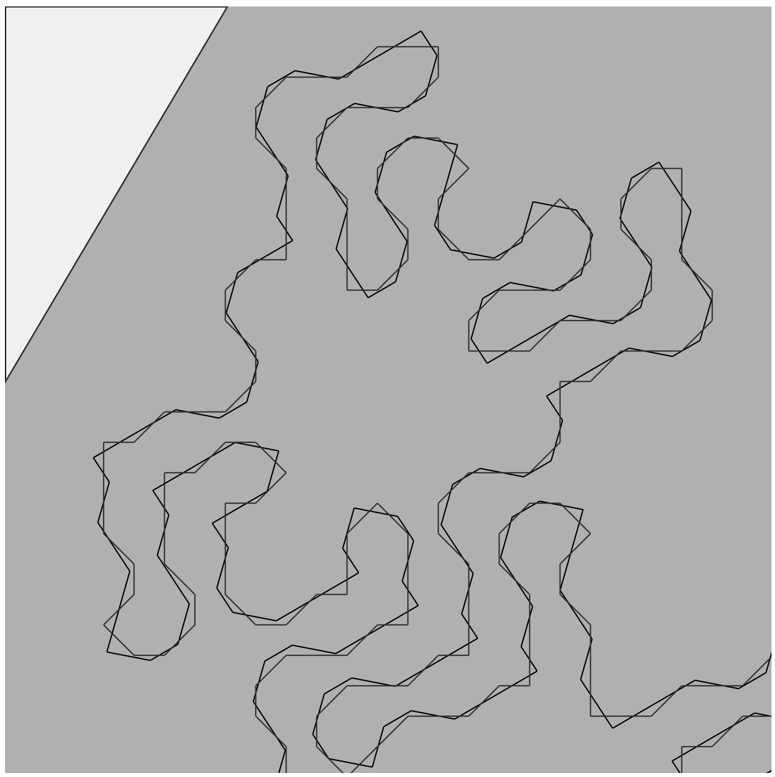

Figure 5.3: Superimposed arithmetic graphs.

There is a unique affine involution I such that $I(FR_1) = FR_1$. The direction fixed by I is parallel to the long diagonal of the arithmetic kite. Figure 5.3 shows a closeup of the superposition

$$I(\Gamma') \cup \Gamma'; \qquad \Gamma' = \Gamma(73/139) \cap FR(73/139).$$

We can see that the symmetry is fundamentally approximate in nature, but somehow quite close. The symmetry must be approximate, because the involution I does not respect the lattice $\boldsymbol{Z}^2$. In this section, we will give a rough explanation of this symmetry. We could make the discussion more precise with some effort, but we don't want to make the effort. We shall not use the near bilateral symmetry for any purpose, but it seems worthwhile to let the curious reader know why it exists. It actually took us about a year to figure out the explanation.

Let $M(m,n) = \xi \in \boldsymbol{R}_+$ denote the point corresponding to the lattice point (m,n). Here M is the fundamental map associated to the parameter p/q. The forwards ψ orbit of $(m,n,1)$ winds halfway around the kite and then encounters a point of the form

$$(2Am' + 2n', \pm 1) \in R_- \times \{-1,1\}.$$

The lattice point (m,n) lies above the baseline of $\Gamma(p/q)$ and the lattice point (m',n') lies below the baseline. Given the geometry of the return map Ψ, as discussed in detail in §11, the points (m,n) and (m',n') lie about the same distance from the baseline. Indeed, a calculation shows that there is a uniformly small constant C such that

$$\|(m',n') - J(m,n)\| < C,$$

where J is the affine involution fixing the baseline and mapping the lines extending $R(p/q)$ to themselves. We could probably take $C = 2$.

We define $J^*(m,n) = (m',n')$. We might have based J^* on the orbit of $(2Am+2n,-1)$, but the result would be uniformly close. In any case, J^* maps components of $\widehat{\Gamma}$ above the baseline to components of $\widehat{\Gamma}$ below the baseline, in such a way that the corresponding components are uniformly close to having the same affine shape. The map J^* does not quite induce a combinatorial isomorphism on each component. J^* would map two vertices (m_1,n_1) and (m_2,n_2) to the same point if

$$(m_1,n_1,1) \to ... \to (m',n',1); \qquad (m_2,n_2,1) \to ... \to (m;,n',-1)$$

under iteration of the map ψ.

Letting ι denote the 180 degree rotation about the point β from Equation 33, the map $\beta \circ J^*$ maps $\Gamma(p/q)$ to itself and is uniformly close to the affine involution I discussed in connection with Figure 5.3. Indeed, up to a tiny translational discrepancy, we have $I = \beta \circ J$.

In terms of Figure 5.1, the maps β and J generate the dihedral symmetry group of the grid, fixing the center point. The maps β and J^* generate an order 4 group of permutations of the components of $\widehat{\Gamma}(p/q)$.

5.5 Stability

Let $O_2(m,n)$ denote the orbit of the point $p_0 = M_\alpha(m,n)$. Let $\widehat{\Gamma}(m,n)$ denote the component of $\widehat{\Gamma}$ that contains (m,n).

Lemma 5.2 *A periodic orbit $O_2(m,n)$ is stable iff $\widehat{\Gamma}(m,n)$ is a polygon.*

Proof: Let K be the period of Ψ on p_0. Tracing out $\widehat{\Gamma}(m,n)$, we get integers (m_k,n_k) such that

$$\Psi^k(p_0) - p_0 = (2m_kA + 2n_k, 2\epsilon_k); \qquad k = 1,...,K. \tag{35}$$

Here $\epsilon_k \in \{0,1\}$, and $\epsilon_k = 0$ iff $m_k + n_k$ is even. The integers (m_k,n_k) are determined by the combinatorics of a finite portion of the orbit. Hence, Equation 35 holds true for all nearby parameters A.

If $\widehat{\Gamma}(m,n)$ is a closed polygon, then $(m_K,n_K) = 0$. But then $\Psi^k(p_0) = p_0$ for all parameters near A. If $O_2(m,n)$ is stable then $(m_K,n_K) = (0,0)$. Otherwise, the equation $m_KA + n_K = 0$ would force $A = -n_K/m_K$. ♠

Proof of Theorem 1.6, Even Case: We want to show that all special orbits are stable. By Lemma 5.2, it suffices to show that all components of $\widehat{\Gamma}(p/q)$ are polygons. Suppose that $\widehat{\Gamma}(m,n)$ is not a polygon. Let R denote reflection in the x-axis. We have

$$R\Psi R^{-1} = \Psi^{-1}; \qquad R(M(m,n)) = M(m+q, n-p). \tag{36}$$

From this equation we see that translation by $(q,-p)$ preserves $\widehat{\Gamma}$ but reverses the orientation of all components. But then $(m,n) + (q,-p) \not\in \widehat{\Gamma}(m,n)$.

Since all orbits are periodic, $(m,n) + k(p,-q) \in \widehat{\Gamma}(m,n)$ for some integer $k \geq 2$. Let γ be the arc of $\widehat{\Gamma}(m,n)$ connecting (m,n) to $(m,n) + k(q,-p)$. By the Embedding Theorem, γ and $\gamma' = \gamma + (q,-p)$ are disjoint. But this situation violates the Jordan Curve Theorem.

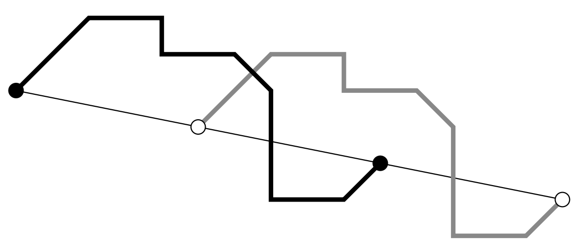

Figure 5.4: γ and $\gamma + (q,-p)$.

5.6 Topological Confinement

The Embedding Theorem tells us that $\widehat{\Gamma}_\alpha(A)$ is an disjoint union of embedded polygons and infinite embedded polygonal arcs. This fact gives us a topological method for producing periodic orbits. There are two methods.

1. Suppose that (m,n) is a lattice point and $\widehat{\Gamma}(m',n')$ is a polygon that winds once around (m,n). Then $\widehat{\Gamma}(m,n)$ is also a polygon.

2. Say that $(m,n) \in \boldsymbol{Z}^2$ is *low* if the baseline separates $(m,n-1)$ from (m,n). If $\widehat{\Gamma}(m',n')$ is a component with two low vertices, then $\widehat{\Gamma}(m,n)$ is a closed polygon for any point (m,n) that lies beneath $\widehat{\Gamma}(m',n')$, above the baseline, and between the two low vertices.

Figure 5.5 illustrates each of the two methods of confinement. The big polygon is $\Gamma(18/41)$. This polygon confines each of two smaller grey polygons in one of the ways just mentioned.

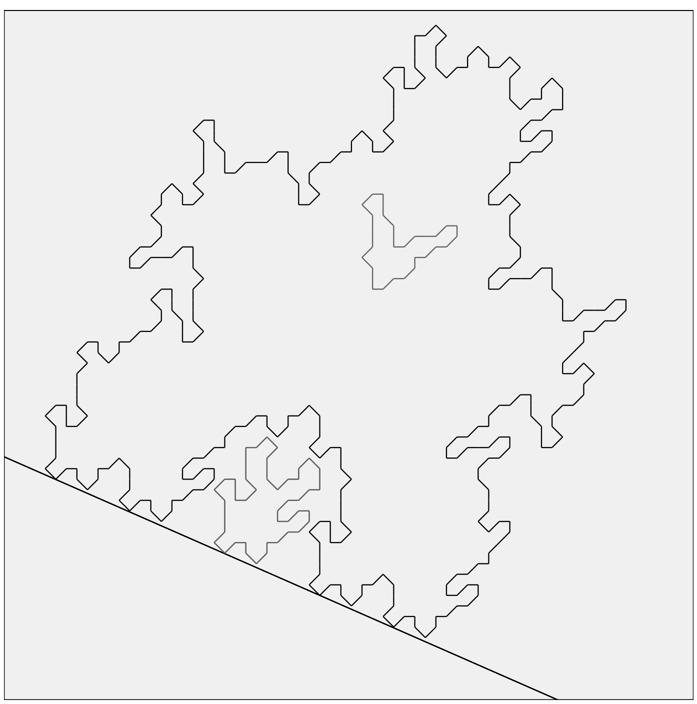

Figure 5.5: $\Gamma(18/41)$ confines two small polygons

5.7 Distinguishing Orbits

We use the next result in our proof of Theorem 1.3. Our proof ignores the case $p = 1$ because the proof is a bit different and we don't need this case.

Lemma 5.3 *Let p/q be any odd rational. Then $O_2(3/q, 1) \neq O_2(1/q, \pm 1)$.*

Proof: We assume that $p > 1$. Suppose $O_2(3/q, 1) = O_2(1/q, 1)$. Then $\Psi^k(3/q, 1) = (1/q, 1)$ for some k. But then

$$(2/q, 0) = (3/q, 1) - \Psi^k(3/q, 1) = (2m(p/q) + 2n, 0); \qquad m + n \text{ even}$$

The parity result comes from Equation 8. Equating the first coordinates and clearing denominators, we get $1 = mp + nq$. This is impossible for p/q odd.

Suppose $O_2(3/q, 1) = O_2(1/q, -1)$. Let's say that $(3/q, 1)$ is in the forwards orbit of $(1/q, -1)$. The backwards case has a similar treatment. Let $A = p/q$. A calculation (when $p > 1$) shows that

$$\Psi(s/q, \pm 1) - (s/q, \pm 1) = (1 \mp A, 0); \qquad s = 1, 3. \tag{37}$$

Let $(q', -p')$ be such that $q' \in (0, q)$ and $M(q', -p') = (3/q, 1)$. In fact p'/q' is the largest continued fraction approximation to p/q that is less than p/q. Equations 37 and the first half of Equation 36 together imply that $\Gamma(p/q)$ first traces out the arc $(-1, 1) \to (0, 0) \to (1, 1)$ and then the arc $(q', -p') + (1, 1) \to (q', -p') \to (q', -p') + (-1, 1)$.

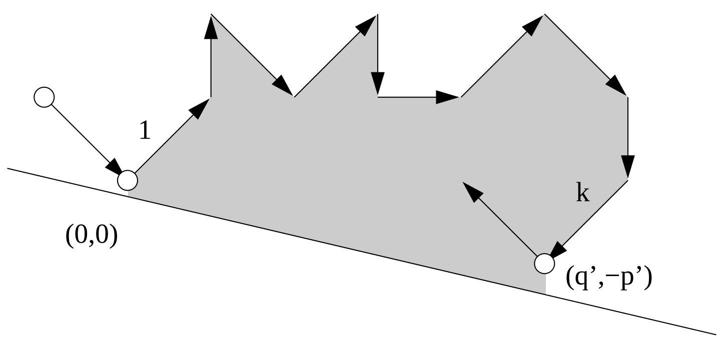

Figure 5.6: $\Gamma(p/q)$ confines itself.

But then $\Gamma(p/q)$ confines itself, as shown in Figure 5.6. The curve cannot escape from the shaded area because it is embedded, and the two vertices at either end are low vertices. This is a contradiction. ♠

6 Odd Rational Approximation

6.1 A Canonical Sequence

For the general parameter $A \in (0,1)$, we do not have $A \in \Delta(2)$. In this chapter, we will produce a canonical sequence $\{p_n/q_n\}$ of odd rationals approximating $A \in (0,1)$. This sequence will not satisfy Equation 16, but it will have sufficiently good approximation properties to let us push through the proof of Theorem 1.1 in general.

Given any p/q, there are unique rationals p_+/q_+ and p_-/q_- such that

$$q_+, q_- \in (0,q); \qquad \frac{p_-}{q_-} < \frac{p}{q} < \frac{p_+}{q_+}; \qquad |pq_\pm - qp_\pm| = 1. \tag{38}$$

We have the general relation

$$\frac{p}{q} = \frac{p_+ + p_-}{q_+ + q_-}. \tag{39}$$

Our canonical sequence $\{p_n/q_n\}$ is characterized by the following equation.

$$\frac{p_{n-1}}{q_{n-1}} = \frac{|(p_n)_- - (p_n)_+|}{|(q_n)_- - (q_n)_+|} \tag{40}$$

More precisely,

$$\frac{p_{n-1}}{q_{n-1}} = \frac{(p_n)_- - (p_n)_+}{(q_n)_- - (q_n)_+} \iff \frac{p_{n-1}}{q_{n-1}} < \frac{p_n}{q_n} \iff (q_n)_+ < (q_n)_-$$

$$\frac{p_{n-1}}{q_{n-1}} = \frac{(p_n)_+ - (p_n)_-}{(q_n)_+ - (q_n)_-} \iff \frac{p_{n-1}}{q_{n-1}} > \frac{p_n}{q_n} \iff (q_n)_+ > (q_n)_- \tag{41}$$

The first rational is odd iff the second one is. When A is an odd rational, the finite sequence approximating A is inductively determined by Equation 41. In general, we need to take a limit. Similar to what one does for ordinary continued rationals, we will give a hyperbolic geometry interpretation of our sequence that makes the convergence clear.

Our sequence is similar to the sequence of continued fraction approximants to A, but somewhat different. The terms of our approximation are not necessarily continued fraction approximants to A. For instance, if $A = [0; 4, 4, 4, ...]$, then the continued fraction approximants to A are all even. See [**Da**] for the theory of continued fractions.

6.2 Constructing the Sequence

The hyperbolic plane is the upper halfplane $\boldsymbol{H}^2 \subset \boldsymbol{C}$. The group $SL_2(\boldsymbol{R})$ of real 2×2 matrices acts isometrically by linear transformations. The geodesics are vertical rays or semicircles centered on $\boldsymbol{R}$. See [**B**] for details.

The *Farey graph* is a tiling of $\boldsymbol{H}^2$ 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 probably the most famous picture in hyperbolic geometry.

We modify the Farey graph by erasing all the lines that connect even fractions to each other. The remaining edges partition $\boldsymbol{H}^2$ into an infinite union of ideal squares. 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 Equation 38 when ordered so that $q_1 < q_2$. We call p_1/q_1 the *tail* of the square and p_2/q_2 the *head* of the square. We draw an arrow in each odd square that points from p_1/q_1 to 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.

The general form of a left biased square is

$$\frac{a_1}{b_1}; \qquad \frac{a_1+a_2}{b_1+b_2}; \qquad \frac{a_1+2a_2}{b_1+2b_2}; \qquad \frac{a_2}{b_2}. \tag{42}$$

The leftmost vertex in a left-biased square is the tail, and the rightmost vertex in a right-biased square is the tail. One gets the equation for a right-biased square just by reversing Equation 42.

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 approximating sequence, we list the tails of the encountered squares and weed out repeaters.

We let κ_n be the largest integer such that

$$\left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| < \frac{2}{\kappa_n q_n^2}. \tag{43}$$

We call κ_n the *diophantine constant* for (A_n, A_{n+1}). The sequence $\{\kappa_n\}$ is akin to the continued fraction expansion of A and the sequence $\{p_n/q_n\}$ is the desired sequence of odd rational approximants.

6.3 Local Structure of the Sequence

Now suppose that $\{A_n = p_n/q_n\}$ is the canonical odd sequence approximating some irrational $A \in (0,1)$. Let $(A_n)_\pm = (p_n)_\pm/(q_n)_\pm$.

Lemma 6.1 *Either $A_n < A < (A_n)_+$ or $(A_n)_- < A < A_n$.*

Proof: We will consider the case when $A_n < A_{n-1}$. The other case is similar. Let γ be the vertical geodesic to A. At some point γ encounters the basic square with vertices

$$(A_n)_- < A_n < (A_n)_+ < A_{n-1}.$$

If $A_{n+1} < A_n$, then γ exits S between $(A_n)_-$ and A_n. Hence $(A_n)_- < A < A_n$. If $A_{n+1} > A_n$, then γ exits S to the right of A_n. If γ exits S to the right of $(A_n)_+$, then γ next encounters a basic square S' with vertices

$$(A_n)_+ < O < E < A_{n-1},$$

where O and E are odd and even rationals. But then A_n would not be the term in our sequence after A_{n-1}. The term after A_{n-1} would lie in the interval $[O, A_{n-1})$. This is a contradiction. Hence γ exits S between A_n and $(A_n)_+$. ♠

Lemma 6.2 *The following is true for any n.*

1. *$A_{n-1} < A_n < A_{n+1}$ then $(q_n)_+ < (q_n)_-$ and $\kappa_n \equiv 1$ mod 2 and*

$$2(q_{n+1})_+ = (\kappa_n + 1)(q_n)_+ + (\kappa_n - 1)(q_n)_-.$$

2. *If $A_{n-1} > A_n < A_{n+1}$ then $(q_n)_+ > (q_n)_-$ and $\kappa_n \equiv 0$ mod 2 and*

$$2(q_{n+1})_+ = (\kappa_n + 0)(q_n)_- + (\kappa_n - 2)(q_n)_+.$$

3. *If $A_{n-1} > A_n > A_{n+1}$ then $(q_n)_+ > (q_n)_-$ then $\kappa_n \equiv 1$ mod 2 and*

$$2(q_{n+1})_- = (\kappa_n + 1)(q_n)_- + (\kappa_n - 1)(q_n)_+.$$

4. *If $A_{n-1} < A_n > A_{n+1}$ then $(q_n)_+ < (q_n)_-$ then $\kappa_n \equiv 0$ mod 2 and*

$$2(q_{n+1})_- = (\kappa_n + 0)(q_n)_+ + (\kappa_n - 2)(q_n)_-.$$

Proof: The first implication, in all cases, is contained in Equation 41. For the other implications, Cases 3 and 4 follow from Cases 1 and 2 by symmetry. For ease of exposition, we will just treat Case 1.

The vertical geodesic γ to A passes through the basic square S with vertices
$$\frac{p_{n-1}}{q_{n-1}} < \frac{(p_n)_-}{(q_n)_-} < \frac{p_n}{q_n} < \frac{(p_n)_+}{(q_n)_+}.$$
Since $A_{n+1} > A_n$, the geodesic γ next crosses through the geodesic α_n connecting p_n/q_n to $(p_n)_+/(q_n)_+$. Following this, γ encounters the basic squares S'_k for $k = 0, 1, 2...$ until it crosses a geodesic that does not have p_n/q_n as a left endpoint. By Equation 42 and induction, we get

$$S'_k: \ \frac{p_n}{q_n} < \frac{(k+1)p_n + (p_n)_+}{(k+1)q_n + (q_n)_+} < \frac{(2k+1)p_n + 2(p_n)_+}{(2k+1)q_n + 2(q_n)_+} < \frac{kp_n + (p_n)_+}{kq_n + (q_n)_+}. \quad (44)$$

Here S'_k is a left-biased square. But then there is some k such that

$$\frac{p_{n+1}}{q_{n+1}} = \frac{(2k+1)p_n + 2(p_n)_+}{(2k+1)q_n + 2(q_n)_+}; \qquad \frac{(p_{n+1})_+}{(q_{n+1})_+} = \frac{kp_n + (p_n)_+}{kq_n + (q_n)_+} \quad (45)$$

Since $(q_n)_+ < (q_n)_-$, we have $2(q_n)_+ < q_n$. Since $2(q_n)_+ < q_n$, we have

$$\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} = \frac{2}{(2k+1)q_n^2 + 2q_n(q_n)_+} \in \left(\frac{2}{(2k+2)q_n^2}, \frac{2}{(2k+1)q_n}\right). \quad (46)$$

Hence $\kappa_n = (2k+1) \equiv 1 \bmod 2$. Finally,

$$2(q_{n+1})_+ = 2kq_n + 2(q_n)_+ =$$
$$2k((q_n)_+(q_n)_-) + 2(q_n)_+ =$$
$$(\kappa_n + 1)(q_n)_+ + (\kappa_n - 1)(q_n)_-.$$

This completes the proof. ♠

Remark: When it comes time in Part IV to prove our Decomposition Theorem, we will use an inductive argument that exploits the Copy Theorems from the next chapter and the structure established in the previous lemma.

6.4 Approximation Properties

Lemma 6.3 *$\kappa_n \geq 2$ infinitely often.*

Proof: We can sort the indices of our sequence into 4 types, depending on which case holds in Lemma 6.2. We distinguish between two kinds of approximating sequences. 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, eventually n has constant type, say Type 1. (The Type 3 case has a similar treatment.) Looking at the formula in Case 1 of Lemma 6.2, we see that the sequence $\{(q_n)_+\}$ is eventually constant. But then

$$r = \lim_{n \to \infty} \frac{(q_n)_+ p_n}{q_n}$$

exists. Since $(q_n)_+ p_n \equiv -1 \bmod q_n$ and $q_n \to \infty$, we must have $r \in \boldsymbol{Z}$. But then $\lim p_n/q_n \in \boldsymbol{Q}$, and we have a contradiction. ♠

Lemma 6.4 *If $\kappa_n \geq 2$ then*

$$\left| A - \frac{p_n}{q_n} \right| < \frac{2}{q_n^2}$$

Proof: Suppose that n has Type 2. Then $2(q_n)_+ > q_n$ and $A_n < A < (A_n)_+$. Hence

$$|A_n - A| < |A_n - (A_n)_+| = \frac{1}{q_n (q_n)_+} < \frac{2}{q_n^2}.$$

The proof when n has Type 4 is similar.

Suppose that n has Type 1. Then $\kappa_n \geq 3$. Combining Lemma 6.1 and Equation 44, we have

$$|A_n - A| < \left| \frac{kp_n + (p_n)_+}{kq_n + (q_n)_+} - \frac{p_n}{q_n} \right| = \frac{1}{q_n (kq_n + (q_n)_+)} < \frac{1}{kq_n^2}.$$

Recall that $\kappa_n = 2k+1$. Hence $k \geq 1$. The proof when n has Type 3 is similar. This completes the proof. ♠

Lemma 6.5 $q_{n+1} - \kappa_n q_n \to \infty$ *as* $n \to \infty$.

Proof: We will consider Cases 1 and 2 of Lemma 6.2. The other cases have similar proofs. Realizing that the fraction on the right hand size of Equation 45 is in lowest terms, we have

$$q_{n+1} = (2k+1)q_n + 2(q_n)_+.$$

In Case 1, we have $\kappa_n = 2k+1$, and we get

$$q_{n+1} - \kappa_n q_n = 2(q_n)_+.$$

Since $(p_n)_+/(q_n)_+ \to A$, we see that the right hand side of the last equation tends to ∞.

In Case 2, we have $\kappa_n = 2k+2$ and $q_n = (q_n)_+ - (q_n)_-$. Therefore

$$q_{n+1} = (\kappa_n - 1)q_n + 2(q_n)_+ = \kappa_n q_n - (q_n - (q_n)_-) + (q_n)_+ = \kappa_n q_n + q_{n-1}.$$

In short,

$$q_{n+1} - \kappa_n q_n = q_{n-1}.$$

Again, the right hand side tends to ∞ with n. ♠

7 Period Copying

7.1 Three Linear Functionals

Let p/q be an odd rational. Let $V = (q, -p)$ be the vector from Equation 11. In terms of the arithmetic kite, $V = v_7$. Our period copying theorems will be stated in terms of the following 3 linear functionals.

$$F(m,n) = (p,q) \cdot (m,n). \tag{47}$$

$$G(m,n) = \left(\frac{q-p}{p+q}, \frac{-2q}{p+q}\right) \cdot (m,n). \tag{48}$$

$$H = \left(\frac{-p^2+4pq+q^2}{(p+q)^2}, \frac{2q(q-p)}{(p+q)^2}\right) \cdot (m,n). \tag{49}$$

The linear functional F is easy to describe. Its fibers are parallel to the baseline of the arithmetic graph $\Gamma(p/q)$. Put another way, the fibers of F are parallel to the short diagonal of $Q(p/q)$, the arithmetic kite. The other two linear functionals are also adapted to the arithmetic kite.

Lemma 7.1 $G(V) = q$. *Moreover, the fibers of G are parallel to the top left edge of the arithmetic kite.*

Proof: The first statement is a computation. Referring to the arithmetic kite in Figure 4.1, we compute that $G(v_6) = G(v_3)$. Thus, G takes on the same values on the top left edge of $Q(p/q)$. ♠

Lemma 7.2 $H(V) = q$. *Moreover, the fibers of H are parallel to the top right edge of the arithmetic kite.*

Proof: The first statement is a computation. Referring to the arithmetic kite in Figure 4.1, we compute that $H(v_7) = H(v_3)$. Thus, H takes on the same values on the top left right of $Q(p/q)$. ♠

Provided that $r < s$, the set of points $(m,n) \in \boldsymbol{R}^2$ such that $F(m,n) \geq 0$ and $G(m,n) \geq r$ and $H(m,n) \leq s$ forms a triangle.

7.2 The Main Result

Now suppose that p_1/q_1 and p_2/q_2 are two odd rationals. Let

$$\lambda_1 = \frac{(q_1)_+}{q_1}; \qquad p_1(q_1)_+ \equiv -1 \bmod q; \qquad (q_1)_+ \in (0, q_1). \tag{50}$$

In §6, we have already defined $(q_1)_+$ as the denominator of $(p_q)_+/(q_1)_+$, but the equation above is meant to gather together the salient properties of $(q_1)_+$.

In these results, we let X_1 stand for any object X associated to p_1/q_1. We have one result for the case $p_1/q_1 < p_2/q_2$ and a parallel result for the case $p_1/q_1 > p_2/q_2$. We prove these results in Part IV.

Theorem 7.3 (Copy I) *Suppose that $\kappa \geq 1$ is an integer such that*

$$0 < \frac{p_2}{q_2} - \frac{p_1}{q_1} < \frac{2}{\kappa q_1^2}.$$

1. *If κ is odd and $\lambda_1 < 1/2$ we set $K = (\kappa+1)/2 + \lambda_1$.*
2. *If κ is even and $\lambda_1 > 1/2$ we set $K = \kappa/2 + \lambda_1$.*
3. *If $\kappa \geq 2$ but doesn't satisfy Conditions 1 or 2, we set $K = \text{floor}(\kappa/2)$.*

Then Γ_1 and Γ_2 agree on any lattice point (m, n) such that $F_1(m, n) \geq 0$ and

$$G_1(m, n), H_1(m, n) \in [-q_1 + 3, Kq_1 - 3].$$

Theorem 7.4 (Copy II) *Suppose that $\kappa \geq 1$ is an integer such that*

$$0 < \frac{p_1}{q_1} - \frac{p_2}{q_2} < \frac{2}{\kappa q_1^2}.$$

1. *If κ is odd and $\lambda_1 > 1/2$ we set $K = (\kappa+1)/2 + (1 - \lambda_1)$.*
2. *If κ is even and $\lambda_1 < 1/2$ we set $K = \kappa/2 + (1 - \lambda_1)$.*
3. *If $\kappa \geq 2$ but doesn't satisfy Conditions 1 or 2, we set $K = \text{floor}(\kappa/2)$.*

Then Γ_1 and Γ_2 agree on any lattice point (m, n) such that $F_1(m, n) \geq 0$ and

$$G_1(m, n), H(m, n) \in [-Kq_1 + 3, q_1 - 3].$$

Figure 7.1 illustrates the Copy Theorem I for

$$p_1/q_1 = 3/7; \qquad p_2/q_2 = 11/25.$$

In this case, we have

$$\kappa = 3; \qquad \lambda = 2/7; \qquad K = 2 + 2/7.$$

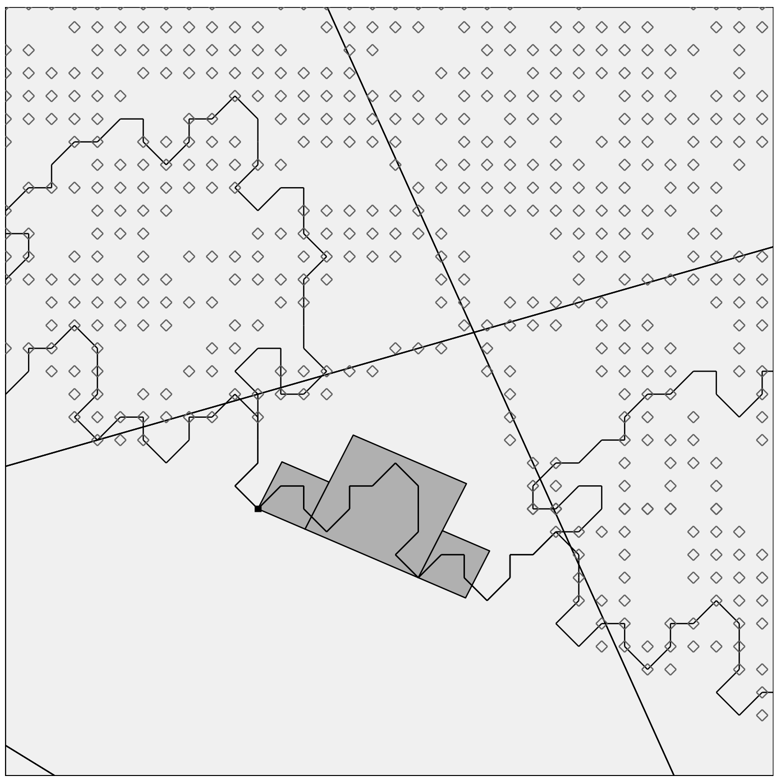

Figure 7.1: The Copy Theorem I in action.

In Figure 7.1 we plot several objects.

- Γ_1 and Γ_2.
- The lines $F_1^{-1}(0)$ and $G_1^{-1}(-q_1)$ and $H_1^{-1}(Kq_1)$.
- The set $SR_1 \cup FR_1 \cup (SR_1 + V_1)$. Note that these grey parallelograms sit symmetrically inside the triangle bounded by our three lines.
- The set of points where $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ disagree. Notice that these points do not penetrate very far into the triangle.

7.3 Proof of Lemma 2.6

We will deduce Lemma 2.6 from The Copy Theorem I. By hypothesis, we have $\kappa = 4$. So, we can take $K = 2$. The vertices of the parallelogram $R_1 = R(p_1/q_1)$ are given in Equation 12. We certainly have $F_1 \geq 0$ on R_1. To show that all vertices in R_1 satisfy the hypotheses of the Copy Theorem I, we just need to evaluate F_1 and G_1 on these vertices. For the bottom vertices, we have

$$G_1(0) = 0; \qquad H_1(0) = 0.$$

$$G_1(V_1) = q; \qquad H_1(V_1) = q.$$

$$G_1(W_1) = -\frac{q^2}{p+q} >^* -q+3.; \qquad H_1(W_1) = \frac{q^2}{p+q} \in (0,q).$$

$$G_1(V_1+W_1) = \frac{pq}{p+q} \in (0,q); \qquad H_1(V_1+W_1) = \frac{q(p+2q)}{p+q} <^* 2q-3$$

The starred inequalities hold true for p sufficiently large. These calculations show that all vertices in R_1 satisfy the hypotheses of the Copy Theorem I once p is large.

We have the easy gradient bounds

$$\|\nabla G_1\| < 8; \qquad \|\nabla H_1\| < 8; \tag{51}$$

which hold for any parameter. From this, and from our first two calculations, we see that any vertex within $q/16$ units of the bottom two vertices of R_1 satisfies the hypotheses of the Copy Theorem I. So, we can take $\epsilon = 1/16$. This proves Lemma 2.6.

7.4 The Main Lemma

Here we state the result we actually use in our proof of our main theorems.

Given any parallelogram X, let

$$A_\epsilon X = X \cup B_{\epsilon q}(x_1) \cup B_{\epsilon q}(x_2). \tag{52}$$

Here $B_{\epsilon q}(x)$ is the ball of radius ϵq around x. The points x_1 and x_2 are the bottom two vertices of X. (The parallelograms of us are situated so that the notion of *bottom* makes sense. The bottom vertices will be the ones near the baseline of the arithmetic graph.)

Let R^* denote the rectangle obtained by shrinking $R = R(p/q)$ by a factor of 2 about the origin. We introduce the following additional notation.

$$\Gamma^{1+\epsilon} = \Gamma \cap (A_\epsilon SR \cup A_\epsilon FR). \qquad \Gamma^{1/2+\epsilon} = \Gamma \cap A_\epsilon R^* \tag{53}$$

$$\sqcap\Gamma^{1+\epsilon} = \widehat{\Gamma} \cap (A_\epsilon SR \cup A_\epsilon FR). \qquad \sqcap\,\Gamma^{1/2+\epsilon} = \widehat{\Gamma} \cap A_\epsilon R^* \tag{54}$$

We write $p_1/q_1 \rightarrow p_2/q_2$ if these rationals satisfy at least one of the hypotheses of at least one of the Copy Theorems. If these rationals are consecutive terms in the sequence constructed in §6, then they have this property by Lemma 6.2.

Corollary 7.5 *There are universal constants C and $\epsilon > 0$ such that the following holds for any pair of odd rationals $p_1/q_1 \rightarrow p_2/q_2$ as long as $p_1 > C$.*

1. *In all cases,*
$$\sqcap\Gamma_1^{1/2+\epsilon} \subset \sqcap\Gamma_2^{1/2+\epsilon}.$$

2.
$$0 < \frac{p_2}{q_2} - \frac{p_1}{q_1} < \frac{2}{2q_1^2}; \quad (q_1)_+ > (q_1)_- \qquad \Longrightarrow \qquad \sqcap\Gamma_1^{1+\epsilon} \subset \sqcap\Gamma_2^1.$$

3.
$$0 < \frac{p_2}{q_2} - \frac{p_1}{q_1} < \frac{3}{2q_1^2}; \quad (q_1)_+ < (q_1)_- \qquad \Longrightarrow \qquad \sqcap\Gamma_1^{1+\epsilon} \subset \sqcap\Gamma_2^1.$$

Proof: If $p_1/q_1 \rightarrow p_2/q_2$ then we can take $K = 1$ in the relevant Copy Theorem. This is to say that Γ_1 and Γ_2 agree on any lattice point (m,n) such that $F_1(m,n) \geq 0$ and

$$G_1(m,n), H_1(m,n) \in [-q_1 + 3, q_1 - 3].$$

We finish the proof in this case using essentially the same calculation as in the proof of Lemma 2.6 above.

Statement 2 corresponds to the case $\kappa = 2$ and $\lambda_1 > 1/2$ in the Copy Theorem I. In this case, we have $K = 1 + \lambda_1$. The left vertices of $FR_1 \cup SR_1$ are the same as the left vertices of R_1, and we have already computed G_1

and H_1 on these. The bottom right vertex is also the same. Finally, the top right vertex of SR_1 is $V + W/2$. We compute

$$G_1(V_1 + W_1/2) = \frac{q_1(2p_1 + q_1)}{2(p_1 + q_1)} > 0$$

$$H_1(V_1 + W_1/2) = \frac{q(2p + 3q)}{2(p + q)} < \frac{3q_1}{2} - 3 < (1 + \lambda_1)q - 3 = Kq_1 - 3.$$

The rest of the proof in this case is essentially the same as for Lemma 2.6.

Statement 3 corresponds to the first case in the Copy Theorem I. Here we have $K \geq 2$ and the proof is the same as for Lemma 2.6. ♠

Remarks:
(i) Just like Lemma 2.6, Corollary 7.5 is biased towards the Copy Theorem I. There is a similar result, biased towards the Copy Theorem II, but we do not state it.
(ii) The hypothesis $p_1 > C$ is not necessary in Corollary 7.5. The Induction Lemma from §22 is a sharper result that eliminates this hypothesis.

8 The Main Results

8.1 Proof of Theorem 1.1

Let $A \in (0,1)$ be an arbitrary irrational. Let $\{A_n = p_n/q_n\}$ be the sequence constructed in §6. By Lemma 6.3 one (or both) of two cases occurs.

1. $\kappa_n \geq 2$ and $A_n < A_{n+1}$ for infinitely many n.
2. $\kappa_n \geq 2$ and $A_n > A_{n+1}$ for infinitely many n.

For ease of exposition, we will consider only the first case. Corollary 7.5 is adapted to this case, and the unstated version of Corollary 7.5 mentioned in Remark (i) at the end of the last chapter would handle the other case in the same way. We call n a *superior index* if $\kappa_n \geq 2$ and $A_n < A_{n+1}$. Such indices have Types 1 or 2, in the sense of Lemma 6.2. Let ϵ be as in Corollary 7.5.

Lemma 8.1 $\Gamma_n^{1+\epsilon} \subset \Gamma_{n+1}^{1/2+\epsilon}$ *for any sufficiently large superior index.*

Proof: Let $v_n = p_n/q_n$ and $R_n = R(p_n/q_n)$. Let H_n be the halfspace above the baseline of Γ_n. From either Statement 2 or Statement 3 of Corollary 7.5 we have $\Gamma_n^{1+\epsilon} \subset \Gamma_{n+1}^1$. Now we make three observations.

1. Since n is a superior index, we have $q_{n+1} > 2q_n$. By Lemma 6.5, the difference $q_{n+1} - 2q_n$ tends to ∞ with n.
2. The slope of each edge of R_n differs from the slope of the corresponding edge of $R_{n+1}/2$ by $O(q_n^{-2})$, whereas the edges of R_n have length $O(q_n)$.
3. R_n and $R_{n+1}/2$ share the bottom left corner.

