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 areas as interval exchange maps, piecewise isometric maps, and area-preserving dynamics. There is a lot left to learn about these kinds of dynamical systems, and a deep 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 the general scope of the sort of phenomena we uncovered for the Penrose kite.

In this monograph we will give a more 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 a convex quadrilateral with an axis of bilateral symmetry that goes through two opposite vertices. 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.

One novel feature of this investigation is its close connection to computer experimentation. I discovered every single result in this monograph by experimenting with my computer program, Billiard King, a graphical user interface I wrote for the purpose of solving the Moser-Neumann problem. 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 and play around with it. Billiard King relates to this monograph much in the way that a cooked meal relates to a recipe. Billiard King is a well-documented java program, available from my website, that runs on many platforms.

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. This result was one of the guiding inspirations for my work here. It may seem wierd to relate outer billiards to hyperbolic 3-manifolds, but there was a strangely persistent analogy that kept me going.

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.

Note: This monograph is perhaps not yet in its final form. I would appreciate any comments or suggestions you have about the text. I would especially like to hear about typos, misprints, and inaccuracies of any kind.

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 polygon exchange maps. See [**T1**] and [**DT**] for an exposition of the subject 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.

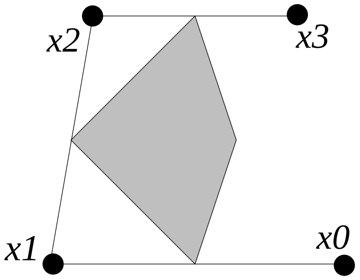

Figure 1.1: Outer Billiards

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 a very 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 [1] and also regular polygons. (For kites, quasi-rational implies rational.) 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 has uncountably many unbounded orbits when defined relative to the Penrose kite, thereby answering the Moser-Neumann question in the affirmative. The Penrose kite is the convex quadrilateral that arises in the Penrose tiling. We also worked out some of the finer dynamical structure of these orbits.
- D. Dolgopyat [**Do**] has very recently produced a preprint that shows that outer billiards on the semidisk has unbounded orbits. The semidisk is the convex hull of a semicircle. This is a second affirmative solution to the Moser-Neumann problem.

The purpose of this monograph is to give a robust affirmative answer to the Moser-Neumann problem, and to develop the theory of outer billiards on kites. Our work here is an outgrowth of [**S**]. 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.

---

[1] Outer billiards is affinely invariant, and so we call a polygon rational if some affine image has all rational vertices.

1.2 Main Results

A kite K is *rational* iff it is affinely equivalent to a polygon with rational vertices, and otherwise *irrational*. We define the kite $K(A)$ to be the one with vertices

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

Every kite is affinely equivalent to some $K(A)$. The Penrose kite is $K(\sqrt{5}-2)$. The kite $K(A)$ is rational iff A is rational. We call p/q an *odd rational* if pq is odd, and otherwise an *even rational*.

Say that an outer billiards orbit on K is *erratic* if both the forwards and backwards directions of the orbit are unbounded, and also return infinitely often to every neighborhood of the vertex set of the kite. On a large scale, an erratic orbit oscillates in the widest possible sense.

Theorem 1.1 (Erratic Orbits) *Outer billiards on any irrational kite has uncountably many erratic orbits.*

The Erratic Orbits Theorem is one of three main results in the monograph. The other two results, the Hexagrid Theorem from §3 and the Master Picture Theorem from §10, cannot be stated precisely without a build-up of terminology.

Informally speaking, the Hexagrid Theorem states that the arithmetic structure of a certain class of outer billiards orbits on odd rational kites is controlled by a multigrid made from 6 infinite families of parallel lines. In naming our result, we have in mind an analogy with De Bruijn's famous pentagrid construction of the Penrose tilings [**De**]. The Hexagrid Theorem has one corollary that is easy to state.

Corollary 1.2 *Let* $A = p/q$ *be an odd rational. Then the orbit of the point* $(1/q, 1)$*, relative to outer billiards on* $K(A)$*, has diameter at least* $q/2$.

A version of Corollary 1.2 holds for even rational kites, but this case does not contribute to our proof of the Erratic Orbits Theorem, so we ignore it.

We think of the Master Picture Theorem, some version of which ought to hold for polygonal outer billiards in general, as a good step towards a comprehensive theory of polygonal outer billiards. We will discuss it below in §1.6, and then in great detail in §10.

1.3 The Main Construction

Let $\boldsymbol{Z}_{\rm odd}$ denote the set of odd integers. Order 2 rotation in any of the vertices of $K(A)$ preserves the set $\boldsymbol{R} \times \boldsymbol{Z}_{\rm odd}$. Therefore, the outer billiards map preserves $\boldsymbol{R} \times \boldsymbol{Z}_{\rm odd}$. We restrict our attention to orbits on $\boldsymbol{R} \times \boldsymbol{Z}_{\rm odd}$. We could probably develop our theory for the general orbit, but this seems quite a bit more complicated.

We prefer to work with the square of the outer billiards map, which we denote by ψ. This map is much nicer. For all $p \in \boldsymbol{R}^2$ on which ψ is defined, we have

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

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.

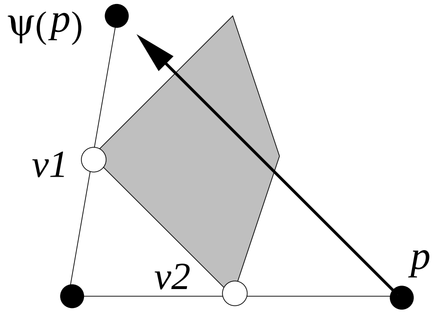

Figure 1.2: The square of the outer billiards map

Given any $\alpha \in \boldsymbol{R}$, the set

$$\Omega_\alpha = (\alpha + 2\boldsymbol{Z}[A]) \times \boldsymbol{Z}_{\rm odd} \tag{3}$$

is invariant under ψ. The *offset* α is chosen so that the outer billiards map is well-defined on all points of Ω_α.