It follows from these facts that

$$R_n \cap H_{n+1} \subset R_{n+1}/2; \qquad B_{\epsilon q_n} \cap H_{n+1} \subset R_{n+1}/2 \cup B_{\epsilon q_{n+1}}(v_{n+1}/2), \quad (55)$$

once n is sufficiently large. Since $\Gamma_n^{1+\epsilon} \subset \Gamma_{n+1}$, we have $\Gamma_n^{1+\epsilon} \subset H_{n+1}$. Therefore,

$$\Gamma_n^{1+\epsilon} \subset \Gamma_{n+1} \cap H_{n+1} \cap \Big(R_n \cup B_{\epsilon q_n}(v_n)\Big) \subset$$

$$\Gamma_{n+1} \cap \Big(R_{n+1}/2 \cup B_{\epsilon q_{n+1}}(v_{n+1}/2)\Big) = \Gamma_{n+1}^{1/2+\epsilon}.$$

This completes the proof. ♠

Lemma 8.2 $\Gamma_n^{1/2+\epsilon} \subset \Gamma_{n+1}^{1/2+\epsilon}$ *for any sufficiently large index.*

Proof: This time we use Statement 1 of Corollary 7.5, and we use $R_n/2$ and $v_n/2$ in place of R_n and v_n. The rest of the proof is about the same as in the previous lemma. ♠

If necessary, we chop the beginning of our sequence so that at least one of the two results just proved holds for each index. We also choose ϵ small enough so that

$$\Gamma_n^{1/2+\epsilon} \subset \Gamma_n^1. \tag{56}$$

As an immediate consequence of the two preceding results, we have

$$\Gamma_n^{1+\epsilon} \subset \Gamma_m^1; \qquad \forall m > n, \tag{57}$$

provided that n is a superior index.

Now we pass to the subsequence of superior indices. Once we do this, Equation 57 holds for all indices. Now we are exactly in the situation analyzed in §4. The argument there, done here word for word, finishes the proof.

8.2 Local Stability of Orbits

The results in this section are useful for the proofs of Theorems 1.2 and 1.3.

Given and $A \in (0,1)$ and $\epsilon > 0$, let $\Sigma_\epsilon(A) \subset (0,1)^2$ denote those pairs (s, A') where $s \in (0,\epsilon)$ and $|A' - A| < \epsilon$. Let $O(s,1;A')$ denote the outer billiards orbit of $(s,1)$ relative to $K(A')$.

Lemma 8.3 *Suppose that $A \in (0,1)$ is irrational. For any N there is some $\epsilon > 0$ with the following property. The first N iterates of $O(s,1;A')$, forwards and backwards, are well defined provided that $(s,A') \in \Sigma_\epsilon(A)$.*

Proof: Inspecting the proof of Lemma 2.1, we draw the following conclusion. If $O(s,1;A')$ is not defined after N iterates, then $s = 2A'm + 2n$ for integers $m,n \in (-N',N')$. Here N' depends only on N. Rearranging this equation, we get

$$|A' - \frac{m}{n}| < \frac{s}{2m}.$$

For s sufficiently small and A' sufficiently close to A, this is impossible. ♠

Lemma 8.4 *Suppose that $A \in (0,1)$ is irrational. For any N there is some $\epsilon > 0$ with the following property. The combinatorics of the first N forward iterates of $O(s,1;A')$ is independent of the choice of point $(s,A') \in \Sigma_\epsilon(A)$. The same goes for the first N backwards iterates.*

Proof: This follows from the fact that the combinatorial type, a discrete piece of data, varies continuously over any region where all N iterates are defined. ♠

Corollary 8.5 *There is a divergent sequence $\{n_k\}$ with the following property. If $|A-A'| < 1/k$ and $s,s' \in (0,1/k)$, then the first n_k forwards (or backwards) iterates of $O(s,1;A)$ have the same combinatorial structure as the first n_k forwards (or backwards) iterates of $O(s',1;A')$.*

Let Ψ denote the first return map relative to the parameter A. Likewise define Ψ'. The following result is just a consequence of the existence of the return map. See the Return Lemma from §2.

Corollary 8.6 *There is a divergent sequence $\{N_k\}$ with the following property. If $|A-A'| < 1/k$ and $s,s' \in (0,1/k)$, then the first N_k iterates of Ψ applied to $(s,1)$ have the same combinatorial structure as the first N_k iterates of Ψ' applied to $(s',1)$. This holds in both the forwards and backwards directions.*

The *Hausdorff distance* between two compact sets $S_1, S_2 \subset \boldsymbol{R}^2$ is the infimal $d = d(S_1,S_2)$ such that S_1 us contained in the d tubular neighborhood of S_2, and *vice versa*. A sequence $\{S_n\}$ of closed sets in $\boldsymbol{R}^2$ converges in the Hausdorff topology if there is a closed subset S such that $d(S_n \cap K, S \cap K) \to 0$ for every compact K.

Lemma 8.7 (Rigidity) *Let $\widehat{\Gamma}_n$ denote any sequence of arithmetic graphs corresponding to and sequence $\{A_n\}$ of parameters converging to an irrational parameter A. Let η_n denote a lattice point that lies above the baseline L_n of Γ_n. Suppose that the distance from η_n to L_n converges to 0. Then the sets $\widehat{\Gamma}_n(\beta_n) - \beta_n$ converge in the Hausdorff topology. The limit Γ_0 only depends on A. Indeed, Γ_0 is the Hausdorff limit of the graphs $\Gamma(p_n/q_n)$, where $\{p_n/q_n\}$ is any sequence of rationals that converge to A.*

Proof: The existence of a universal limit Γ_0 is just a reformulation of the preceding corollary. the point is that the arithmetic graph exactly captures the combinatorial structure of the return map. Since the limit Γ_0 is independent of which sequence of graphs/points we choose, we let $\Gamma_n = \Gamma(p_n/q_n)$ and we let $\beta_n = (0,0)$ for all n. ♠

Lemma 8.8 *For any n there is a vertex v of Γ_0 such that $\|v\| > n$ and v is within $1/n$ of the line of slope $-A$ through 0.*

Proof: Let $\{p_n/q_n\}$ denote our final sequence considered in §7. We have $\Gamma_n^1 \subset \Gamma_m^1$ for all $m > n$. Taking the limit, we see that $v_n = (q_n, -p_n) \in \Gamma_0$ for all n. These points serve our purpose. (In the case not treated in our proof of Theorem 1.1, we would have $\Gamma_n^0 \subset \Gamma_m^0$ for all $m > n$. Here $\Gamma_n^0 = \Gamma_n^1 - v_n$. In this case, the points $-v_n$ would all belong to Γ_0.) ♠

8.3 Proof of Theorem 1.2

Recall that $\Xi = \boldsymbol{R}_+ \times \{-1, 1\}$. Let $\xi \in \Xi$ be a point whose orbit is erratic.

Suppose, for the sake of contradiction, that the orbit of $\zeta \in \Xi$ is bounded in (w.l.o.g.) the forwards direction and also aperiodic. Let $\widehat{\Gamma}$ be the arithmetic graph corresponding to ζ. Let Γ' be the half-component corresponding to the forwards orbit of ζ.

By construction, Γ' is contained in some infinite strip parallel to the baseline of $\widehat{\Gamma}$. Since this strip has a finite width, we know that Γ' travels unboundedly far either to the left or to the right. We will suppose that Γ' travels infinitely far to the right. Let L be the baseline of $\widehat{\Gamma}$. Since L has irrational slope, we can find a sequence of lattice points (m_k, n_k) such that the distance from (m_k, n_k) to L tends to 0. We can assume that these points travel to the right.

Taking a Hausdorff limit of the graphs $\widehat{\Gamma}(m_k, n_k) - (m_k, n_k)$, we get the only limit possible, by the Rigidity Lemma. But this limit rises unboundedly far from the baseline, by Theorem 1.1. But this shows, for k large, that the component $\widehat{\Gamma}(m_k, n_k)$ starts out beneath Γ' and rises up through the strip containing Γ'. But then the component passing through the strip would intersect the component contained in the strip, contradicting the Embedding Theorem. ♠

8.4 Proof of Theorem 1.3: Main Argument

We define the *depth* of a lattice point (above the baseline) to be the distance from this lattice point to the baseline of $\widehat{\Gamma}$. We define the depth of a polygon in $\widehat{\Gamma}$ to be the minimum of the depth of its vertices. In this section we will prove the following result.

Lemma 8.9 *Suppose that $\widehat{\Gamma}_\infty$ contains a sequence $\{\gamma_k\}$ of polygon components such that the depth of γ_k tends to 0. Then the set of special periodic orbits relative to $K(A)$ is dense in the set of all special periodic orbits.*

Let $|\gamma_k|$ denote the supremal value of d such that there are two vertices of γ_k, at least d apart, having depth less than $1/d$.

Lemma 8.10 *$|\gamma_k| \to \infty$ as $k \to \infty$.*

Proof: Let (m_k, n_k) be the vertex of γ_k of lowest depth. By the Rigidity Lemma, the component $\gamma_k - (m_k, n_k)$ converges to Γ_0, and one direction or the other connects (m_k, n_k) to a far away and low-depth vertex (m_k', n_k'). ♠

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 8.11 *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 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 fundamental map associated to $\widehat{\Gamma}$. Then $M^{-1}(U_k)$ is an open strip, parallel to L. Since L has irrational slope, there is some constant N_k so that every point of L is within N_k of some point of $M^{-1}(U_k) \cap \boldsymbol{Z}^2$. 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 8.1. Any vertex V below Y corresponds to a periodic orbit, by our orbit confinement result of §5.5.

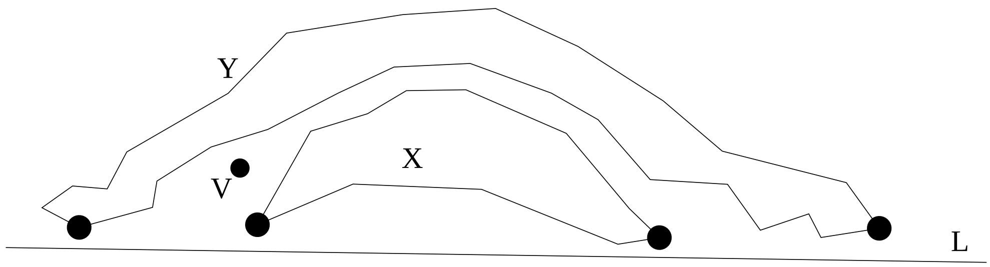

Figure 8.1: One polygon overlaying another.

Now we pass to a subsequence so that

$$|\gamma_{k+1}| > 10(N_k + |\gamma_k|). \tag{58}$$

Equation 58 has the following consequence. For any integer N, we can find components γ_j of S_j, for $j = N, ..., 2N$ such $\gamma_N \bowtie ... \bowtie \gamma_{2N}$. 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.

Lemma 8.12 *The set $M(\boldsymbol{Z}^2 \cap \Lambda_N)$ 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 upwards at least N units before escaping from underneath γ_{2N}. This is an application of the pideonhole princple. 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}. ♠

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

8.5 Manufacturing the Periodic Sequence

It remains to show that and irrational A satisfies the hypotheses of Lemma 8.9.

Let Γ'_k denote the component of $\widehat{\Gamma}_k = \widehat{\Gamma}(p_k/q_k)$ that contains the vertex corresponding to the orbit of $(3/q_k, 1)$.

Lemma 8.13 Γ'_k *is a polygon.*

Proof: Note that $\Gamma'_k \neq \Gamma_k$, by Lemma 5.3. Theorem 1.6 now implies that $O(3/q_k, 1)$ is stable. Hence Γ'_k is a polygon. Alternatively, Γ'_k contains a low vertex and hence lies beneath Γ_k. Hence, Γ_k confines Γ'_k. ♠

To finish the proof of Theorem 1.3, we prove the following result.

Lemma 8.14 $\widehat{\Gamma}_\infty$ *contains a translate of* Γ'_k *for infinitely many indices* k. *Moreover, the depth of the translates converges to* 0 *as* k *tends to* ∞.

As above, we give the proof assuming that $A_n < A_{n+1}$ infinitely often when $\kappa_n \geq 2$. (These are the superior indices.) The other case has essentially the same proof.

Our first result is not necessary *per se* for our proof of Lemma 8.14. However, it serves as a template for a later result, Lenna 8.16, that is necessary. Lemma 8.16 has the same proof as Lemma 8.15, once the relevant definitions are in place, but the proof of Lemma 8.15 in isolation is more transparent.

Lemma 8.15 Γ_∞ *contains translates of* Γ^1_n *for all superior* n.

Proof: Referring to the proof of Lemma 3.4, the portion of the arc γ'_{n_0} starting at the point μ_n and travelling to the right contains a translate of Γ^1_k for some choice of superior index k that tends to ∞ with n. Also Γ^1_k contains Γ^1_{k-1}. Hence $\widehat{\Gamma}_1$ contains translates of Γ^1_n for all superior n. ♠

There is a translation τ_n such that

$$\tau_n(\Gamma^1_n) \subset \Gamma_\infty. \tag{59}$$

Our proof of Lemma 8.15 gives us the additional piece of information that the distance from $\tau_n(0,0)$ to the baseline of Γ_∞ converges to 0 with n.

We are also interested in other components of $\widehat{\Gamma}_\infty$ besides Γ_∞. Similar to Equation 54, we define

$$\sqcap\Gamma_n^1 = \widehat{\Gamma} \cap (SR_n \cup FR_n); \qquad \sqcap\Gamma_n = \bigcup_{j=-\infty}^{\infty} \sqcap\Gamma_n^1. \tag{60}$$

Thus $\sqcap\Gamma_n^1$ is a single, distinguished period of $\sqcap\Gamma_n$.

Lemma 8.16 *$\widehat{\Gamma}_\infty$ contains translates of $\sqcap\Gamma_n^1$ for all n.*

Proof: The same argument as in §4 gives us the system

- $\sqcap\Gamma_n^1 \subset \sqcap\Gamma_{n+1}^1$ for all n.
- $\sqcap\Gamma_n^2 = \sqcap\Gamma_n^1 + v_{n+1} \subset \sqcap\Gamma_m^1$ for all (m,n) such that $m \geq n+2$.

In reference to the translation lemma, we can define the arc $\sqcap\gamma_{nk}$ to be the set of edges of $\sqcap\Gamma_{2n+1}^1$ that lie on the same vertical lines as points of γ_{nk}. With this definition in place, the Translation Lemma goes through, with $\sqcap X$ in place of each object X. Indeed, all the constructions in §4 go through with $\sqcap X$ in place of X. Finally, the proof of Lemma 8.15 goes through essentially word for word. ♠

Let L_k be the baseline for Γ_k and let L_∞ denote the baseline for Γ_∞. Let L_k^1 denote a single period of L_k, starting at $(0, -1/(2q))$.

Lemma 8.17 *$\tau_k(L_k^1)$ lies in the $\epsilon/4$ neighborhood of L_∞ for k large.*

Proof: Let a and b be the two endpoints of L_k^1. By construction, $\tau_k(a)$ is within $\epsilon/4$ of L_∞ for large k. The slope of L_k is within $O(q_k^{-2})$ of the slope of L_∞, and the length of L_k is $O(q_k)$. So, taking k large, we can arrange that $|d(\tau(a), L_\infty) - d(\tau(b), L_\infty) < \epsilon/4$. This does it. ♠

Since Γ_k' has a low vertex and Γ_k, we have $\Gamma_k' \subset \sqcap\Gamma_k$. By periodicity, we can assume that $\Gamma_k' \subset \sqcap\Gamma_k^1$. We set $\gamma_k = \tau_k(\Gamma_k')$. By construction $\gamma_k \subset \widehat{\Gamma}_\infty$ for each superior index k.

The depth of Γ_k' relative to the baseline L_k tends to 0. Combining this fact with Lemma 8.17, and the fact that depth of $\tau_k(0,0)$ tends to 0, we see that the depth of γ_k relative to the baseline of $\widehat{\Gamma}_\infty$ tends to 0 as well.

This completes the proof of Lemma 8.14 and, hence, Theorem 1.3.

9 Further Results and Conjectures

In this chapter we will state some conjectures, and also a few additional results. Mainly we include the results to motivate the conjectures and in some case provide support for them. Our proofs of these results will be pretty sketchy. The skeptical reader can just ignore these results. We think that we have only uncovered part of what must be a very beautiful structure underlying outer billiards on kites.

9.1 Return Dynamics

In this section we will consider the full outer billiards orbit, and not just the orbit under the square map.

Let $X = (0,2) \times \{\pm 1\}$. The orbits in X seem to have a particularly nice structure. Given a Cantor set $C \subset X$, let C' denote the set of endpoints of the components of $X - C$.

Conjecture 9.1 *Let $A \in (0,1)$ be an irrational parameter. Then there is a Cantor set $C_A \subset X$ with the following properties relative to outer billiards on $K(A)$.*

1. *An orbit of X is well-defined iff it lies in $X - C'_A$.*
2. *All the orbits in $X - C_A$ are periodic. Orbits within the same component of $X - C_A$ have the same combinatorial structure as each other.*
3. *Each orbit of $C_A - C'_A$ is unbounded, and intersects C_A in a dense set.*
4. *All but one orbit of $C_A - C'_A$ is erratic in the forwards direction and all but one orbit of $C_A - C'_A$ is erratic in the backwards direction.*

In [**S**] we essentially proved Statements 3 and 4 for the Penrose kite parameter $A = \phi^{-3}$.

We don't know much about the orbits that don't start in X, but here is one conjecture. Let U be union of all points that have unbounded orbits.

Conjecture 9.2 *Any unbounded special orbit is dense in U. In particular, at least one of the forwards or backwards direction of any unbounded orbit is erratic.*

9.2 Low Components

Recall that a lattice point (m, n) is *low* with respect to the parameter p/q if the baseline of $\Gamma(p/q)$ separates $(m, n-1)$ from (m, n). Such vertices correspond to points in X. We say that a *low* component of $\widehat{\Gamma}(p/q)$ is a component that contains a low vertex. The basic idea behind Conjecture 9.1 is that the low components have a structure that (conjecturally) is built up inductively from the continued fraction expansion of the parameter. In this section we explore this structure.

Now that we are talking about the arithmetic graph again, we revert back to the study of the square of the outer billiards map.

Conjecture 9.3 *Let p/q be any rational. Two low vertices (m_1, n_1) and (m_2, n_2) lie on the same component of $\widehat{\Gamma}(p/q)$ only if $m_1 + n_1$ and $m_2 + n_2$ have the same parity.*

We will sketch the proof in the odd case.

Lemma 9.4 *Let $A \in (0, 1)$ be any parameter, and let $s_0, s_1 \in (0, 2)$ be any points on which the outer billiards orbits are defined. Then*

$$O_2(s_0, 1) \neq O_2(s_1, -1).$$

Proof: (sketch) Let Γ be the arithmetic graph corresponding to the common orbit $O_2(s_0, 1) = O_2(s_1, -1)$. Let (m_j, n_j) denote the vertex corresponding to s_j. Then (m_j, n_j) is a low vertex. The line segment σ_j connecting $(m_j, n_j - 1)$ to (m_j, n_j) crosses the baseline of Γ. As Γ' passes (m_j, n_j), the segment σ_j is either on the right or the left, relative to the canonical orientation on Γ'. An explicit calculation shows that σ_j is on the right for $j = 0$ and on the left for $j = 1$. But then we get the same topological contradiction that we got in the proof of Lemma 5.3. ♠

The first half of the proof of Lemma 5.3 combines with Lemma 9.4 to establish Conjecture 9.3 in the odd case. Lemma 9.4 generalizes the argument we gave in the second half of the proof of Lemma 5.3

We don't know how to prove Conjecture 9.3 in the even case, but we will state another conjecture below that implies it.

Let p/q be an odd rational. Define

$$p' = \min(p_+, p_-); \qquad q' = \min(q_+, q_-). \tag{61}$$

Then p'/q' is one of the two even rationals p_+/q_+ or p_-/q_-. Let p^*/q^* be the rational such that

$$\frac{p^*}{q^*} = \frac{|(p)_+ - (p)_-|}{|(q)_+ - (q)_-|}. \tag{62}$$

Note that p^*/q^* is the last term on the (finite) odd rational sequence approximating p/q. We now set

$$\frac{p_2}{q_2} = \frac{p}{q}; \qquad \frac{p_1}{q_1} = \frac{p'}{q'}; \qquad \frac{p_0}{q_0} = \frac{(p^*)'}{(q^*)'}. \tag{63}$$

Then p_2/q_2 is an odd rational and p_0/q_0 and p_1/q_1 are even rationals.

Recall that

$$\Gamma_2^1 \subset FR_2 \cup SR_2,$$

according to the Decomposition Theorem. All the low components of $\widehat{\Gamma}_2$ other than Γ_2 are polygons. The following conjecture (and periodicity) says that all these other low components are inherited from simpler rationals.

Conjecture 9.5 *The set of low components of $\widehat{\Gamma}_2 \cap FR_2$ is a translate of the set of low components of $\widehat{\Gamma}_1 \cap FR_2$. The set of low components of $\widehat{\Gamma}_2 \cap SR_1$ is a translate of the set of low components of $\widehat{\Gamma}_0 \cap SR_1$.*

In case $p_1/q_1 = 0/1$ or $1/1$, the graph $\widehat{\Gamma}_1$ is interpreted to be the empty set. The same goes for p_0/q_0.

Proof of Conjecture 9.3: Each even rational p_1/q_1 participates in a triple as described above, and each low component of $\widehat{\Gamma}_1$ is translation equivalent to one of the low components of $\widehat{\Gamma}_2 \cap FR_2$. Thus, we can embed the "even picture" inside the "odd picture" and deduce the even case of Conjecture 9.3 from the odd case.

The inheritence of low components cuts both ways. Let p_2/q_2 be an even rational. There is a unique odd rational p_1/q_1 such that $q_1 < q_2$ and $|p_1q_2 - p_2q_1| = 1$.

Conjecture 9.6 *The set of low components of $\widehat{\Gamma}_2$ lying underneath Γ_2 is a translate of the set of low components of $\widehat{\Gamma}_1$ that are contained in one of FR_1 or SR_1. The choice depends on whether $2q_1 < q_2$ or $2q_1 > q_2$.*

Conjecture 9.7 *One period's worth of the set of low components of $\widehat{\Gamma}_2$ not lying underneath Γ_2 is a translate of the set of low lying components of $\widehat{\Gamma}_1$ that are contained in one of SR_1 or FR_1. The choice depends on whether $2q_1 < q_2$ or $2q_1 > q_2$.*

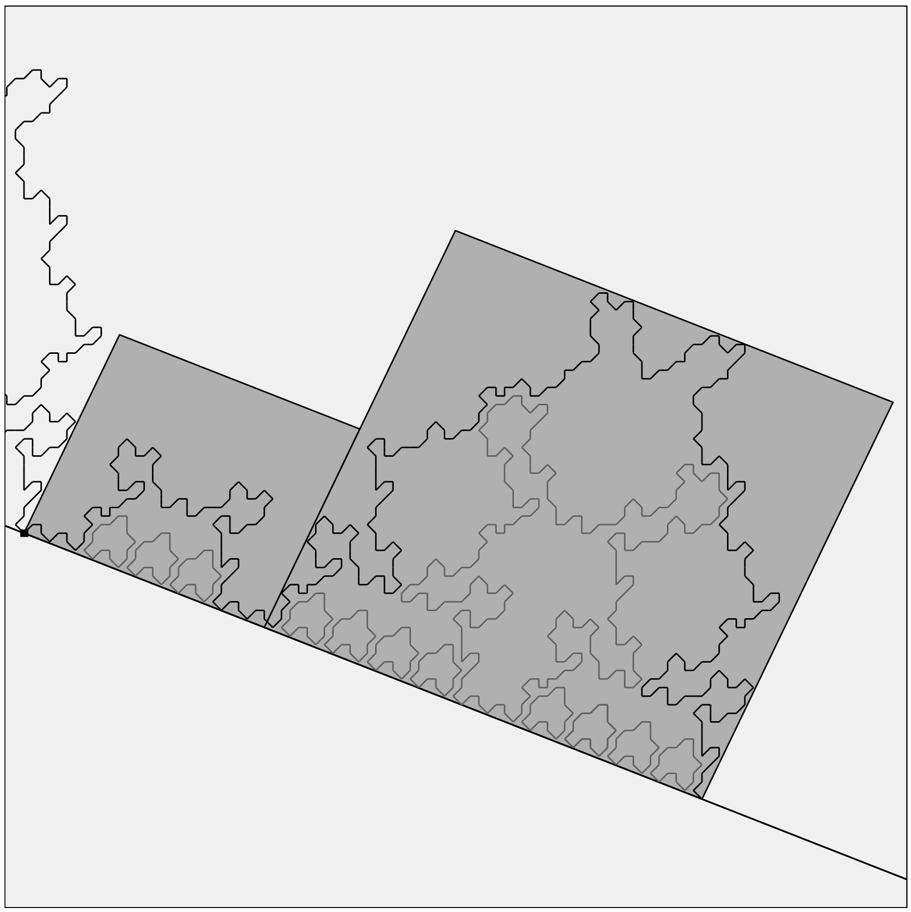

Figure 9.1: Low components of $\Gamma(31/79)$

Figure 9.1 illustrates Conjecture 9.5 for $p_2/q_2 = 31/79$. In this case, $p_1/q_1 = 11/28$ and $p_0/q_0 = 2/5$. We have drawn the low polygons in grey. There are 3 such components in $SR(31/79)$, and these are all translates of $\Gamma(2/5)$. There are $9 = 4 + 1 + 4$ such components in FR_2, eight of which are translates of $\Gamma(2/5)$. The central component is a translate of $\Gamma(11/28)$.

9.3 The Modular Limit Phenomenon

Most of the analysis Part I of the monograph deals with sequences of rationals $\{p_n/q_n\}$ where the complexity of the rational tends to ∞. By this we mean e.g. that the length of the continued fraction expansion tends to ∞. In this section will study a different kind of sequence $\{p_n/q_n\}$ in which the length of the continued fraction expansion stays bounded but nonetheless the denominators tend to ∞. As in the "unbounded complexity" case, we will take our limits in a very controlled way.

Our construction is based on an odd rational $p/q \in (0,1)$ and an even ratonal $r/s \in (0,1)$, where r is odd and s is even. There is an element $T \in SL_2(\boldsymbol{Z})$ such that $T(\infty) = p/q$. Here we mean that T acts as a linear fractional transformation. The choice of T is not unique, but Billiard King always makes the choice

$$T = \begin{bmatrix} p & q_- \\ q & q_- \end{bmatrix}. \tag{64}$$

We define

$$A_n = T(n + \frac{r}{s}). \tag{65}$$

Finally, we let $\Gamma_n = \Gamma(A_n)$. Notice that $A_n \to p/q$.

Conjecture 9.8 (Modular Limit) *Restricting n to either the odd integers or the even integers, the rescaled Hausdorff limit*

$$\lim_{n \to \infty} \frac{1}{n} \Gamma_n$$

exists and is an infinite periodic polygonal curve.

The reader in the right frame of mind will recognize the similarity between the Modular Limit Conjecture and the phenomenon that the geometric limit of a sequence of Kleinian groups can be larger than the associated algebraic limit. Both are examples of "renormalization phenomena".

Figure 9.2 illustrates this phenomenon for the choices $p/q = 1/3$ and $r/s = 1/4$ and $n = 20$. The shape $\Gamma(31/121)$ very closely conforms to a pretty simple polygonal curve, with many small fluctuations about this curve. In the rescaled limit, as $n \to \infty$, these fluctuations disappear.

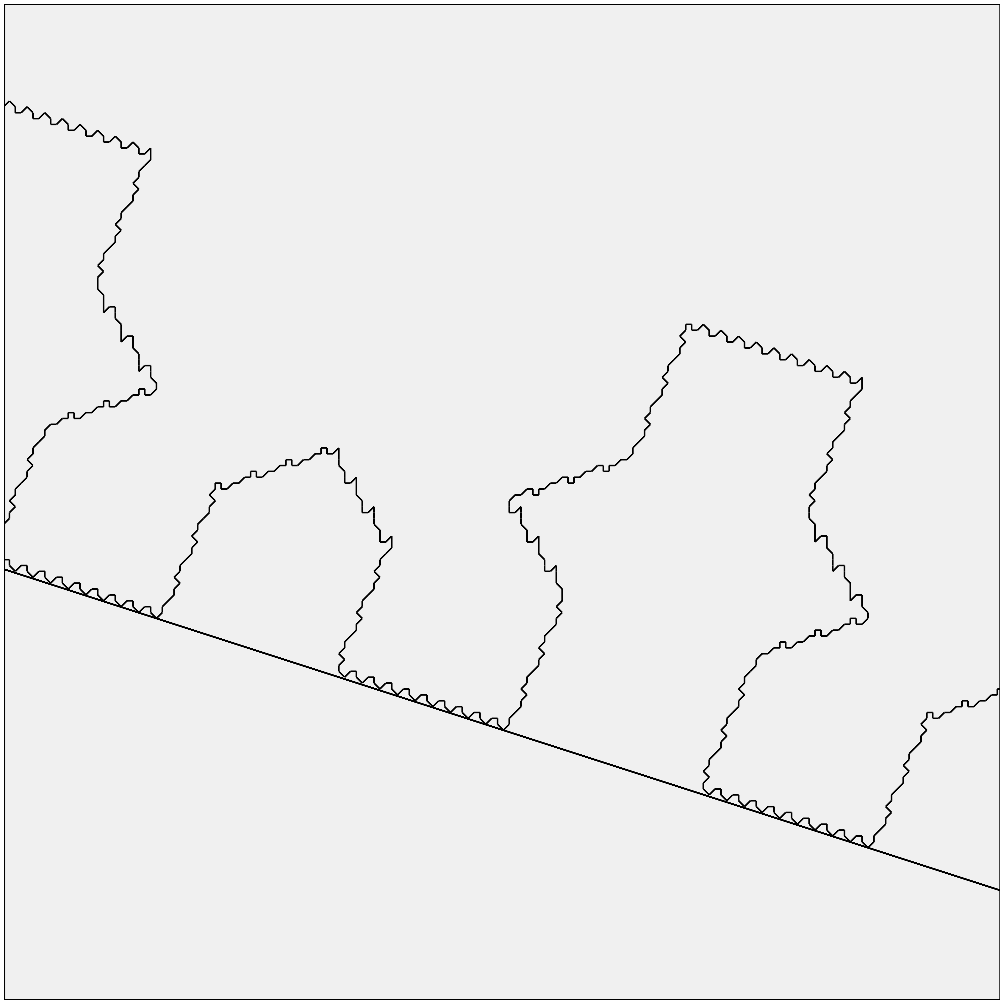

Figure 9.2: $\Gamma(39/121)$.

Here is a weaker but related result that we can prove.

Lemma 9.9 *When n is restricted to either the odd or even integers, the rescaled Haussdorff limit*

$$\lim_{n\to\infty} \frac{\log(n)}{n} \Gamma_n$$

exists and is the union of 2 rays. One of the rays is $\pm \lim V_n$ and the other ray is $\lim W_n$. Here $V_n = V(p_n/q_n)$ and W_n are as in Equation 11.

Proof: (sketch) Our sketch will use the "big O notation".

The sign of $p_n/q_n - p/q$ depends on the parity of n. Fixing the parity keeps the sign constant. We consider the case when $p/q < p_n/q_n$. The Copy Theorem I forces the forwards direction of Γ_n to copy $O(n)$ periods of $\Gamma(p/q)$. When this portion is rescaled at a sublinear rate, it converges to the forward direction of the baseline, namely $\lim V_n$.

Recall that the behavior of Γ_n is controlled not just by the hexagrid G_n but also a rotated copy of G_n. See the discussion in §5.3. It turns out that there are 2 line segments σ_1 and σ_2 with the following properties.

- σ_1 and σ_2 are both parallel to W_n and have length $O(n)$.
- σ_1 and σ_2 are $O(1)$ units apart from each other.
- The first $O(n)$ steps of the backwards portion of Γ_n are contained between σ_1 and σ_2.

σ_1 is the left edge of the rectangle $R_n = R(p_n/q_n)$ and σ_2 is the parallel segment containing the point $(-q,p)/2$. So, the backwards direction of Γ_n travels up $O(n)$ steps up a uniformly narrow strip. When rescaled at a sublinear rate, the result converges to $\lim W_n$. ♠

Both Billiard King and the interactive guide to the monograph are set up so that the reader can explore this phenomenon and variants. One interesting variant arises when we take p/q to be even. In this case, the limit above seems to contain solid triangular pieces. Figure 9.3 below shows an example of this, for the parameter $101/200$, a very close approximation to $1/2$.

Here is a weaker but more concrete formulation, which we have no idea how to prove.

Conjecture 9.10 *The limit*

$$\lim_{n\to\infty} \frac{\log(n)}{n}\Gamma_n$$

exists and is the union of two cones in $\boldsymbol{R}^2$.

When p/q is extremely near an even rational, $\Gamma(p/q)$ exhibits some plane-filling tendencies. We have no idea why this happens. All we can say that Γ_n can't copy a full period of $\Gamma(p/q)$ because Γ_n is embedded and $\Gamma(p/q)$ is a closed polygon. So, something else is forced to happen.

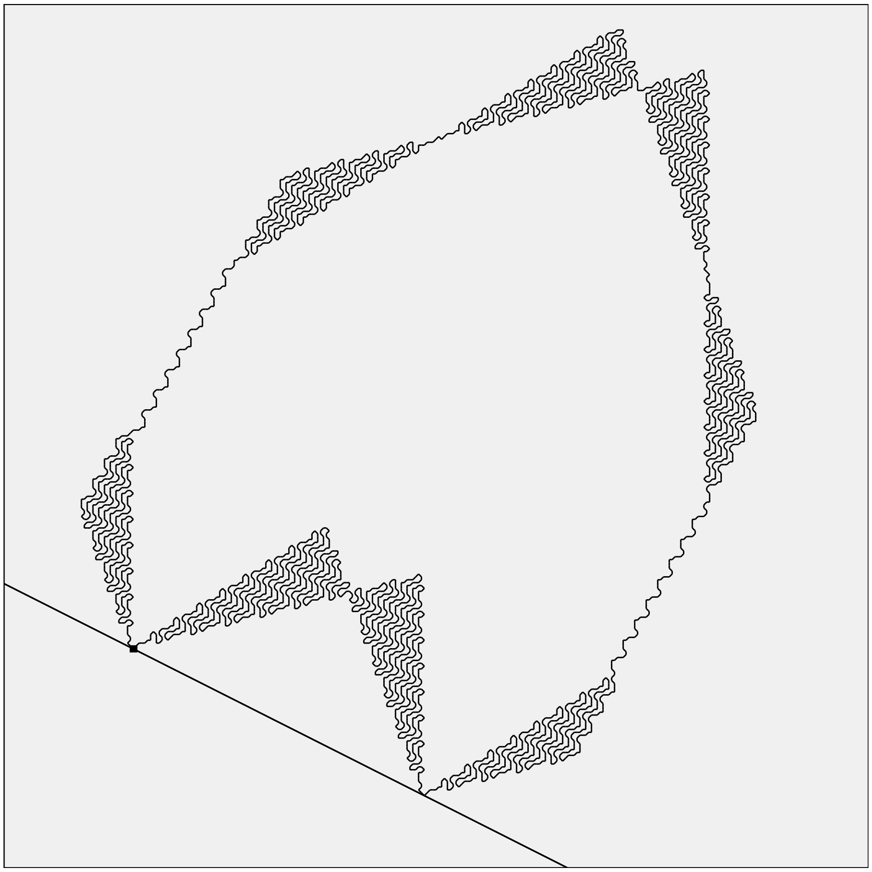

Figure 9.3: $\Gamma(101/200)$.

9.4 General Orbits

In this section we briefly discuss what seems to happen for the general outer billiards orbit. First of all, as we mentioned in the introduction, Theorems 1.2 and 1.3 seem to be true in this case.

Here is our first conjecture.

Conjecture 9.11 *The orbit $O_2(x,0)$ is periodic for any $x>0$. This orbit winds once around the origin.*

We could probably prove this conjecture using a result like our Master Picture Theorem from Part II, but we haven't tried to do this. The general outer billiards orbits interpolate, in some sense, between the special orbits and the orbits from Conjecture 9.11.

We can study the general outer billiards orbit in a way that is similar to what we did for the special orbits. For any $\beta \in (-1,1)$, the set

$$S_\beta = \boldsymbol{R} \times (\boldsymbol{Z}_{\rm odd} + \beta) \tag{66}$$

is preserved by the square of the outer billiards map. For this reason, every general orbit lies on some S_β. The special orbits lie on S_0. The case $\beta = \pm 1$ corresponds to the orbits in Conjecture 9.11.

To study the orbits on S_β we introduce the set

$$\Xi_\beta = \boldsymbol{R}_+ \times \{\beta - 1, \beta + 1\}. \tag{67}$$

We can then define an arithmetic graph similar to what we did for the special orbits. We choose some $\alpha > 0$ and then follow the orbits having the form

$$\Big(2Am + 2n + 2\alpha, (-1)^{m+n+1} + \beta\Big). \tag{68}$$

We then form the arithetic graph $\widehat{\Gamma}_\alpha(A;\beta)$ as in §2. It seems that the basic topological features of this graph hold for all β. For instance, the graph is a lattice graph whose edges connect nearest neighbors. Also, the Embedding Theorem seems to hold.

The Hexagrid Theorem no longer holds. Also, the stability classification changes. It seems that there are both unstable and stable general orbits both in the even and odd cases. The period copying seems to work roughly in the same way, but we have not seriously investigated it.

When β is near 0, the picture looks pretty much the same as what we see for for the special orbits – i.e. $\beta = 0$. However, as $\beta \to \pm 1$ we see a new phenomenon emerge that throws a new light on the whole business. Figure 9.4 illustrates the parameter $A = 11/25$ and $\beta = 5/6$. It seems that the orbits follow along 4 infinite families of lines, and then swerve to avoid each other at intersections between these lines. The lines governing the orbits are all parallel to the lines in the door grid $DG(p/q)$. These lines are something like "guides" for the orbits.

We think of β as a kind of "temperature". As $\beta \to \pm 1$ the picture seems to "freeze": The spacing of the $\boldsymbol{Z}^2$ lattice is tiny in comparison to the spacing between the guiding lines, and the guiding lines stand out very clearly. Very few orbits are nontrivial, and those few nontrivial orbits travel for a long time in nearly straight lines before having any "interactions" with other orbits.

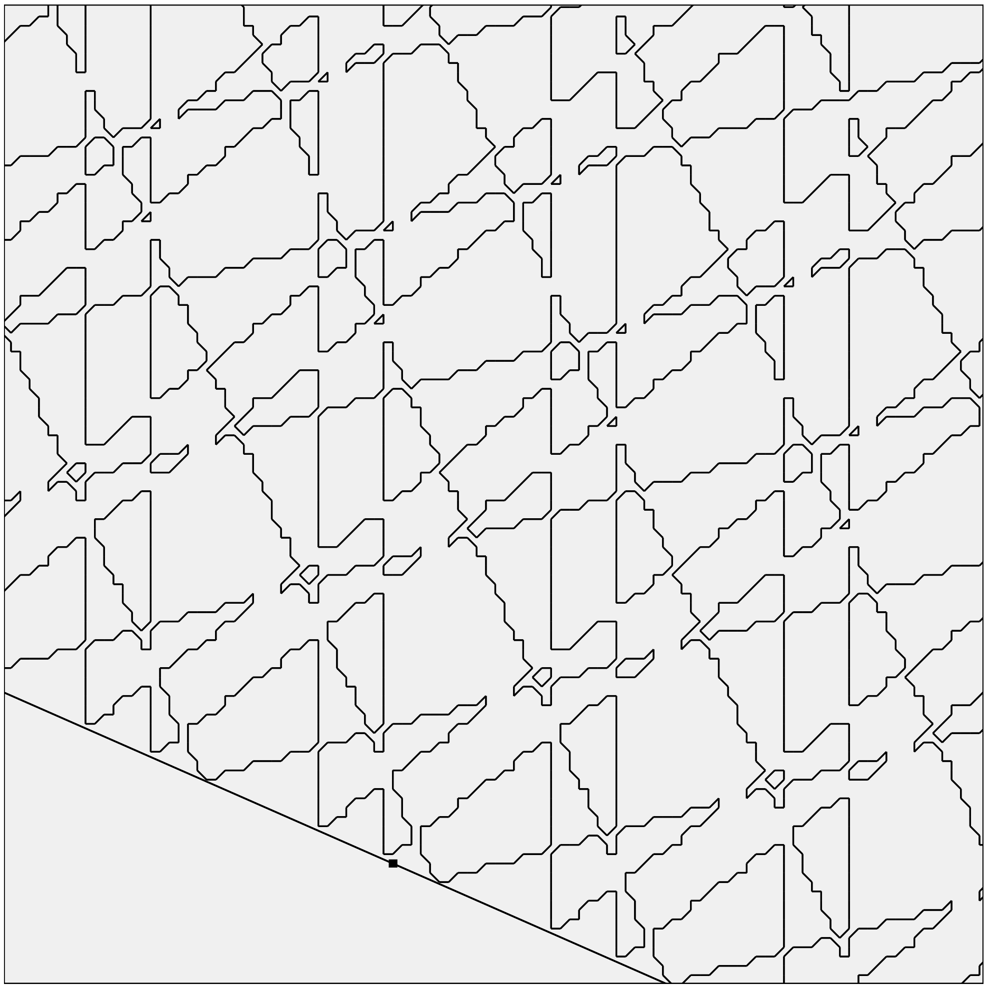

Figure 9.4: $\widehat{\Gamma}(11/25; 5/6)$

As β approaches 0 the picture heats up, so to speak. The spacing of the the $\boldsymbol{Z}^2$ lattice approaches the scale of the guiding lines, and these guiding lines are obscured. There are many more nontrivial orbits, but these orbits interact with each other much more frequently. We think of the Hexagrid Theorem as a sort of remnant of this guiding line structure, a ghost that survives to the boiling point.

The reader can see all this phenomena using Billiard King. The above account is really just an impressionistic view of what might be going on. We don't really understand the general picture very well and we hope to revisit it at a later time.

Part II:

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

- In §10 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, $\boldsymbol{R}^3/\Lambda$, together with a partition of $\boldsymbol{R}^3/\Lambda$ into polyhedra. Here Λ is a certain 3-dimensional lattice that depends on the parameter.

- In §11, we will prove the Pinwheel Lemma, a key technical step along the way to our 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 §12 we prove the Torus Lemma, another key result. The Torus Lemma implies that there exists some partition of our 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 §13 we verify, with the aid of symbolic manipulation, certain functional identities that arise in connection with the Torus Lemma. These function identities are the basis for our analysis of the singular set.

- In §14 we combine the Torus Lemma with the functional identities to prove the Master Picture Theorem.

- in §15 we will explain how one actually makes computations with the Master Picture Theorem. §15.2 will be very important for Part IV of the monograph.

10 The Master Picture Theorem

10.1 Coarse Formulation

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

$$\Xi_+ = \bigcup_{k=0}^{\infty}(2k,2k+2)\times\{(-1)^k\}; \qquad \Xi_- = \bigcup_{k=1}^{\infty}(2k,2k+2)\times\{(-1)^{k-1}\}. \quad (69)$$

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

Define

$$R_A = [0, 1+A] \times [0, 1+A] \times [0,1] \quad (70)$$

R_A is a fundamental domain for the action of a certain lattice Λ_A. We have

$$\Lambda_A = \begin{bmatrix} 1+A & 1-A & -1 \\ 0 & 1+A & -1 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{Z}^3 \quad (71)$$

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

We define $\mu_+ : \Xi_+ \to R_A$ and $\mu_- : \Xi_- \to R_A$ by the equations

$$\mu_\pm(t,*) = \left(\frac{t-1}{2}, \frac{t+1}{2}, \frac{t}{2}\right) \pm \left(\frac{1}{2}, \frac{1}{2}, 0\right) \mod \Lambda. \quad (72)$$

The maps only depend on the first coordinate. In each case, we mean to map t into $\boldsymbol{R}^3$ 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 the problem with boundary points, as usual with fundamental domains.) However, if $t \not\in 2\boldsymbol{Z}[A]$, this situation does not happen. 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 Ψ be the first return map.

Theorem 10.1 *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 $\Psi(\xi_1) - \xi_1 = \Psi(\xi_2) - \xi_2$.*

10.2 The Walls of the Partitions

In order to make Theorem 10.1 precise, we need to describe the nature of the partitions $(\mathcal{P}_A)_\pm$, and also the rule by which the polygon in the partition determines $\Psi(\xi) - \xi$. 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 $t = -1 + A, A, 1 + A, 2 + A$.

The complements of the union of these planes are the open polyhedra in the partitions.

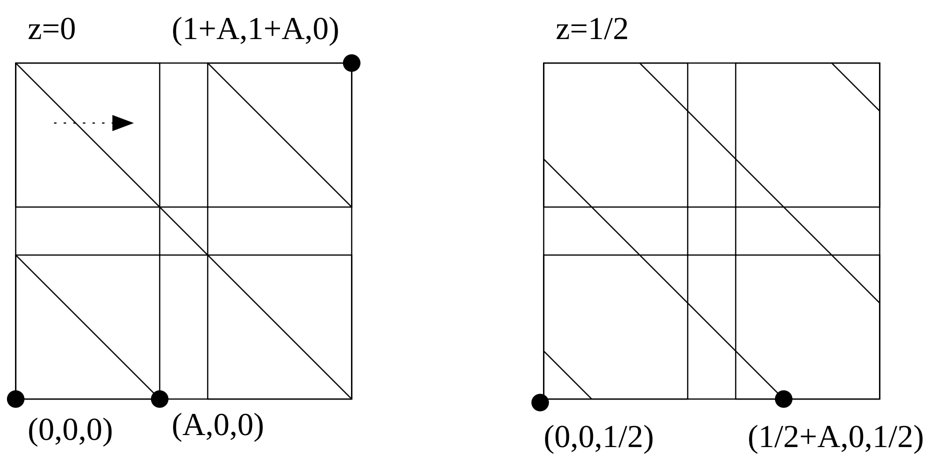

Figure 10.1: Two slices of the partition for $A = 2/3$.

Figure 10.1 shows a picture of two slices of the partition for the parameter $A = 2/3$. We have sliced the picture at $z = 0$ and $z = 1/2$. We have labelled several points just to make the coordinate system more clear. The little arrow in the picture indicate the "motion" the diagonal lines would make were we to increase the z-coordinate and show a kind of movie of the partition. The reader can see this partition for any parameter and slice using Billiard King.

10.3 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

$$R = \bigcup_{A \in (0,1)} R_A \times \{A\} \subset \boldsymbol{R}^4. \tag{73}$$

Each 2-plane family discussed above gives rise to a hyperplane family in $\boldsymbol{R}^4$. These hyperplane families are now all defined over $\boldsymbol{Z}$, because the variable A is just the 4th coordinate of $\boldsymbol{R}^4$ 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 $f : \boldsymbol{R}^4 \to \boldsymbol{R}^2$ given by $f(x, y, z, A) = (z, A)$. This fibration in turn gives a fibration of R over the unit square $B = (0, 1)^2$. Figure 10.1 draws the fiber $f^{-1}(3/2, 1/2)$. The base space B has a partition into 4 regions, as shown in Figure 10.2.

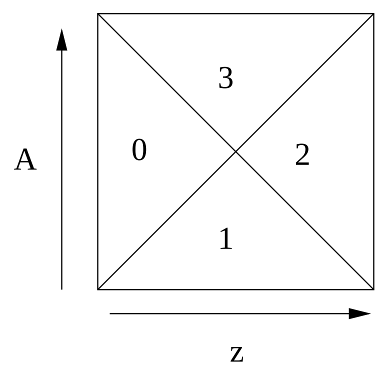

Figure 10.2: The Partition of the Base Space

All the fibers above the same open region in the base space have the same combinatorial structure. Figure 10.3 explains 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

$$\Psi(\xi) - \xi = 2(\epsilon_1^+, \epsilon_2^+, *). \tag{74}$$

The second coordinate, ± 2, is determined by the parity relation in Equation 8. Similarly, we have $(\epsilon_1^-, \epsilon_2^-)$ for $\xi \in \Xi_-$.

Figure 10.3 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^{-}. A light shading indicates a value of $+1$. A dark shading indicates a value of -1. No shading indicates a value of 0.

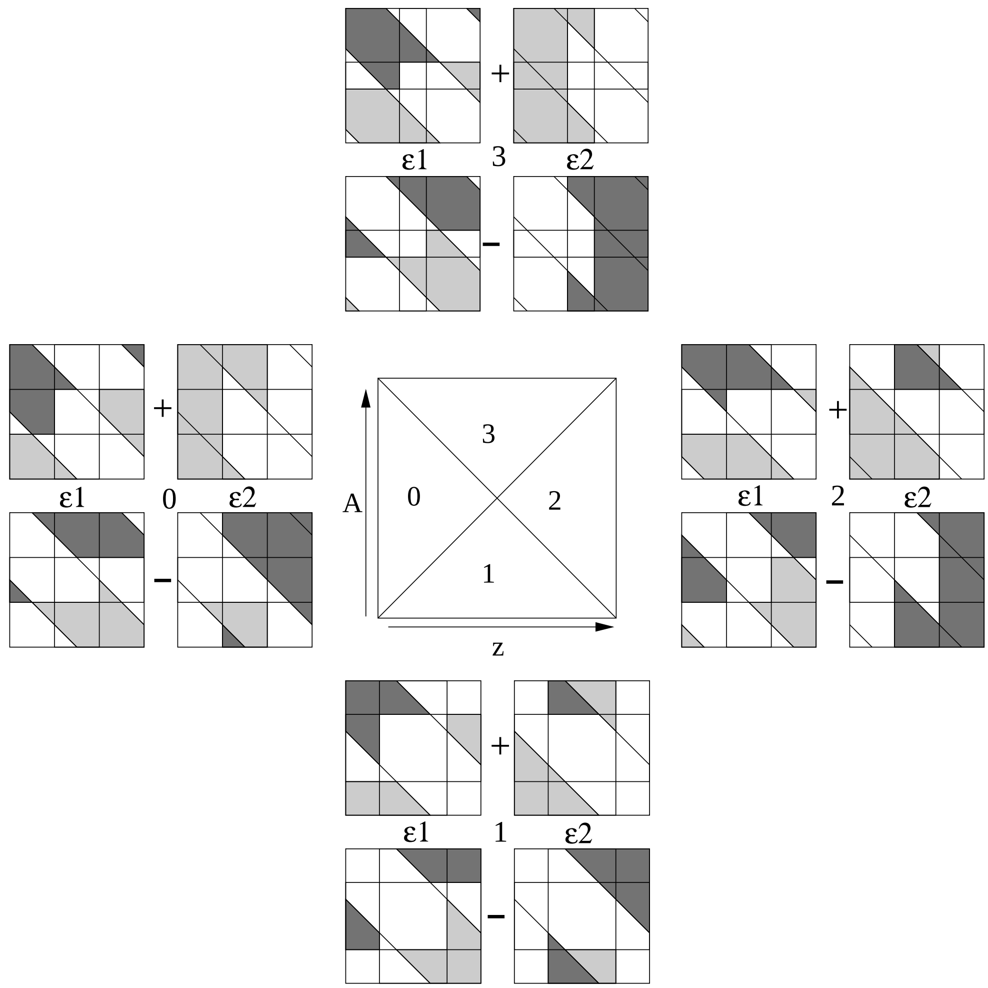

Figure 10.3: The decorated fibers

Given a generic point $\xi \in \Xi_{\pm}$, the image $\mu_{\pm}(\xi)$ lies in some fiber. We then use the coloring scheme to determine $\epsilon_j^{\pm}(\xi)$ for $j = 1, 2$. (See below for examples.) Theorem 10.1, together with the description in this section, constitutes the Master Picture Theorem. In §15 we explain with more traditional formulas how to compute these values. The reader can get a vastly superior understanding of the partition using Billiard King.

10.4 A Typical Example

Here we will explain how the Master Picture Theorem determines the local structure of the arithmetic graph $\Gamma(3/5)$ at the point $(4,2)$. Letting M be the fundamental map associated to $A = 3/5$ (and $\alpha = 1/(2q) = 1/10$).

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

So, $\mu_-(9,-1)$ determines the forwards direction and $\mu_+(9,1)$ determines the backwards direction. (Reflection in the x-axis conjugates Ψ to its inverse.)

We compute

$$\mu_+(9,1) = (\frac{9}{2}, \frac{11}{2}, \frac{9}{2}) \equiv (\frac{1}{10}, \frac{3}{2}, \frac{1}{2}) \bmod \Lambda;$$
$$\mu_-(9,-1) = (\frac{7}{2}, \frac{9}{2}, \frac{9}{2}) \equiv (\frac{7}{10}, \frac{1}{2}, \frac{1}{2}) \bmod \Lambda.$$

(In §15 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 10.3, we find that $(\epsilon_1^+, \epsilon_2^+) = (-1,1)$ and $(\epsilon_1^-, \epsilon_2^-) = (1,0)$. The bottom right of Figure 10 shows the corresponding local picture for the arithmetic graph.

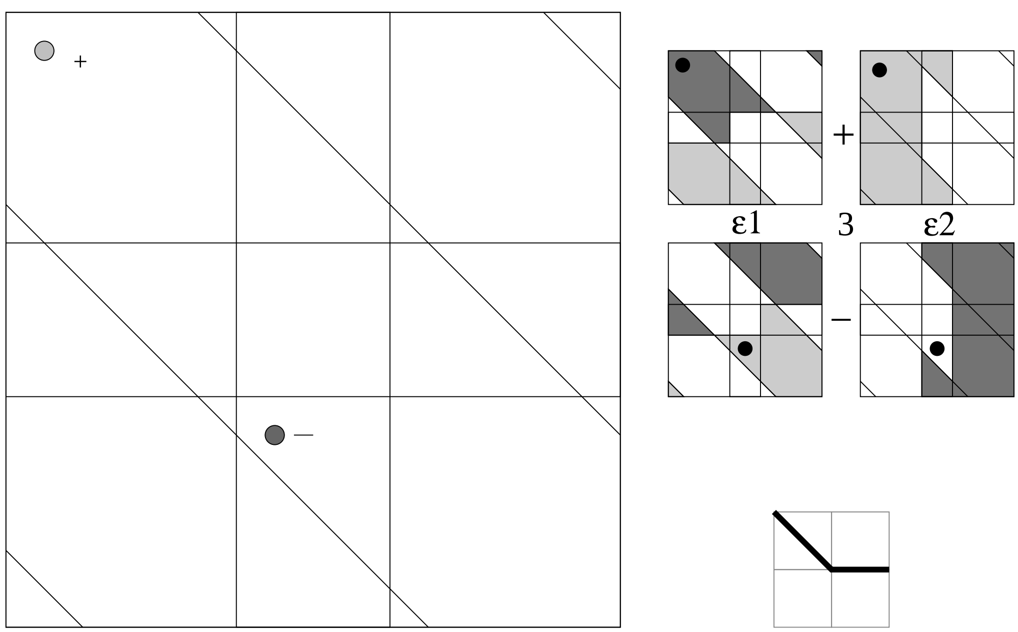

Figure 10.4: Points in the fiber.

10.5 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 these 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.

Our method is to redefine our polygonal regions to include their *lower* edges. A lower edge is an edge first encountered by a line of slope 1. Figure 10.5 shows what we have in mind.

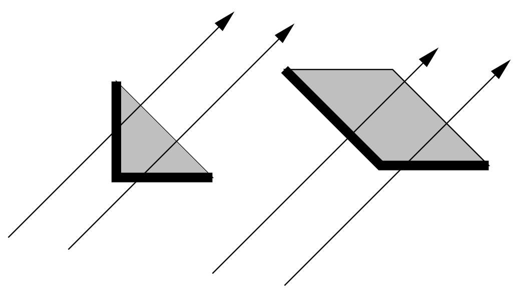

Figure 10.5: Polygons with their lower boundaries included.

We then set $\alpha = 0$ and determine the relevent edges of the arithmetic graph by which *lower borded* polygon contains our points. if it happens that $z \in \{0, A, 1 - A\}$, 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 pictures. 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

$$\mu_+(16, 1) = (8, 9, 8) \equiv (\frac{4}{5}, 1, 0) \text{ mod } \Lambda$$

$$\mu_-(16, -1) = (7, 8, 8) \equiv (0, \frac{7}{5}, 0) \text{ mod } \Lambda.$$

Figure 10.6 draws the relevant fibers. The bottom right of Figure 10.6 shows the local structure of the arithmetic graph. For instance, $(\epsilon_1^+, \epsilon_2^+) = (0, 1)$.

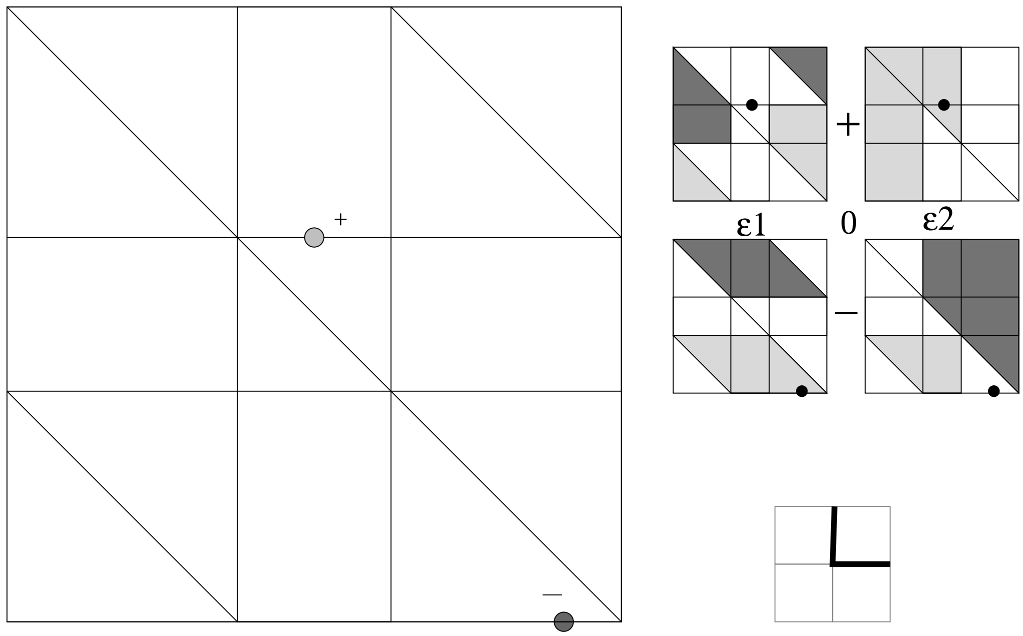

Figure 10.6: Points in the fiber.

The only place where we need to use our special definition of a lower borded 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 our point in the lightly shaded parallelogram, because our point lies in the lower border of this parallelogram.

There is one exception to our construction that requires an explanation. Referring to the lower right fiber, suppose that the bottom point actually was the bottom right vertex, as shown in Figure 10.7. 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 moves the point into the region on the left.

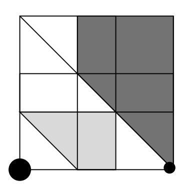

Figure 10.7: An exceptional case.

10.6 The Integral Structure

10.6.1 An Affine Action

We can describe Figure 10.3, and hence the Master Picture Theorem, in a different way. Let **Aff** denote the 4 dimensional affine group. We define a discrete affine group action $\Lambda \subset \textbf{Aff}$ on the infinite slab $\widetilde{R} = \boldsymbol{R}^3 \times (0,1)$. 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. Explicitly, we have

$$\gamma_1 \begin{bmatrix} x \\ y \\ z \\ A \end{bmatrix} = \begin{bmatrix} x+1+A \\ y \\ z \\ A \end{bmatrix}$$

$$\gamma_2 \begin{bmatrix} x \\ y \\ z \\ A \end{bmatrix} = \begin{bmatrix} x+1-A \\ y+1+A \\ z \\ A \end{bmatrix};$$

$$\gamma_3 \begin{bmatrix} x \\ y \\ z \\ A \end{bmatrix} = \begin{bmatrix} x-1 \\ y-1 \\ z+1 \\ A \end{bmatrix}. \tag{75}$$

These are all affine maps of $\boldsymbol{R}^4$. The quotient $\widetilde{R}/\Lambda$ is naturally a fiber bundle over $(0,1)$. Each fiber $(\boldsymbol{R}^3 \times \{A\})/\Lambda$ is isomorphic to $\boldsymbol{R}^3/\Lambda_A$.

The region R, from Equation 73, is a fundamental domain for the action of Λ. Note that R is naturally an *integral polytope*. That is, all the vertices of R have integer coordinates. R has 16 vertices, and they are as follows.

$$(\epsilon_1, \epsilon_2, \epsilon_3, 0); \qquad (2\epsilon_1, 2\epsilon_2, \epsilon_3, 1); \qquad \epsilon_1, \epsilon_2, \epsilon_3 \in \{0,1\}. \tag{76}$$

10.6.2 Integral Polytope Partitions

Inplicit in Figre 10.3 is the statement that the regions R_+ and R_- are partitioned into smaller convex polytopes. The partition is defined by the 4 families of hyperplanes discussed above. An alternate point of view leads to a simpler partition.

For each pair $(\epsilon_1, \epsilon_2) \in \{-1, 0, 1\}$, we let $R_+(\epsilon_1, \epsilon_2)$ denote the closure of the union of regions that assign (ϵ_1, ϵ_2). It turns out that $R(\epsilon_1, \epsilon_2)$ if a finite union of convex integral polytopes. There are 14 such polytopes, and they give an integral partition of R_+. We list these polytopes in §15.4.

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

$$\iota(x, y, z, A) = (1 + A - x, 1 + A - y, 1 - z, A). \tag{77}$$

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

$$R_-(-\epsilon_1, -\epsilon_2) = \iota(R_+(\epsilon_1, \epsilon_2)). \tag{78}$$

Thus, the partition of R_- is a mirror image of the partition of R_+. (See Example 15.5 for an example calculation.)

We use the action of Λ to extend the partitions of R_+ and R_- to two integral polytope tilings of $\widetilde{R}$. (Again, see §15.5 for an example calculation.) These 4 dimensional tilings determine the structure of the special orbits.

10.6.3 Notation

Suppose that $\widehat{\Gamma}$ is an arithmetic graph. Let M be the fundamental map associated to $\widehat{\Gamma}$. We define

$$M_+ = \mu_+ \circ M; \qquad M_- = \mu_- \circ \rho \circ M. \tag{79}$$

Here ρ is reflection in the x-axis. Given a point $p \in \boldsymbol{Z}^2$, the polytope of R_+ containing $M_+(p)$ determines the forward edge of $\widehat{\Gamma}$ incident to p, and the polytope of R_- containing $M_-(p)$ determines the backward edge of $\widehat{\Gamma}$ incident to p. Concretely, we have

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

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

$$s = Am + n + \alpha. \tag{80}$$

As usual, α is the offset value. Note that μ_+ and μ_- only depend on the first coordinate, and this first coordinate is not changed by ρ. The map ρ is present mainly for bookkeeping purposes, because $\rho(\Xi_+) = \Xi_-$, and the domain of $\mu_\pm$ is $\Xi_\pm$.

11 The Pinwheel Lemma

11.1 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 $V_- = -V_+$.

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

$$E(p) := p + nV_+ \in \Sigma. \tag{81}$$

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.

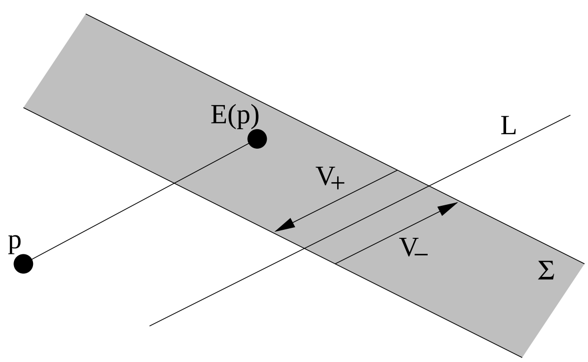

Figure 11.1: A strip map

Figure 11.2 shows 4 strips we associate to our kite. To describe the strips in Figure 11.2 write $(v_1, v_2, v_3)^t$ (a column vector) to signify that $L = \overline{v_2 v_3}$ and $\partial\Sigma = \overline{v_1 v_2} \cup I(\overline{v_1 v_2})$, where I is the order 2 rotation fixing v_3. Here is the data for the strip maps E_1, E_2, E_3, E_4.

$$\begin{bmatrix} (-1,0) \\ (0,1) \\ (0,-1) \end{bmatrix}; \quad \begin{bmatrix} (A,0) \\ (0,-1) \\ (-1,0) \end{bmatrix}; \quad \begin{bmatrix} (0,1) \\ (A,0) \\ (-1,0) \end{bmatrix}; \quad \begin{bmatrix} (-1,0) \\ (0,-1) \\ (0,1) \end{bmatrix}. \tag{82}$$

We set $\Sigma_{j+4} = \Sigma_j$ and $V_{j+4} = -V_j$. Then $\Sigma_{j+4} = \Sigma_j$. The reader can also reconstruct the strips from the information given in Figure 11.2. Figure 11.2 shows the parameter $A = 1/3$, but the formulas in the picture are listed for general A. In particular, the point $(3,0)$ is independent of A. Here is an explicit formula for the vectors involved.

$$V_1 = (0,4); \quad V_2 = (-2,2); \quad V_3 = (-2-2A,0); \quad V_4 = (-2,-2) \quad (83)$$

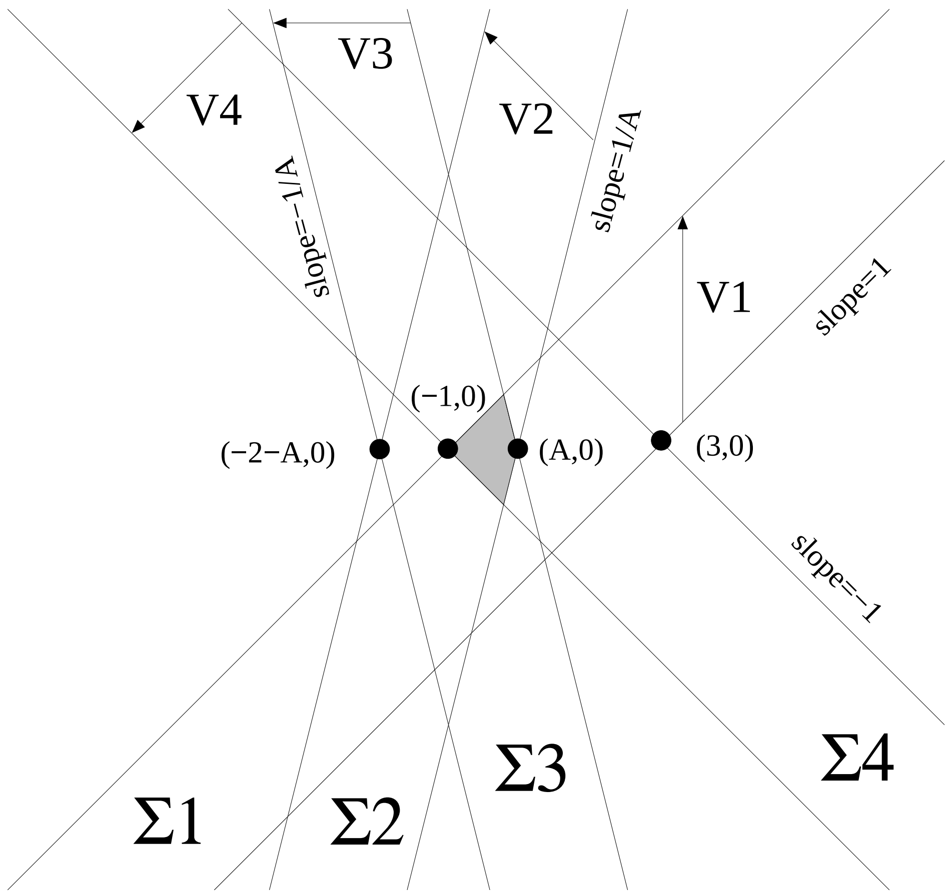

Figure 11.2: The 4 strips for the parameter $A = 1/3$.

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

$$\chi(x, 4n \pm 1) = (x, \pm 1) \quad (84)$$

Lemma 11.1 (Pinwheel) Ψ *exists for any point of* Ξ *having a well-defined outer billiards orbit. In all cases,* $\Psi = \chi \circ (E_8 ... E_1)$.

We call the map in the Pinwheel Lemma the *pinwheel map*. In §15.1 we give concrete formulas for this map.

11.2 Some Corollaries

Before we prove the Pinwheel Lemma, we list two corollaries.

Corollary 11.2 *The parity equation in Equation 8 is true.*

Proof: The Pinwheel Lemma tells us that

$$\Psi(x,1) - (x,1) = 2(\epsilon_1 A + \epsilon_2, \epsilon_3); \qquad (\epsilon_1, \epsilon_2, \epsilon_3) \in \boldsymbol{Z}^2 \times \{-1,0,1\}. \quad (85)$$

Given Equation 83, we see that the sum of the integer coefficients in each vector V_j is divisible by 4. (For instance, $-2-2A$ yields $-2-2=-4$.) Hence $\epsilon_1+\epsilon_2+\epsilon_3$ is even. ♠

The Pinwheel Lemma gives a formula for the quantities in Equation 8.

For $j=0,...,7$ we define points p_{j+1} and integers n_j by the following equations.

$$p_{j+1} = E_{j+1}(p_j) = p_j + n_j V_{j+1}. \quad (86)$$

Given the equations

$$V_1 = (0,4); \quad V_2 = (-2,2); \quad V_3 = (-2-2A, 0); \quad V_4 = (-2,-2) \quad (87)$$

we find that

$$\epsilon_1 = n_2 - n_6; \qquad \epsilon_2 = n_1 + n_2 + n_3 - n_5 - n_6 - n_7; \quad (88)$$

We call $(n_1,...,n_7)$ the *length spectrum* of p_0.

The precise bound in Equation 8 follows from the Master Picture Theorem, but here we give a heuristic explanation. If we define

$$m_1 = n_7; \qquad m_2 = n_6; \qquad m_3 = n_5; \quad (89)$$

then we have

$$\epsilon_1(p) = n_2 - m_2; \qquad \epsilon_2(p) = (n_1 - m_1) + (n_2 - m_2) + (n_3 - m_3). \quad (90)$$

The path with vertices $p_0, p_1, ..., p_7, p_8, \chi(p_8)$ uniformly close to an octagon with dihedral symmetry. See Figure 11.3 below. For this reason, there is a universal bound to $|n_i - m_i|$. This is a heuristic explanation of the bound in Equation 8.

11.3 The Simplest Case

Here we prove the Pinwheel for points of Ξ far from K. Figure 11.3 shows a decomposition of $\boldsymbol{R}^2 - K'$ into 8 regions, $S_0, ..., S_7$. Here K' is a suitably large compact set. Let $V_1, ..., V_4$ we the vectors associated to our special strip maps. We set $V_{4+j} = -V_j$. A calculation shows that

$$x \in S_j; \qquad \Longrightarrow \qquad \psi(x) - x = V_j. \tag{91}$$

One can easily see this using Billiard King or else our interactive guide to the monograph.

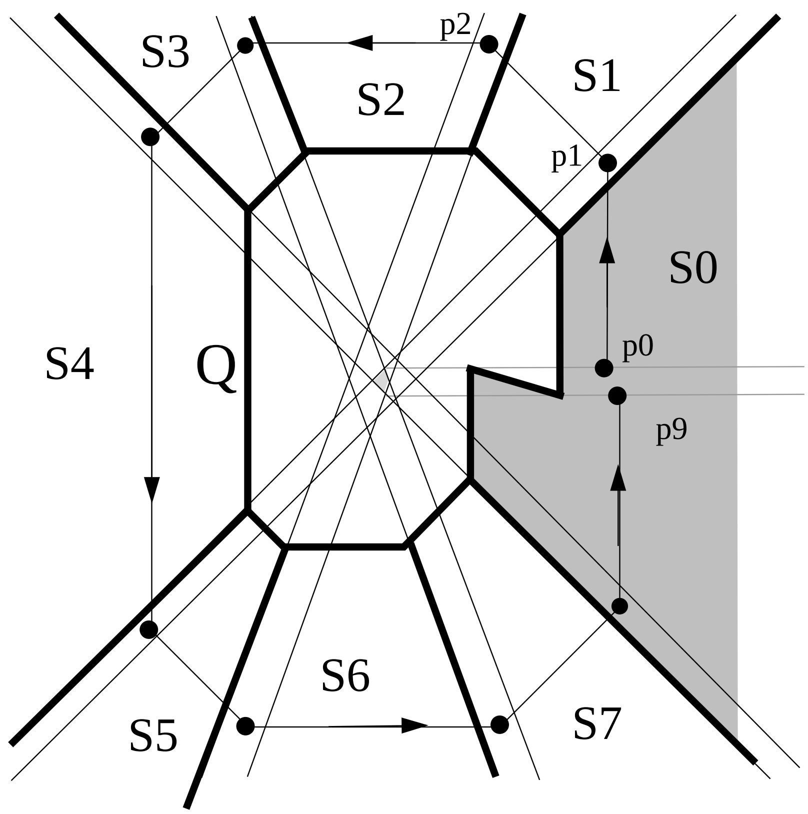

Figure 11.3: The Simplest Sequence of Regions

Equation 91 tells the whole story for points of Ξ far away from K. As above, let $p_{j+1} = E_{j+1}(p_j)$ for $j = 0, ...7$. let $p_9 = \chi(p_8)$. Here we have set $E_{j+4} = E_j$. By induction and Equation 91, p_{j+1} lies in the forward orbit of p_j for each $j = 0, ..., 8$. But then $p_9 = \Psi(p_1) = \chi \circ E_8 ... E_1(p_0)$.

11.4 Discussion of the General Case

As we have just seen, the Pinwheel Lemma is a fairly trivial result for points that are far from the origin. For points near the origin, the Pinwheel Lemma is a surprising and nontrivial result. In fact, it only seems to work because of a lucky accident. The fact that we consider the Pinwheel Lemma to be an accident probably means that we don't yet have a good understanding of what is going on.

Verifying the Pinwheel Lemma for any given parameter A is a finite calculation. We just have to check, on a fine enough mesh of points extending out sufficiently far away from $K(A)$, that the equation in the Pinwheel Lemma holds. The point is that all the maps involved are piecewise isometries for each parameter. We took this approach in [**S**] when we proved the Pinwheel Lemma for $A = \phi^{-3}$.

Using Billiard King, we computed 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 tiny number ϵ is included to make sure that the outer billiards orbit is actually defined for all the points we sample. This calculation does not constitute a proof of anything. However, we think that it serves as a powerful sanity check that the Pinwheel Lemma is correct. We have fairly well carpeted the region of doubt about the Pinwheel Lemma with instances of its truth.

Our proof of the Pinwheel Lemma essentially boils down to finding the replacement equation for Equation 91. We will do this in the section. As the reader will see, the situation in general is much more complicated. There is a lot of information packed into the next section, but all this information is easily seen visually on Billiard King. We have programmed Billiard King so that the reader can see pictures of all the regions involved, as well as their interactions, for essentially any desired parameter.

We think of the material in the next section as something like a written description of a photograph. The written word is probably not the right medium for the proof of the Pinwheel Lemma. To put this in a different way, Billiard King relates to the proof given here much in the same way that an ordinary research paper would relate to one that was written in crayon.

11.5 A Partition of the Plane

Let $\psi = \psi_A$ be the square of the outer billiards map relative to $K(A)$. For each $x \in \boldsymbol{R}^2 - K$ on which ψ is defined, there is a vector v_x such that

$$\psi(x) - x = v_x.$$

This vector is twice the difference between 2 vertices of K, and therefore can take on 12 possible values. It turns out that 10 of these values occur. We call these vectors V_j, with $j = 1, 2, 3, 4, 4^\sharp, 5, 6^\flat, 6, 7, 8$. With this ordering, the argument of V_j increases monotonically with j. Compare Figure 11.4. For each of our vectors V, there is an open region $R \subset \boldsymbol{R}^2 - K$ such that $x \in R$ if and only if $\psi(x) - x = V$. The regions $R_1, ..., R_8$ are unbounded. The two regions $R_4^\sharp$ and $R_6^\flat$ are bounded.

One can find the entire partition by extending the sides of K in one direction, in a pinwheel fashion, and then pulling back these rays by the outer billiards map. To describe the regions, we use the notation $\overrightarrow{q_1}, p_1, ..., p_k, \overrightarrow{q_2}$ to indicate that

- The two unbounded edges are $\{p_1 + tq_1 | \; t \geq 0\}$ and $\{p_k + tq_2 | \; t \geq 0\}$.
- $p_2, ..., p_{k-1}$ are any additional intermediate vertices.

To improve the typesetting on our list, we set $\lambda = (A-1)^{-1}$. Figure 10.3 shows the picture for $A = 1/3$. The reader can see any parameter using Billiard King.

$V_1 = (0, 4)$.	$R_1 : \overrightarrow{(1,-1)}, (1,-2), \overrightarrow{(1,1)}$.
$V_2 = (-2, 2)$.	$R_2 : \overrightarrow{(1,1)}, (1,-2), (0,-1), \overrightarrow{(A,1)}$.
$V_3 = (-2-2A, 0)$	$R_3 : \overrightarrow{(A,1)}, (2A,1), \lambda(2A^2, -1-A), \overrightarrow{(-A,1)}$.
$V_4 = (-2, -2)$	$R_4 : \overrightarrow{(-A,1)}, \lambda(2A, A-3), \overrightarrow{(-1,1)}$.
$V_{4^\sharp} = (-2A, -2)$	$R_{4^\sharp} : (A, 0), (2A, 1), \lambda(2A^2, -1-A))$
$V_5 = (0, -4)$	$R_5 : \overrightarrow{(-1,1)}, \lambda(2A, A-3), (-A, 2), \lambda(2A, 3A-1), \overrightarrow{(-1,-1)}$
$V_{6^\flat} = (2A, -2)$	$R_{6^\flat} : (0, 1), (-A, 2), \lambda(2A, 3A-1)$
$V_6 = (2, -2)$	$R_6 : \overrightarrow{(-1,-1)}, \lambda(2, A+1), \overrightarrow{(-A,-1)}$
$V_7 = (2+2A, 0)$	$R_7 : \overrightarrow{(-A,-1)}, \lambda(2, A+1), (-2,-1), \overrightarrow{(A,-1)}$
$V_8 = (2, 2)$	$R_8 : \overrightarrow{(A,-1)}, (-2,-1), (-1,0), \overrightarrow{(1,-1)}$.