At least when A is irrational, the set Ω_α is naturally bijective with $\boldsymbol{Z}^3$. Thus, ψ induces a more "discrete" dynamical system on $\boldsymbol{Z}^3$. Roughly speaking, we choose a section $\boldsymbol{Z}^2 \subset \boldsymbol{Z}^3$ and consider the return map $\Psi : \boldsymbol{Z}^2 \to \boldsymbol{Z}^2$. The slice $\boldsymbol{Z}^2 \subset \boldsymbol{Z}^3$ corresponds roughly to the two rays $\boldsymbol{R}_+ \times \{-1, 1\} \subset \boldsymbol{R}^2$. Roughly, we join each $p \in \boldsymbol{Z}^2$ to $\Psi(p) \in \boldsymbol{Z}^2$, in order to produce a kind of graph $\widehat{\Gamma}_\alpha(A)$. See §2.4 for the precise construction. The graph $\widehat{\Gamma}_\alpha(A)$ encodes the orbits of Ψ on $\boldsymbol{Z}^2$. We call this graph the *arithmetic graph*.

When $A = p/q$ is an odd rational, we set $\alpha = 1/q$. This is our canonical choice. It turns out that we can still carry through the above constructions, although points in $\boldsymbol{Z}^3$ do not correspond uniquely to points of $\Omega_{1/q}$. We still produce a graph $\widehat{\Gamma}(p/q)$, whose vertices lie in $\boldsymbol{Z}^2$. It turns out that $\widehat{\Gamma}(p/q)$ is always invariant under translation by the vector $(q, -p)$. The map ψ permutes the intervals of length $2/q$ centered at the points $(a/q, b)$ where $a, b \in \boldsymbol{Z}_{\text{odd}}$. We call these intervals *basic intervals*. $\widehat{\Gamma}(p/q)$ encodes the return dynamics of the basic intervals to the two rays $\boldsymbol{R}_+ \times \{-1, 1\}$.

Let $\Gamma(p/q)$ denote the connected component of $\widehat{\Gamma}(p/q)$ that contains $(0, 0)$. We set things up so that $\Gamma(p/q)$ records the way the orbit $O(p/q)$ of the interval

$$I_A = (0, 2/q) \times \{1\} \tag{4}$$

returns to $\boldsymbol{R}_+ \times \{-1, 1\}$. We call $O(p/q)$ the *fundamental orbit*. Corollary 1.2 concerns the fundamental orbit.

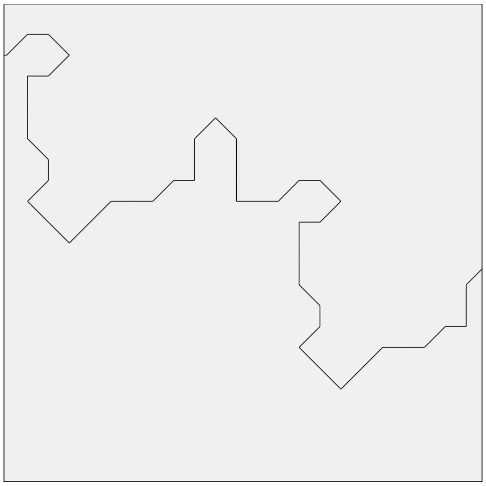

Figure 1.3: The graph $\Gamma(7/13)$.

$\Gamma(p/q)$ is an infinite embedded polygonal curve that is invariant under addition by the vector $(q, -p)$. See the Embedding Theorem from §2.4. We say that the *baseline* of $\Gamma(A)$ is the line of slope $-A$ through the origin. When $A = p/q$, the baseline is invariant under addition by the vector $(-q, p)$. See Figure 1.3. It turns out that $\Gamma(p/q)$ lies entirely in the halfplane above its baseline. The distance from any given vertex of $\Gamma(p/q)$ to the baseline is proportional to the distance from the corresponding basic interval in $\boldsymbol{R}_+ \times \{-1, 1\}$ to the origin. Thus, the higher $\Gamma(A)$ rises up above its baseline, the farther the fundamental orbit wanders from the origin.

When p/q is an even rational, $\Gamma(p/q)$ seems to be an embedded closed polygon. We ignore the even case because it does not contribute to our proof of the Erratic Orbits Theorem.

Our general goal is to take a sequence $\{p_n/q_n\}$ of odd rationals approximating A in the best possible sense, and consider a geometric limit of the corresponding sequence $\{\Gamma_n\}$ of arithmetic graphs. We will take a suitable geometric limit Γ_∞, and deduce the Erratic Orbits Theorem from the properties of Γ_∞.

We will illustrate our method with some pictures. 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 Copy Theorem from §5 gives a precise result along these lines.

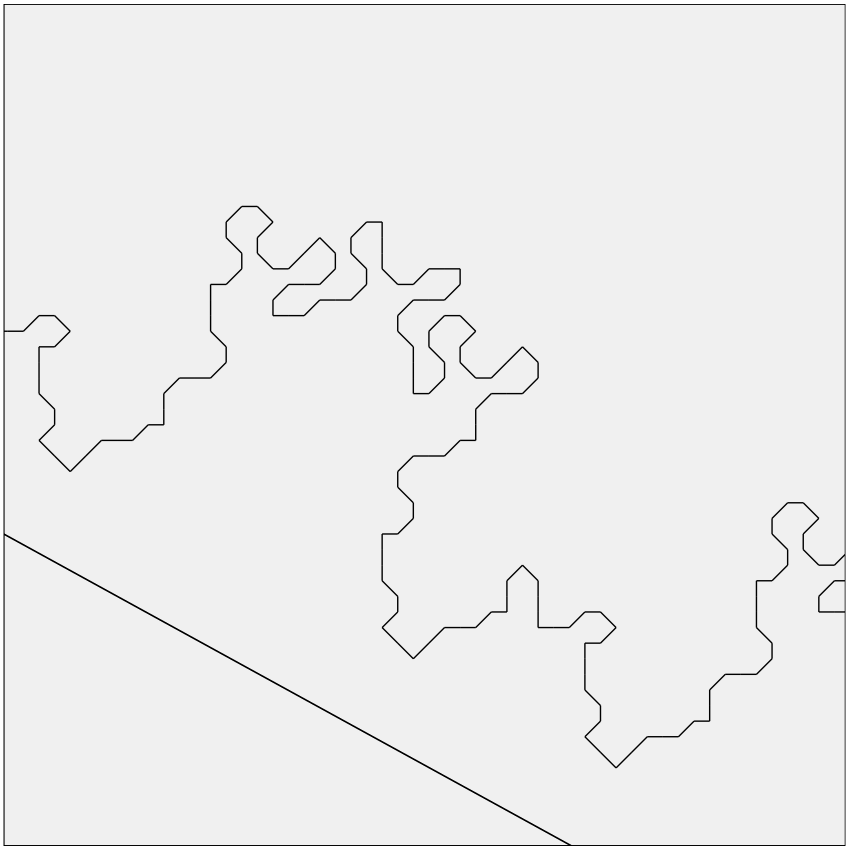

Figure 1.4: The graph $\Gamma(19/35)$.