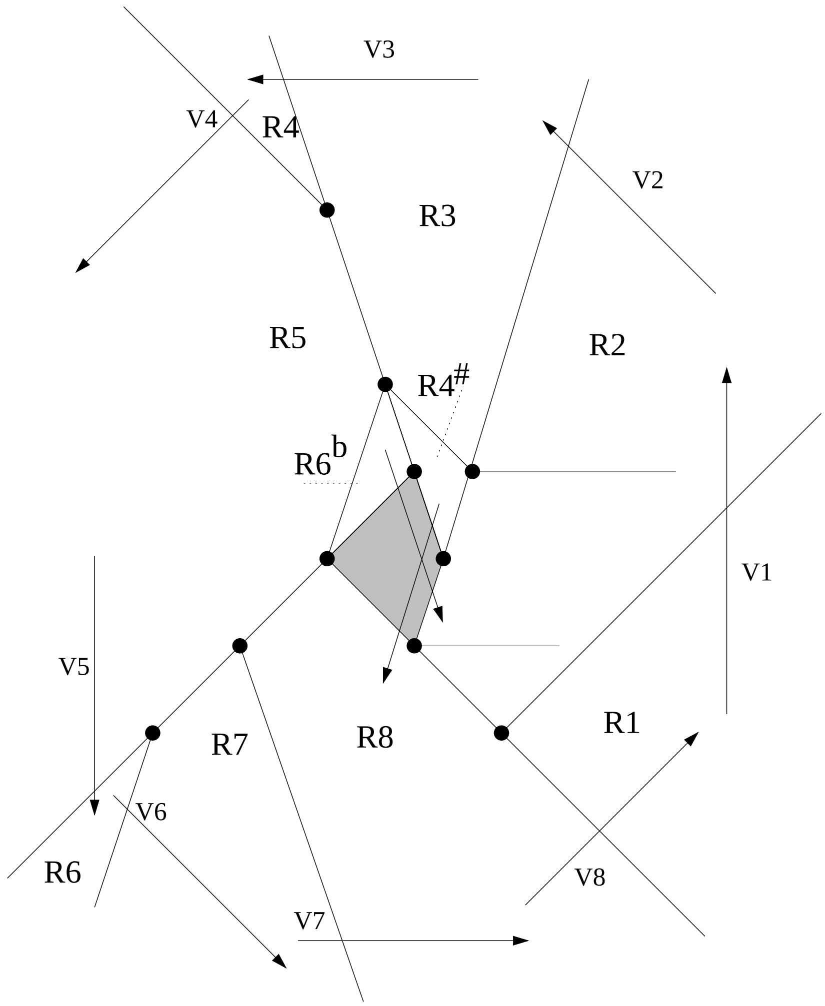

Figure 11.4: The Partition for $A = 1/3$.

It is convenient to set

$$\widehat{R}_a = R_a + V_a = \{p + V_a|\ p \in R_a\}. \tag{92}$$

One symmetry of the partition is that reflection in the x-axis interchanges $\widehat{R}_a$ with R_{10-a}, for all values of a. (To make this work, we set $R_9 = R_1$, and use the convention $4^\sharp + 6^\flat = 10$.)

We are interested in *transitions* between one region R_a, and another region R_b. If $\widehat{R}_a \cap R_b \neq \emptyset$ for some parameter A it means that there is some $p \in R_a$ such that $\psi_A(p) \in R_b$. (We think of our regions as being open.) We create a *transition matrix* using the following rules.

- A 0 in the (ab)th spot indicates that $\widehat{R}_a \cap R_b = \emptyset$ for all $A \in (0,1)$.
- A 1 in the (ab)th spot indicates that $\widehat{R}_a \cap R_b \neq \emptyset$ for all $A \in (0,1)$.
- A t^+ in the (ab)th spot indicates that $R_a \cap R_b \neq \emptyset$ iff $A \in (t,1)$.
- A t^- in the (ab)th spot indicates that $R_a \cap R_b \neq \emptyset$ iff $A \in (0,t)$.

$$
\begin{array}{ccccccccccc}
 & R_1 & R_2 & R_3 & R_4 & R_{4\sharp} & R_5 & R_{6\flat} & R_6 & R_7 & R_8 \\
\widehat{R}_1 & 1 & 1 & (\frac{1}{3})^+ & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\widehat{R}_2 & 0 & 1 & 1 & (\frac{1}{3})^- & 1 & 1 & 1 & 0 & 0 & 0 \\
\widehat{R}_3 & 0 & 0 & 1 & 1 & (\frac{1}{2})^+ & 1 & 0 & 0 & 0 & 0 \\
\widehat{R}_4 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\widehat{R}_{4\sharp} & 0 & 0 & 0 & 0 & 0 & (\frac{1}{3})^+ & (\frac{1}{3})^+ & 0 & 0 & 1 \\
\widehat{R}_5 & 0 & 0 & 0 & 0 & 0 & 1 & (\frac{1}{3})^+ & 1 & 1 & 1 \\
\widehat{R}_{6\flat} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & (\frac{1}{2})^+ & 1 \\
\widehat{R}_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & (\frac{1}{3})^- \\
\widehat{R}_7 & (\frac{1}{3})^+ & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\widehat{R}_8 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0
\end{array}
\tag{93}
$$

We have programmed Billiard King so that the interested reader can see each of these relations at a single glance. Alternatively, they can easily be established using routine linear algebra. For example, interpreting $\widehat{R}_3$ and R_2 as projectivizations of open convex cones $\widehat{C}_3$ and C_2 in $\boldsymbol{R}^3$, we easily verifies that the vector $(-1, A, -2-A)$ has positive dot product with all vectors in $\widehat{C}_3$ and negative dot product with all vectors in C_2. Hence $R_2 \cap \widehat{R}_3 = \emptyset$.

We can relate all the nonempty intersections to our strips. As with the list of intersections, everything can be seen at a glance using Billiard King, or else proved using elementary linear algebra. First we list the intersections that comprise the complements of the strips.

- $\widehat{R}_2 \cap R_2$ and $\widehat{R}_6 \cap R_6$ are the components of $\boldsymbol{R}^2 - (\Sigma_1 \cup \Sigma_2)$.
- $\widehat{R}_4 \cap R_4$ and $\widehat{R}_8 \cap R_8$ are the components of $\boldsymbol{R}^2 - (\Sigma_3 \cup \Sigma_4)$.
- $\widehat{R}_3 \cap (R_3 \cup R_{4\sharp})$ and $(\widehat{R}_{6\flat} \cup \widehat{R}_7) \cap R_7$ are the components of $\boldsymbol{R}^2 - (\Sigma_2 \cup \Sigma_3)$.
- $\widehat{R}_1 \cap R_1$ and $(\widehat{R}_{4\sharp} \cup \widehat{R}_5) \cap (R_5 \cup R_{6\flat})$ are the components of $\boldsymbol{R}^2 - (\Sigma_1 \cup \Sigma_4)$.

Now we list the intersections that are contained in single strips. To make our typesetting nicer, we use the term *u-component* to denote an unbounded connected component. We use the term b-component to denote a bounded connected component.

- $\widehat{R}_1 \cap R_2$ and $\widehat{R}_5 \cap R_6$ are the two u-components of $\Sigma_1 - (\Sigma_2 \cup \Sigma_4)$.
- $\widehat{R}_8 \cap R_1$ and $\widehat{R}_4 \cap R_5$ are the two u-components of $\Sigma_4 - (\Sigma_1 \cup \Sigma_3)$.
- $\widehat{R}_3 \cap R_4$ and $(\widehat{R}_{6^\flat} \cup \widehat{R}_7) \cap R_8$ are the two u-components of $\Sigma_3 - (\Sigma_2 \cup \Sigma_4)$.
- $\widehat{R}_6 \cap R_7$ and $\widehat{R}_2 \cap (R_3 \cup R_{4^\sharp})$ are the two u-components of $\Sigma_2 - (\Sigma_1 \cup \Sigma_3)$.
- $\widehat{R}_{6^\flat} \cap R_7$ is contained in the b-component of $\Sigma_1 - (\Sigma_2 \cup \Sigma_3)$.
- $\widehat{R}_3 \cap R_{4^\sharp}$ is contained in the b-component of $\Sigma_4 - (\Sigma_2 \cup \Sigma_3)$.
- $\widehat{R}_{4^\sharp} \cap (R_5 \cup R_{6^\flat})$ is contained in the b-component of $\Sigma_3 - (\Sigma_1 \cup \Sigma_4)$.
- $(\widehat{R}_{4^\sharp} \cup \widehat{R}_5) \cap R_{6^\flat}$ is contained in the b-component of $\Sigma_2 - (\Sigma_1 \cup \Sigma_4)$.

Now we list the intersections of regions that are contained in double intersections of strips. In this case, all the components are bounded: Any two strips intersect in a bounded region of the plane.

- $\widehat{R}_1 \cap R_3$ and $\widehat{R}_5 \cap R_7$ are the components of $(\Sigma_1 \cap \Sigma_2) - (\Sigma_3 \cup \Sigma_4)$.
- $\widehat{R}_7 \cap R_1$ and $\widehat{R}_3 \cap R_5$ are the components of $(\Sigma_3 \cap \Sigma_4) - (\Sigma_1 \cup \Sigma_2)$.
- $\widehat{R}_2 \cap R_4$ and $\widehat{R}_6 \cap R_8$ are bounded components of $(\Sigma_2 \cap \Sigma_3) - (\Sigma_1 \cup \Sigma_4)$.
- $\widehat{R}_8 \cap R_2 = (\Sigma_1 \cap \Sigma_4) - (\Sigma_2 \cup \Sigma_3)$.

Now we list all the intersections of regions that are contained in triple intersections of strips.

- $\widehat{R}_2 \cap (R_5 \cup R_{6^\flat}) = \Sigma_2 \cap \Sigma_3 \cap \Sigma_4 - \Sigma_1$.
- $\widehat{R}_8 \cap (R_3 \cup R_{4^\sharp}) = \Sigma_1 \cap \Sigma_2 \cap \Sigma_4 - \Sigma_3$.
- $(\widehat{R}_{4^\sharp} \cup \widehat{R}_5) \cap R_8 = \Sigma_1 \cap \Sigma_2 \cap \Sigma_3 - \Sigma_4$.
- $(\widehat{R}_{6^\flat} \cup \widehat{R}_7) \cap R_2 = \Sigma_1 \cap \Sigma_3 \cap \Sigma_4 - \Sigma_2$.

Here we list a bit more information about the two regions $R_{4^\sharp}$ and $R_{6^\flat}$ some of the information is redundant, but it is useful to have it all in one place.

- $R_{4^\sharp} \subset \Sigma_4 - \Sigma_3$.
- $R_{4^\sharp} + V_3 = \Sigma_3 - (\Sigma_2 \cup \Sigma_4)$.
- $R_{6^\flat} \subset \Sigma_2 - \Sigma_1$.
- $R_{6^\flat} + V_5 \subset \Sigma_1 - \Sigma_2$.
- $R_{6^\flat} + V_5 - V_6 = \Sigma_2 - (\Sigma_1 \cup \Sigma_3)$.

Finally, we mention two crucial relations between our various vectors:

- $V_3 - V_4 + V_5 = V_{4^\sharp}$.
- $V_5 - V_6 + V_7 = V_{6^\flat}$.

These two relations are responsible for the lucky cancellation that makes the Pinwheel Lemma hold near the kite.

We will change our notation slightly from the simplest case considered above. Given any point $z_1 \in \Xi$, we can associate the *sequence of regions*

$$R_{a_1} \to \ldots \to R_{a_k} \tag{94}$$

through which the forwards orbit of z_1 transitions until it returns as $\Psi(z_1)$. The simplest possible sequence is the one where $a_j = j$ for $j = 1, \ldots, 9$. See Figure 11.2. We already analyzed this case in §11.2. We let z_j denote the first point in the forward orbit of z_1 that lies in R_{a_j}.

To prove the Pinwheel Lemma in general, we need to analyze all allowable sequences and see that the equation in the Pinwheel Lemma always holds. We will break the set of all sequences into three types, and then analyze the types one at a time. Here are the types.

1. Sequences that do not involve the indices $4^\sharp$ or $6^\flat$.
2. Sequences that involve $4^\sharp$ but not $6^\flat$.
3. Sequences that involve $6^\flat$.

11.6 No Sharps or Flats

Lemma 11.3 *If $j < k$ then $\widehat{R}_j \cap R_k \subset \Sigma_j \cap ... \cap \Sigma_{k-1}$.*

Proof: This is a corollary of the the intersections listed above. ♠

Suppose by induction we have shown that

$$z_j = E_{a_j - 1} E_{a_j - 2} ... E_1(z_1). \tag{95}$$

By construction and Lemma 11.3,

$$z_{j+1} = E_{a_j}(z_j) \in \widehat{R}_{a_j} \cap R_{a_{j+1}} \subset \Sigma_{a_j} \cap ... \cap \Sigma_{a_{j+1}-1}.$$

Therefore, $E_{a_j}, ..., E_{a_{j+1}-1}$ all act trivially on z_{j+1}, forcing

$$z_{j+1} = E_{a_{j+1}-1} E_{a_{j+1}-2} ... E_1(z_1).$$

Hence, Equation 95 holds true for all indices j.

By the Intersection Lemma, we eventually reach either a point z_9 or z_{10}. (That is, we wrap all the way around and return either to $R_9 = R_1$ or else to $R_{10} = R_2$.) We will consider these two cases one at a time.

Case 1: If we reach $z_9 = (x_9, y_9) \in R_9$ then we have

$$z_9 = E_8 ... E_1(z_1); \qquad x_9 > 0; \qquad y_9 \leq 1. \tag{96}$$

From this we get that $\Psi(z_1) = \chi \circ (E_4 ... E_1)^2(z_1)$, as desired. The last inequality in Equation 96 requires explanation. By the Intersection Lemma, the point preceding z_9 on our list must lie in R_a for some $a \in \{6^\flat, 6, 7, 8\}$. However, the distance between any point on $\boldsymbol{R}_+ \times \{3, 5, 7...\}$ to any point in R_a exceeds the length of vector V_a.

Case 2: If we arrive at $z_{10} = (x_{10}, y_{10})$, then the Intersection Lemma tells us that the point preceding z_{10} lies in V_a for $a = \{6^\flat, 6, 7, 8\}$ and $z_{10} \in \Sigma_9$. Hence $E_9(z_{10}) = z_{10}$. That is

$$z_{10} = E_8 ... E_1(z_1); \qquad x_{10} > 0; \qquad y_{10} < 3.$$

The last inequality works just as in Case 1. All points in R_{10} have y-coordinate at least -2. Hence $y_{10} = \pm 1$. Hence $\chi(z_{10}) = z_{10}$. Putting everything together gives the same result as Case 1.

11.7 Dealing with Four Sharp

In this section we will deal with orbits whose associated sequence has a $4^\sharp$ in it, but not a $6^\flat$. The following result is an immediate consequence the intersections discussed above.

Lemma 11.4 *The following holds for all parameters.*

$\widehat{R}_{4^\sharp} \cap R_{4^\sharp} = \emptyset;\quad R_{4^\sharp} \subset \Sigma_4 - \Sigma_3;\quad R_{4^\sharp} + V_3 \in \Sigma_3 - \Sigma_4;\quad \widehat{R}_{4^\sharp} \cap R_8 \subset \Sigma_1 \cap \Sigma_2 \cap \Sigma_3$

Let z be the first point in the forward orbit of z_1 such that $z \in R_{4^\sharp}$. Using Lemma 11.3 and the same analysis as in the previous section, we get

$$\exists n \in \boldsymbol{N} \cup \{0\} \qquad z = E_2E_1(z_1) + nV_3, \tag{97}$$

From Lemma 11.3 and Item 1 of Lemma 11.4, the next point in the orbit is

$$w = z + V_{4^\sharp} \in R_5 \cup R_8. \tag{98}$$

Items 2 and 3 of Lemma 11.4 give

$$E_3E_2E_1(z_1) = E_3(z) = z + V_3; \qquad E_4E_3(z) = z + V_3 - V_4.$$

Figure 11.5 shows what is going on. Since $V_3 - V_4 + V_5 = V_{4^\sharp}$,

$$w = z + V_{4^\sharp} = z + V_3 - V_4 + V_5 = E_4E_3(z) + V_5 = E_4E_3E_2E_1(z_1) + V_5. \tag{99}$$

The rest of the analysis is as in the previous section. We use Item 4 of Lemma 11.4 as an *addendum* to Lemma 11.3 in case $w \in R_8$.

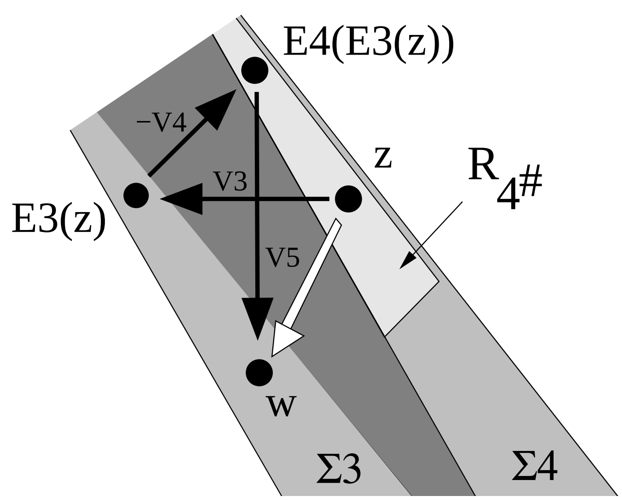

Figure 11.5: The orbit near $R_{4^\sharp}$.

11.8 Dealing with Six Flat

Here is another immediate consequence of the intersections listed above.

Lemma 11.5 *The following is true for all parameters.*

$$V_{6^\flat} \subset \Sigma_2 - \Sigma_1; \quad V_{6^\flat} + R_5 \subset \Sigma_1 - \Sigma_2;$$

$$\widehat{R}_{6^\flat} \cap R_2 \subset \Sigma_3 \cap \Sigma_4 \cap \Sigma_1; \quad \widehat{R}_2 \cap R_{6^\flat} \subset \Sigma_2 \cap \Sigma_3 \cap \Sigma_4;$$

Let z be the first point in the forwards orbit of z_1 such that $z \in R_{6^\flat}$ and let $w = \psi(z)$. The same arguments as in the previous section give

$$z = E_4E_3E_2E_1(z_1) + nV_5; \qquad w = z + V_{6^\flat} \in R_7 \cup R_2. \tag{100}$$

Here $n \in \boldsymbol{N} \cup \{0\}$. (The possibility of $w \in R_{6^\flat}$ is ruled out by Item 1 of Lemma 11.4 and the reflection symmetry.) Items 2 and 3 of Lemma 11.5 give

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

Figure 11.6 shows what is going on. Since $V_5 - V_6 + V_7 = V_{6^\flat}$,

$$w = E_6E_5E_4E_3E_2E_1(z) + V_7. \tag{101}$$

The rest of the analysis is as in the previous cases. We use Item 3 of Lemma 11.5 as an *addendum* to Lemma 11.3 in case $w \in R_2$.

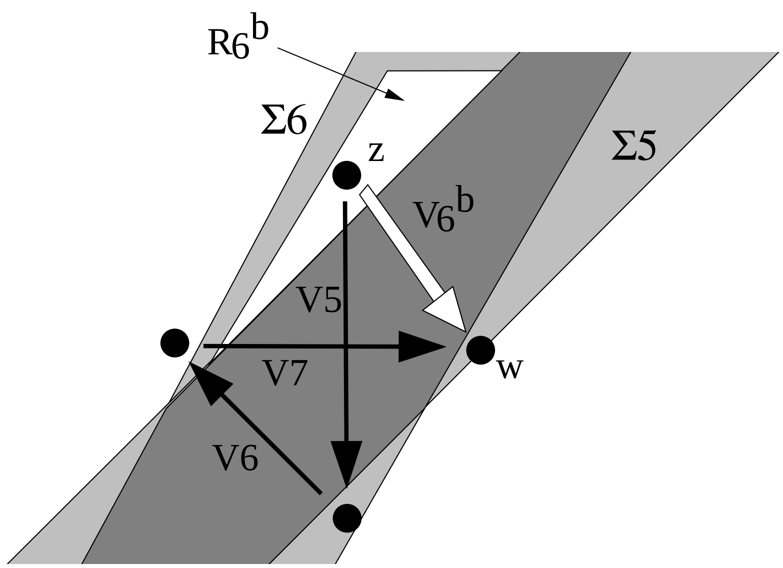

Figure 11.6: The orbit near $R_{6^\flat}$.

12 The Torus Lemma

12.1 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 $T^4 = \widetilde{R}/\Lambda$, the 4 dimensional quotient discussed in §10.6. Topologically, T^4 is the product of a 3-torus with $(0,1)$. Let $(\mu_+)_A$ denote the map μ_+ as defined for the parameter A. We now define $\mu_+ : \Xi_+ \times (0,1) \to T^4$ by the obvious formula $\mu_+(p,A) = ((\mu_+)_A(p), A)$. We are just stacking all these maps together.

The Pinwheel Lemma tells us that $\Psi(p) = \chi \circ E_8...E_1(p)$ whenever both maps are defined. This map involves the sequence $\Sigma_1, ..., \Sigma_8$ of strips. We are taking indices mod 4 so that $\Sigma_{j+4} = \Sigma_j$ and $E_{4+j} = E_j$. Let $p \in \Xi_+$. We set $p_0 = p$ and indctively define

$$p_j = E_j(p_{j-1}) \in \Sigma_j. \tag{102}$$

We also define

$$\theta(p) = \min \theta_j(p); \qquad \theta_j(p) = \text{distance}(p_j, \partial\Sigma_j). \tag{103}$$

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 12.1 (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 $(\epsilon_1(q^*), \epsilon_2(q^*)) = (\epsilon_1(p), \epsilon_2(p))$.*

Remark: My proof of the Torus Lemma owes a big intellectual debt to many sources. I discovered the Torus Lemma experimentally, but I got some inspiration for its proof by reading [**T2**], an account of unpublished work by Chris Culter about the existence of periodic orbits for polygonal outer billiards. Culter's proof is closely related to ideas in [**K**]. The paper [**GS**] implicitly has some of these same ideas, though they are treated from a different point of view. If all these written sources aren't enough, I was also influenced by some conversations with John Smillie.

12.2 Input from the Torus Map

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

$$\widehat{\lambda}_j = \lambda_0 \times ... \times \lambda_j; \qquad \lambda_j = \frac{\text{Area}(\Sigma_{j-1} \cap \Sigma_j)}{\text{Area}(\Sigma_j \cap \Sigma_{j+1})}; \qquad j = 1, ..., 7. \tag{104}$$

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

$$\mu_+(q) - \mu_+(p) = (t, t, t) \mod \Lambda; \qquad t = \frac{y - x}{2}. \tag{105}$$

Lemma 12.2 *If* $\text{dist}(\mu_+(x), \mu_+(y)) < \delta$ *in* T^3*, then there is an integer* I_k *such that* $t\widehat{\lambda}_k$ *is within* ϵ *of* I_k *for all* k*,*

Proof: We compute

$$\text{Area}(\Sigma_0 \cap \Sigma_1) = 8; \qquad \text{Area}(\Sigma_1 \cap \Sigma_2) = \frac{8 + 8A}{1 - A};$$

$$\text{Area}(\Sigma_2 \cap \Sigma_3) = \frac{2(1 + A)^2}{A}; \qquad \text{Area}(\Sigma_3 \cap \Sigma_4) = \frac{8 + 8A}{1 - A}. \tag{106}$$

This leads to

$$\widehat{\lambda}_0 = \widehat{\lambda}_4 = 1; \quad \widehat{\lambda}_1 = \widehat{\lambda}_3 = \widehat{\lambda}_5 = \widehat{\lambda}_7 = \frac{1 - A}{1 + A}; \quad \widehat{\lambda}_2 = \widehat{\lambda}_6 = \frac{4A}{(1 + A)^2}. \tag{107}$$

The matrix

$$H = \begin{bmatrix} \frac{1}{1+A} & \frac{A-1}{(1+A)^2} & \frac{2A}{(1+A)^2} \\ 0 & \frac{1}{1+A} & \frac{1}{1+A} \\ 0 & 0 & 1 \end{bmatrix} \tag{108}$$

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 $\boldsymbol{Z}^3$. We compute

$$H(t, t, t) = \left(\frac{4A}{(1 + A)^2}, \frac{2}{1 + A}, 1 \right) t = (\widehat{\lambda}_2, \widehat{\lambda}_1 - 1, 1) t. \tag{109}$$

Equations 107 and 109 now finish the proof. ♠

12.3 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

$$p_2 - p_1 = nV_2; \qquad \alpha = \frac{\text{area}(B)}{\text{area}(S_1 \cap S_2)}; \qquad \sigma_j = \frac{\|p_j - p_j'\|}{\|V_2\|} \tag{110}$$

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

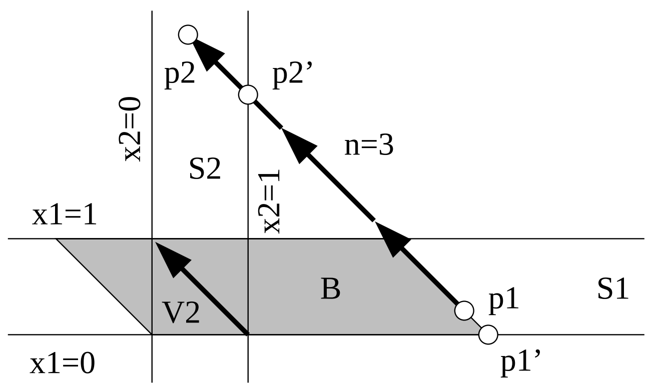

Figure 12.1: Strips and associated objects

Figure 12.1 shows what we call the *standard pair* of strips, where Σ_j is the strip bounded by the lines $x_j = 0$ and $x_j = 1$. To get a better picture of the quantities we have defined, we consider them on the standard pair. We have a

$$\alpha = p_{11}+p_{12} = p_{21}+p_{22}; \quad \sigma_1 = p_{12}; \quad \sigma_2 = 1-p_{22}; \quad n = \text{floor}(p_{11}). \tag{111}$$

Here p_{ij} is the jth coordinate of p_i. These equations lead to the following affine invariant relations.

$$n = \text{floor}(\alpha - \sigma_1); \qquad \sigma_2 = 1 - [\alpha_1 - \sigma_1] \tag{112}$$

Here $[x]$ denotes the fractional part of x. Again, the relations in Equation 112 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 12.3 *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 only depends on ϵ and the distance from $\sigma(p_1)$ and $\sigma(p_2)$ to $\{0, 1\}$.*

Proof: If δ is small enough then $[\alpha(p) - \sigma(p_1)]$ and $[\alpha(q) - \sigma(q_1)]$ are very close, and relatively far from 0 or 1. Equation 112 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 112. ♠

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 $p_3 = E_3(p_2)$. Define,

$$\alpha_j = \alpha(S_j, S_{j+1}, V_{j+1}, p_j, p_{j+1}); \qquad j = 1, 2; \qquad \lambda = \frac{\text{Area}(S_1 \cap S_2)}{\text{Area}(S_2 \cap S_3)}. \tag{113}$$

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

Lemma 12.4 *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, as above, so that Equation 111 holds. Then

$$p_2 = (1 - \sigma_2, \alpha_1 + \sigma_2 - 1). \tag{114}$$

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

$$T(x, y) = \begin{pmatrix} a & \lambda \\ -1 & 0 \end{pmatrix} (x, y) + (b, 1) = (ax + b + \lambda y, 1 - x). \tag{115}$$

Here a and b are constants depending on $S_2 \cap S_3$. Setting $q = T(p_2)$, Equation 111 gives $\alpha = q_1 + q_2$. Hence

$$\alpha_2 = a(1 - \sigma_2) + b + \lambda(\alpha_1 + \sigma_2 - 1) + \sigma_2 = \lambda\alpha_1 + C\sigma_2 + D. \tag{116}$$

This completes the proof. ♠

12.4 Single Parameter Proof

We are still working under the assumption, in the Torus Lemma, that $A = A^*$. Our main argument relies on the Equation 88, 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.

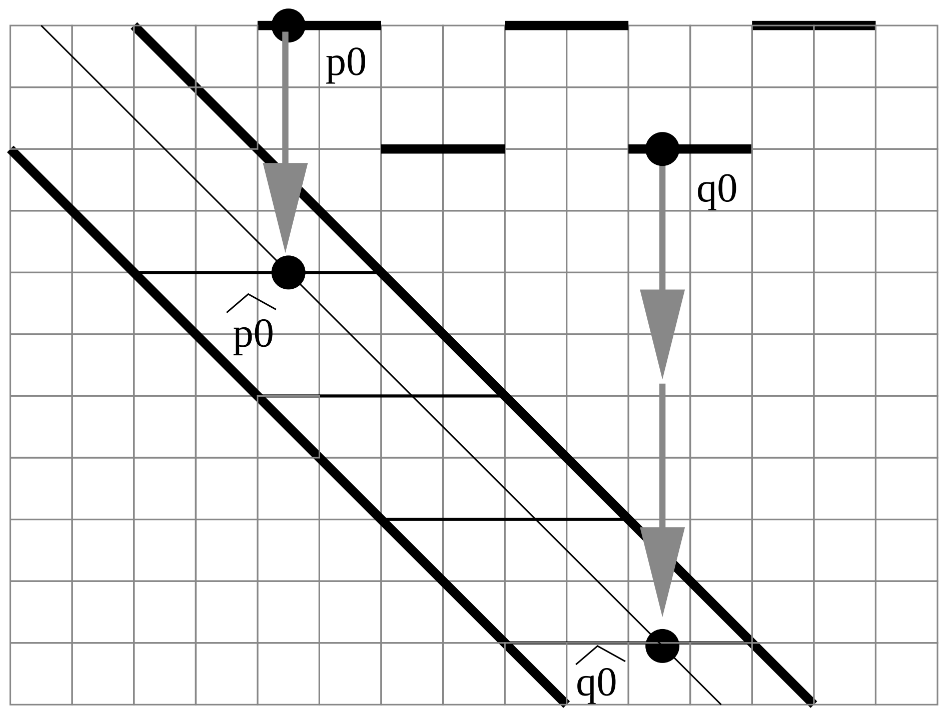

Figure 12.2: The points $\widehat{p}_0$ and $\widehat{q}_0$.

We would like to apply Lemmas 12.2, 12.3, and 12.4 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

$$\widehat{p}_0 = E_0(p_0); \qquad \widehat{q}_0 = E_0(q_0). \tag{117}$$

We have $\widehat{p}_0, \widehat{q}_0 \in \Sigma_0$. Let t be the near-integer from Lemma 12.2. Looking at Figure 12.4, we see that $|\sigma(\widehat{q}_0) - \sigma(\widehat{p}_0)|$ tends to 0 as η tends to 0.

We define

$$\alpha_k(p) = \alpha(\Sigma_k, \Sigma_{k+1}, V_{k+1}, p_k, p_{k+1}) \tag{118}$$

It is also convenient to write

$$\sigma_k(p) = \sigma(p_k); \qquad \Delta\sigma_k = \sigma_k(q) - \sigma_k(p). \tag{119}$$

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.

Lemma 12.5 *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 12.2, 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 us $\alpha_0(q) - \alpha_0(p) = t$. Applying Lemma 12.4 inductively, we find that

$$\alpha_k = \alpha_0\widehat{\lambda}_k + \sum_{i=1}^{k} \xi_i \sigma_i + C_k. \tag{120}$$

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

$$\alpha_k(q) - \alpha_k(p) = t\widehat{\lambda}_k + \sum_{i=1}^{k} \xi_i \Delta\sigma_i; \qquad k = 1, ..., 7 \tag{121}$$

By Lemma 12.2, the term $t\lambda_k$ is near an integer for all k. By Lemma 12.3 and induction, the remaining terms on the right hand side are near 0. This lemma now follows from Lemma 12.3. ♠

Combining our last result with Equation 107, we see that

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

$$n_2(q) - n_2(p) = n_6(q) - n_6(p). \tag{122}$$

once η is small enough. Given the dependence of constants in Lemma 12.3, the necessary bound on η only depends on $\min(A, 1-A)$ and $\theta(p)$. Equation 88 now tells us that $\epsilon_j(p) = \epsilon_j(q)$ for $j = 1, 2$ once η is small enough.

12.5 A Generalization of Lemma 12.3

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 put a $(*)$ by each quantity associated to the second triple.

Lemma 12.6 *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 only depends on 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 12.3 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 12.6. Let $[x]$ denote the image of $x \in \boldsymbol{R}/\boldsymbol{Z}$.

Lemma 12.7 *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 $[\sigma(p_2)]$ and $[\sigma(q_2)^*]$ in $\boldsymbol{R}/\boldsymbol{Z}$ is less than ϵ. $|\sigma(p_2) - \sigma(q_2^*)| < \epsilon$ and $N = n(q^*) - n(p)$. The number δ only depends on ϵ and the distance from $\sigma(p_1)$ to $\{0, 1\}$.*

Proof: Using the same trick as in Lemma 12.3, we reduce to the single variable case. In this case, we mainly repeat the proof of Lemma 12.3. If δ is small enough then $[\alpha(p) - \sigma(p_1)]$ and $[\alpha(q) - \sigma(q_1)]$ are very close, and relatively far from 0 or 1. Equation 112 now says that $[\sigma(p_2)]$ and $[\sigma(q_2)]$ are close in $\boldsymbol{R}/\boldsymbol{Z}$. ♠

12.6 Proof in the General Case

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 §12.2, this time keeping track of the parameter. Let η be as in the Torus Lemma. We use the big O notation.

Lemma 12.8 *There is an integer I_k such that $|\alpha_0^*\widehat{\lambda}_k^* - \alpha_0\lambda_k - I_k| < O(\eta)$.*

Proof: Let $[V]$ denote the distance from $V \in \boldsymbol{R}^3$ to the nearest point in $\boldsymbol{Z}^3$. Let $p = (x, \pm 1)$ and $q^* = (x^*, \pm 1)$. Recalling the definition of μ_+, the hypotheses in the Torus Lemma imply that

$$\Big[H^*\Big(\frac{x^*}{2}, \frac{x^*}{2} + 1, \frac{x^*}{2}\Big) - H\Big(\frac{x}{2}, \frac{x}{2} + 1, \frac{x}{2}\Big)\Big] < O(\eta) \tag{123}$$

We compute that $\alpha_0 = x/2 + 1/2$, independent of parameter. Therefore

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

The same goes with the starred quantities. Therefore,

$$[(\widehat{\lambda}_2^*, \widehat{\lambda}_1^* - 1, 1)\alpha_0^* - (\widehat{\lambda}_2, \widehat{\lambda}_1 - 1, 1)\alpha_0] =$$

$$[H^*(\alpha_0^*, \alpha_0^*, \alpha_0^*) - H(\alpha_0, \alpha_0, \alpha_0)] < O(\eta) + \|(H^* - H)(-1, 1, -1)\| < O(\eta).$$

Our lemma now follows immediately from Equation 107. ♠

The integer I_k of course depends on (p, A) and (q^*, A^*), but in all cases Equation 107 gives us

$$I_0 = I_4; \qquad I_1 = I_3 = I_5 = I_7; \qquad I_2 = I_6, \tag{124}$$

Lemma 12.9 *As $\eta \to 0$, the pairwise differences between the 3 quantities $\alpha_k^* - \alpha_k$ and $n_k^* - n_k$ and I_k tends to 0 for all k.*

Proof: Here α_k^* stands for $\alpha_k(q^*)$, etc. Equation 120 works separately for each parameter. The replacement for Equation 121 is

$$\alpha_k^* - \alpha_k = W + X + Y; \qquad W = \alpha_0^*\widehat{\lambda}_k^* - \alpha_0\widehat{\lambda}_k \tag{125}$$

$$X = \sum_{i=1}^{k} \xi_i^* \sigma_i^*(q^*) - \sum_{i=1}^{k} \xi_i \sigma_i(p) = \sum_{i=1}^{k} \xi_i(\sigma_i^* - \sigma_i) + O(|A - A^*|); \qquad (126)$$

$$Y = \sum_{i=1}^{k} C_i^* - \sum_{i=1}^{k} C_i = O(|A - A^*|). \qquad (127)$$

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

$$\alpha_k^* - \alpha_k = \left(\alpha_0^* \widehat{\lambda}_k^* - \alpha_0 \lambda_k\right) + \sum_{i=1}^{k} \xi_i(\sigma_i^* - \sigma_i) + O(|A - A^*|). \qquad (128)$$

In light of Lemma 12.8, 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 12.6 used in place of Lemma 12.3. ♠

Similar to the single parameter case, Equations 88 and 124 now finish the proof.

13 The Strip Functions

13.1 The Main Result

The purpose of this chapter is to understand the functions σ_j that arose in the proof of the Master Picture Theorem. We call these functions the *strip functions*.

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

$$W'_k = \bigcup_{j=0}^{k-1} W_j; \qquad S'_k = \bigcup_{j=0}^{k-1} S_j; \qquad k = 1, ..., 7. \tag{129}$$

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

$$S = S_0 \cup ... \cup S_7. \tag{130}$$

If $p \in \Xi_+ - W'_k$ then the points $p = p_0, ..., p_k$ are defined. Here, as in the previous chapter, $p_j = E_j(p_{j-1})$. The functions $\sigma_1, ..., \sigma_k$ and $\alpha_1, ..., \alpha_k$ 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 $[p_{k+1}]$ is well defined in the cylinder $\boldsymbol{R}^2/\langle V_{k+1}\rangle$. The corresponding function $\sigma_{k+1}(q) = \sigma(q_{k+1})$ is well defined as an element of $\boldsymbol{R}/\boldsymbol{Z}$.

Let $\pi_j : \boldsymbol{R}^4 \to \boldsymbol{R}$ be the jth coordinate projection. Let $[x]$ denote the image of x in $\boldsymbol{R}/\boldsymbol{Z}$. The following identities refer to the $(+)$ case. We discuss the $(-)$ case at the end of the chapter.

$$\sigma_1 = \left[\frac{2-\pi_3}{2}\right] \circ \mu_+ \qquad \text{on } \Xi_+ \tag{131}$$

$$\sigma_2 = \left[\frac{1+A-\pi_2}{1+A}\right] \circ \mu_+ \qquad \text{on } \Xi_+ - W'_1 \tag{132}$$

$$\sigma_3 = \left[\frac{1+A-\pi_1}{1+A}\right] \circ \mu_+ \qquad \text{on } \Xi_+ - W'_2 \tag{133}$$

$$\sigma_4 = \left[\frac{1+A-\pi_1-\pi_2+\pi_3}{2}\right] \circ \mu_+ \qquad \text{on } \Xi_+ - W'_3 \tag{134}$$

In the next chapter we deduce the Master Picture Theorem from these identities and the Torus Lemma. In this chapter, we prove the identities.

13.2 Continuous Extension

Let $g = \sigma_j$ for $j = 0, ..., k+1$. since the image $\mu_+(\Xi \times (0,1))$ is dense in $R - S'_k$, we define

$$\widetilde{g}(\tau) := \lim_{n \to \infty} g(p_n, A_n); \qquad \tau \in R - S'_k. \tag{135}$$

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.