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 $\Gamma(45/83)$ oscillates on 3 scales. What makes this work is that 19/35 is a good approximation of 45/83.

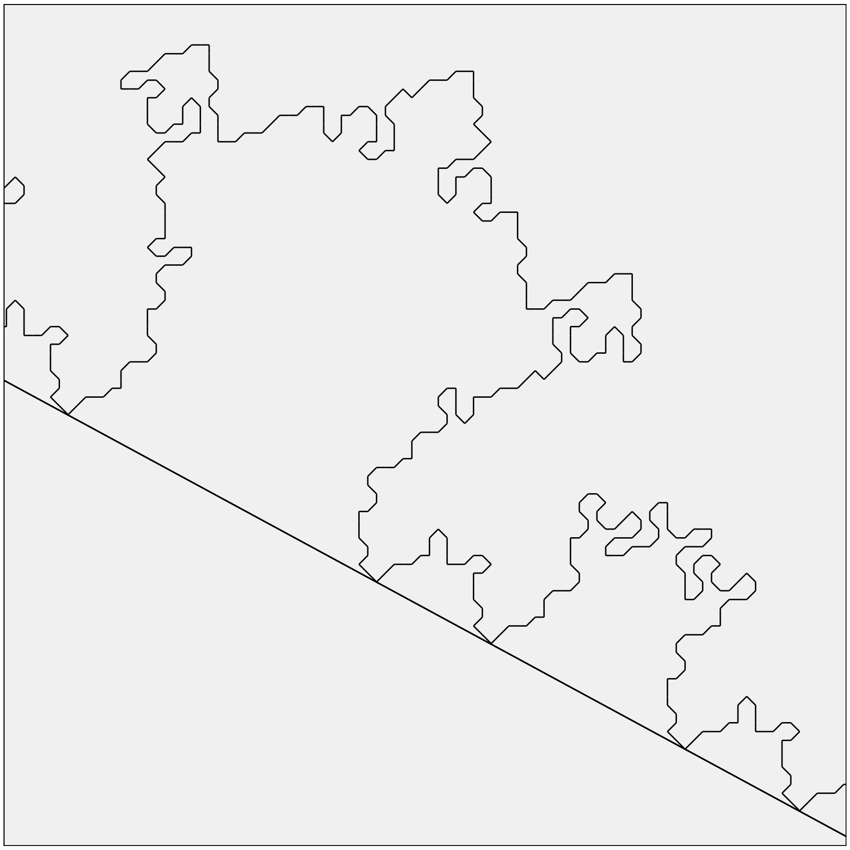

Figure 1.5: The graph $\Gamma(45/83)$.

Continuing in this way, and taking a limit, we produce a graph $\Gamma(A)$ that oscillates on all scales and hence rises unboundedly far from its baseline. We devote a whole chapter, §4, to discussing the subtleties of the above limit. The main problem, in addition to the one of producing an unbounded limit, is producing a limit that actually describes a well-defined orbit on the limiting dynamical system. We solve this problem by forcing the limiting graph $\Gamma(A)$ comes close to the baseline along a set of points that form a kind of large scale Cantor set. It turns out that we need to produce uncountably many erratic orbits just to find one unbounded orbit!

1.4 The Case of the Penrose Kite

First of all, we note that

$$\sqrt{5}-2=\phi^{-3}; \qquad \phi=\frac{1+\sqrt{5}}{2}. \tag{5}$$

Here ϕ is the golden ratio.

In [**S**],we proved that outer billiards on $K(\phi^{-3})$ has erratic orbits by considering the arithmetic graph $\Gamma(\phi^{-3})$. Our offset value is $\alpha=\phi^{-2}$. We showed that $\Gamma(\phi^{-3})$ oscillates on infinitely many scales by showing that $\Gamma(\phi^{-3})$ lies within 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**].

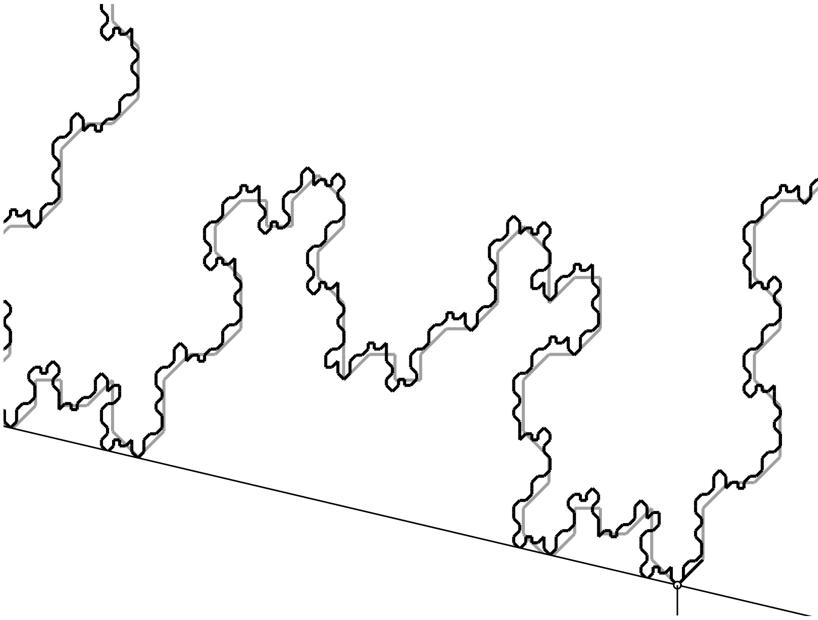

Figure 1.6: The graph $\Gamma(\phi^{-3})$ superimposed over $\phi^3\Gamma(\phi^{-3})$.

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. In general, we do not get such a symmetric picture, and the scales of oscillation of $\Gamma(A)$ depend on the Diophantine properties of the parameter A.