Lemma 13.1 *The functions $\widetilde{\sigma}_1, ..., \widetilde{\sigma}_{k+1}$, considered as $\boldsymbol{R}/\boldsymbol{Z}$-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 $k = 2$. 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. $\widetilde{\sigma}_1$ is well defined and continuous on all of R, by Equation 131.

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 $\min(A_n, 1 - A_n)$. Note that $[\alpha_1(p_n, A_n)] = [\pi_3(\mu_+(p_n, A_n)]$. Hence $\{[\alpha_1(p_n, A_n)]\}$ is a Cauchy sequence in $\boldsymbol{R}/\boldsymbol{Z}$.

Lemma 12.7 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 12.7 implies that $\{\sigma_2(\tau_m, A_m)\}$ forms a Cauchy sequence in $\boldsymbol{R}/\boldsymbol{Z}$. Hence, $\widetilde{\sigma}_2$ 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 all sufficiently large n. As in our proof of the General Torus Lemma, Equation 128 now says that shows that $\{\alpha_2(p_n, A_n)\}$ forms a Cauchy sequence in $\boldsymbol{R}/\boldsymbol{Z}$. We now repeat the previous argument to see that $\{\sigma_3(\tau_m, A_m)\}$ forms a Cauchy sequence in $\boldsymbol{R}/\boldsymbol{Z}$. Hence, $\widetilde{\sigma}_3$ is well defined on $R - S'_2$, and continuous. ♠

Implicit in our proof above is the function

$$\beta_k = [\alpha_k] \in \boldsymbol{R}/\boldsymbol{Z}. \tag{136}$$

This function will come in handy in our next result.

13.3 Quality of the Extension

Let $X = R - \partial R \subset \boldsymbol{R}^4$. Note that X is open and convex

Lemma 13.2 *Suppose $X \subset R - S'_k$. Then $\widetilde{\sigma}_{k+1}$ is locally affine on X_A.*

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

We already know that $\widetilde{\sigma}_1, ..., \widetilde{\sigma}_{k+1}$ are all defined and continuous on X. We already remarked that Equation 131 is true by direct inspection. As we already remarked in the previous proof, $\beta_0 = \pi_3 \circ \mu_+$. Thus, we define $\widetilde{\beta}_0 = [\pi_3]$. Let $\widetilde{\beta}_0 = [\pi_3]$. Both $\widetilde{\sigma}_0$ and $\widetilde{\beta}_0$ are locally affine on X_A.

Let $m \leq k$. The second half of Equation 112 tells us that $\widetilde{\sigma}_m$ is a locally affine function of $\widetilde{\sigma}_{m-1}$ and $\widetilde{\beta}_{m-1}$. Below we will prove that $\widetilde{\beta}_m$ is defined on X_A, and locally affine, provided that $\widetilde{\sigma}_1, ..., \widetilde{\sigma}_m$ are locally affine. Our lemma follows from this claim and induction.

Now we prove the claim. All the addition below is done in $\boldsymbol{R}/\boldsymbol{Z}$. Since $\mu_+(\Xi_+)$ is dense in X_A, we can at least define $\widetilde{\beta}_m$ on a dense subset of X_A. Define

$$p = (x, \pm 1); \quad p' = (x', \pm 1); \quad \tau = \mu_+(p); \quad \tau' = \mu_+(p'); \quad t = \frac{x' - x}{2}. \tag{137}$$

We choose p and p' so that the pinwheel map is entirely defined.

From Equation 121, we have

$$\widetilde{\beta}_m(\tau') - \widetilde{\beta}_m(\tau) = [t\widehat{\lambda}_k] + \sum_{j=1}^{m}[\xi_j \times (\widetilde{\sigma}_j(\tau') - \widetilde{\sigma}_j(\tau))]. \tag{138}$$

Here $\xi_1, ..., \xi_m$ are constants that depend on A. Let H be the matrix in Equation 108. We have $H(t,t,t) \equiv H(\tau' - \tau) \bmod \boldsymbol{Z}^3$ because $(t,t,t) \equiv \tau' - \tau \bmod \Lambda$. Our analysis in §12.2 shows that

$$[t\widehat{\lambda}_k] = [\pi \circ H(t,t,t) - \epsilon t] = [(\pi - \epsilon\pi_3) \circ H(\tau' - \tau)]. \tag{139}$$

Here $\epsilon \in \{0, 1\}$ and π is some coordinate projection. The choice of ϵ and π depends on k. We now see that

$$\widetilde{\beta}_m(\tau') = \widetilde{\beta}_m(\tau) + (\pi + \epsilon_3\pi) \circ H(\tau' - \tau) + \sum_{j=1}^{m}[\xi_j \times (\widetilde{\sigma}_j(\tau') - \widetilde{\sigma}_j(\tau))]. \tag{140}$$

The right hand side is everywhere defined and locally affine. Hence, we define $\widetilde{\beta}_m$ on all of X_A using the right hand side of the last equation. ♠

Lemma 13.3 *Suppose $X \subset R - S_k'$. Then σ_{k+1} is analytic on X.*

Proof: The constants ξ_j in Equation 138 vary analytically with A. Our argument in Lemma 13.2 therefore shows that the linear part of σ_{k+1} varies analytically with A. We just have to check the linear term. Since X_A is connected we can compute the linear 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 $A \to \sigma_k(p, A)$ varies analytically follows from the fact that our strips vary analytically. ♠

Remark: We have $S_k \subset \widetilde{\sigma}_{k+1}^{-1}([0])$. Given Equation 131, we see that $X \subset R - S_1'$. Hence σ_2 is defined on X. Hence σ_2 is analyic on X and locally affine on each X_A. We use these two properties to show that Equation 132 is true. But then $X \subset R - S_2'$. etc. So, we will know at each stage of our verification that Lemmas 13.2 and 13.3 apply to the function of interest.

Equations 132, 133, and 134 are formulas for $\widetilde{\sigma}_2$, $\widetilde{\sigma}_3$, and $\widetilde{\sigma}_4$ respectively. Let $f_{k+1} = \widetilde{\sigma}_{k+1} - \sigma_{k+1}'$, where $k = 2, 3, 4$. Here σ_{k+1}' is the right hand side of the identity for $\widetilde{\sigma}_{k+1}$. Our goal is to show that $f_{k+1} \equiv [0]$ for $k = 1, 2, 3$. Call a parameter A *good* if $f_{k+1} \equiv [0]$ on X_A. Call a subset $S \subset (0, 1)$ *substantial* if S is dense in some open interval of $(0, 1)$. By analyticity, $f_{k+1} \equiv [0]$ provided that a substantial set of parameters is good.

In the next section we explain how to verify that a parameter is good. If f_{k+1} was a locally affine map from X_A into $\boldsymbol{R}$, 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 $\boldsymbol{R}/\boldsymbol{Z}$, we have to work a bit harder.

13.4 Irrational Quintuples

We will give a construction in $\boldsymbol{R}^3$. When the time comes to use the construction, we will identify X_A as an open subset of a copy of $\boldsymbol{R}^3$.

Let $\zeta_1, ..., \zeta_5 \in \boldsymbol{R}^3$ be 5 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 the there is no rational relation

$$\sum_{i=1}^{5} c_j \zeta_j = 0; \qquad c_j \in \boldsymbol{Q}; \qquad c_1 c_2 c_3 c_4 c_5 = 0. \tag{141}$$

If we allow all the constants to be nonzero, then there is always a relation.

Lemma 13.4 *Let C be an open convex subset of $\boldsymbol{R}^3$. Let $f : C \to \boldsymbol{R}/\boldsymbol{Z}$ 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 $F : C \to \boldsymbol{R}$. But then F is affinc on C, and we can extend F to be an affine map from $\boldsymbol{R}^3$ to $\boldsymbol{R}$. By construction $F(\zeta_i) - F(\zeta_j) \in \boldsymbol{Z}$ for all i, j. Adding a constant to F, we can assume that F is linear. There are several cases.

Case 1: Suppose that $F(\zeta_j)$ is independent of j. In this case, all the points lie in the same plane, and all volumes are zero. This violates the irrationality condition.

Case 2: Suppose we are not in 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 $\{F(\zeta_j)\}$ only has a total of 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's 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 in 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 $F(\zeta'_j) \in \boldsymbol{Z} - \{0\}$ 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 $v''_j / v'_j \in \boldsymbol{Q}$ 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 between the areas of the 4 triangles defined by 4 points in the plane. Hence, there is a rational relation between $v''_2, v''_3, v''_4, v''_5$. But then there is a rational relation between v_2, v_3, v_4, v_5. This contradicts the irrationality condition. ♠

13.5 Verification in the Plus Case

Proceeding somewhat at random, we define

$$\phi_j = \left(8jA + \frac{1}{2j}, 1\right); \qquad j = 1, 2, 3, 4, 5. \tag{142}$$

We check that $\phi_j \in \Xi_+$ for A near $1/2$. Letting $\zeta_j = \mu_+(\phi_j)$, we check that $f_{k+1}(\zeta_j) = [0]$ for $j = 1, 2, 3, 4, 5$.

Example Calculation: Here is an example of what we do automatically in Mathematica. Consider the case $k = 1$ and $j = 1$. When $A = 1/2$, the length spectrum for ϕ_1 starts out $(1, 1, 2, 1)$. 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

$$(x, y) \to (x, y, 1) \cdot \frac{(-1, A, A)}{2 + 2A} \tag{143}$$

takes on the value 0 on the line $x = Ay + A$ and 1 on the line $x = Ay - 2 - A$. These are the two edges of Σ_2. (See §15.1.) 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

$$\mu_+(\phi_1) = \frac{1}{4}(-7 + 24A, 1 + 4A, -7 + 16A),$$

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 $f_2(\zeta_2) = [0]$ for all A near $1/2$. The verifications for the other pairs (k, j) are similar.

Checking Irrationality: It only remains to check that the points $(\zeta_1, ..., \zeta_5)$ form an irrational quintuple for a dense set of parameters A. In fact this will true in the complement of a countable set of parameters.

The 5 volumes associated to our quintuple are as follows.

- $v_5 = 5/24 - 5A/12 + 5A^2/24$.
- $v_4 = 71/40 + 19A/20 - 787A^2/120 - 4A^3$.
- $v_3 = 119/60 + 7A/60 - 89A^2/15 - 4A^3$
- $v_2 = -451/240 - 13A/40 + 1349A^2/240 + 4A^3$
- $v_1 = -167/80 - 13A/40 + 533A^2/80 + 4A^3$.

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 on 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 curve $A \to (v_5, v_4, v_3, v_2)$ lies in a proper linear subspace of $\boldsymbol{R}^4$.

We evaluate this curve at $A = 1, 2, 3, 4$ and see that the resulting points are linearly independent in $\boldsymbol{R}^4$. Hence, there is no global rational relation. Hence, on a dense set of parameters, there is no rational relation between the first 4 volumes listed. A similar argument rules out rational relations amongst any other 4-tuple of these volumes.

13.6 The Minus Case

In the $(-)$ case, Equations 132 and 133 do not change, except that μ_- replaces μ_+ and all the sets are defined relative to Ξ_- and μ_-. Equations 131 and 134 become

$$\sigma_1 = \left[\frac{1-\pi_3}{2}\right] \circ \mu_- \qquad \text{on } \Xi_-. \tag{144}$$

$$\sigma_4 = \left[\frac{A-\pi_1-\pi_2+\pi_3}{2}\right] \circ \mu_- \qquad \text{on } \Xi_- - S_3'. \tag{145}$$

Lemma 13.2 and Lemma 13.3 have the same proof in the $(-)$ case. We use the same method as above, except that we use the points

$$\phi_j + (2,0); \qquad j = 1,2,3,4,5. \tag{146}$$

These points all lie in Ξ_- for A near $1/2$. The rest of the verification is essentially the same as in the $(+)$ case.

14 Proof of the Master Picture Theorem

14.1 The Main Argument

Let S be the singular set defined in Equation 130. Let $\widetilde{S}$ denote the union of hyperplanes listed in §10.2. let d denote distance on the polytope R. In this chapter we will prove

Lemma 14.1 (Hyperplane) *$S \subset \widetilde{S}$ and $\theta(p, A) \geq d(\mu_+(p, A), \widetilde{S})$.*

We finish the proof of the Master Picture Theorem assuming the Hyperplane Lemma.

Say that a *ball of constancy* in $R - \widetilde{S}$ 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 14.2 *Any point τ of $R - \widetilde{S}$ 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 14.3 *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 $R - \widetilde{S}$. Then the return pair for (p_0, A_0) equals the return pair for (p_1, A_1).*

Proof: Let $L \subset R - \widetilde{S}$ be a path joining points $\tau_0 = \mu_+(p_0, A_0)$ and $\tau_1 = \mu_+(p_1, A_1)$. 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 $R - \widetilde{S}$. Here is an example calculation that does the job. For each $\alpha = j/16$ for $j = 1, ..., 15$, we plot the image

$$\mu_A(2\alpha + 2n); \qquad n = 1, ..., 2^{15}; \tag{147}$$

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 only proved the half of the Master Picture Theorem that deals with Ξ_+ and μ_+. 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 144 replace Equations 131 and 134. We omit the details in the $(-)$ case.

14.2 The First Four Singular Sets

Our strip function identites make short work of the first four pieces of the singular set.

- Given Equation 131,

$$S_0 \subset \{z = 0\} \cup \{z = 1\}. \tag{148}$$

- Given Equation 132,

$$S_1 \subset \{y = 0\} \cup \{y = 1 + A\}. \tag{149}$$

- Given Equation 133,

$$S_2 \subset \{x = 0\} \cup \{x = 1 + A\}. \tag{150}$$

- Give Equation 134,

$$S_3 \subset \{x + y - z = 1 + A\} \cup \{x + y - z = -1 + A\}. \tag{151}$$

14.3 Symmetry

We use symmetry to deal with the remaining pieces. Suppose we start with a point $p \in \Xi_+$. We define $p_0 = p$ and $p_j = E_j(p)$. As we go along in our analysis, these points will be defined for increasingly large values of j. However, for the purposes of illustration, we assume that all points are defined.

Let ρ denote reflection in the x-axis. Then

$$\rho(\Sigma_9 - j) = \Sigma_j; \qquad q_j = \rho(p_{9-j}); \qquad j = 1, 2, 3, 4. \tag{152}$$

Figure 14.1 shows a picture. The disk in the center is included for artistic purposes, to cover up some messy intersections. In the picture, we have included the coordinates for the vectors $-V_1$ and $-V_2$ 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.

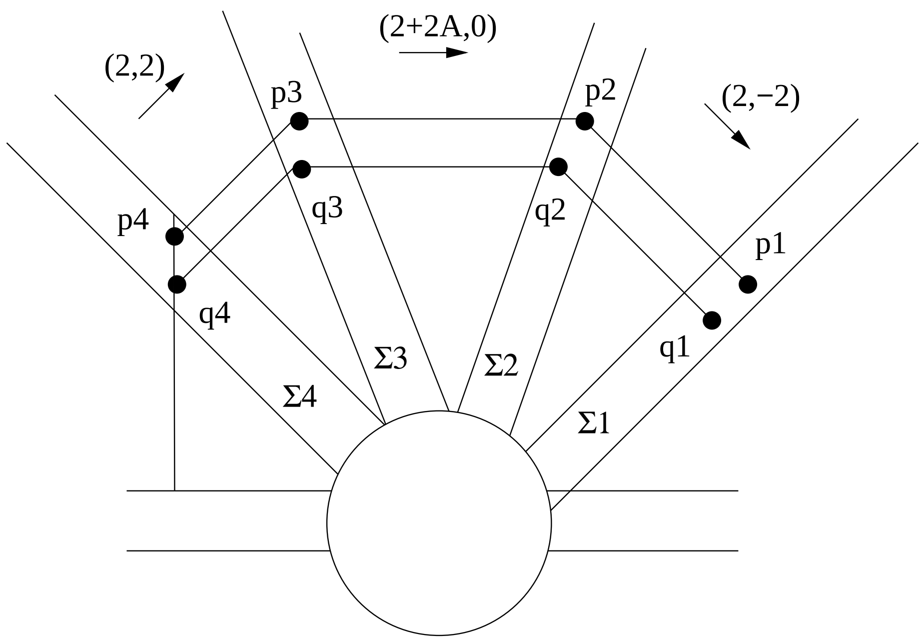

Figure 14.1: Reflected points

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 one lies in the interior.

14.4 The Remaining Pieces

14.4.1 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 our points are odd integers, we have $p_4 - q_4 = (0,2) + k(0,4)$ for some $k \in \boldsymbol{Z}$. 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 seqment 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 134, we get

$$S_4 \subset \{x + y - z = A\} \cup \{x + y - z = 2 + A\}.$$

14.4.2 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

$$p_3 - q_3 = \epsilon(0,2) + k(2,2); \qquad \epsilon \in \{-1,1\}; \qquad k \in \boldsymbol{Z}.$$

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 either $p_3 = q_3 - (2,0)$ or $p_3 = q_3 + (0,2)$. 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 133 we get

$$S_5 \subset \{x = A\} \cup \{x = 1\}.$$

14.4.3 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

$$p_2 - q_2 = (p_3 - q_3) + k(2 + 2A, 0). \qquad (153)$$

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

$$p_2 - q_2 = X_j + \epsilon(2A, 2); \qquad \epsilon \in \{-1, 0, 1\}; \qquad j \in \{1, 2, 3, 4\}. \tag{154}$$

Note that the vector $(2A, 2)$ is parallel to the boundary of Σ_2. Hence, for the purposes 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 got for $\sigma_3(p)$ in the previous section. Given Equation 132 we get

$$S_6 \subset \{y = A\} \cup \{y = 1\}.$$

14.4.4 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

$$p_1 - q_1 = (p_2 - q_2) + k(-2, 2). \tag{155}$$

Note that the vector $(2, 2)$ is parallel to Σ_1. For the purposes of finding $\sigma_1(p)$, we can do our computation modulo $(2, 2)$. For instance, $(2, -2) \equiv (0, 4)$ mod $(2, 2)$. Given Equation 154, we have

$$p_1 - q_1 = \epsilon_1(0, 2) + \epsilon_2(2A, 2) + k(0, 4) \mod (2, 2). \tag{156}$$

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

$$\left[\frac{k}{2} + 2\epsilon A\right]; \qquad \epsilon \in \{-1, 0, 1\}; \qquad k \in \boldsymbol{Z}.$$

Given Equation 131 we see that $S_7 \subset \{z = A\} \cup \{z = 1 - A\}$.

14.5 Proof of The Second Statement

Our analysis above establishes the first statement of the Hyperplane Lemma. For the second statement, suppose that $d(\mu_+(p, A), \widetilde{S}) = \epsilon$. Given Equations 131, 132, 133, and 134, 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 it is for the index $9 - j$.

15 Some Formulas

15.1 Formulas for the Pinwheel Map

In this section we explain how to implement the pinwheel map. We define

$$V_1 = (0,4); \qquad V_2 = (-2,2);$$
$$V_3 = (-2-2A, 0); \qquad V_4 = (-2,-2). \tag{157}$$

Next, we define vectors

$$W_1 = \frac{1}{4}(-1,1,3); \qquad W_2 = \frac{1}{2+2A}(-1,A,A);$$
$$W_3 = \frac{1}{2+2A}(-1,-A,A); \qquad W_4 = \frac{1}{4}(-1,-1,3); \tag{158}$$

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

$$F_j(p) = W_j \cdot (p_1, p_2, 1). \tag{159}$$

$F(j,p)$ 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.

Example: Let $p = (2A, 1)$ and $q = (-2, 1)$, we compute that

$$F_2(p) = \frac{1}{2+2A}(-1,A,A) \cdot (2A,1,1) = 0.$$
$$F_2(q) = \frac{1}{2+2A}(-1,A,A) \cdot (-2,1,1) = 1.$$

This checks out, because p lies in one component of $\partial\Sigma_2$ and q lies in the other component of $\partial\Sigma_2$.

Here is a formula for our strip maps.

$$E_j(p) = p - \text{floor}(F_j(p))V_j. \tag{160}$$

If we set $V_{4+j} = -V_j$ and $F_{j+4} = -F_j$ then we get the nice formulas

$$dF_j(V_j) = dF_j(V_{j+1}) = 1. \tag{161}$$

with indices taken mod 8.

15.2 The Reduction Algorithm

Let $A \in (0,1)$ and $\alpha \in \boldsymbol{R}_+$ and $(m,n) \in \boldsymbol{Z}^2$ be a point above the baseline of $\Gamma_\alpha(A)$. In this section we describe how we compute the points

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

This algorithm will be important when we prove the Copy Theorems in Part IV of the monograph.

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

$$\mu_-(M_\alpha(m,n)) = \begin{pmatrix} x-(1+A)X \\ y-(1+A)Y \\ z-Z \end{pmatrix} \tag{162}$$

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

$$y = z + Z + 1. \tag{163}$$

Example: Referring to §10.4, consider the case when $A = 3/5$ and $\alpha = 1/10$ and $(m,n) = (4,2)$. We get

$$z = \frac{9}{2}; \qquad Z = 4; \qquad y = \frac{17}{2}; \qquad Y = \text{floor}\left(\frac{17/2}{8/5}\right) = 5.$$

$$x = \frac{17}{2} - 5(\frac{2}{5}) - 1 = \frac{11}{2}; \qquad X = \text{floor}\left(\frac{11/2}{8/5}\right) = 3.$$

$$\mu_-(M(4,2)) = \left(\frac{11}{2} - 3(\frac{8}{5}), \frac{17}{2} - 5(\frac{8}{5}), \frac{9}{2} - 4\right) = \left(\frac{7}{10}, \frac{1}{2}, \frac{1}{2}\right).$$

15.3 Computing the Partition

Here we describe how Billiard King applies the Master Picture Theorem.

15.3.1 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 doesnt depend on n_0. It turns out that this polyhedron is empty unless $n_4 \in \{-2, -1, 0, 1, 2\}$.

15.3.2 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 $n_1 \in \{3, 0\}$.
  - 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 $n_1 \in \{0, 1\}$.
  - 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 $n_1 \in \{1, 2\}$.
  - 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 $n_1 \in \{2, 3\}$.
  - 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 $\epsilon_2(n) = 0$.

15.3.3 Step 3

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

- Let $(a_\pm, b_\pm, c_\pm) = \mu_\pm(A,m,n)$. See §15.2.
- 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 $(m,n) + (\epsilon_1^\pm, \epsilon_2^\pm)$.

15.4 The List of Polytopes

Referring to the simpler partition from §10.6, we list the 14 polytopes that partition R_+. In each case, we list some vectors, followed by the pair (ϵ_1, ϵ_2) that the polytope determines.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

15.5 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 10.6. $R_+(0,0)$ 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.

Part III

In this part of the monograph we use the Master Picture Theorem to prove most of the results we quoted in Part I of the monograph.

- In §16 we prove the Embedding Theorem.
- In §17 we prove some results about the symmetries of the arithmetic graph and the hexagrid.
- In §18 we prove Statement 1 of the Hexagrid theorem, namely that the arithmetic graph does not cross any floor lines. We treat the odd case, which is important for our main theorems.
- In §19 we prove Statement 2 of the Hexagrid theorem, namely that the arithmetic graph only crosses the walls near the doors. Again, we consider the odd case, which is important for our main theorems. The two statements of the Hexagrid Theorem have similar proofs, though Statement 2 has a more elaborate proof.
- In §20 we will prove the even case of the Hexagrid Theorem. The proof is almost identical to the odd case.

Many of the proofs in this part of the monograph 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 only involve a small amount of integer arithmetic. An energetic mathematician could do them all by hand in an afternoon. To help make the proofs surveyable, we will include extensive computer pictures of 2 dimensional slices of our polytopes. These pictures, all reproducible on Billiard King, serve as sanity checks for the computer calculations.

16 Proof of the Embedding Theorem

Let $\widehat{\Gamma} = \widehat{\Gamma}_\alpha(A)$ be the arithmetic graph for a parameter A and some number $\alpha \not\in 2\boldsymbol{Z}[A]$. In this chapter we prove that $\widehat{\Gamma}$ is a disjoint union of embedded polygons and infinite polygonal arcs. This is the Embedding Theorem.

16.1 Step 1

We will first prove that every nontrivial vertex of $\widehat{\Gamma}$ has valence 2. Each point $p \in \widehat{\Gamma}$ is connected to two points q_+ and q_-. Hence, each non-trivial vertex has valence either 1 or 2. The following two cases are the only cases that lead to valence 1 vertices:

- $p = q_+$ and $q_+ \neq q_-$.
- $q_+ = q_-$ and $q_\pm \neq p$.

The following lemma rules out the first of these cases.

Lemma 16.1 *If $p = q_+$ or $p = q_-$ then $p = q_+ = q_-$.*

Proof: Our proof refers to §10.6. Recall that $R_+(0,0)$ consists of 2 convex integer polytopes. Likewise $R_-(0,0)$ consists of 2 convex integer polytopes. It suffices to show that

$$(t, t+1, t) \in R_+(0,0) \qquad \Longleftrightarrow \qquad (t-1, t, t) \in R_-(0,0). \tag{164}$$

This is equivalent to the statement that

$$R_-(0,0) + (1,1,0,0) \subset \Lambda R_+(0,0)$$

Here $\Lambda R_+(0,0)$ is the orbit of $R_+(0,0)$ under the action of Λ. Let ι be the involution from Equation 77. Recall that $R_-(0,0) = \iota(R_+(0,0))$. Hence, Equation 164 equivalent to the statement that

$$\iota(R_+(0,0)) + (1,1,0,0) \subset \Lambda R_+(0,0). \tag{165}$$

Let P_1 and P_2 denote the two polytopes comprising $R_+(0,0)$, as listed at the end of §10. Let γ_2 be the element of Λ described in §10.6. We compute that

$$\iota(P_1) + (1,1,0,0) = P_1; \qquad \iota(P_2) + (1,1,0,0)) = \gamma_2(P_2). \tag{166}$$

We did the second calculation in §15.5, and the first computation is similar. This does it for us. ♠

16.2 Step 2

Our next goal is to rule out the possibility that $p \neq q_\pm$, but $q_+ = q_-$. This situation happens iff there is some $(\epsilon_1, \epsilon_2) \in \{-1, 0, 1\}$ such that

$$\Lambda R_+(\epsilon_1, \epsilon_2) \cap (R_-(\epsilon_1, \epsilon_2) + (1, 1, 0, 0)) \neq \emptyset. \tag{167}$$

A visual inspection and/or a compute computer search – we did both – reveals that at least one of the two sets above is empty unless (ϵ_1, ϵ_2) is one of

$$(1, 1); \qquad (-1, -1); \qquad (1, 0); \qquad (-1, 0). \tag{168}$$

To rule out Equation 167 for each of these pairs, we need to consider all possible pairs (P_1, P_2) of integral convex polytopes such that

$$P_1 \subset \Lambda R_+(\epsilon_1, \epsilon_2); \qquad P_2 \subset (R_-(\epsilon_1, \epsilon_2) + (1, 1)) \tag{169}$$

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

$$\Lambda'_{10} = \{a_1\gamma_1 + a_2\gamma_2 |\ |a_1|, |a_2| \leq 10\}. \tag{170}$$

Lemma 16.2 *Let $\gamma \in \Lambda - \Lambda'$. Suppose that*

$$P_1 = \gamma(Q_1); \qquad Q_1 \subset R_+(\epsilon_1, \epsilon_2); \qquad P_2 \subset R_-(\epsilon_1, \epsilon_2) + (1, 1, 0, 0).$$

Then P_1 and P_2 have disjoint interiors.

Proof: The third coordinates of points in P_1 lies between n and $n + 1$ for some $n \neq 0$ whereas the third coordinates of points in P_2 lie in $[0, 1]$. ♠

Lemma 16.3 *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: Q_1 is contained in the ball of radius 4 about P_2, but γ moves this ball entirely off itself. ♠

The last two results leave us with a finite problem. Given a pair (ϵ_1, ϵ_2) from our list above, and

$$\gamma \in \Lambda'_{10}; \quad P_1 = \gamma(Q_1); \quad Q_1 \subset R_+(\epsilon_1, \epsilon_2); \quad P_2 \subset R_-(\epsilon_1, \epsilon_2) + (1, 1, 0, 0),$$

we produce a vector

$$w = w(P_1, P_2) \in \{-1, 0, 1\}^4 \tag{171}$$

such that

$$\max_{v \in \mathrm{vtx}(P_1)} v \cdot w \leq \min_{v \in \mathrm{vtx}(P_2)} v \cdot w. \tag{172}$$

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 integer arithmetic. It is a bit surprising to us that such a simple vector works in all cases, but that is how it works out.

Using Billiard King, the interested reader can draw arbitrary (z, A) slices of the sets $\Lambda R_+(\epsilon_1, \epsilon_2)$ and $\Lambda R_-(\epsilon_1, \epsilon_2) + (1, 1, 0, 0)$, and see that the interiors of the polygons from the first set are disjoint from the interiors of the polygons from the second set. We will illustrate this with pictures in §16.4.

16.3 Step 3

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 situation like the one shown in Figure 16.1.

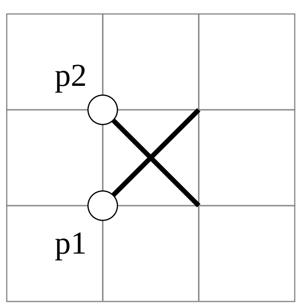

Figure 16.1: Embedding Failure

Let M_+ and M_- be the maps from §10.6.3. Given the Master Picture Theorem, this situation arises only in the following 4 cases:

- $M_+(p_1) \in \Lambda R_+(1,1)$ and $M_+(p_2) \in \Lambda R_+(1,-1)$.
- $M_-(p_1) \in \Lambda R_-(1,1)$ and $M_-(p_2) \in \Lambda R_-(1,-1)$.
- $M_-(p_1) \in \Lambda R_-(1,1)$ and $M_+(p_2) \in \Lambda R_+(1,-1)$.
- $M_+(p_1) \in \Lambda R_+(1,1)$ and $M_-(p_2) \in \Lambda R_-(1,-1)$.

Note that $p_2 = p_1 + (0,1)$ and hence

$$M_\pm(p_2) = M_\pm(p_1) + (1,1,1,0) \bmod \Lambda. \tag{173}$$

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 inspect the picture and see that this situation never occurs for the types $(1,1)$ and $(1,-1)$. Hence, Cases 1 and 2 do not occur. More inspection shows that there are $R_+(1,-1) = \emptyset$. Hence, Case 3 does not occus. This leaves Case 4, the only nontrivial case.

Case 4 leads to the statement that

$$(t,t,t,A) + (0,1,0,0) \in \Lambda R_+(1,1);$$

$$(t,t,t,A) - (1,0,0,0) + (1,1,1,0) = (t,t,t,A) + (0,1,1,0) \in \Lambda R_-(1,-1). \tag{174}$$

Setting p equal to the first of the two points above, we get

$$p \in \Lambda R_+(1,1); \qquad p + (0,0,1,0) \in \Lambda R_-(1,-1). \tag{175}$$

Letting $\gamma_3 \in \Lambda$ be as in Equation 75, we have

$$p - (1,1,0,0) = \gamma_3(p + (0,0,1,0)) \in \Lambda R_-(1,-1). \tag{176}$$

For any subset $S \subset \widetilde{R}$, we have

$$(\Lambda S) + (a,b,c,0) = \Lambda(S + (a,b,c,0)). \tag{177}$$

The point here is that Λ acts as a group of translations on each set of the form $R^3 \times \{A\}$, and addition by $(x,y,z,0)$ commutes with this action on every such set. Equations 176 and 177 combine to give

$$p - (1,1,0,0) \in \Lambda(R_-(1,-1) - (1,1,0,0)) \tag{178}$$

Now we see that

$$\Lambda R_+(1,1) \cap \Lambda(R_-(1,-1) - (1,1,0,0)) \neq \emptyset.$$

Since the whole picture is Λ-equivariant, we have

$$\Lambda R_+(1,1) \cap (R_-(1,-1) - (1,1,0,0)) \neq \emptyset. \tag{179}$$

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. In §16.4 we illustrate this with a convincing picture.

16.4 A Visual Tour

The theoretical part of our proof amounts to reducing the Embedding Theorem to the statement that finitely many pairs of polytopes have disjoint interiors. The computer-aided part of the proof amounts to verifying the disjointness finitely many times. Our verification used a very fragile disjointness test. We got a lucky, because many of our polytope pairs share a 2-dimensional face. Thus, a separating hyperplane has to be chosen very carefully. Needless to say, if our simple-minded approach did not work, we would have used a more robust disjointness test.

If we could write this monograph on 4-dimensional paper, we could simply replace the computer-aided part of the proof with a direct appeal to the visual sense. Since we don't have 4-dimensional paper, we need to rely on the computer to "see" for us. In this case, "seeing" amounts to finding a hyperplane that separates the interiors of the two polytopes. In other words, we are getting the computer to "look" at the pair of polytopes in such a way that one polytope appears on one side and the other polytope appears on the other side.

We do not have 4 dimensional paper, but we can draw slices of all the sets we discussed above. The interested user of Billiard King can see any desired slice. We will just draw typical slices. In our pictures below, we will draw the slices of R_+ with dark shading and the slices of R_- with light shading. in our discussion, the *base space* B refers to the (z, A) square over which our picture fibers. Let B_j denote the jth component of B, as determined by the characteristic n_1 discussed in §15.3.

In reference to Step 2, our pictures for the pair $(-\epsilon_1, -\epsilon_2)$ look like rotated versions of the pictures for the pair (ϵ_1, ϵ_2). Accordingly, we will just draw pictures for $(1,1)$ and $(1,0)$.

Figure 16.2 shows a slice of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,1)+(1,1,0,0))$ over B_0. Both slices are nonempty over B_1 as well, and the picture is similar.

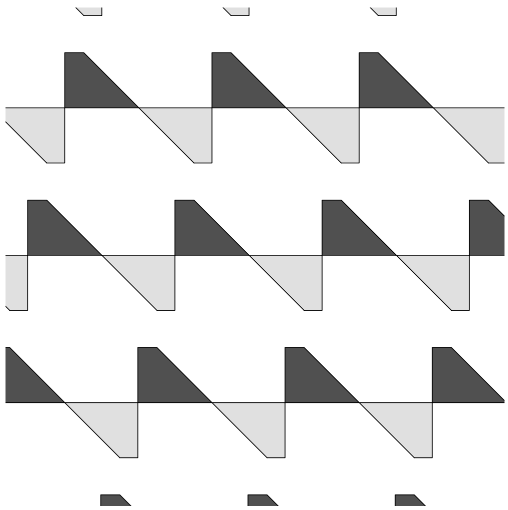

Figure 16.2: A slice of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,1)+(1,1,0,0))$

Figure 16.3 shows a slice of $\Lambda R_+(1,0)$ and $\Lambda(R_-(1,0)+(1,1,0,0))$ over B_0. The picture over B_1 is similar. Figure 16.4 shows a slice of $\Lambda R_+(1,0)$ and $\Lambda(R_-(1,0)+(1,1,0,0))$ over B_2. The picture over B_3 is similar.

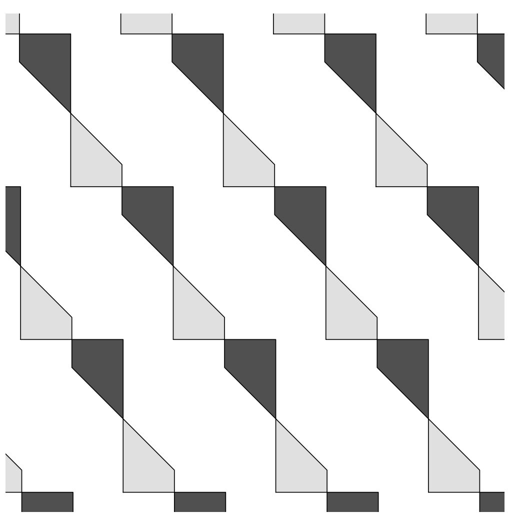

Figure 16.3: A slice of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,-1)-(1,1,0,0))$.

Figure 16.4 shows a slice of $\Lambda R_+(1,0)$ and $\Lambda(R_-(1,0)+(1,1,0,0))$ over B_2. The picture over B_3 is similar.

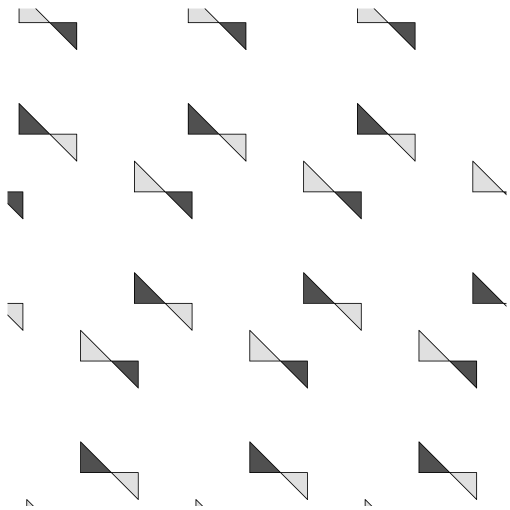

Figure 16.4: A slice of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,-1)-(1,1,0,0))$.

Figure 16.5 shows a slice of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,-1)-(1,1,0,0))$ over B_2. The picture looks similar over B_3 and otherwise at least one of the slices is empty.

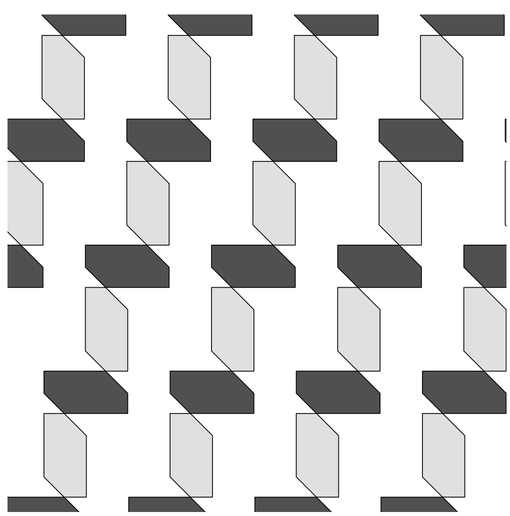

Figure 16.5: A slice of $\Lambda R_+(1,1)$ and $\Lambda(R_-(1,-1)-(1,1,0,0))$.

17 Extension and Symmetry

17.1 Translational Symmetry

Referring to §10.6.3, the maps M_+ and M_- are defined on all of $\boldsymbol{Z}^2$. This gives the extension of the arithmetic graph to all of $\boldsymbol{Z}^2$.

Lemma 17.1 *The extended arithmetic graph does not cross the baseline.*

Proof: By the Pinwheel Lemma, the arithmetic graph describes the dynamics of the pinwheel map, Φ. Note that Φ is generically defined and invertible on $\boldsymbol{R}_+ \times \{-1, 1\}$. Reflection in the x-axis conjugates Φ to Φ^{-1}. By the Pinwheel Lemma, Φ maps $\boldsymbol{R}_+ \times \{-1, 1\}$ into itself. By symmetry the same goes for Φ^{-1}. Hence Φ and Φ^{-1} also map $\boldsymbol{R}_- \times \{-1, 1\}$ into itself. If some edge of $\widehat{\Gamma}$ crosses the baseline, then one of Φ or Φ^{-1} would map a point of $\boldsymbol{R}^+ \times \{-1, 1\}$ into $\boldsymbol{R}_- \times \{-1, 1\}$. This is a contradiction. ♠

The above result also follows from Statement 1 of the Hexagrid Theorem, which we prove in the next chapter.