1.5 Diophantine Approximation

The Copy Theorem, the result that lets us produce the large scale Cantor set structure, depends on the Diopantine properties of the numbers involved.

Say that $A \in \boldsymbol{R}$ belongs to the set $\Delta(k)$ if there is a monotone sequence $\{p_n/q_n\}$ of odd rationals converging to A such that

$$\left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| < \frac{1}{kq_n^2}; \qquad \forall n. \tag{6}$$

For any k, the set $\Delta(k)$ has full measure. Curt McMullen pointed out to me that every irrational number belongs to $\Delta(1)$.

We will prove the following results:

- If $A \in \Delta(3)$ then Γ_{n+1} copies 2 periods of Γ_n for all n.
- If $A \in \Delta(2)$, then Γ_{n+1} copies a bit more than 1 period of Γ_n for all n.

See the Weak Copy Theorems from §5, and their corollaries. We will first prove the Erratic Orbits Theorem when $A \in \Delta(3)$, where the argument is maximally clean, and then we scramble a bit to handle $A \in \Delta(2)$.

After we prove Erratic Orbits Theorem for $A \in \Delta(2)$ we turn to the general case. In general, we need to choose the odd rationals in our sequence very carefully to get the enough period copying. To help us, we develop a little theory of odd rational approximation that is similar to the theory of continued fractions. See §7. I'm sure that nothing in this theory is new. The only thing that would prevent the results we state from being well known is their idiosyncratic nature.

The period copying phenomenon is quite closely related to the theory of odd rational approximation. The reader should compare Lemma 7.2, a structural result about odd rational approximation, with Lemma 24.1, a strikingly similar result about period copying. It seems that outer billiards on kites is closely related to the modular group $SL_2(\boldsymbol{Z})$, or more exactly the level 2 congruence subgroup. We will rigorously develop this connection so far as it serves the proof of the Erratic Orbits Theorem. However, after proving the Erratic Orbits Theorem we will discuss an even deeper connection between the two topics, something we call the *Modular Limit Phenomenon*. See §9.4.

1.6 The Master Picture Theorem

We would like establish properties of $\Gamma(A)$ in a way that does not require us to keep referring back to the dynamical properties of outer billiards on $K(A)$. The content of our Master Picture Theorem is that the local structure of $\Gamma(A)$ is controlled by the following data:

- A 3 dimensional torus $\boldsymbol{R}^3/\Lambda_A$, where Λ_A is lattice given by an upper-triangular matrix defined over $\boldsymbol{Z}[A]$. See Equation 74.
- A certain fundamental domain $R_A = (0, 1 + A)^2 \times (0, 1)$ for the action of Λ_A, together with a partition of R_A into finitely many polyhedra.
- A pair of maps $M_+, M_- : \boldsymbol{Z}^2 \to R_A$. The local structure of $\Gamma(A)$ at the point (m, n) is determined by which polyhedra contain the points $M_+(m, n)$ and $M_-(m, n)$.

When we consider the union

$$R = \bigcup_{A \in (0,1)} R_A \times \{A\} \subset \boldsymbol{R}^4. \tag{7}$$

we get a convex polytope that is partitioned into smaller convex polytopes. The entire partition is defined over $\boldsymbol{Z}$, in the sense that the equations describing the walls of the partition are linear, with integer coefficients. We call this 4 dimensional polytope partition the *master picture*. The master picture fibers over the interval of parameters, and each fiber has the form R_A. In short, our Master Picture Theorem relates the structure of the arithmetic graph $\Gamma(A)$ to the master picture.

One consequence of the Master Picture Theorem is that first return map to $\boldsymbol{R}_+ \times \{-1, 1\}$, for the parameter A, is controlled by a certain polyhedron exchange map defined on two copies of R_A. The return map itself is an infinite interval exchange map, and we are saying that there is a certain compactification of $\boldsymbol{R}_+ \times \{-1, 1\}$ that wraps the whole picture up as a dense and compatible subset a union of two 3-tori.

Putting all these polyhedron exchange maps together, we produce a single polytope exchange map defined on two copies of the polytope R. This 4 dimensional polytope exchange map simultaneously controls the return dynamics for outer billiards on all kites.

John Smillie and I think that this kind of master picture should persist for general polygonal outer billiards. We hope to eventually work this out.

1.7 Computational Issues

As we mentioned above, we discovered all the structure of outer billiards by experimenting with Billiard King. In places, our proof is fairly computational. Ultimately we are trying to verify the structure we noticed on the computer. Our proof here uses considerably less computation than our argument in [**S**], but we still use a computer-aided proof in several places. Mainly, we use the computer to check a small finite number of special cases by hand.

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.

1.8 Organization of the Monograph

This monograph comes in 4 parts. In Part I we prove Erratic Orbits Theorem modulo 5 auxilliary theorems:

- The Return Lemma. (Stated in §2.4 and proved in part III.)
- The Embedding Theorem. Stated in §2.5 and proved in part III.)
- The Hexagrid Theorem. Stated in §3.4 and proved in part III.)
- The Copy Theorem. Stated in §8.1 and proved in Part IV.)
- The Decomposition Theorem (Stated in §8.2 and proved in Part IV.)

These results all follow from the Master Picture Theorem, which we state and prove in Part II. In Parts III and IV, we prove all the auxilliary results, as indicated in our list above.

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

Part I

In this part of the monograph we prove the main result, the Erratic Orbits Theorem, modulo the 5 auxilliary results mentioned in §1.8. Here is an overview of the chapters.

- In §2 we give some background material on polygonal outer billiards, and then define the arithmetic graph. The arithmetic graph is our main object of study. Along the way, we state the Return Lemma and the Embedding Theorem, two of our basic results about the arithmetic graph.
- In §3 we state the Hexagrid Theorem, our main structural result about the arithmetic graph. We also deduce Corollary 3.3, which gives lower bounds on the amount of oscillation of the arithmetic graph for odd rational parameters.
- In §4 we informally discuss how to take a geometric limit of rational arithmetic graphs, using the forced oscillation property to produce an unbounded arithmetic graph in the irrational limit. The upshot of §4 is that we need to combine the oscillation estimate with a period copying phenomenon.
- In §5 we state the Weak Copy Theorem, a fairly weak result about the period copying phenomenon. This result, combined with Corollary 3.3, suffices to prove the Erratic Orbits Theorem for all $A \in \Delta(2)$.
- In §6 we prove the Erratic Orbits Theorem for $A \in \Delta(2)$.
- In §7 we develop a little theory about approximating irrational numbers by odd rational numbers. This theory parallels the ordinary theory of continued fraction approximants.
- In §8 we state the Copy Theorem and the Decomposition Theorem, two results which give more precise information about the nature of the period copying phenomenon.
- In §9 we prove the Erratic Orbits Theorem in general. We also prove several additional results, and present some conjectures based on experimentally observed phenomena.

2 The Arithmetic Graph

In this chapter we give some background material on outer billiards, and then proceed to the definition of the arithmetic graph. The arithmetic graph is the main object of study in this monograph.

2.1 Polygonal Outer Billiards

Let P be a convex polygon. We always denote the outer billiards map on P by ψ'. Our handedness convention is that a person walking from p and $\psi'(p)$ would see the shape P on the right. Thus, ψ' is defined using the left tangent line, from the point of view of the point to which ψ' is applied.

For any point $p \in \boldsymbol{R}^2 - P$ that lies on a line extending a side of P, one of the two maps ψ' or $(\psi')^{-1}$ is not defined. If we start pulling back these lines by the dynamics then we produce a countably infinite set $\boldsymbol{L}$ of line segments with the following property. $p \in \boldsymbol{L}$ if and only if some power of ψ' is not defined on p. This set $\boldsymbol{L}$ is frequently called the *limit set*.

We let $\psi = (\psi')^2$ denote the square of the outer billiards map. We usually prefer to work with ψ because it is a piecewise translation of $\boldsymbol{R}^2 - P$. By convention, when we say that ψ is defined, we mean that ψ' is also defined. Thus, ψ and all its powers are defined away from $\boldsymbol{L}$.

One basic property of outer billiards is that it is affinely natural. Let T be any affine transformation of the plane. Since T preserves ratios of distances on each individual line, T carries an outer billiards orbit on P to an outer billiards orbit on $T(P)$. For this reason, we can always normalize P so that 3 of its vertices are the vertices of a square. The remaining vertices of P lie in the square.

If P is a kite, then P is affinely equivalent to the kite $K(A)$ with vertices

$$(-1,0); \qquad (0,1) \qquad (0,-1) \qquad (A,0); \qquad A \in (0,1). \tag{8}$$

We deliberately ignore the case $A = 1$, which corresponds to the square. Thus, the space of all kites is parametrized by the interval $(0,1)$.

As we mentioned in the introduction, D. Genin [**G**] worked out the picture of outer billiards on trapezoids. Modulo affine transformations, the trapezoids also form a 1-parameter family. After the trapezoids, the kites form the next simplest family of examples.

2.2 Rational Outer Billiards

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 a locally finite tiling of the complement of the polygon. Each tile is a convex tile. Figure 2.1 shows the picture for the kite $K(1/3)$.

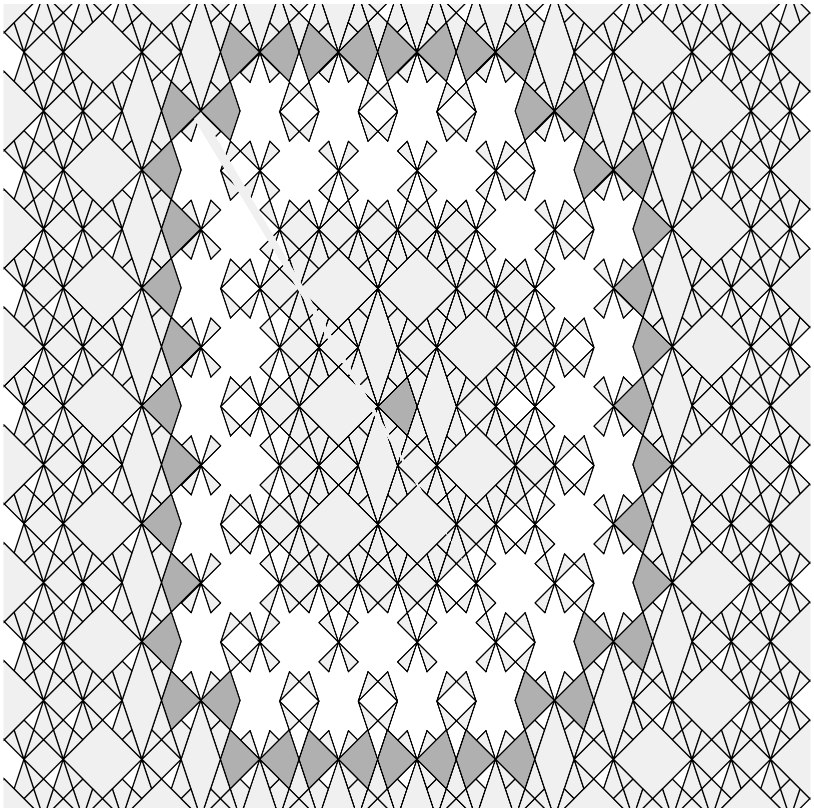

Figure 2.1: *Part of the Tiling for $K(1/3)$.*

This is the simplest tiling we see amongst all the kites. We have only drawn part of the tiling. The grey orbit consists of a necklace of tiles, each one an isometric copy of $K(1/3)$. These necklaces exist in general, and were discovered by olodziej in **K** *. Outside the necklace, the pattern repeats in some sense. The interested reader can draw many pictures like this using Billiard ing.*

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 the first really nontrivial pictures.