Recall from §5.1 that Θ is the lattice of translations generated by

$$V = (q, -p) \qquad\qquad V' = \left(0, \frac{(p+q)^2}{4}\right) \tag{180}$$

Lemma 17.2 *The arithmetic graph $\widehat{\Gamma}(p/q)$ is invariant under Θ.*

Proof: As we remarked in §5.1 we just have to show that $\widehat{\Gamma}(p/q)$ is invariant under translation by V'. By the Master Picture Theorem, it suffices to prove that $(t, t, t) \in \Lambda$ when t is the second coordinate of V'. Here Λ is as in Equation 71.

We have $(t, t, t) \equiv (2t, 2t, 0)$ mod Λ because t is an integer. Setting

$$a = pq; \qquad b = \frac{pq + q^2}{2}, \tag{181}$$

We compute that

$$a \begin{bmatrix} 1 + A \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 1 - A \\ 1 + A \\ 0 \end{bmatrix} = \begin{bmatrix} 2t \\ 2t \\ 0 \end{bmatrix}. \tag{182}$$

This completes the proof. ♠

Lemma 17.3 *When $A = p/q$ is an odd rational, the hexagrid is invariant under the action of* Θ.

Proof: Let $G(p/q)$ denote the hexagrid. For ease of notation, we write $X = X(p/q)$ for various objects X that depend on p/q.

By construction, the hexagrid is invariant under the action of V. We just have to see what happens for V'. Let

$$W = \left(\frac{pq}{p+q}, \frac{pq}{p+q} + \frac{q-p}{2}\right)$$

be the vector from the definition of the hexagrid G. It suffices to prove that 6 lines of G contain V'. We compute that

$$V' = -\frac{p}{2}V + \frac{p+q}{2}W. \tag{183}$$

The second coefficient is an integer. Given that the room grid RG is invariant under the lattice $\boldsymbol{Z}[V/2, W]$, we see that 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 DG is only invariant under $\boldsymbol{Z}[V]$, 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 the origin certainly contains V'. This line extends the bottom left edge of Q and hence belongs to DG.

- Let L_4 be the line containing V' and point $-(p+q)V/2 \in \boldsymbol{Z}[V]$. We compute that the slope of L_4 coincides with the slope of the top left edge of Q. The origin contains a line of DG parallel to the top left edge of Q, and hence every point in $\boldsymbol{Z}[V]$ contains such a line. Hence L_4 belongs to DG. To avoid a repetition of words below, we call our argument here the *translation principle*.

- Let L_5 be the line containing V' and point $-pV \in \boldsymbol{Z}[V]$. We compute that the slope of L_5 coincides with the slope of the bottom right edge of Q. The translation principle shows that L_5 belongs to DG.

- Let L_6 be the line containing V' and point $(q-p)V/2 \in \boldsymbol{Z}[V]$. We compute that the slope of L_6 coincides with the slope of the top right edge of Q. The translation principle shows that L_6 belongs to DG.

(The reader can see these lines, for any desired parameter, using Billiard King.) ♠

17.2 Rotational Symmetry

Here we will establish the rotational symmetry of the arithmetic graph that we discussed in §5.2. We will assume that p/q is an odd rational, as usual. Let p_+/q_+ be as in §5.2. Let ι denote 180 degree rotation about the point $(q_+, -p_+)/2$. We have

$$\iota(m,n) = (q_+, p-_+) - (m,n). \tag{184}$$

Here is the main result of this section.

Lemma 17.4 $\iota(\widetilde{\Gamma}) = \widetilde{\Gamma}$.

Proof: Let M_+ and M_- be as in §10.6.3. As usual, we take $\alpha = 1/(2q)$. We will first compare $M_+(m,n)$ with $M_-(\iota(m,n))$. We have

$$M_+(m,n) = (t, t+1, t) \bmod \Lambda; \qquad \frac{pm}{q} + n + \frac{1}{2q} \tag{185}$$

Next, we have

$$M_-(\iota(m,n)) = (t'-1, t', t') \bmod \Lambda;$$
$$t' = \left(\frac{ap}{q} - b\right) - \left(\frac{pm}{q} + n\right) + \frac{1}{2q} =$$
$$\left(\frac{ap-bq}{q}\right) - \left(\frac{pm}{q} + n\right) + \frac{1}{2q} =$$
$$\frac{-1}{q} - \left(\frac{pm}{q} + n\right) + \frac{1}{2q} = -\left(\frac{pm}{q} + n\right) - \frac{1}{2q} = -t.$$

In short

$$M_-(\iota(m,n)) = (-t-1, -t, -t) \bmod \Lambda. \tag{186}$$

Recall that R_A is the fundamental domain for the action of $\Lambda = \Lambda_A$. We mean to equate Λ with the $\boldsymbol{Z}$ span of its columns. There is some $v \in \Lambda$ such that

$$(s_1, s_2, s_3) = (t, t+1, t) + (v_1, v_2, v_3) \in R_A \tag{187}$$

Given Equation 71, we have $(2+A, A, 1) \in \Lambda$. Hence

$$w = (-v_1 + 2 + A, -v_2 + A, -v_3 + 1) \in \Lambda. \tag{188}$$

We compute that

$$(-t-1,-t,-t)+w=(1+A,1+A,1)-(s_1,s_2,s_3). \tag{189}$$

So, we have

$$M_+(m,n)=\rho\circ M_-(\iota(m,n)), \tag{190}$$

where ρ is reflection through the midpoint of the space R_A. A similar calculation shows

$$M_-(m,n)=\rho\circ M_+(\iota(m,n)), \tag{191}$$

But now we just verify by inspection that our partition of R_A is symmetric under ρ, and has the labels appropriate to force the type determined by

$$\rho\circ M_+(m,n),\ \ \rho\circ M_-(m,n)$$

to be the 180 degree rotation of the type forced by

$$M_-(m,n),\ \ M_+(m,n)$$

Indeed, we can determine this with an experiment performed on any rational large enough such that all regions are sampled. A little inspection of the picture of $\widetilde{\Gamma}(3/7)$, for instance, suffices to finish the proof. Compare Figure 3.6. ♠

Since $\widehat{\Gamma}$ is also invariant under the lattice Θ of translations, we see that there are actually infinitely many points of order 2 symmetry.

17.3 The Structure of the Doors

Say that a *wall line* is a line of positive slope in the room grid. The doors are the intersection points of lines in the door grid with the wall lines. Let L_0 be the wall line through $(0,0)$. Let L_1 be the wall line through $V/2$.

Lemma 17.5 *Any two wall lines are equivalent mod* Θ*.*

Proof: We check explicitly that the vector

$$V'+\frac{p+1}{2}V\in\Theta\cap L_1$$

Hence L_0 and L_1 are equivalent mod Θ. But any other wall line is obtained from one of L_0 or L_1 by adding a suitable integer multiple of V. ♠

Lemma 17.6 *The first coordinate of any door is an integer.*

Proof: Any wall line is equivalent mod Θ to L_0. Since Θ acts by integer translations, it suffices to door lies on L_0. Such a door is an integer multiple of the point v_3 in Figure 4.1. That is, our door has coordinates

$$\frac{k}{2q}(2pq, (p+q)^2 - 2p^2). \tag{192}$$

The first coordinate here is certainly an integer. ♠

It could happen that the second coordinate of a door is an integer. Call such a door exceptional.

Lemma 17.7 *Modulo the action of* Θ*, there are only two exceptional doors.*

Proof: The point $(0,0)$ gives rise to two exceptional doors with (the same) integer coordinates. One of these doors is associated to the wall above $(0,0)$, and one of these doors is associated to the door below. Hence, it suffices to show that any door with integer coordinates lies in Θ.

As in the preceding result, it suffices to consider doors on L_0. Given Equation 192, we see that

$$k\frac{(p+q)^2 - 2p^2}{2q} \in \boldsymbol{Z}$$

for an exceptional door. Expanding this out, and observing that q divides both q^2 and pq, we get that

$$k\frac{q^2 - p^2}{2q} \in \boldsymbol{Z}.$$

But q and $q^2 - p^2$ are relatively prime. Hence $k = jq$ for some $j \in \boldsymbol{Z}$. But

$$qv_3 = 2V' + V \in \Theta.$$

Hence $jqv_3 = kv_3 \in \Theta$ as well. ♠

Here is a related result.

Lemma 17.8 *Any lattice point on a wall line is equivalent to* $(0,0)$ *mod* Θ.

Proof: By symmetry, it suffices to consider the cases when $(m,n) \in L_0$.

Looking at Figure 4.1, we see that any point on L_0 has the form

$$sv_5 = \frac{s}{2(p+q)}(2pq, (p+q)^2 - 2p^2). \tag{193}$$

In order for this point to lie in $\boldsymbol{Z}^2$, the first coordinate must be an integer. Since p and q are relatively prime, pq and $p+q$ are relatively prime. Hence, the first coordinate is an integer only if $s = k(p+q)$ for some $k \in \boldsymbol{Z}$. Hence (m,n) is an integer multiple of the point

$$(p+q)v_5 = \left(pq, \frac{(p+q)^2}{2} - p^2\right) = 2V' + pV \in \Theta.$$

Here V and V' are the vectors generating Θ, as in Equation 180. ♠

The vertical lines in the door grid have the form $x = kq$ for $k \in \boldsymbol{Z}$. Say that a *Type 1 door* is the intersection of such a line with a wall line.

Lemma 17.9 *Let* (kq, y) *be a Type 1 door. Then* $py \in \boldsymbol{Z}$.

Proof: The group Θ acts transitively, by integer translations, on the vertical lines of the door grid. Hence, is suffices to prove this lemma for the case $k = 0$. In other words, we need to show that $py \in \boldsymbol{Z}$ if $(0, y)$ lies on a wall line.

We order the wall lines according to the order in which they intersect the line of slope $-A = -p/q$ through the origin. Let y_n be such that $(0, y_n)$ lies on the kth wall line. The sequence $\{y_n\}$ is an arithmetic progression. Hence, it suffices to prove our result for two consecutive values of n. Note that $(0,0)$ is a type A door. We might as well normalize so that $y_0 = 0$. Then $(0, y_1)$ lies on the wall line L_1 through $(-q, p)$. Referring to Equation 11, two points on L_1 are $-V$ and $-V + W$. These points are given by

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

From this information, we compute that $y_1 = (p+q)^2/2p$. Since $p+q$ is even, $py_1 = (p+q)^2/2 \in \boldsymbol{Z}$. ♠

Recall that $\underline{y}$ is the greatest integer less than y.

Corollary 17.10 *Suppose that (x, y) is a door of type 1, then $y - \underline{y} \neq 1/2$.*

Proof: $p(y - \underline{y}) = p/2$ is an integer, by the previous result. But $p/2$ is not an integer. This is a contradiction. ♠

Say that a *Type 2* door is a door on L_0 that is not of Type 1. One obtains a Type 2 door by intersecting L_0 with a line of the door grid that is parallel to the top left (or right) edge of the arithmetic kite.

Lemma 17.11 *The Type 2 doors are precisely the points on L_0 of the form (kp, y_k), where $k \in \boldsymbol{Z}$ and y_k is a number that depends on k.*

Proof: Referring to Figure 4.1, two consecutive doors on L_0 are $(0, 0)$ and $v_3 = (p, y_1)$. Our lemma now follows from the fact that the sequence of doors on L_0 forms an arithmetic progression. ♠

18 The Odd Hexagrid Theorem I

18.1 The Key Result

We fix an odd rational parameter $A = p/q$.

Say that a *floor line* is a negatively sloped line of the floor grid. Say that a *floor point* is a point on a floor line. Such a point need not have integer coordinates. Let M_+ and M_- denote the maps from Equation 10.6.3.

Lemma 18.1 *If (m,n) is a floor point, then $M_-(m,n)$ is equivalent mod Λ to a point of the form $(\beta,0,0)$.*

Proof: The map M_- is constant when restricted to each floor line, because these lines have slope $-A$. Hence, it suffices to prove this result for one point on each floor line. The points

$$(0,t); \qquad t = \frac{k(p+q)}{2}; \qquad k \in \boldsymbol{Z}. \tag{194}$$

form a sequence of floor points, one per floor line. 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 our reduction algorithm from §15.2. We have

1. $z = t$.
2. $Z = \text{floor}(t) = t$, because t is an integer.
3. $y = 2t = k(p+q) = kq(1+A)$.
4. $Y = \text{floor}(t/(1+A)) = kq$.
5. $x = y - (1-A)Y - 1 = \beta$..
6. $X = 0$.

Hence

$$M_-(0,t) = (x-(1+A)X, y-(1+A)Y, z-Z) = (\beta,0,0), \tag{195}$$

for some number $\beta \in \boldsymbol{R}$ that depends on A and k. ♠

18.2 Two Special Planes

Let $\Pi_- \subset \boldsymbol{R}^3$ denote the plane given by $y = z$. We can think of Π_- as the plane through the origin generated by the vectors $(1,0,0)$ and $(1,1,1)$. In particular, the vector $(1,1,1)$ is contained in Π_-. Let $\Pi_-(0)$ denote the line through the origin parallel to $(1,0,0)$. Then $\Pi_-(0)$ is a line in Π_-. Define

$$\Pi_+ = \Pi_- + (1,1,0); \qquad \Pi_+(0) = \Pi_-(0) + (1,1,0). \tag{196}$$

Lemma 18.2 *If (m,n) is a floor point, then $M_\pm(m,n)$ is equivalent mod Λ to a point in $\Pi_\pm(0)$.*

Proof: The $(-)$ case of this result is just a restatement of Lemma 18.1. The $(+)$ case follows from the $(-)$ case and symmetry. That is, we just translate the $(-)$ case by the vector $(1,1,0)$ to get the $(+)$ case. ♠

Define

$$\Pi_\pm(r) = \Pi_\pm(0) + (r,r,r). \tag{197}$$

Let $\Pi_\pm(r,s)$ denote the open infinite strip that is bounded by $\Pi_\pm(r)$ and $\Pi_\pm(s)$. In the case of interest to us, we will have $r = 0$ and $s = \lambda > 0$.

For each pair $(\epsilon_1,\epsilon_2) \in \{-1,0,1\}^2$, let $\Sigma(\epsilon_1,\epsilon_2)$ denote the set of lattice points (m,n) such that (m,n) and $(m,n)+(\epsilon_1,\epsilon_2)$ are separated by some floor line. The set $\Sigma(\epsilon_1,\epsilon_2)$ is obtained by intersecting $\boldsymbol{Z}^2$ with an infinite union of evenly spaced infinite strips, each of which has a floor line as one boundary component. For our purposes, it suffices to consider the pairs

$$(-1,0); \qquad (-1,-1); \qquad (0,-1); \qquad (1,-1). \tag{198}$$

For these pairs, the floor lines are the lower boundaries of the strips. We define

$$\lambda(\epsilon_1,\epsilon_2) = -(A\epsilon_1 + \epsilon_2). \tag{199}$$

Lemma 18.3 *Let $\lambda = \lambda(\epsilon_1,\epsilon_2)$. Suppose that $(m,n) \in \Sigma(\epsilon_1,\epsilon_2)$. Then $M_\pm(m,n) \in \Pi_\pm(0,\lambda)$.*

Proof: We consider the case of M_- and the pair $(-1,0)$. The other cases have essentially the same proof. If $(m,n) \in \Sigma(-1,0)$, Then there is some $x \in (m-1,m)$ such that (x,n) is a floor point. Then $M_+(x,n)$ is Λ-equivalent to a point p in $\Pi(0)$. But then $M_+(m,n)$ is Λ-equivalent to $p + (m-x)(A,A,A) \in \Pi(0,A) = \Pi(0,\lambda(-1,0))$. ♠

18.3 Critical Points

Say that a point $v \in \Sigma(\epsilon_1, \epsilon_2)$ is *critical for* (ϵ_1, ϵ_2) if the arithmetic graph contains the edge joining (m, n) to $(m + \epsilon_1, n + \epsilon_2)$. Statement 1 of the Hexagrid Theorem says, in particular, that there are no such points like this.

Lemma 18.4 *There are no critical points.*

Proof: Let $\mathcal{R}_+$ denote the tiling of $\boldsymbol{R}^3$ by polyhedra, according to the Master Picture Theorem. Let $\mathcal{P}_+$ denote the intersection of $\mathcal{R}_+$ with the plane Π. We make the same definitions in the $(-)$ case. If (m, n) is critical for (ϵ_1, ϵ_2), then one of two things is true.

1. $\Pi_+(0, \lambda)$ nontrivially intersects a polygon of $\mathcal{P}_+$ labelled by (ϵ_1, ϵ_2).
2. $\Pi_-(0, \lambda)$ nontrivially intersects a polygon of $\mathcal{P}_-$ labelled by (ϵ_1, ϵ_2).

Here we have set $\lambda = \lambda(\epsilon_1, \epsilon_2)$. Considering the 4 pairs of interest to us, and the 2 possible signs, we have 8 conditions to rule out. We check, in all cases, that the relevant strip is disjoint from the relevant polygons.

We can check the disjointness for all parameters at once. The union

$$S_{\pm}(\epsilon_1, \epsilon_2) := \bigcup_{A \in (0,1)} \Big(\Pi_{\pm}(\epsilon_1, \epsilon_2; A) \times \{A\} \Big)$$

is a polyhedral subset of $\boldsymbol{R}^4$. To get an honest polyhedron, we observe that S is invariant the action of the lattice element γ_1 from Equation 75, and we take a polyhedron whose union under translates by γ_3 tiles S. In practice, we simply restrict the x-coordinate to lie in $[0, 2]$.

We check that $S_{\pm}(\epsilon_1, \epsilon_2)$, or rather the compact polyhedron replacing it, is disjoint from all Λ-translates of the polytope $P_{\pm}(\epsilon_1, \epsilon_2)$, the polytope listed in §15.3. In practice, most translates are very far away, and we only need to check a small finite list. This is a purely algebraic calculation. ♠

Rather than dwell on the disjointness calculation, which gives no insight into what is going on, we will draw pictures for the parameter $A = 1/3$. The combinatorial type changes with the parameter, but not the basic features of interest to us. The interested reader can see the pictures for any parameter using Billiard King.

To draw pictures, we identify the planes $\Pi_\pm$ with $\boldsymbol{R}^2$ using the projection $(x, y, z) \to (x, (y+z)/2)$. Under this identification, all the polygons in question are rectangles! The coordinates of the rectangle vertices are small rational combinations of 1 and A, and can easily be determined by inspection. The whole picture is invariant under translation by $(1+A, 0)$. The thick line in the first picture corresponds to $\Pi_-(0)$. In terms of $\boldsymbol{R}^2$ coordinate, this is the x-axis. The black dot is $(0,0)$. dot is $(4/3, 0) = (1 + A, 0)$.

We explain by example the notation on the right hand side of the fiture. The label $\lambda(-1,-1)$ denotes the line $\Pi(\lambda)$, where $\lambda = \lambda(-1,-1)$. In each case, the relevant strip lies below the relevant shaded piece. While the combinatorics of the picture changes as the parameter changes, the basic disjointness stays the same.

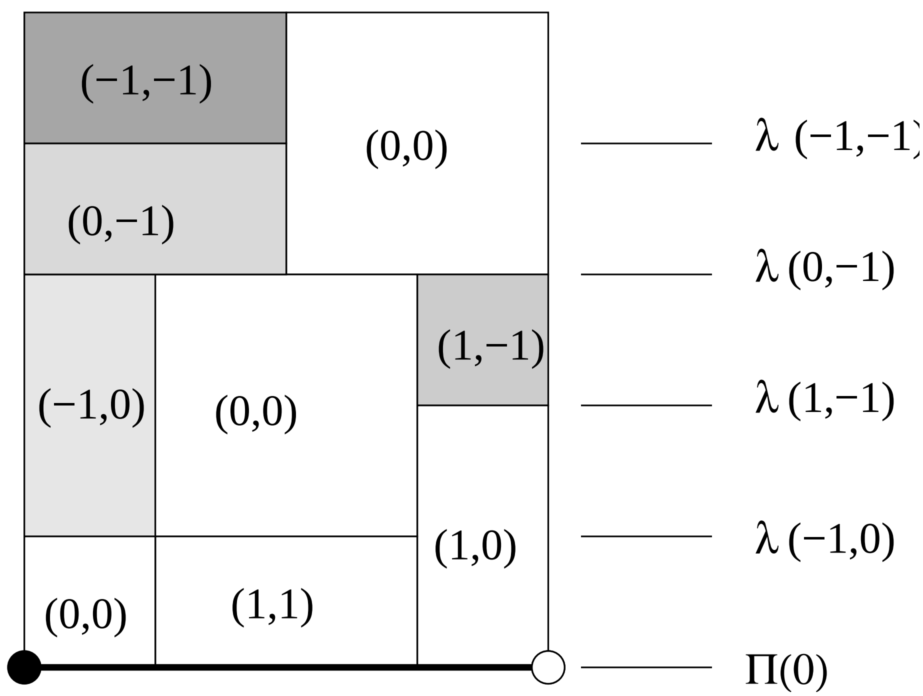

Figure 19.1: The $(-)$ picture for $A = 1/3$.

Figure 19.2 shows the same thing for the $(+)$ case. This time the black dot is $(1/2, 1/2)$ and the white dot is $(1/2, 1/2) + (1 + A, 0)$. The thick line represents $\Pi_+(0)$. In $\boldsymbol{R}^2$ coordinates, this is the line $y = 1/2$. In the $(+)$ case is isn't even a close call.

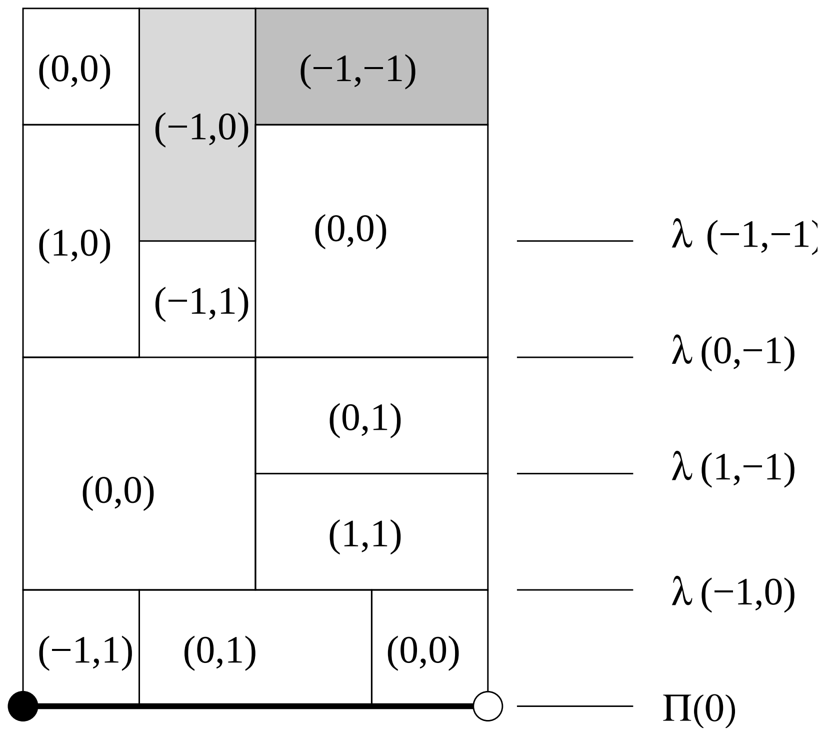

Figure 19.2: The (+) picture for $A = 1/3$.

18.4 The End of the Proof

Now we know that there are no critical points. The only other way that the arithmetic graph could cross a floor line would be at a floor point that was also a lattice point. It might happen that one edge emanating from such a floor point lies above the floor line, and the other lies below.

Define

$$\zeta_k = \left(0, \frac{k(p+q)}{2}\right); \qquad k \in \boldsymbol{Z}. \tag{200}$$

Lemma 18.5 *Modulo the symmetry group Θ, the only lattice floor points are the ones listed in Equation 200.*

Proof: If (m,n) is a lattice floor point, then $2Am + 2n \in \boldsymbol{Z}$. But his means that q divides m. Subtracting off a suitable multiple of $V = (q, -p) \in \Theta$, we can arrange that the first coordinate of our lattice floor point is 0. But, now we must have one of the points in Equation 200. ♠

The slices as shown in Figure 10.3 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 §10.5, 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 our partitions. Figure 18.1 does for these slices what Figure 10.3 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.

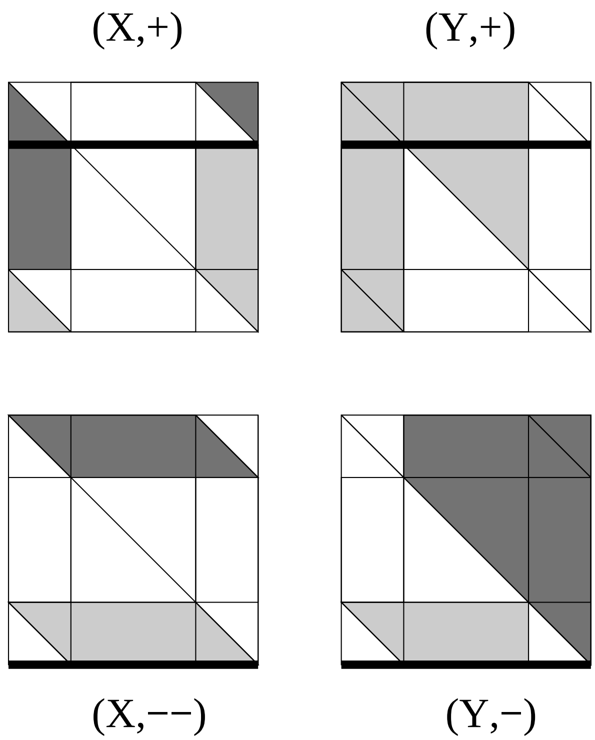

Figure 18.1: The $(0, A)$ slices.

From this picture 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 halfplane above the relevant floor line. This what we wanted to establish.

19 The Odd Hexagrid Theorem II

19.1 The Basic Definitions

It turns out that the secret to proving Statement 2 of the Hexagrid Theorem is to use variants of the maps M_+ and M_- from Equation 10.6.3. Let $A \in (0,1)$ be any parameter. Let Λ the lattice from the Master Picture Theorem. Let $\Pi \subset \boldsymbol{R}^3$ be the plane defined by the relation $x+y=A$.

For $(m,n) \in \boldsymbol{R}^2$ we define $\Delta_+(m,n)=(x,y,z)$, where

$$x = 2A(1-m+n)-m; \qquad y = A-x; \qquad z = Am. \tag{201}$$

We also define

$$\Delta_-(m,n) = \Delta_+(m,n) + (-A,A,0). \tag{202}$$

Note that $\Delta_\pm(m,n) \in \Pi$. Indeed, Δ is an affine isomorphism from $\boldsymbol{R}^2$ onto Π.

Lemma 19.1 *Suppose that $(m,n) \in \boldsymbol{Z}^2$. 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, $v_1 = (1+A,0,0)$ and $v_1 = (1-A,1+A,0)$ and $v_3 = (-1,-1,1)$. Let

$$c_1 = -1+2m; \qquad c_2 = 1-m+2n; \qquad c_3 = n.$$

We compute directly that

$$M_+(m,n) - \Delta_+(m,n) = c_1v_1 + c_2v_2 + c_3v_3.$$

$$M_-(m,n) - \Delta_-(m,n) = c_1v_1 + (c_2-1)v_2 + c_3v_3.$$

This completes the proof. ♠

We introduce the vector

$$\zeta = (-A,A,1) \in \Lambda. \tag{203}$$

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.

19.2 Interaction with the Hexagrid

Now we will specialize to the case when $A = p/q$ is an odd rational. The results above hold, and we can also define the hexagrid. We will see how the maps Δ_+ and Δ_- interact with the Hexagrid. Let L_0 denote the wall line through the origin.

Lemma 19.2 *$\Delta_\pm(L_0)$ is parallel to ζ and contains $(-2A, A, 0)$.*

Proof: We refer to the points in Figure 4.1. The points v_5 and v_1 both lie on L_0. We compute

$$\Delta_+(v_5) - \Delta_+(v_1) = \frac{p^2}{p+q}\zeta.$$

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,

$$\Delta_+(0,0) \subset \Pi(2A); \qquad \Delta_-(0,0) \subset \Pi(A). \tag{204}$$

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 lattice points (m, n) such that L_0 separates (m, n) from $(m+\epsilon_1, n+\epsilon_2)$. Now we define constants

$$\lambda(0,1) = 2A \qquad \lambda(-1,-1) = 1 - A^2;$$

$$\lambda(-1,0) = 1 + 2A - A^2 \qquad \lambda(-1,1) = 1 + 4A - A^2 \tag{205}$$

Lemma 19.3 *Let (ϵ_1, ϵ_2) be any of the 4 pairs listed above. Let $\lambda = \lambda(\epsilon_1, \epsilon_2)$. The following 3 statements are equivalent.*

1. *$(m, n) \in \Sigma(\epsilon_1, \epsilon_2)$.*
2. *$\Delta_+(m, n)$ is congruent mod Λ to a point in the interior of $\Pi(2A-\lambda, 2A)$.*
3. *$\Delta_-(m, n)$ is congruent mod Λ to a point in the interior of $\Pi(A-\lambda, A)$.*

Proof: The formula $\Delta_- = \Delta_+ + (-A, A, 0)$ immediately implies the equivalence of the second and third statements. So, it suffices to prove the equivalence of the first two statements. We will consider the pair $(-1, 0)$. The other cases have the same treatment. The set $\Sigma(-1, 0)$ is the intersection of $\boldsymbol{Z}^2$ with the interior of some infinite strip, one of whose boundaries is L_0. To find the image of this strip under Δ_+, we just have to see what Δ_+ does to two points, one per boundary component of the strip. We choose the points $(0, 0)$ and $(1, 0)$. We already know that $\Delta_+(0, 0) \subset \Pi(2A)$. We just have to compute $\Delta_+(1, 0)$. We compute

$$\Delta(1, 0) = (1, A-1, A) \subset \Pi(1 - A^2).$$

This gives us $\lambda(-1, 0) = 1 + 2A - A^2$. Our lemma follows from this fact, and from the fact that Δ_+ is an affine isomorphism from $\boldsymbol{R}^2$ to Π. ♠

19.3 Determining the Local Picture

A crossing cell can consist of either 1 edge or 2, depending on whether or not a vertex of the cell lies on a wall line. According to Lemma 17.8, the only crossing cells with one edge are equivalent mod Θ to the one whose center vertex is $(0, 0)$. For these *special* crossing cells, Statement 2 of the Hexagrid Theorem is obvious. The door is just the central vertex.

The remaining crossing cells are what we call *generic*. Each generic crossing cell has one vertex in one of our sets $\Sigma(\epsilon_1, \epsilon_2)$, for one of the 4 pairs considered above. We call $v \in \Sigma(\epsilon_1, \epsilon_2)$ a *critical for* (ϵ_1, ϵ_2). v and $v + (\epsilon_1, \epsilon_2)$ are the two vertices of a crossing cell. To prove Statement 2 of the Hexagrid Theorem, we need to understand the critical vertices. This means that we need to understand the local picture of the arithmetic graph in terms of the maps Δ_+ and Δ_-.

We want to draw pictures as in the previous chapter, but here we need to be more careful. In the previous chapter, our plane Π contained the vector $(1, 1, 1)$. Thus, we could determine the structure of the arithmetic graph just by looking at the intersection $\Pi \cap \mathcal{R}$. Here $\mathcal{R}$ is the polyhedron partition for the given parameter. The situation here is different. The vector $(1, 1, 1)$ is transverse to the plane Π. What we really need to do is to understand the way that the plane Π_α intersects the our partition. Here Π_α is the plane satisfying the equation $x + y = A + 2\alpha$. We think of α an infinitesimally small but positive number. More formally, we take the geometric limit of the set $\Pi_\alpha \cap \mathcal{R}$ as $\alpha \searrow 0$.

We say that a subset $S \subset \Pi$ is *painted* $(\epsilon_1, \epsilon_2, +)$ if $\Delta_+(m,n) \in S$ implies that $\Delta_+(m,n)$ determines the pair (ϵ_1, ϵ_2). This is to say that S is contained in the Hausdorff limit of $\Pi_\alpha \cap \mathcal{R}_+(\epsilon_1, \epsilon_2)$ as $\alpha \to 0$. We make the same definition with $(+)$ in place of $(-)$. We think of $(\epsilon_1, \epsilon_2, \pm)$ as a kind of color, because these regions are assigned various colors in Billiard King. For instance $(0,1,\pm)$ is green. There is essentially one *painting* of Π for $(+)$ and one for $(-)$.

To visualize the painting, we identify Π with $\boldsymbol{R}^2$ using the map $(x,y,z) \to (x,z)$. We just drop the second coordinate. The vector ζ maps to the $(-A,1)$. Thus, our whole painting is invariant under translation by this vector. Each wall of $\mathcal{R}$ intersects Π in a line segment whose image in $\boldsymbol{R}^2$ is either horizontal or vertical. The endpoints of each such segment have coordinates that are simple rational combinations of 1 and A. For this reason, we can determine the intersection we seek just by inspecting the output from Billiard King. In practice, we take $\alpha = 10^{-5}$, examine the resulting picture, and then adjust the various vertices slightly so that their coordinates are small rational combinations of 1 and A.

19.4 An Extended Example

We consider the pair $(0,1)$ in detail. We will draw pictures for the parameter $A = 1/3$, though the same argument works for any parameter. There is no polyhedron $\mathcal{R}_-(0,1)$, so no points are painted $(0,1,-)$. The interesting case is $(0,1,+)$. First of all, we only care about points in our strip $\Sigma(0,1)$. So, we only need to understand the portion of our painting that lies in our strip $\Pi(0,2A)$. In $\boldsymbol{R}^2$ (considered as the xz plane), our strip is bounded by the lines $x = -zA$ and $x = -zA + 2A$.

We will first study the picture when $z = 0$. Referring to Figure 19.1, the shaded triangles correspond to $\mathcal{R}(0,1)$. The thick line corresponds to the intersection of Π with our fiber. The black dot is the point $(A,0,0,A)$. Moving away from the black dot, the white dots are

$$(0,A,0,A); \quad (-A,2A,0,A); \quad (-1,1+A,0,A.$$

It we move the thick line an infinitesimal amount in the direction of $(1,1)$, we see that it crosses though a shaded region whose diagonal edge is bounded by the points $(0, A)$ and $(A, 0)$. The only tricky part of the analysis is that the point $(A, 0)$ determines the pair $(0, 0)$ and the point $(0, A)$ determines the pair $(-1, 1)$.

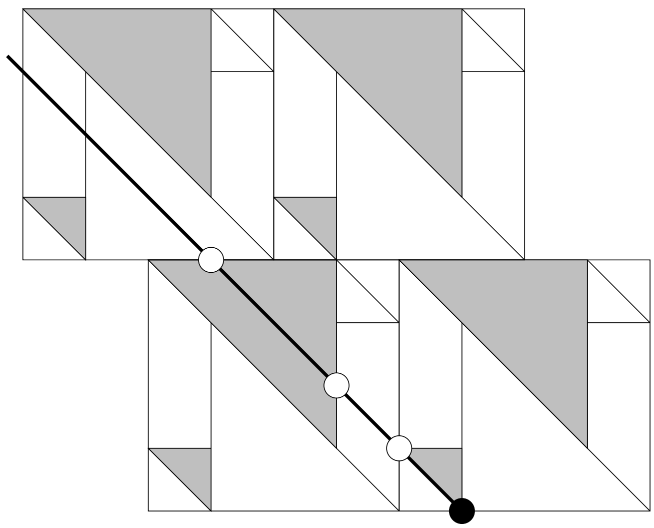

Figure 19.1: Slicing the 0 fiber.

From this discussion, we conclude that $(0, A) \times \{0\}$ is painted $(0, 1, +)$. For later use, we remark that $(0, 0)$ is painted $(-1, 1, +)$ and $(A, 0)$ is painted $(0, 0, +)$. Looking at the picture, we also see that $(-1, -A) \times \{0\}$ is painted $(0, 1, +)$. Notice, however, that this set lies outside our strip. It is irrelevant.

Figure 19.2 shows the picture for a typical parameter $z \in (0, 1)$. We choose $z = 1/6$, though the features of interest are the same for any choice of z. The interested reader can see essentially any slice (and in color) using Billiard King.

The black dot and the white dots have the same coordinates as in Figure 19.1. Notice that the point $(0, A, z, A)$ lies at the bottom corner of a shaded region. This remains true for all z. We conclude that the open line segment $\{0\} \times (0, 1)$ is painted $(0, 1, +)$. Similarly, the rectangle $(-1, -A) \times [0, 1]$ is painted $(0, 1, +)$. However, this rectangle is disjoint from the interior of our strip. Again, it is irrelevant.

Recalling that our painting is invariant under translation by $(0, 1, +)$, we can now draw the portion of plane painted $(0, 1, +)$ that is relevant to our analysis. To give the reader a sense of the geometry, we also draw one copy of the irrelevant rectangle. Again, we draw the picture for the parameter

$A = 1/3$. The interested reader can see the picture for any parameter using Billiard King.

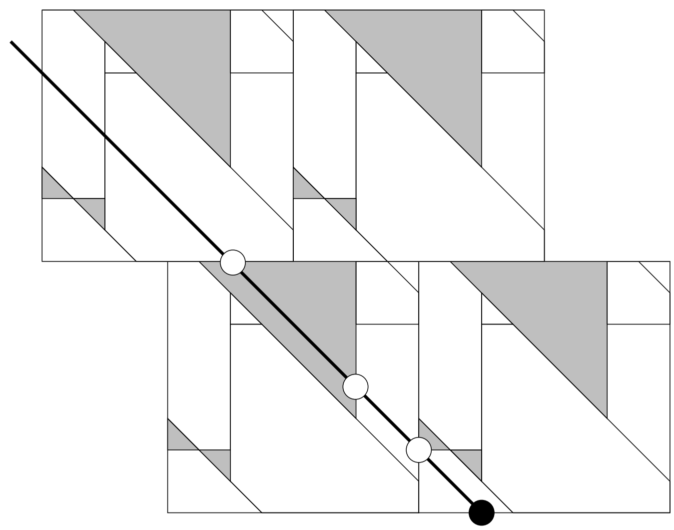

Figure 19.2: Slicing a typical fiber.

In Figure 19.3, the arrow represents the vector $(-A, 1)$. The black dot is $(0, 0)$ and the white dot is $(A, 0)$. The thick zig-zag, which is meant to go on forever in both directions, is the relevant part of the painting. The lightly shaded region is the strip of interest to us.

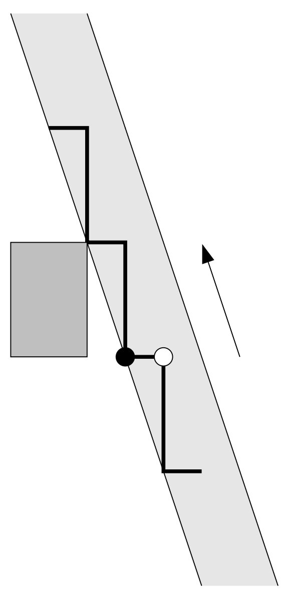

Figure 19.3: The relevant part of the $(0, 1, +)$ painting.