2.3 Algebraic Dynamics and the Arithmetic Graph

This section is meant to be an informal discussion. We first consider the general situation when we have a map $\psi : \boldsymbol{R}^n \to \boldsymbol{R}^n$*, perhaps defined only on a large subset of* $\boldsymbol{R}^n$*. Suppose that* ψ *is given locally by equations defined over some finite dimensional* $\boldsymbol{Z}$*-module* $R \approx \boldsymbol{Z}^k$*. Here the symbol* $\approx$ *denotes an isomorphism of* $\boldsymbol{Z}$*-modules. The map* ψ *induces an action on*

$$\boldsymbol{Z}^{nk} \quad (\boldsymbol{Z}^{\ k})^n \approx R^n \subset \boldsymbol{R}^n, \qquad (\ )$$

or at least on some subset of $\boldsymbol{Z}^{nk}$*. If we are lucky,* ψ *will be defined on enough of* $\boldsymbol{Z}^{nk}$ *for us to extract some useful information. In particular, if we are looking for unbounded orbits of* ψ*, we might as well first look in* $\boldsymbol{Z}^{\ nk}$*.*

Our construction above produces a more discrete dynamical system defined on a higher dimensional lattice. We trade low dimensionality for discreteness. One advantage of having a dynamical system defined on a lattice is that it might allow us to introduce methods of a combinatorial nature to study the system. One might say that we are pulling apart the dynamics in order to take a closer look. The reader can imagine that sometimes this construction would be helpful and sometimes not.

The situation we discussed above typically happens for polygonal outer billiards, because the $\boldsymbol{Z}$*-module generated by the coordinates of the vertices is finite dimensional. For the kite* $K(A)$*, we have* $R \quad \boldsymbol{Z}(A)$*. When* A *is irrational, we have* $\boldsymbol{Z}\ A\ \approx \boldsymbol{Z}^2$*. Thus, outer billiards on an irrational kite naturally gives rise to a certain dynamical system defined (or at least partially defined) on* $\boldsymbol{Z}^4$*.*

In trading $\boldsymbol{R}^2$ *for* $\boldsymbol{Z}^4$*, we have gained some discreteness, but we have lost the ability to visualize things. Fortunately, it turns out that outer billiards on* $K(A)$ *preserves the 1-dimensional subset* $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$*. Here* $\boldsymbol{Z}_{\text{odd}}$ *is the set of odd integers. This gets us down to a dynamical system on* $\boldsymbol{Z}^{\ 3}$*. It turns out that there is a certain subset* $\boldsymbol{Z}^{\ 2} \subset \boldsymbol{Z}^3$ *that serves as a section for the dynamics. By considering the return map to this section, we get to draw pictures on* $\boldsymbol{Z}^2$*. Roughly speaking, the arithmetic graph is the record, in* $\boldsymbol{Z}^{\ 2}$*, of the return dynamics we have been discussing.*

Our discussion above works best when A *is irrational. When* A *is rational, the discussion above is slightly misleading, and we need to define the arithmetic graph with some care. Needless to say, the reader should pay more attention to our formal definitions below than to the informal discussion given in this section.*

2.4 Kites and Fundamental Orbits

Until the last result in this section the parameter A is an odd fraction, p/q. Say that a **basic interval** *is an open horizontal interval of length $2/q$ centered at a point of the form $(a/q, b)$ where $a, b \in \boldsymbol{Z}_{\text{odd}}$. Here $\boldsymbol{Z}_{\text{odd}}$ is the set of odd integers.*

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

Proof: *We note first that the order 2 rotations about the vertices of $K(A)$ send the point (x, y) to the point:*

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

Let ψ_1 denote the outer billiards map on $K(A)$. The map ψ_1 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 $\subset \boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$ denote the set of points of the form $(2a \quad 2bA, 2c \quad 1)$, with $a, b, c \in \boldsymbol{Z}$. The complementary set $^c \quad \boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}} -$ is the union of the basic intervals.

Looking at Equation 7, we see that $\psi_1(x) \in {}^c$ provided that $x \in {}^c$ and ψ_1 is defined on x. To prove this lemma, it suffices to show that ψ_1 is defined on any point of ^{c}.

To find the points of $\boldsymbol{R} \times \boldsymbol{Z}_{\text{odd}}$ where ψ_1 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.

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

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

More generally, the same proof gives:

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

When A is irrational, the set $\boldsymbol{Z} \quad A$ is dense in $\boldsymbol{R}$. However, it is always just a countable set and, from our point of view, easy to avoid.

2.5 The Return Map

Let $A \in (\ , 1)$ be any parameter. As in §1 we mean to dissuss outer billiards on the kite $K(A)$. We know from Lemma 2.2 that the entire outer billiards orbit of $(\alpha, 1)$ is defined provided that $\alpha \notin 2\boldsymbol{Z}\ A\ $. We have the following basic result. The proof is a bit long, and we prefer to prove it later on.

Lemma 2.3 (Return) The forwards and backwards ψ orbits of $p\quad (x, 1)$ eventually return to the set $\boldsymbol{R}_+ \times \{-1, 1\}$. The same goes for $(x, -1)$.

Proof: *This is a consequence of the Pinwheel Lemma, proved in §11.* ♠

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 $\boldsymbol{R}_+ \times \{-1, 1\}$.*
(ii) Once we develop the picture a bit, the reader will see that the Return Lemma is obvious for points $(x, \pm 1)$, where x is sufficiently large. The tricky part of the result is for fairly small values of x.
(iii) Reflection in the x-axis conjugates the map ψ to the map $\psi^{\ 1}$. Thus, once we understand the orbit of the point $(x, 1)$ we automatically understand the orbit of the point $(x, -1)$.

We let $(x, \pm 1)$ denote the first point in the forwards orbit of ψ that lies on $\boldsymbol{R}_+ \times \{-1, 1\}$. Given the nature of the maps in Equation 7 comprising ψ, we see that

$$(x, \pm 1) - (x, \pm 1) \in 2\boldsymbol{Z}\ A\ \times \{-2,\ , 2\}. \qquad (\ )$$

In fact, it turns out that

$$(x, \pm 1) - (x, \pm 1)\quad 2(A\epsilon_1\quad \epsilon_2, \epsilon_3); \quad |\epsilon_j| \in \{-1,\ , 1\}; \quad {}^{3}_{j=1}\ \epsilon_j \equiv\quad \text{mod } 2. \qquad (\ )$$

The Spiral Lemma, proved in §12, establishes the bound $|\epsilon_j| \leq 1$. The parity result is an immediate consequence of the Pinwheel Lemma, proved in §11. On a nuts-and-bolts level, this monograph concerns how to determine the triple (ϵ_1, ϵ_2) given $(x, \pm 1)$. The pair (ϵ_1, ϵ_2) and the parity condition determine ϵ_3.