19.5 The Rest of the Painting

We determine the rest of the painting using the same techniques. The interested reader can see everything plotted on Billiard King. The left side of igure 19.4 shows the relevant part of the $(+)$ painting. The right side shows the relevant part of the $(-)$ painting. The dots are exceptional points in the painting. The two grey dots at the endpoints correspond to the right endpoints of the special crossing cells. We have shown a "fundamental domain" for the paintings. The whole painting is obtained taking the orbit under the group $\langle\zeta\rangle$. In our picture, ζ acts as translation by the vector $(-A, 1)$, because we are leaving off the y coordinate. In particular, the two endpoints of the L are identified when we translate by this group.

The small double-braced labels, such as $((0, 1))$, indicate the paint colors. The large labels, such as $(0, 0)$, indicate the coordinates in the plane. Note that the point (x, z) in the plane actually corresponds to $(x, A - z, y)$ in Π. The grey vertices on the left corresponds to $\Delta_+(0, 0)$. The grey vertices correspond to the various images of points on the special crossing cell. These vertices are not relevant to our analysis of the points that are critical relative to our 4 pairs.

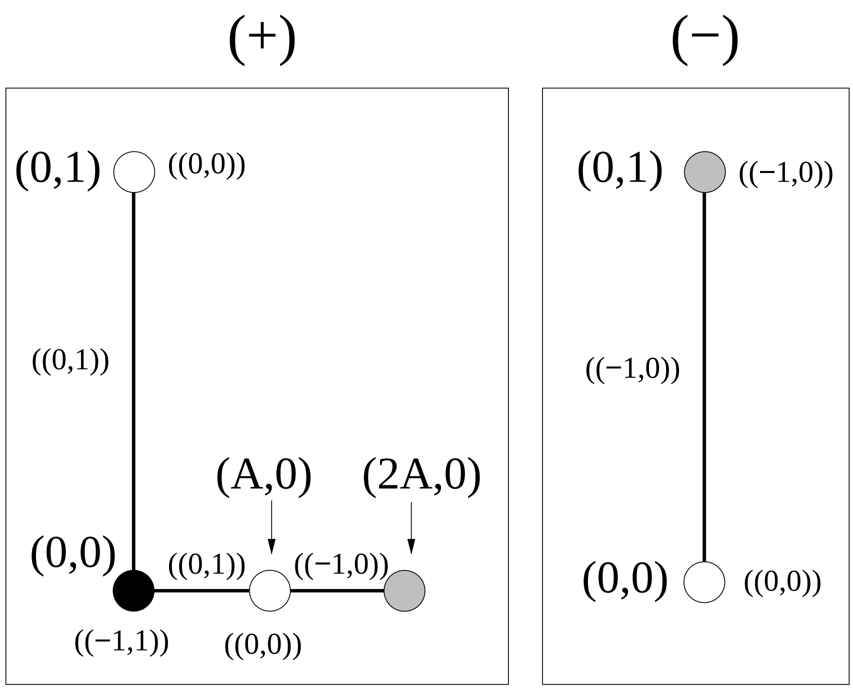

Figure 19.4: The relevant part of the $(+)$ painting

Say that a vertex $v = (m, n)$ is *critical* if it either lies in Θ or else is critical for one of our strips. The point $(0,0)$ is the center vertex of a special crossing cell. Hence, By Lemma 17.8, the critical vertices are in bijection with the crossing cells. Given our analysis above, we see that $v \in \boldsymbol{Z}^2$ is *critical* if and only if it satisfies the following criterion. *Modulo the action of* Λ*, the point* $\Delta_+(v)$ *(respectively the point* $\Delta_-(v)$*) lies in one of the colored parts of the painting on the left (respectively right) in Figure 19.4.*

Recalling that $\Delta_+ = \Delta_- + (A, -A, 0)$, we can eliminate Δ_- from our discussion. We translate the right hand side of Figure 19.4 by $(A, 0)$ and then superimpose it over the left hand side. (This translation does not reflect the way the two halves of Figure 19.4 are related to each other on the page.) See Figure 19.5. The result above has the following reformulation.

Lemma 19.4 (Critical) *A vertex* v *is critical if and only of* $\Delta_+(v)$ *is equivalent mod* Λ *to a point colored portion of Figure 19.5.*

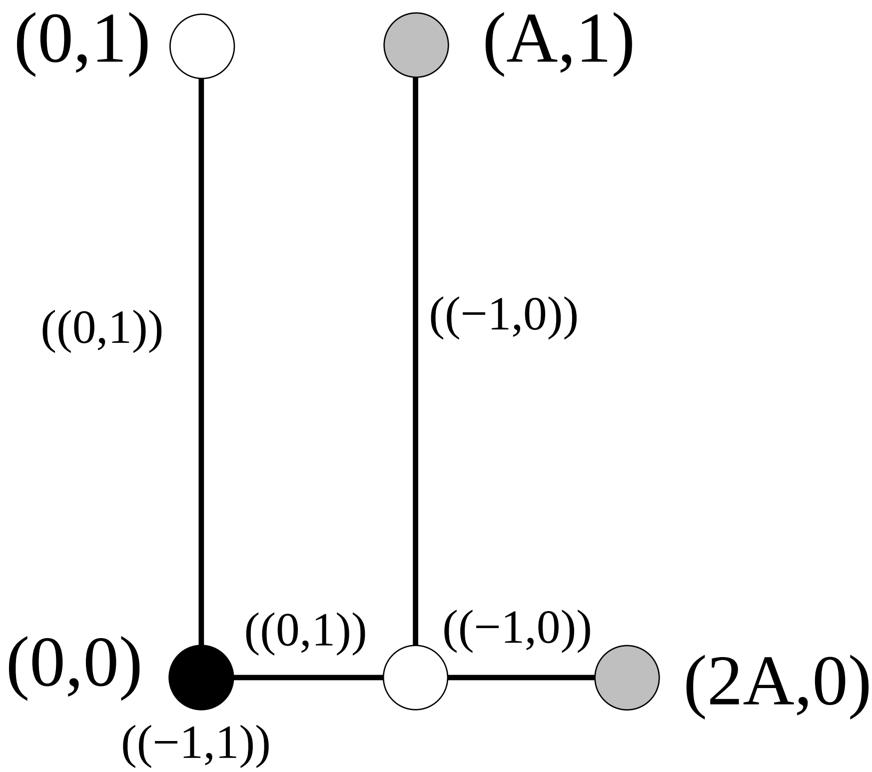

Figure 19.5: Superimposed paintings

Our drawing of Figure 19.5 somewhat hides the symmetry of our picture. In Figure 19.6, we show several translates of this fundamental domain at the same time, without the labels. We also show the strip $\Pi(0, 2A)$. The pattern is meant to repeat endlessly in both directions. The line on the left is $\Pi(0)$ and the line on the right is $\Pi(2A)$. Again, we are drawing the picture for the parameter $A = 1/3$. The combinatorial pattern is the same for any A.

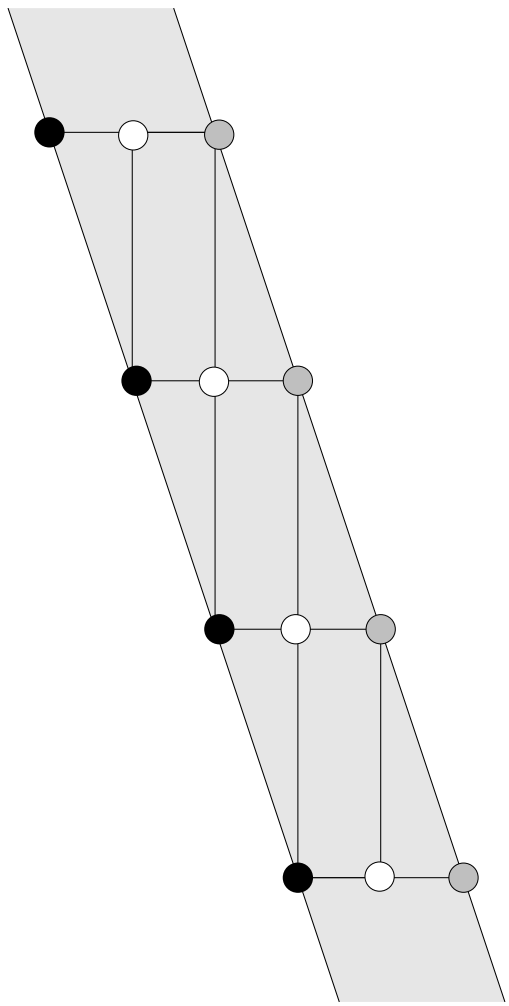

Figure 19.6: Superimposed paintings

To prove the hexagrid theorem, it only remains to identify the lattice points in the Critical Lemma with the doors from the Hexagrid Theorem.

19.6 The End of the Proof

Now we interpret the Critical Lemma algebraically. A vertex $v \in \boldsymbol{Z}^2$ is *critical* if and only if $\Delta_+(v)$ is equivalent mod Λ to one of the following kinds of points.

1. $(2A, -A, 0)$.
2. $(0, A, 0)$.
3. $(x, A - x, 0)$, where $x \in (0, 2A) - \{A\}$.
4. $(0, A, z)$, where $z \in (0, 1)$.

As we point out in our subsection headings, each case corresponds to a different feature of our painting in Figures 19.5 and 19.6

19.6.1 Case 1: The Grey Dots

Note that $\Delta_+(0,0) = (2A, -A, 0)$. Moreover, $\Delta_+(v) \equiv \Delta_+(v')$ mod Λ iff $M_+(v) \equiv M_-(v)$ mod Λ iff $v \equiv v'$ mod Θ. Hence, Case 1 above corresponds precisely to the special crossing cells. The door associated to v is precisely v. In this case, the door is associated to the wall above it.

19.6.2 Case 2: The Black Dots

Note that $\Delta_+(0, -1) = (0, A, 0)$. Hence the second case occurs iff v is equivalent mod Θ to $(0, -1)$. But $(0, -1)$ is the vertex of a crossing cell whose other vertex is $(-1, 0)$. The door associated to $(0, -1)$ is $(0, 0)$. In this case, the door is associated to the wall below it.

19.6.3 Case 3: Horizontal Segments

We are going to demonstrate the bijection between the Type 1 doors not covered in Cases 1 and 2 and the critical points that arise from Case 3 above.

Let v be a critical point. Using the symmetry of Θ, we can arrange that our point v is closer to L_0 than to any other wall line. In this case, $\Delta_+(v)$ lies in the strip $\Pi(0, 2A)$. Hence $v \in \Sigma(0, 1)$. Hence L_0 separates v from $v + (0, 1)$. Let $y \in (n, n + 1)$ be such that $(m, y) \in L_0$.

The third coordinate of $\Delta_+(v)$ is an integer. Setting $v = (m, n)$, we see that $Am \in \boldsymbol{Z}$. Hence q divides m. Hence $v = (kq, n)$ for some $k \in \boldsymbol{Z}$. Hence (kq, y) is a Type 1 door.

For the converse, suppose that the point (kq, y) is a Type 1 door and that $n = \underline{y}$. Let $v = (kq, n)$. We want to show that v is critical. By construction $(kq, n) \in \Sigma(0, 1)$. Hence $\Delta_+(kq, n) \in \Pi(0, 2A)$. But the third coordinate of $\Delta_+(kq, n)$ is an integer. Hence $\Delta_+(kq, n)$ is equivalent mod Λ to a point of the form $(x, A - x, 0)$. Here $x \in (0, 2A)$.

If $x = A$ then v lies on the centerline of the strip $\Sigma(0, 1)$. But then $y - \underline{y} = 1/2$. This contradicts Lemma 17.10. Hence $A \neq x$.

Now we know that $\Delta_+(kq, n)$ satisfies Case 3 above. Hence (kq, n) is critical, either for $(0, 1)$ or for $(-1, 0)$. These are the relevant labellings in Figure 19.5. Note that L_0 has positive slope greater than 1. Hence

$$(kq, n) \in \Sigma_+(0, 1) \cap \Sigma_+(-1, 0).$$

If $\Delta'_+(v)$ is colored $(0, 1)$, then v is critical for $(0, 1)$. If $\Delta'_+(v)$ is colored $(-1, 0)$, then v is critical for $(-1, 0)$. So, v is always vertex of a crossing cell.

19.6.4 Case 4: Vertical Segments

We are going to demonstrate the bijection between Type 2 doors and the critical points that arise from Case 4 above.

We use the symmetry of Θ to guarantee that our critical point is closer to L_0 than to any other wall line. As in Case 3, the point $\Delta_+(v) \in \Pi(0, A)$. Hence $v \in \Sigma(0, 1)$ and L_0 separates v from $v + (0, 1)$. We define y as in Case 3. We want to show that $(m, y) \in L$ is a Type 2 door.

Since we are in Case 4, the first coordinate of $\Delta_+(v)$ lies in $A\boldsymbol{Z}$. The idea here is that $\Delta_+(v)$ is equivalent mod $(-A, A, 1)$ to a point whose first coordinate is either 0 or A. Hence

$$x = 2A(1 - m + n) - m \in A\boldsymbol{Z}.$$

Hence $x/A \in \boldsymbol{Z}$. Hence $m/A \in \boldsymbol{Z}$. Hence $m = kp$. By Lemma 17.11, the point (x, y) is a door.

Conversely, suppose that the point (kp, y) is door contained in L_0. Let $n = \underline{y}$. Then $(kp, n) \in \Sigma(0, 1)$ and the first coordinate of $\Delta_+(kp, n)$ lies in the set $A\boldsymbol{Z}$. Also, $\Delta_+(kp, n) \in \Pi(0, 2A)$. Hence, (kp, n) satisfies Case 4 above. Hence (kp, n) is critical for either $(0, 1)$ of $(-1, 0)$. In either case, (kp, n) is a vertex of a crossing cell.

19.7 The Pattern of Crossing Cells

Our proof of the Hexagrid Theorem is done, but we can say more about the nature of the crossing cells. First of all, there are two crossing cells consisting of edges of slope ± 1. These crossing cells correspond to the black and grey corner dots in Figure 19.6.

The remaining crossing cells involve either vertical or horizontal edges. These crossing cells correspond to the interiors of the segments in Figures 19.5 and 19.6. Let $v = (m, n)$ be the critical vertex associated to the door (m, y). Then v is critical either for $(0, 1)$ or $(-1, 0)$. In the former case, the crossing cell associated to v is vertical, and in the latter case it is horizontal. Looking at the way Figure 19.5 is labelled, we see that

- The crossing cell is vertical if $y - \underline{y} > 1/2$.
- The crossing cell is horizontal if $y - \underline{y} < 1/2$.

The case $y - \underline{y} = 1/2$ does not occur, by lemma 17.10.

There are exactly $p + q$ crossing cells mod Θ. These cells are indexed by the value of $y - n$. The possible numbers are

$$\{0, \frac{1}{p}, ..., \frac{p-1}{p}, \frac{1}{q}, ... \frac{q-1}{q}, 1\}.$$

In all cases we have $y - n = y - \underline{y}$, except when $n = y - 1$.

Figure 19.7 shows, for the case $p/q = 3/5$, the images of the critical vertices, on one fundamental domain for Figure 19.6. (The fundamental domain here is nicer than the one in Figure 19.5.) We have labelled the image points by the indices of the corresponding crossing cells. 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. The lines inside the dots show the nature of the crossing cell.

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 19.7 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.

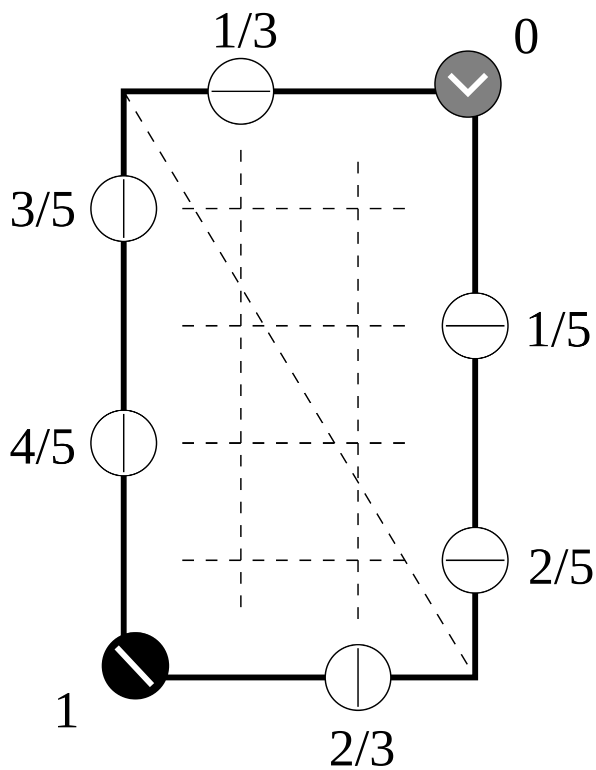

Figure 19.7: Images of the critical points

Figure 19.7 is representative of the general case. It is meant to suggest the general pattern. We hope that the pattern is clear.

20 The Even Hexagrid Theorem

20.1 Proof of Statement 1

The only place where we used the fact that p/q is odd was in Lemma 18.1. We needed to know that the number t in Equation 18.1 was odd. This no longer works when $p+q$ is odd. However, 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 18.1 is an even integer. Hence, for the floor lines in the even case, the number t is an integer. The rest of the proof of Lemma 18.1 works word for word. The rest of the proof of Statement 1 goes through word for word.

20.2 Symmetries Revisited

Before we get to Statement 2 of the Hexagrid Theorem, we need to investigate some symmetries. When p/q is an even rational, we redefine Θ to be the lattice generated by

$$V = (q, -p) \qquad\qquad V' = (0, 2(p+q)^2). \tag{206}$$

The even version of V' is 4 times the odd version. Essentially the same proof as in the odd case shows that both the arithmetic graph and the hexagrid are invariant under Θ in the even case.

20.3 The Door Structure Revisited

Lemma 20.1 *Any two wall lines are equivalent mod* Θ.

Proof: This is easy in the even case. Translation by V maps each wall line to the adjacent one. ♠

Lemma 20.2 *The first coordinate of any door is an integer.*

Proof: The first door on L_0 is the same in the even case as in the odd case. The rest of the proof is the same as in the odd case. ♠

Lemma 20.3 *Modulo the action of* Θ*, there are only two exceptional doors.*

Proof: As in the odd case, we just have to show that any door with integer coordinates is equivalent to $(0,0)$ mod Θ. As in the odd case, the doors on L_0 have the form kv_3. As in the odd case, this leads to the statement that

$$k\frac{q^2-p^2}{2q} \in \boldsymbol{Z}.$$

Now the proof is a bit different. Here $2q$ and q^2-p^2 are relatively prime. Hence $k = 2jq$ for some $j \in \boldsymbol{Z}$. But

$$2qv_3 = V' + 2V \in \Theta.$$

Hence $2jqv_3 = kv_3 \in \Theta$ as well. ♠

Lemma 20.4 *Any lattice point on a wall line is equivalent to* $(0,0)$ *mod* Θ*.*

Proof: As in the proof of Lemma 17.8, we see that

$$sv_5 = \frac{s}{2(p+q)}(2pq, (p+q)^2 - 2p^2) \in \boldsymbol{Z}. \tag{207}$$

As in the odd case, we look at the first coordinate and deduce the fact that $s = k(p+q)$ for some $k \in \boldsymbol{Z}$. This is not enough for us in the even case. Looking now at the second coordinate, we see that

$$\frac{k(p+q)}{2} - kp^2 \in \boldsymbol{Z}.$$

Hence k is even. Hence (m,n) is an integer multiple of the point

$$2(p+q)v_5 = (2pq, (p+q)^2 - 2p^2) = V' + 2pV \in \Theta.$$

♠

We don't repeat the proof of Lemma 17.9 because we don't need it. We only need the even version of Corollary 17.10. In the even case, we have simply forced Corollary 17.10 to be true by eliminating the crossings for which it fails.

We say that a *Type 2* door is a door on L_0 that is not of Type 1, and also is not one of the crossings we have eliminated. Once we make this redefinition, we have the following result

Lemma 20.5 *The Type 2 doors are precisely the points on L_0 of the form (kp, y_k), where $k \in \boldsymbol{Z}$ is not an odd multiple of q, and y_k is a number that depends on k.*

Proof: The Type two doors are as in the odd case, except that we eliminate the points (kp, y_k) where k is an odd multiple of q. ♠

20.4 Proof of Statement 2

At this point, all the constructions in the previous chapter go through word for word.

It appears that we used the fact that p/q is odd in Cases 3 and 4 at the end of the last chapter, but this isn't so. We only used the fact that Corollary 17.10 was true, and that Lemma 17.11 was true. At the time, we had only proved these results in the odd rational case. However, since these results hold in the even case, the arguments for Cases 3 and 4 go through word for word.

A final remark on Case 3: Case 3 required us to use Corollary 17.10 to rule out the possibility that the point (m, n) is equivalent mod Λ to the points $(A, 0, 0)$. This can happen in the even case, and indeed it happens when (m, y) is a crossing of the kind we are no longer calling a door. In other words, this does not happen for a *door* because we have forced the situation.

20.5 The Even Case of Theorem 1.4

Referring to the plane containing the arithmetic graph, let S_0 be the line segment connecting the origin to v_3, the very tip of the arithmetic kite. Then S_0 is bounded by two consecutive doors on L_0. The bottom endpoint of S_0 is $(0, 0)$, one of the vertices of $\Gamma(0, 0)$. We know already that $\Gamma(p/q)$ is a closed polygon. By the hexagrid Theorem, $\Gamma(p/q)$, cannot cross S_0 except within 1 unit of the door v_3. Hence, $\Gamma(p/q)$ must engulf all but the top 1 unit of S_0.

Essentially the same calculation as in the odd case now shows that $\Gamma(p/q)$ rises up at least $(p + q)$ units from the baseline when $p > 1$. When $p = 1$ the same result holds, but the calculation is a bit harder. The reason why we get an extra factor of 2 in the even case is that v_3 is twice as far from the baseline as is the door d_0 from Equation 13.

Part IV

In this part of the monograph we use the Master Picture Theorem to prove the Copy Theorems and the Decomposition Theorem, two technical results left over Part I. The Copy Theorems were discussed in §7. The Decomposition Theorem was discussed in §5.3.

- In §21 we prove the Copy Theorems.
- In §22 we derive a technical consequence of the Copy Theorems, which we call the Induction Lemma. This result parallels Lemma 6.2, a structural result about the sequence of odd rationals we constructed in §6.
- In §23 we prove the Decomposition Theorem for parameters of the form $(q-2)/q$. This is the base case of what will be an inductive argument.
- In §24 we prove the Decomposition Theorem. Our proof is inductive, and the Induction Lemma provides the key step.

21 Proof of the Copy Theorems

21.1 The Goal

For ease of exposition, we will prove the Copy Theorem I. The Copy Theorem II has an essentially identical proof. Recall from §6 that $\lambda = q_+/q$, where $q_+ \in (0, q)$ is such that $pq+ \equiv -1 \bmod q$. Recall also that.

$$F(m,n) = (p,q).(m,n). \tag{208}$$

$$G(m,n) = \Big(\frac{q-p}{p+q}, \frac{-2q}{p+q}\Big) \cdot (m,n). \tag{209}$$

$$H = \Big(\frac{-p^2+4pq+q^2}{(p+q)^2}, \frac{2q(q-p)}{(p+q)^2}\Big) \cdot (m,n). \tag{210}$$

Theorem 21.1 (Copy I) *Suppose that $\kappa \geq 1$ is an integer such that*

$$0 < \frac{p_2}{q_2} - \frac{p_1}{q_1} < \frac{2}{\kappa q_1^2}.$$

1. *If κ is odd and $\lambda_1 < 1/2$ we set $K = (\kappa+1)/2 + \lambda_1$.*
2. *If κ is even and $\lambda_1 > 1/2$ we set $K = \kappa/2 + \lambda_1$.*
3. *If $\kappa \geq 2$ but doesn't satisfy Conditions 1 or 2, we set $K = \mathrm{floor}(\kappa/2)$.*

Then Γ_1 and Γ_2 agree on any lattice point (m,n) such that $F_1(m,n) \geq 0$ and

$$G_1(m,n), H_1(m,n) \in [-q_1+3, Kq_1-3].$$

We will use the method in §10.5 to compute the arithmetic graphs Γ_1 and Γ_2. This means that we have $\alpha = 0$ in the equations for the maps M_+ and M_- in §10.6.3. Alternatively, we can think of α as being an infinitesimally small positive number. The point is that any choice of $\alpha < \min(1/q_1, 1/q_2)$ gives rise to the same arithmetic graphs.

The Master Picture Theorem implies that the local picture of $\widehat{\Gamma}$ at each point (m,n) depends on the locaton of the points $M_+(m,n)$ and $M_-(m,n)$ in the polyhedron $R_A = (0, 1+A)^2 \times (0,1)$. Our goal is to show that the two maps $M_\pm(m,n)$ land in the same polyhedron, relative to $A_1 = p_1/q_1$ and $A_2 = p_2/q_2$, when (m,n) satisfies the hypotheses of the Copy Theorem I.

The following 10 step outline gives a list of what we must accomplish for a given point (m,n). The list is derived from the reduction algorithm in §15, and also the list of hyperplanes in §10.2. In our outline, ϵ_1 and ϵ_2 stand for integers in $\{-1,0,1\}$. Except for Step 4, rur outline works the same way for both M_+ and M_-.

1. Let $z_j = A_j m + n$.
2. Let $Z_j = \text{floor}(x_j)$.
3. We prove that $Z_1 = Z_2$. Let $Z = Z_1 = Z_2$. We also prove that the statement
$$z_j - Z < \epsilon_1 A_j + \epsilon_2$$
has the same truth value independent of $j \in \{1,2\}$. We mean this to hold for any relevant choice of ϵ_1 and ϵ_2.
4. Let $y_j = z_j + Z_j$ in the $(-)$ case and $y_j = z_j + Z_j + 1$ in the $(+)$ case.
5. Let $Y_j = \text{floor}(y_j/(1+A_j)$.
6. We prove that $Y_1 = Y_2$. Let $Y = Y_1 = Y_2$. We also prove that the statement
$$y_j - Y(1+A_j) < \epsilon_1 A_j + \epsilon_2$$
has the same truth value independent of $j \in \{1,2\}$. We mean this to hold for any relevant choice of ϵ_1 and ϵ_2.
7. Let $x_j = y_j - Y(1-A_j) - 1$.
8. Let $X_j = \text{floor}(x_j/(1+A_j))$.
9. We prove that $X_1 = X_2$. Let $X = X_1 = X_2$. We also prove that
$$x_j - X(1+A_j) < \epsilon_1 A_j + \epsilon_2$$
has the same truth value independent of $j \in \{1,2\}$. We mean this to hold for any relevant choice of ϵ_1 and ϵ_2.
10. We want to see that all of the statements
$$(x_j - X(1+A_j)) + (y_j - Y(1+A_j)) - (z_j - Z) < h + A_j; \qquad h \in \boldsymbol{Z}$$
has the same truth value independent of $j \in \{1,2\}$.

21.2 Good Integers

In this section we will state 3 technical lemmas that imply the Copy Theorem I.

Let $A_j = p_j/q_j$ for $j = 1, 2$. We say that an integer μ is *good* if

$$\text{floor}(\mu A_1) = \text{floor}(\mu A_2). \tag{211}$$

Lemma 21.2 (Good Integer) *Suppose (m, n) satisfies the hypotheses of the Copy Theorem I, and X_1 and Y_1 are the integers that arise when we perform the reduction algorithm above to (m, n). For any $\epsilon \in \{-1, 0, 1\}$, the integers*

$$m - \epsilon; \qquad m - Y_1 - \epsilon; \qquad m - X_1 - \epsilon; \qquad m + Y_1 - X_1 + \epsilon$$

are all good integers.

Proof: See §21.4. ♠

Lemma 21.3 *Let $d \geq 0$ be an integer. Suppose $\mu, \eta \in \boldsymbol{Z}$. Let N_j be the integer such that*

$$N_j(dA_j + 1) < \mu A_j + \nu + \alpha < (N_j + 1)(dA_j + 1).$$

Suppose $\mu - dN_1$ and $\mu - dN_2$ are both good integers. Then $N_1 = N_2$.

Proof: Suppose first that $N_1 < N_2$. In this case we set $N = N_2$ and we note that

$$\mu A_1 + \nu < N(dA_1 + 1); \qquad N(dA_2 + 1) > \mu A_2 + \nu.$$

But then

$$(\mu - dN)A_1 < N - \nu < (\mu - dN)A_2.$$

This contradicts the fact that $\mu - dN_2$ is good. If $N_1 > N_2$ we set $N = N_1$ and get the same equations with the inequalities reversed. This contradicts the fact that $\mu - dN_1$ is good. ♠

Lemma 21.4 *Let $d \geq 0$ be an integer and. Suppose $\mu, \eta \in \boldsymbol{Z}$. Let N be an integer. The truth of the statement*

$$(\mu A_j + \eta + \alpha) - N(dA_j + 1) < \epsilon_1 A_1 + \epsilon_2$$

is independent of j provided that $\mu - dN - \epsilon_1$ is good.

Proof: The proof is almost the same as in Lemma 21.3. If the above statement is true for $j = 1$ and false for $j = 2$ then we get the inequalities

$$(\mu - dN - \epsilon_1)A_1 < \epsilon_2 + N - \nu < (\mu - dN - \epsilon_1)A_2,$$

and this contradicts the fact that $\mu - dN - \epsilon_1$ is good. We get a similar contradiction, with the inequalities reversed, if the statement is true for $j = 1$ and false for $j = 2$. ♠

21.3 The Main Argument

We will carry out the 10 step outline for M_-. The argument for M_+ is essentially identical.

- We have $z_j = mA_j + n$. To see that $Z_1 = Z_2$ we apply the Good Integer Lemma and Lemma 21.3 to the case

  $$(\mu, d, N_j) = (m, 0, Z_j).$$

  The relevant good integer is m. Let $Z = Z_1 = Z_2$.

- To see that the truth of the statement $z_j - Z < \epsilon_1 A_j + \epsilon_2$ is true independent of j, we apply the Good Integer Lemma and Lemma 21.4 to the case

  $$(\mu, d, N) = (m, 0, Z).$$

  The relevant good integer is $m - \epsilon_1$.

- We have $y_j = mA_j + n'$ for some $n' \in \boldsymbol{Z}$. To see that $Y_1 = Y_2$ we apply the Good Integer Lemma and Lemma 21.3 to the case

  $$(\mu, d, N_j) = (m, 1, Y_j).$$

  The relevant good integers are $m - Y_1$ and $m - Y_2$. We set $Y = Y_1 = Y_2$.

- To see that the truth of the statement $y_j - Y(A_1+1) < \epsilon_1 A_j + \epsilon_2$ is independent of j, we apply the Good Integer Lemma and Lemma 21.4 to the case
$$(\mu, d, N) = (m, 1, Y).$$
The relevant good integer is $m - Y - \epsilon_1$.

- We have
$$x_j = (m+Y)A_j + n''$$
for some $n'' \in \boldsymbol{Z}$. To see that $X_1 = X_2$ we apply the Good Integer Lemma and Lemma 21.3 to the case
$$(\mu, d, N_j) = (m+Y, 1, X_j).$$
The relevant good numbers are $m + Y - X_1$ and $m + Y - X_2$. We set $X = X_1 = X_2$.

- To see that the truth of the statement $x_j - (1+A_j)X < \epsilon_1 A_j + \epsilon_2$ is independent of $j = 1, 2$ we apply the Good Integer Lemma and Lemma 21.4 to the case
$$(\mu, d, N) = (m+Y, 1, X).$$
The relevant good number is $m + Y - X - \epsilon_1$.

- 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 $n''' \in \boldsymbol{Z}$. Let $h \in \boldsymbol{Z}$ be arbitrary. To see that the truth of the statement $\sigma_j < A_j + h$ is independent of j we apply the Good Integer Lemma and Lemma 21.4 to
$$(\mu, d, N) = (m - X, 1, 0).$$
The relevant good number is $m - X - 1$.

This proves that $M_-(m,n)$ lands in the same polyhedron for each parameter A_1 and A_2. As we said above, essentially the same argument works for M_+. Modulo the Good Integer Lemma, this completes the proof of the Copy Theorem I. Again, the proof of the Copy Theorem II is almost identical.

It only remains to prove the Good Integer Lemma.

21.4 Proof of the Good Integer Lemma

21.4.1 Step 1

First we prove the following result.

Lemma 21.5 *If $\mu \in (-q_1, Kq_1) \cap \boldsymbol{Z}$ then μ is a good integer.*

We will prove this result in two stages, one for the lower bound for one for the upper bound.

Lemma 21.6 *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

$$\mu A_1 = M + \frac{j}{q_1}; \qquad |j| < \frac{q_1}{2}. \tag{212}$$

Note that

$$|A_2 - A_1| < 2/q_1^2 \tag{213}$$

in all cases. If this result is false, then there is some integer N such that

$$\mu A_2 < N \leq \mu A_1. \tag{214}$$

Referring to Equation 212, we have

$$\frac{|j|}{q_1} < \mu A_1 - N \leq \mu A_1 - \mu A_2 < \frac{2|\mu|}{q_1^2} < \frac{2}{q_1}. \tag{215}$$

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.

$$\mu \in (-q_1, 0) \quad \text{and} \quad \mu p_1 \equiv 1 \bmod q_1 \qquad \Longrightarrow \qquad \mu = -\lambda_1 q_1. \tag{216}$$

Equation 212 implies

$$\frac{\mu p_1}{q_1} - \frac{1}{q_1} \in \boldsymbol{Z}.$$

But then $\mu p_1 \equiv 1 \bmod q_1$. Equation 216 now tells us that $\mu = -\lambda_1 q_1$. Hence $|\mu| < q_1/2$. But now Equation 215 is twice as strong and gives $|j| = 0$. This is a contradiction. ♠

Lemma 21.7 *If $\mu \in (0, Kq_1)$ then μ is good.*

Proof: We argue by contradiction. There is an integer N such that

$$\mu A_1 < N \leq \mu A_2. \tag{217}$$

Suppose first that $\mu \leq \kappa q_1/2$. Referring to Equation 212, we have

$$\frac{|j|}{q_1} \leq N - \mu A_1 \leq \mu(A_2 - A_1) < \frac{2|\mu|}{\kappa q_1^2} \leq \frac{1}{q_1}. \tag{218}$$

Hence $j = 0$ in Equation 212. But then, $\mu A_1 \in \boldsymbol{Z}$. Since $\mu A_1 < N$, we must have

$$1 \leq N - \mu A_1 \leq \mu A_2 - \mu A_1 \leq \frac{1}{q_1},$$

a contradiction. This argument handles Case 3 in the Copy Theorem I.

Now consider Case 1. Here κ is odd and $\lambda_1 < 1/2$. It suffices to take

$$\mu \in \left(\frac{\kappa q_1}{2}, \frac{(\kappa+1)q_1}{2} + \lambda_1 q_1\right). \tag{219}$$

We again have Equation 217. As in the proof of the lower bound, the version of Equation 218 gives us $|j| \leq 1$. If $j = 0$ then q_1 divides μ. But this contradicts Equation 219. If $j = 1$ then $|N - \mu A_1| \leq (q_1 - 1)/q_1$, which is false. Hence $j = -1$. But then Equation 212 tells us that $\mu q_1 + 1 \equiv 0 \bmod q_1$. Referring to the definition of λ_1, we see that $\mu \equiv (q_1)_+ \bmod q_1$. In other words, there is an integer K' such that

$$\mu = (K' + \lambda_1)q_1. \tag{220}$$

Since κ is odd, the interval in in Equation 219 contains no such point.

Now we consider Case 2. Here κ is even and $\lambda_1 > 1/2$. It suffices to take

$$\mu \in \left(\frac{\kappa q_1}{2}, \frac{\kappa q_1}{2} + \lambda_1 q_1\right). \tag{221}$$

The same argument as in Case 2 rules out this situation. ♠

21.4.2 Step 2

The Good Integer Lemma only involves the numbers p_1, q_1, X_1, etc. For ease of notation, we set $p = p_1$, etc.

The Good Integer Lemmma involves 4 statements, one per listed number. Statement 1 follows immediately from Lemma 21.5. By construction, we have $X \in [-1, Y]$. Hence $m - X \in [m - Y, m + 1]$. Hence, Statement 2 implies Statement 3. Our next two results deal respectively with Statements 2 and 4. In these results, ϵ can be any of the three integers in $\{-1, 0, 1\}$. The reduction algorithm from §21.1 must be performed for the map M_- and for the map M_+.

Lemma 21.8 *$G(m,n) \in (-q_1 + 2, Kq_1 - 3)$ then $m - Y_1 - \epsilon$ is good.*

Proof: We will treat the $(+)$ case. In the $(-)$ case, the only difference is that $Y + 1$ replaces Y. We will use the terminology *Reduce k* to refer to the kth step in our outline from §21.1. Given that we are working with M_-, Reduce 4 gives us

$$y \in [2z - 1, 2z] \tag{222}$$

Reduce 5 gives us

$$Y \in \left[\frac{y}{1+A} - 1, \frac{y}{1+A}\right] \subset \left[\frac{2z}{1+A} - 1 - \frac{1}{1+A}, \frac{2z}{1+A}\right]. \tag{223}$$

Using $z = Am + n$ and $A = p/q$, we get

$$m - Y \in \left[G_-(m,n), G_-(m,n) + 1 + \frac{1}{1+A}\right]. \tag{224}$$

If $G_-(m,n) > -q + 1$ then

$$m - Y - \epsilon > -q + 1.$$

If $G_-(m,n) < Kq - 3$ then

$$m - Y - \epsilon < Kq - 2 + 1 + \frac{1}{1+A} < Kq.$$

Since $m - Y - \epsilon$ is an integer, we have $m - Y - \epsilon < Kq$. All in all, we have $m - Y - \epsilon \in (-q, Kq)$. Lemma 21.5 now applies. ♠

Lemma 21.9 *If $H(m,n) \in (-q_1+3, Kq_1-3)$ then $m+Y_1-X_1-\epsilon$ is good.*

Proof: Our proof works the same in the $(+)$ and $(-)$ cases. Reduce 5 and Reduce 8 give us

$$Y \in \left[\frac{y}{1+A}-1, \frac{y}{1+A}\right]; \qquad X \in \left[\frac{x}{1+A}-1, \frac{x}{1+A}\right].$$

Hence

$$Y-X \in \left[\frac{y-x}{1+A}-1, \frac{y-x}{1+A}+1\right] \tag{225}$$

Reduce 7 gives us

$$\frac{y-x}{1+A} = Y\frac{1-A}{1+A} + \frac{1}{1+A} \tag{226}$$

Equation 223 gives us

$$Y\frac{1-A}{1+A} \in \left[2z\frac{1-A}{(1+A)^2}-\eta, 2z\frac{1-A}{(1+A)^2}\right]; \qquad \eta = \left(1+\frac{1}{1+A}\right) \times \frac{1-A}{1+A}.$$

Observing that

$$\frac{1}{1+A} - \eta = \frac{A^2+2A-1}{(1+A)^2} \in (-1, \frac{1}{2}),$$

we see that

$$m + \frac{y-x}{1+A} \in \left(2z\frac{1-A}{(1+A)^2}-1, 2z\frac{1-A}{(1+A)^2}+\frac{1}{2}\right) =$$

$$\left(G+(m,n)-1, G+(m,n)+\frac{1}{2}\right). \tag{227}$$

Combining this last result with Equation 225, we have

$$m+Y-X \in \left(G+(m,n)-2, G+(m,n)+\frac{3}{2}\right). \tag{228}$$

The rest of the proof is as in Lemma 21.8. ♠

21.5 An Addendum

In the coming chapters, we will use the Copy Theorems to prove the Decomposition Theorem. As stated, the Copy Theorems are not quite strong enough for our purposes. Here we deal with a few special cases not covered by our formulation above.