2.6 The Arithmetic Graph

Define $M \quad M_{A,\alpha} : \boldsymbol{Z}^2 \to \boldsymbol{R}$ *by the equation*

$$M_{A,\alpha}(x,y) \quad 2Ax \quad y \quad 2\alpha. \qquad (1\ )$$

For ease of notation, set $M \quad M_{A,\alpha}$*. Note that* M *is never injective if we extend it to all of* $\boldsymbol{R}^2$*. However, this is not something we wish to do.*

When A *is irrational,* M *is injective. In the rational case, we have* $M(m,n) \quad M(m \quad q, n-p)$ *when* $A \quad p/q$*, but at least* M *is injective on disks of radius less than*

$$r \quad p\ \overline{{}^{2} \quad q\ {}^{2}}. \qquad (11)$$

We set $r \quad \infty$ *when* A *is irrational.*

Given $p_1, p_2 \in \boldsymbol{Z}^2$ *we join* p_1 *to* p_2 *by a line segment, and write* $p_1 \leftrightarrow p_2$*, if and only if the following conditions are satisfied:*

- $x_1 \quad M(p_1)$ *and* $x_2 \quad M(p_2)$ *are positive.*
- x_2 *is the first coordinate of* (x_1, ϵ) *for some* $\epsilon \in \{-1, 1\}$ *and furthermore* $\|p_2 - p_1\| < r$*.*

The second condition is symmetric in p_1 *and* p_2 *because reflection in the* x*-axis conjugates to* ^{1}*. In other words,* $p_1 \leftrightarrow p_2$ *iff* $p_2 \leftrightarrow p_1$*. The* **furthermore** *condition of course is only necessary in the rational case.*

We call the resulting graph the **arithmetic graph** *and denote it by* ${}_{\alpha}(A)$*. When* $A \quad p/q$ *we have the canonical choice* $\alpha \quad 1/(2q)$ *and then we set* $(A) \quad {}_{1/(2q)}(p/q)$*. We say that the* **baseline** *of* (A) *is the line of slope* $-A$ *through the origin. One important structural result we will establish is:*

Theorem 2.4 (Embedding) **For any** $A \in (\ , 1)$ **and and** $\alpha \notin \boldsymbol{Z}\ A$ **the graph** ${}_{\alpha}$ **is a disjoint union of embedded polygons and embedded infinite polygonal curves. Each edge of** ${}_{\alpha}(A)$ **has length at most** $\sqrt{2}$**.**

Proof: *See Part III of the monograph.* ♠

We are generally not interested in all of , but only in the component that contains $(\ ,\ )$*. When* A *is rational, this component encodes the structure of the fundamental orbit* $O(A)$*. It is easy to check that* (A) *is* *nontrivial and contains the arc* $(-1,1) \to (\;,\;) \to (1,1)$ *for all* $A \in (\;, 1)$. *Sometimes we abuse our terminology and call* (A) *the arithmetic graph, when technically* (A) *is only one component of it.*

We experimentally observe the following dichotomy:

1. *When* p/q *is odd,* (p/q) *is an infinite embedded polygonal curve, invariant under translation by* $(q,-p)$*. Figures 1.2-1.4 illustrate this. See Corollary 3.3 in the next chapter.*
2. *When* p/q *is an even,* (p/q) *is an embedded polygon.*

We shall not discuss the even case, because it does not contribute to our proof of the Erratic Orbits Theorem. The interested reader can explore the structure in the even case using Billiard King. Just for fun, we include one picture in the even case. The parameter we have chosen is one of the continued fraction approximants to the Penrose parameter $\sqrt{5}-2$.

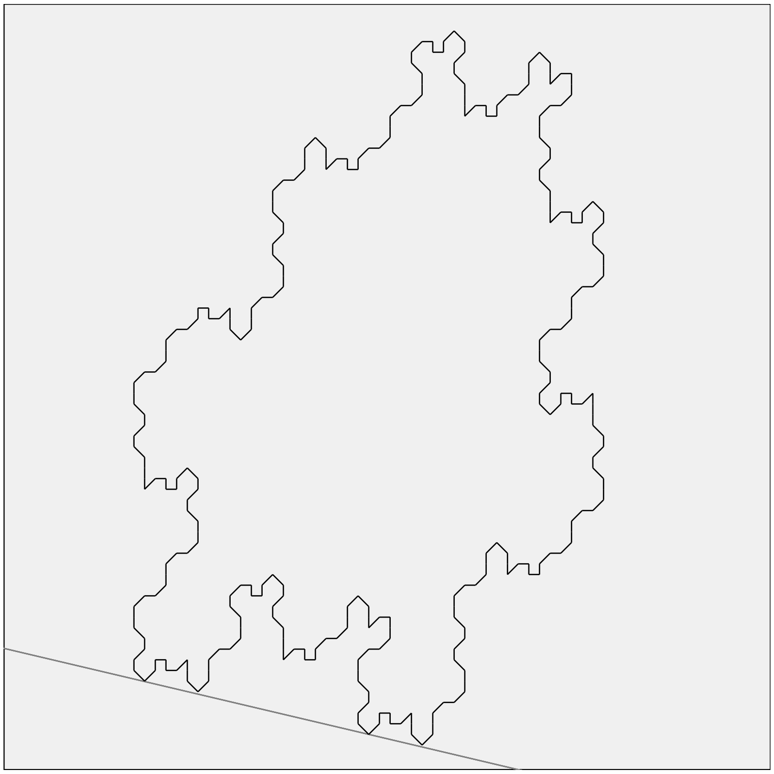

Figure 2.3: *The graph* $(17/72)$.

2.7 Producing Erratic Orbits

From the way we have set things up, the outer billiards orbit of the point $(\alpha, \pm 1)$, relative to the kite $K(A)$, is erratic provided that the following five properties hold.

1. *The orbit of $(\alpha, \pm 1)$ is entirely defined.*
2. *The forwards direction of $_\alpha(A)$ rises unboundedly far away from its baseline. Recall that the baseline is the line of slope $-A$ through $(\;,\;)$.*
3. *The forwards direction of $_\alpha(A)$ comes arbitrarily close to the line of slope $-A$ through the point $(\;, -\alpha)$. This line is parallel to the baseline.*
4. *The backwards direction of $_\alpha(A)$ rises unboundedly far away from its baseline. Recall that the baseline is the line of slope $-A$ through $(\;,\;)$.*
5. *The backwards direction of $_\alpha(A)$ comes arbitrarily close to the line of slope $-A$ through the point $(\;, -\alpha)$. This line is parallel to the baseline.*

The **forwards direction** *of is defined to be the half-infinite portion of that starts at $(\;,\;)$, continues to $(1, 1)$, and so on. The* **backwards direction** *of is defined to be the half-infinite portion of that starts at $(\;,\;)$, continues to $(-1, 1)$, and so on.*

3 The Grid Structure

The purpose of this chapter is to detail our main structural result about the arithmetic graph, the Hexagrid Theorem.