Lemma 21.10 *Case 2 of the Copy Theorem I is true for* $p_1 = 1$ *and*

$$(m,n) = \left(1, \frac{q-1}{2}\right) \tag{229}$$

Proof: We have $K > 1$. We check by hand that the Good Integer Lemma remains true for this point. For M_-, we compute that $X = 0$ and $Y = q-2$ (and $Z = (q-1)/2$.) In this case, all the quantities in the Good Integer Lemma lie in $[-q+2, q-1] \subset (-q, Kq)$. For M_+ we the only change is that $Y = q-1$. In this case, all the quantities lie in $[-q+1, q] \subset (q, Kq)$. ♠

Lemma 21.11 *Case 1 of the Copy Theorem II is true for* $p_1 = 1$ *and*

$$\left(-1, \frac{q+1}{2}\right). \tag{230}$$

Proof: For the Copy Theorem II, the replacement for Lemma 21.5 is that $\mu \in (-Kq_1, q_1)$. We have $K > (1 + 1/q)$. For M_- we compute that $X = 0$ and $Y = q-2$. In this case, all the quantities in the Good Integer Lemma lie in $[-q_1 - 2, q_1 - 2] \in (-Kq_1, q_1)$. For M_+, we get the same results except that $Y = q-1$. Then all the quantities in the Good Integer Lemma lie in $[-q_1 - 1, q_1 - 1] \in (-Kq_1, q_1)$. ♠

22 The Induction Lemma

22.1 The Main Result

Let p_1/q_1 be an odd rational. Given any set $X \subset \boldsymbol{R}^2$ we define.

$$X^n = X + n(q_1, -p_1). \tag{231}$$

For any integer $n \geq 1$ we define. $\Omega_n(+)$ and $\Omega_n(-)$ for any integer $n \geq 2$.

$$\begin{aligned}
\Omega_{2n}(+) &= FR_1 \cup SR_1 \cup ... \cup SR_1^{n-1} \cup FR_1^n \\
\Omega_{2n}(-) &= FR_1^- \cup SR_1^- \cup ... \cup SR_1^{-n+1} \cup FR_1^{-n} \\
\Omega_{2n+1}(+) &= SR_1 \cup FR_1 \cup ... \cup FR_1^n \cup SR_1^{n+1}. \\
\Omega_{2n+1}(-) &= SR_1^- \cup FR_1^- \cup ... \cup FR_1^{-n} \cup SR_1^{-n-1}.
\end{aligned} \tag{232}$$

The union of Ω_k involves $k-1$ terms in the even case, and k terms in the odd case. The terms alternate father, son, father... and have a palindromic character to them.

Say that a set X is *copied* if Γ_1 and Γ_2 agree in X.

Lemma 22.1 (Induction) *Let κ be the greatest integer such that*

$$\left|\frac{p_2}{q_2} - \frac{p_1}{q_1}\right| < \frac{2}{\kappa q_1^2}.$$

1. *If $A_1 < A_2$ and $\kappa \equiv 1$ mod 2 and $\lambda_1 < 1/2$ then $\Omega_\kappa(+)$ is copied.*
2. *If $A_1 < A_2$ and $\kappa \equiv 0$ mod 2 and $\lambda_1 > 1/2$ then $\Omega_\kappa(+)$ is copied.*
3. *If $A_1 > A_2$ and $\kappa \equiv 1$ mod 2 and $\lambda_1 > 1/2$ then $\Omega_\kappa(-)$ is copied.*
4. *If $A_1 > A_2$ and $\kappa \equiv 0$ mod 2 and $\lambda_1 < 1/2$ then $\Omega_\kappa(-)$ is copied.*

The Induction Lemma is really just a more precise version of Corollary 7.5. Accordingly, our proof here is just a more careful version of the proof given there. For most rationals, the proofs here is exactly the same as the proof there. However, when p_1 is small, we need to keep careful track of the formulas involved in the proof. In a few cases, we really need to look at the picture. Were we able to get a slightly sharper version of the Copy Theorems, we would eliminate some of the tedious work in the special cases.

22.2 Most Rationals

In this section we will prove the Induction Lemma for most rationals. In the sections following this one, we will deal with the remaining cases by hand. We will concentrate on Statements 1 and 2. Statements 3 and 4 lead to exactly the same inequalities, and thus have the same treatment.

Let Λ_κ denote the region given by the Copy Theorem I. Λ_κ is determined by the inequalities

$$F_1 \geq 0; \qquad G_1, H_1 \in (-q_1+3, Kq_1-3). \tag{233}$$

We only need to deal with Cases 1 and 2. So, we have

- $K = (\kappa+1)/2 + \lambda_1$ for κ odd.
- $K = \kappa/2 + \lambda_1$ for κ even.

Statements 1 and 2 of the Induction Lemma follow immediately from the statement that $\Omega_\kappa \subset \Lambda_\kappa$. If we change p_2/q_2 in such a way that κ changes to $\kappa+2$, then the two regions Ω_κ and Λ_κ both grow rightward by 1 period. For this reason, it suffices to establish the following.

$$\Omega_\kappa \subset \Lambda_\kappa; \qquad \kappa = 2, 3. \tag{234}$$

Compare Figure 7.1.

Case 1: For ease of notation, we set $p = p_1$ and $q = q_1$, etc. In this case we have $K = 1 + \lambda$. Let w_1 be the top left vertex of $= \Omega_2 = FR$ and w_2 is the top right vertex. In terms of the Equation 11, we have

$$w_1 = W; \qquad w_2 = W + \lambda V. \tag{235}$$

We have $G(w_1) < H(w_2)$ and $H(w_2) > G(w_1)$. Therefore, to establish Equation 234 in case $\kappa = 2$ it suffices to prove that

$$G(w_1) = -Z \geq -q+3; \qquad H(w_2) = \lambda q + Z \leq (1+\lambda)q - 3.$$

Here we have set

$$Z = \frac{q^2}{p+q}, \tag{236}$$

So, both cases boil down to the inequality $Z \leq q-3$. One expects that both cases boil down to the same inequality because of affine symmetry. A bit of algebra shows that this inequality holds except when $p = 1, 3$ or when $p/q = 5/7$. We check by hand that all *lattice points* in FR lie in Λ_2 when $p/q = 5/7$. Hence, Statement 1 of the Induction Lemma is true except possibly when $p_1 = 1$ or $p_1 = 3$. We will treat these cases below.

Case 2: The region Ω_3 looks like the one in Figure 22.1. In this case, we have $K = 2 + \lambda$ Compare Figure 7.1. In this case, the vertices relevant to the calculation are

$$w_3 = W/2; \quad w_4 = W + \lambda V; \quad w_5 = W + V; \quad w_6 = W/2 + (1+\lambda)V.$$

This time we need to check that

$$G(w_3) = -Z/2 \geq -q + 3; \tag{237}$$

$$G(w_4) = -Z + \lambda q \geq -q + 3; \tag{238}$$

$$H(w_5) = Z + q \leq (2+\lambda)q - 3; \tag{239}$$

$$H(w_6) = -Z/2 + (1+\lambda)q \leq (2+\lambda)q - 3. \tag{240}$$

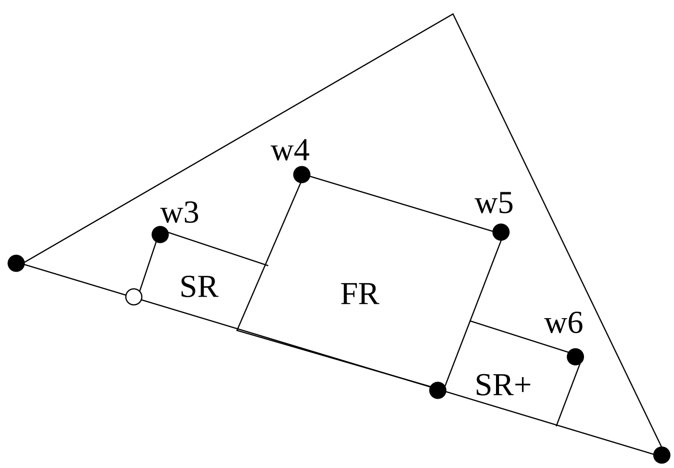

Figure 22.1: SR and FR and SR^+ and $\Lambda(-1, 1+\lambda)$.

All the bounds above are implied by the inequality $Z \leq q - 3$. We check the case 5/7 by hand in the same way as in Case 1. Hence, Statement 2 of the Induction Lemma is true except possibly when $p_1 = 1$ or $p_1 = 3$. We will treat these cases below.

22.3 Strategy

It remains to deal with the cases $p_1 = 1$ and $p_1 = 3$. A slightly sharper version of the Copy Theorem would eliminate the need for what we do here. Our argument has 2 kinds of slack in it. First of all, we don't need to analyze all points in Ω_κ, but just the lattice points. The vertices we considered above are not lattice points, and so we can get some savings by finding the nearby lattice points and re-evaluating. Our argument relies on visual inspection of the picture, but we will include what is a fairly representative picture. The interested reader can see the picture for any parameter using Billiard King.

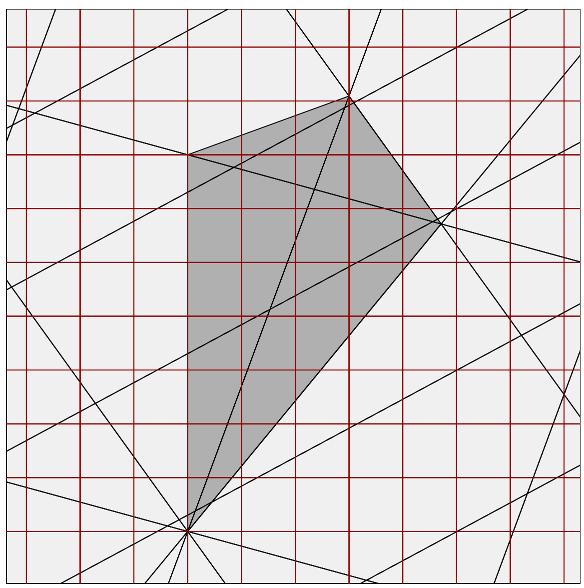

Figure 22.2: The arithmetic kite $Q(3/11)$ and the hexagrid $G(3/11)$.

In all cases, we can study the placement of various lattice points by translating so that we are in a neighborhood of the point where the diagonals of $Q(A)$ cross.

A second kind of slack comes from our proof of the Good Integer Lemma. Call a lattice point *questionable* if it lies in Ω_κ but fails to satisfy the hypotheses of the Copy Theorem I. If we encounter a questionable vertex, we just prove the Good Integer Lemma by hand for that vertex. This is what we did in Lemmas 21.10 and 21.11. All the vertices there were questionable.

For the most part, we will concentrate on Statements 1 and 2. In one case, where it really matters, we will consider Statement 3.

22.4 Revisiting Case 1

22.4.1 Case 1A

First we suppose that $p = 3$. When $q \geq 11$, the lattice points in Ω_2 with smallest G-value and largest H-value respectively are

$$\widetilde{w}_1 = \left(3, \frac{(q+1)}{2}\right); \qquad \widehat{w}_2 = (\lambda q, -2) + \left(2, \frac{q+1}{2}\right). \tag{241}$$

We get $\widetilde{w}_1$ by inspecting the lower right quadrant in the center of $Q(A)$, as in Figure 22.2. For $\widetilde{w}_2$, we note that $(\lambda q, -2)$ asymptotically maps a neighborhood of the lower left quadrant of $Q(A)$ to a neighborhood of the upper right corner of FR, because $2/(\lambda q)$ is an excellent approximation to $3/q$. So, we get $\widetilde{w}_2$ by looking at the lower left quadrant in the center of $Q(A)$.

We compute

$$G(\widehat{w}_1) + q = \frac{5q-9}{q+3} > 3; \quad (1+\lambda)q - H(\widehat{w}_2) = \frac{28 + 75k + 54k^2}{(4+3k)^2} > 3 \tag{242}$$

In the second case we have set $q = 6k + 5$. The only case left is $q = 5$. The only questionable lattice point is $(4, 1)$ and we checked this point by hand.

22.4.2 Case 1B

We suppose $p = 1$ and When $q \geq 7$, the lattice points in Ω_2 with smallest G-value and largest H-value respectively are

$$\widetilde{w}_1 = \left(1, \frac{q-1}{2}\right); \qquad \widetilde{w}_2 = \left(q-3, \frac{q-3}{2}\right) \tag{243}$$

The questionable point $\widetilde{w}_1$ is covered by Lemma 21.10, and for $\widetilde{w}_2$ we compute

$$2q + \lambda - H(\widetilde{w}_2) = \frac{4q-2}{q+1} > 3 \tag{244}$$

for $q \geq 7$. This leaves only $q = 3, 5$. We check the several questionable lattice points by hand.

22.5 Revisiting Case 2

22.5.1 Case 2A

Suppose that $p = 3$. Equations 237 and 240 are implied by the inequality $Z \le 2q - 6$, and this always holds. We have $\lambda = 1$, so Equations 238 and 239 are implied by the inequality $Z \le q - 2$. This inequality holds unless $q = 7$. For $p/q = 3/7$, we check that there are no questionable lattice points.

22.5.2 Case 2B

When $p_1 = 1$ and κ is odd, we always have $p_1/q_1 > p_2/q_2$. Thus, Statement 1 of the Induction Lemma does not occur. However, to give a complete picture of what is going on, we will describe what happens when $1/q_1 > p_2/q_2$ and $\kappa = 3$. This case is covered by Statement 3 of the Induction Lemma. In this case, the picture looks exactly like Figure 22.1, except that the points have different definitions.

$$w_6 = W/2; \qquad w_5 = W - \lambda V; \qquad w_4 = W - V; \qquad w_3 = W/2 - (1 + \lambda)V. \tag{245}$$

This time we want to show that

$$G(w_3), G(w_4) > (-2 - \lambda)q + 3; \qquad H(w_5), H(w_6) < q - 3.$$

These equations lead to the same exact inequalities as for Statement 1.

The inequalities associated to w_3 and w_6 again are implied by $Z \le 2q - 6$. The only cases where this fails is $q = 3, 5$. As above, we check all the questionable points by hand.

Now we deal with w_4. The lattice point closest to $G^{-1}(-2q - \lambda q)$ is

$$\Big(1 - q, \frac{q + 1}{2}\Big).$$

The corresponding inequality comes down to $4q/(1 + q) \ge 3$, and this holds for all $q \ge 3$. We check the questionable points by hand.

For w_5, the lattice point closest to $H^{-1}(q)$ is

$$\Big(\frac{q + 1}{2}, -1\Big).$$

This is the only point for which the corresponding inequality fails. Lemma 21.11 takes care of this point.

23 Decomposition Theorem: The Base Case

23.1 Starting the Induction

The Decomposition Theorem states that $\Gamma \subset SR \cup FR$. Here $\Gamma = \Gamma(p/q)$. We let $S(p/q)$ be the statement that the Decomposition Theorem is true for the parameter p/q.

In the next chapter, we will give an inductive proof of $S(p/q)$. If p_2/q_2 does not lie on the list of odd rationals in Equation 246, then there is a simpler odd rational $p_1/q_1 \in (0,1)$ such that the two rationals satisfy Equation 40 (for $n = 1$.) We will deduce $S(p_2/q_2)$ from $S(p_1/q_1)$ and the Induction Lemma. Here is the list of exceptional odd rationals.

$$A_k = \frac{2k-1}{2k+1}; \qquad k = 1, 2, 3... \tag{246}$$

In this chapter we will prove $S(A_k)$ for all k. Figure 23.1, which is part of a general pattern, shows the result for $A_4 = 7/9$.

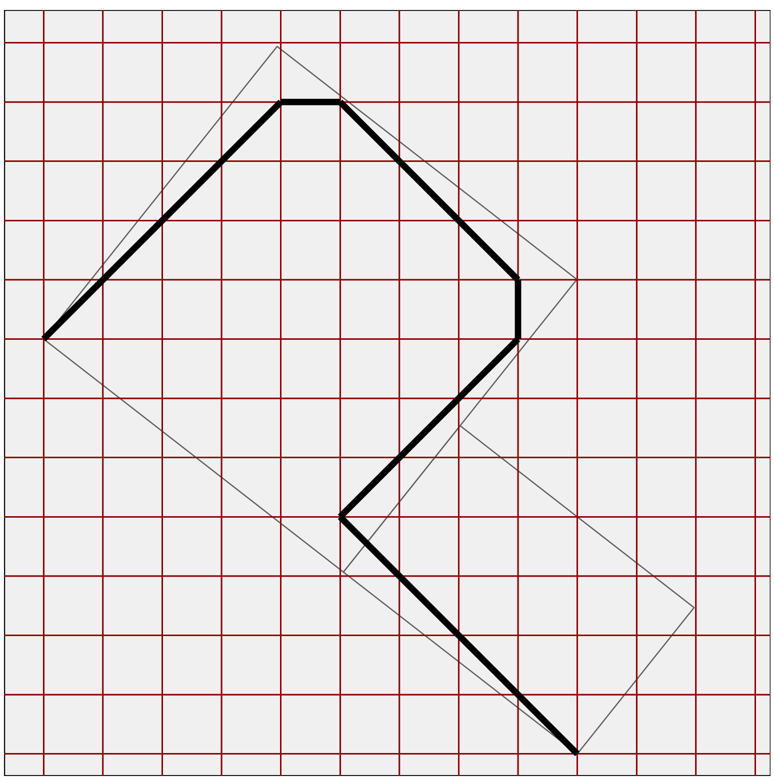

Figure 23.1: $\Gamma(7/9)$ and $FR(7/9)$ and $SR(7/9)$.

23.2 Analyzing the Pattern

Recall that $DR(A)$ is the dividing line between $SR(A)$ and $FR(A)$. In the picture, $S\Gamma(A_4)$ contains the 4 vertices

$$2 \times (4,-4) + (1,1) + j(1,-1); \qquad j = 0,1,2,3,$$

and also contains about half the segment joining $(4+2,-4)$ to $(4+1,-4+1)$. This half segment crosses $DR(A_4)$. The line $DR(A_4)$ contains the point $(4+1,-4)$, and has slope pretty close to 1. We have written things out this way so as to point out the beginnings of a general pattern.

In general, $DR(A_k)$ contains the point $(k+1,-k)$, because, in the sense of §6, we have $p_+/q_+ = k/(k+1)$.

Lemma 23.1 *Suppose* $\Gamma(A_k)$ *contains the* $k+1$ *points*

$$p_j = 2 \times (k,-k) + (1,1) + j(-1,1); \qquad j = 0,...,k. \tag{247}$$

Then the Decomposition Theorem is true for A_k.

Proof: From the Hexagrid Theorem and rotational symmetry, Γ only crosses DR once. (See Lemma 24.1 for more details.) We follow Γ backwards from p_j and continue along the diagonal until it crosses DR between p_k and p_{k+1}. Thus, we have identified the only crossing and shown that $S\Gamma \cap SR$ only rises about 1 unit from the baseline. The rest of Γ lies in FR, because the curve does not cross DR again. ♠

Say that the vertex p_j is *good* if $M_+(p_j)$ lands in the polyhedron corresponding to the vector $(-1,1)$.

Lemma 23.2 *Suppose that the vertices* $p_0,...,p_{k-1}$ *are all good. Then the Decomposition Theorem is true.*

Proof: We know that $p_0 \in \Gamma(p/q)$. Suppose by induction that $p_j \in \Gamma(p/q)$. Since p_j is good, one of the edges emanating from p_j joins the point p_j to the point $p_j + (-1,1) = p_{j+1}$. Hence $p_{j+1} \in \Gamma$. So, by induction, in this way, we find that $p_0,...,p_k$ are all vertices of $\Gamma(A_k)$. Now we apply Lemma 23.1. ♠

23.3 Checking for Good Vertices

Let $A = p/q$ be some odd rational parameter. Recall that

$$M_+(m,n) = (t, t+1, t) \text{ mod } \Lambda; \qquad t = Am + n + \alpha. \tag{248}$$

As usual, α is a tiny positive number. Say that (m,n) has *trivial reduction* relative to A if

$$Am + n + \alpha < A. \tag{249}$$

Lemma 23.3 *Suppose that*

- *p_{k-1} has a trivial reduction relative to A_k;*
- *$M_+(p_{k-1})$ and $M_+(p_1)$ are contained in the polyhedron.*
- *$M_+(p_1)$ lands in a polyhedron that specifies the vector $(-1,1)$.*

Then the vertices $p_0,...,p_{k-1}$ are all good.

Proof: In general, the size of the quantity $Am + n$ depends on the distance from (m,n) to the baseline of the arithmetic graph. The points $p_0,...,p_k$ are ordered in terms of their distance to the baseline. So, if p_{k-1} has trivial reduction, so do $p_0,...,p_{k-1}$. But then $M_+(p_j)$ lies on the line segment connecting $M_+(p_1)$ to $M_+(p_{k-1})$ for $j = 1,...,k-2$. But then $M_+(p_j)$ lands in the same polygon, independent of $j = 1,...,k-1$. But this means that $M_+(p_j)$ specifies the vector $(-1,1)$ for $j = 1,...,k-1$. We also check explicitly that $M_+(p_0) = M_+(0,0)$ specifies $(-1,1)$. Now we apply the previous result. ♠

To finish the proof that the Decomposition Theorem holds for A_k, we just have to verify the 3 hypotheses of Lemma 23.3.

Lemma 23.4 *The first hypothesis of Lemma 23.3 holds.*

Proof: We have $p_{k-1} = (k+2, -k)$. We compute

$$\frac{2k-1}{2k+1}(k+2) - k + \alpha = \frac{2k-2}{2k+1} + \alpha < A_k. \tag{250}$$

This shows that p_k has a trivial reduction. Finally, we compute that

$$M_+(p_{k-1}) = \left(\frac{2k-2}{2k+1}, \frac{4k-1}{2k-1}, \frac{2k-2}{2k+1}\right) + (\alpha, \alpha, \alpha). \tag{251}$$

This verifies the first hypothesis. ♠

Lemma 23.5 *The second hypothesis of Lemma 23.3 holds.*

Proof: Let X_j, Y_j, and Z_j stand for the three coordinates of $M_+(p_j)$.. We observe that

1. $X_{k-1} \in (0, A_k)$.
2. $Y_{k-1} \in (1, 1+A_k)$.
3. $Z_{k-1} \in (1-A_k, A_k)$.
4. $X_{k-1} + Y_{k-1} - Z_{k-1} \in (A_k, 1+A_k)$.

We call these 4 inequalities the 4 *determiners*, because they determine the polyhedron containing $M_+(p_{k-1})$. We compute that

$$M_+(p_1) = M_+(-1,1) = \left(\frac{2}{2k+1}, \frac{2k+3}{2k+1}, \frac{2}{2k+1}\right) + (\alpha, \alpha, \alpha). \tag{252}$$

We now see that the same 4 determiners hold with $j=1$ in place of $j=k-1$. Hence $M_+(p_1)$ and $M_+(p_{k-1})$ lie in the same polyhedron in the partition. ♠

Lemma 23.6 *The third hypothesis of Lemma 23.3 holds.*

Proof: Our 4 determiners specify the same polyhedron independent of the parameter A_k. We just check for a specific choice of parameter, say A_1, that the 4 determiners above specify the polyhedron that in turn specifies the vector $(-1,1)$. ♠

Now we know that the $S(A_k)$ is true.

24 Proof of the Decomposition Theorem

24.1 A Preliminary Division into Regions

Recall that $DR(p/q)$ is the line parallel to the sides of $R(p/q)$ that contains the point $(a,-b)$. Put another way, $DR(p/q)$ is the line extending the edge common to $FR(p/q)$ and $SR(p/q)$.

Lemma 24.1 *$\Gamma(p/q)$ only crosses $DR(p/q)$ once, and the crossing point lies within one unit the baseline of $\Gamma(p/q)$.*

Proof: Let ι denote 180 degree rotation about the point $1/2(q_+,-p_+)$. We proved in §17.2 that $\iota(\widehat{\Gamma}) = (\widehat{\Gamma})$. Let L be the line through the origin that is parallel to $DR(p/q)$. Note that L is the line extending the long diagonal of $Q(A)$, the arithmetic kite. By construction, $DR(p/q) = \iota(L)$.

Let L_k^- denote the line of slope $-A$ through the point kW. So, L_0^- is the baseline of the arithmetic graph, and L_{-1}^- lies below the baseline and L_1^- lies above it. By the Hexagrid Theorem, Γ lies between L_0^- and L_1^-. The image $\iota(\Gamma)$ lies between L_0^1 and L_{-1}^-. The hexagrid is invariant under 180 degree rotation through the origin. (Note that $\widehat{\Gamma}$ is not invariant under this rotation, but this does not bother us.) From this symmetry, we see that no line of the door grid intersects L in a point that lies between the lines L_-^0 and L_-^{-1}. By Statement 2 of the Hexagrid Theorem, $\iota(\Gamma)$ only intersects L within 1 unit of $(0,0)$. (This crossing occurs slightly below the baseline.) Applying ι again, we now see that Γ only intersects $DR(p/q)$ within 1 unit of L_0^-. ♠

As this point, we define

$$F\Gamma = \Gamma \cap FR; \qquad S\Gamma = \Gamma \cap \widehat{SR}; \qquad \widehat{SR} = R - FR. \tag{253}$$

Given Corollary 2.5 and Lemma 24.1, both $F\Gamma$ and $S\Gamma$ are connected arcs. We also have $F\Gamma \subset FR$ and $S\Gamma \subset \widehat{SR}$ by construction. It only remains to show that $S\Gamma \subset SR$. That is, $S\Gamma$ does not rise more than halfway towards the ceiling in $\widehat{SR}$. We will give an inductive proof of this result.

Remark: It is tempting to try to prove this by finding a barrier in $\widehat{SR}$, at most halfway up, that $S\Gamma$ does not cross. This strategy sometimes works, but not always.

24.2 An Example

We will show an example of our induction argument in action. The general case is modelled on this example. Whenever $31/79$ appears as a term in the canonical sequence, the previous term is $9/23$. The Diophantine constant is $\kappa = 3$. The Induction Lemma implies that the two arithmetic graphs agree on $\Omega_3(9/23)$, rhe shaded region in Figure 24.1. We observe first of all that $\Gamma(9/23)$ and $\Gamma(31/79)$ agree on the edge e that crosses the right edge of $\Omega_3(9/23)$. In general, this is a consequence of the Induction Lemma.

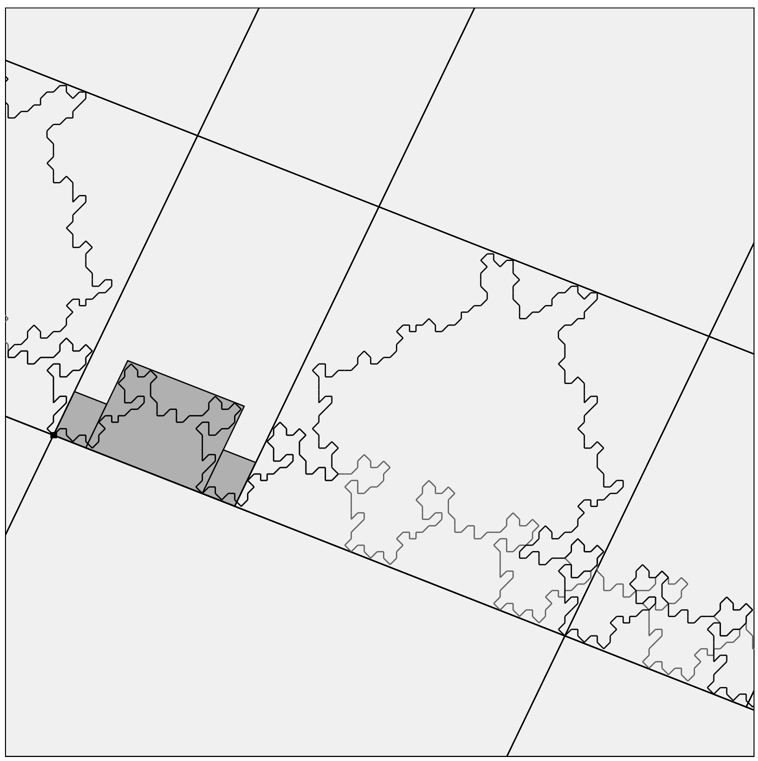

Figure 24.1: $\Gamma(9/23)$ and $\Gamma(31/79)$ and $DR(31/79)$ and $\Omega_3(9/23)$.

The line $DR = DR(31/79)$ is almost identical to the line DR' extending the right edge of $\Omega_3(9/23)$. Here are the two main properties we need.

1. $\Omega_3(9/23)$ lies in the bottom half of $R(31/79)$.

2. One and the same edge e of $\Gamma(31/79)$ crosses both DR and DR'.

As we trace $\Gamma(31/79)$ from $(0,0)$, we stay in $\Omega_3(9/23)$ until we reach the edge e. All this time, we stay in the lower half of $R(31/79)$. Once we reach e, we cross $DR(31/79)$. After crossing $DR(31/79)$, we never cross back. So, $S\Gamma(31/79)$ consists of all the edges of $\Gamma(31/79)$ we encounter before e, and also part of e itself. Therefore $S\Gamma(31/79)$ remains in the bottom half of $R(31/79)$. Since $S\Gamma(31/79)$ also remains in $\widehat{SR}(31/79)$, we now see that $S\Gamma(31/79)$ remains in the bottom half of $\widehat{SR}(31/79)$. This is to say that $S\Gamma(31/79) \subset SR(31/79)$, as desired.

24.3 The Induction Step

We will work with the canonical sequences discssed in §6. Now we work out the general version of the argument given in the example above.

Lemma 24.2 *Each odd rational $p/q \in (0,1)$ appears in a canonical sequence.*

Proof: Starting from p/q one can work backwards, to simpler fractions, as in §6. Then we can work forward more or less arbitrarily. ♠

The first term in any canonical sequence is $1/1$. We ignore this term. The second term always has the form A_k for some k. We have already established the Decomposition Theorem for these terms.

Lemma 24.3 *Suppose $n \geq 1$. If the Decomposition Theorem holds for p_n/q_n, the nth term in a canonical sequence, then the Decomposition Theorem also holds for the next term p_{n+1}/q_{n+1} in the sequence.*

Now we turn to the general proof of Lemma 24.3. There are 4 cases to consider, depending on the 4 cases of Lemma 6.2. Each case of Lemma 6.2 corresponds to the same-numbered case of the Induction Lemma. We will give the proof for Cases 1 and 2. Cases 3 and 4 have essentially the same treatment. For ease of notation, we use the index $n=1$, so that we can write the relevant terms as p_1/q_1 and p_2/q_2.

24.4 Case 1

24.4.1 Property 1

We have $p_2/q_2 > p_1/q_1$ and κ even. We must verify that $\Omega_\kappa(p_1/q_1)$ lies entirely in the lower half of $R(p_2/q_2)$. The line extending the top edges of $\Omega_\kappa(p_1/q_1)$ has equation

$$y = -\frac{p_1}{q_1}x + \frac{p_1+q_1}{2}. \tag{254}$$

Let $L^-_{1/2} = L^-_{1/2}(p_2/q_2)$ denote the line parallel to $L^-_0(p_2/q_2)$ and $L^-_1(p_2/q_2)$, and halfway between them. This line has the equation

$$y = -\frac{p_2}{q_2}x + \frac{p_2+q_2}{4}. \tag{255}$$

Equating the right hand sides of these lines, we find that the two lines intersect at a point p whose first coordinate is

$$x = \frac{q_1q_2}{8} \times (p_2+q_2-2(p_1+q_1)). \tag{256}$$

Since $\kappa \geq 2$ we have $q_2 \geq 2p_1+1$. Since $p_2/q_2 > p_1/q_1$, we have $p_2 > 2p_1+1$. Hence

$$p_2+q_2-2(p_1+q_1) \geq 2.$$

We have equality only if $q_1 - p_1 = 2$. Otherwise we have

$$p_2+q_2-2(p_1+q_1) \geq 4.$$

Hence, except for the case $p_1/q_1 = 1/3$ and $p_2/q_2 = 3/7$ (which we check by hand) we have

$$x > \frac{3q_2}{2}. \tag{257}$$

But we check that all points of $R(p_2/q_2)$ have x-coordinate less than $3q_2/2$. Hence, our two lines cross to the right of $R(p_2/q_2)$, as desired. This proves that all points of $\Omega_\kappa(p_1/q_1)$ lie beneath $L^-_{1/2}$.

24.4.2 Property 2

We have $p_2/q_2 > p_1/q_1$ and κ even and $\lambda_1 < 1/2$. We must verify that the edge e of Γ_1 crosses both the right edge of $\Omega_\kappa(p_1/q_1)$ and $DR(p_2/q_2)$. Consider the edge e with vertices

$$((q_2)_+ + 1, -(p_2)_+); \qquad ((q_2)_+, -(p_2)_+ + 1). \tag{258}$$

Lemma 24.4 *e crosses the line $DR = DR(p_2/q_2)$.*

Proof: Let ι be the rotation from §17.2. We have $e = \iota(e')$, where e' is the edge connecting $(0, -1)$ to $(-1, 0)$. By symmetry e' is an edge of $\widehat{\Gamma}$ that lies below the baseline. Referring to the proof of Lemma 24.1, the edge e' crosses the line L. Hence e is the edge of Γ_2 that crosses the line DR_2 ♠

Let DR' denote the mystery line that extends the right edge of $\Omega_\kappa(p_1/q_1)$.

Lemma 24.5 *e crosses DR'.*

In the analysis to follow, we define the *horizontal width* of a parallelogram to be the difference in x-coordinates of the bottom two vertices of the parallelogram. We take the horizontal width to be positive. Let $h(R)$ denote the horizontal width of a parallelogram R.

Now for the moment of truth. There are two important points. First of all,

$$(q_1)_+ = h(SR_1); \qquad (q_1)_- = h(FR_1). \tag{259}$$

Second of all, the coefficients in the formula for $(q_2)_+$ in Lemma 6.2 (Case 2) count the number of copies of FR_1 and SR_1 that occur in the union that defined $\Omega_\kappa(p_1/q_1)$.

From this general picture, we see that DR' contains a point ζ very near the baseline of $\Gamma(p_1/q_1)$ and having x-coordinate $(q_2)_+$. Given that both p_1/q_1 and $(p_2)_+/(q_2)_+$ are excellent approximations to p_2/q_2, the second coordinate of ζ must be $-(p_2)_+$. In short, $\zeta = ((q_2)_+, -(p_2)_+)$, the same point that is contained in $DR(p_2/q_2)$. Finally, the line DR' has slope

$$\sigma_1 = 1 + \frac{q_1}{2p_1} - \frac{p_1}{2q_1} \in (1, \infty). \tag{260}$$

whereas e has slope -1. Hence, e crosses DR' as well. In fact, the slope of DR' is an extremely good approximation to the slope of DR for large p_1/q_1. ♠

24.5 Case 2

24.5.1 Property 1

We have $p_2/q_2 > p_1/q_1$ and κ odd. If $\kappa \geq 3$ we have the same argument as in the previous case. If $\kappa = 1$ the argument is just about the same, but some of the estimates are a bit more delicate. The main detail involves establishing Equation 257.

Lemma 24.6 *Equation 257 holds when $\kappa = 1$.*

Proof: If $\kappa = 1$ then $\Omega_\kappa(p_1/q_1) = SR(p_1/q_1)$ and the line extending the top of edge of $\Omega_\kappa(p_1/q_1)$ has equation

$$y = -\frac{p_1}{q_1}x + \frac{p_1+q_1}{4}. \tag{261}$$

This time, the intersection of the two lines is

$$x = \frac{q_1 q_2}{8} \times (p_2 + q_2 - (p_1 + q_1)). \tag{262}$$

The case $q_1 \leq 5$ does not occur here. Hence $q_1 \geq 7$. Also $p_2+q_2 \geq p_1+q_1+2$. Finally, $7/8 \times 2 > 3/2$. Hence, Equation 257 holds in this case. ♠

The rest of the proof is the same as in Case 1.

24.5.2 Property 2

The verification of Property 2 is the same as in Case 1, except for the following changes.

1. When $\lambda_1 > 1/2$, we have $(q_1)_+ = h(FR_1)$.
2. When $\lambda_1 > 1/2$, we have $(q_1)_- = h(SR_1)$.

The rest of the proof is the same.

25 References

[**B**] P. Boyland, *Dual Billiards, twist maps, and impact oscillators*, Nonlinearity **9** (1996) 1411-1438

[**De**] N .E. J. De Bruijn, *Algebraic Theory of Penrose's Nonperiodic Tilings*, Nederl. Akad. Wentensch. Proc. **84** (1981) pp 39-66

[**Da**] Davenport, *The Higher Arithmetic: An Introduction to the Theory of Numbers*, Hutchinson and Company, 1952

[**D**], R. Douady, *These de 3-eme cycle*, Universite de Paris 7, 1982

[**DT**] F. Dogru and S. Tabachnikov, *Dual Billiards*, Math Intelligencer vol. 27 No. 4 (2005) 18–25

[**EV**] D. B. A. Epstein and E. Vogt, *A Counterexample to the Periodic Orbit Conjecture in Codimension 3*, Annals of Math **108** (1978) pp 539-552

[**G**] D. Genin, *Regular and Chaotic Dynamics of Outer Billiards*, Penn State Ph.D. thesis (2005)

[**GS**] E. Gutkin and N. Simanyi, *Dual polygonal billiard and necklace dynamics*, Comm. Math. Phys. **143** (1991) 431–450

[**Ke**] R. Kenyon, *Inflationary tilings with a similarity structure*, Comment. Math. Helv. **69** (1994) 169–198

[**Ko**] Kolodziej, *The antibilliard outside a polygon*, Bull. Polish Acad Sci. Math. **37** (1989) 163–168

[**M**] J. Moser, *Stable and Random Motions in Dynamical Systems, with Special Emphasis on Celestial Mechanics*, Annals of Math Studies **77**, Princeton University Press (1973)

[**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*, Journal of Modern Dynamics **3** (2007)

[**T1**] S. Tabachnikov, *Geometry and Billiards*, A.M.S. Math. Advanced Study Semesters (2005)

[**T2**] S. Tabachnikov, *A proof of Culter's theorem on the existence of periodic orbits in polygonal outer billiards*, preprint (2007)

[**VS**] F. Vivaldi, A. Shaidenko, *Global stability of a class of discontinuous dual billiards*, Comm. Math. Phys. **110** (1987) 625–640