3.1 The Arithmetic Kite

Let $A \quad p/q$ be an odd rational. Our construction is based on a certain affine copy $Q(A)$ of the kite $K(A)$.

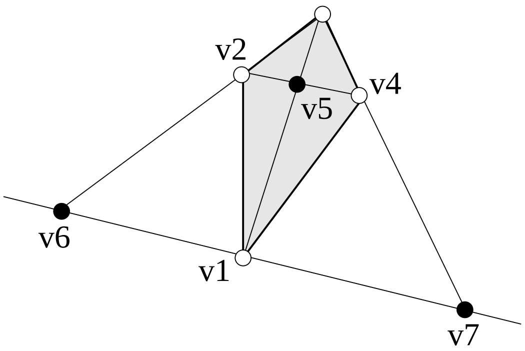

Figure 3.1: *The arithmetic kite*

Figure 3.1 shows a schematic picture of $K(A)$. The vertices are given by the equations.

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

. $v_6 \quad (-q, p).$

7. $v_7 \quad (q, -p).$

Lemma 3.1 *$K(A)$* **and** *$Q(A)$* **are affine equivalent.**

Proof: *This is just a calculation. First of all, we compute that* v_2, v_5, v_4 *are collinear and* $\|v_2 - v_5\| \quad \|v_4 - v_5\|$. *Next, we compute that* v_1, v_5, v_3 *are collinear and* $\|v_3 - v_5| \quad A\|v_1 - v_5\|$. *Hence* v_5 *is the intersection of the diagonals of* $Q(A)$, *and* v_5 *divides these two diagonals into the same proporations as the one sees for* $K(A)$. *This suffices to show that* $Q(A)$ *is affine equivalent to* $K(A)$. ♠

Figure 3.2 shows $Q(A)$ *for* $A \quad 73/13$. *We picked this parameter essentially at random, and noticed that it has nice pictures. We will use this value for all our pictures in this chapter. The interested reader can draw the pictures for essentially any parameter, using Billiard ing.*

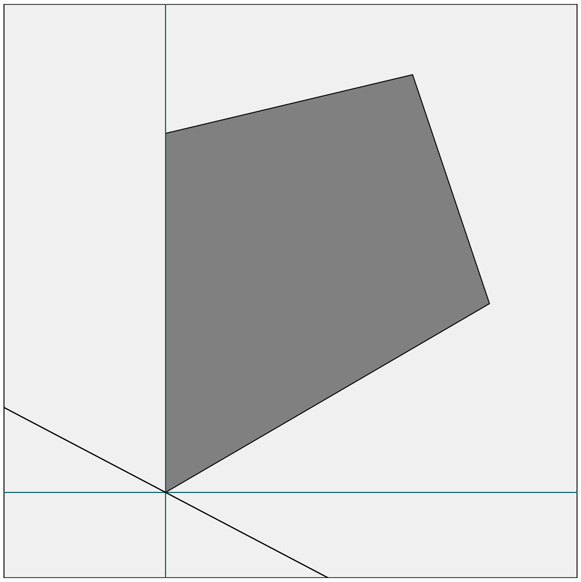

Figure 3.2: *The arithmetic kite* $Q(73/13\ )$.

We single out the vectors

$$V \quad v_7 \quad (q, -p); \qquad W \quad v_5 \quad \frac{pq}{p \quad q}, \frac{pq}{p \quad q} \quad \frac{q-p}{2} \quad . \tag{12}$$

These vectors play a fundamental role in the structure of the arithmetic graph (A).

3.2 The Room Grid

The **room grid** *RG consists of two infinite families of parallel lines. We will give two descriptions, and below we will explain why we give this grid the name* **the room grid***.*

First Description: *The room grid RG(A) consists of the lines obtained by extending the diagonals of Q(A) and then taking the orbit under the lattice* $\boldsymbol{Z}$ $V/2, W$.

Second Description: *One family of lines in RG(A) has slope* $-A$. *We call these lines the* **floors***. We let* L_k *denote the floor that contains the point* kW*. So, The minus sign indicates that these lines have negative slope. The other family of lines in RG(A) has positive slope* $1-$ $(2/A)$ $(A/2)$*. We call these lines the* **walls** *We let* L_k^+ *denote the wall that contains the point* $kV/2$*.*

Figure 3.3 illustrates our construction for the parameter 73/13 .

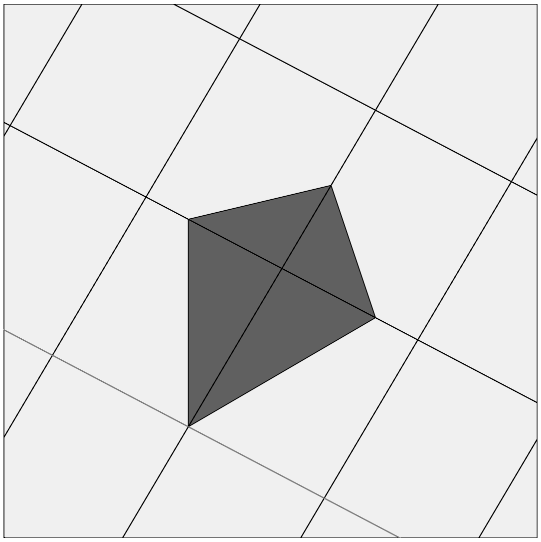

Figure 3.3: *RG(73/13) and Q(73/13).*

The grid RG(p/q) divides $\boldsymbol{R}^2$ *into different connected components which we call* **rooms***. The reason for this terminology will become clear momentarily. Recall that (p/q) consists of some infinite polygonal arcs and some polygons. Figure 3.4 shows some of the infinite polygonal arcs of (73/13) drawn in relation to RG(73/13).*

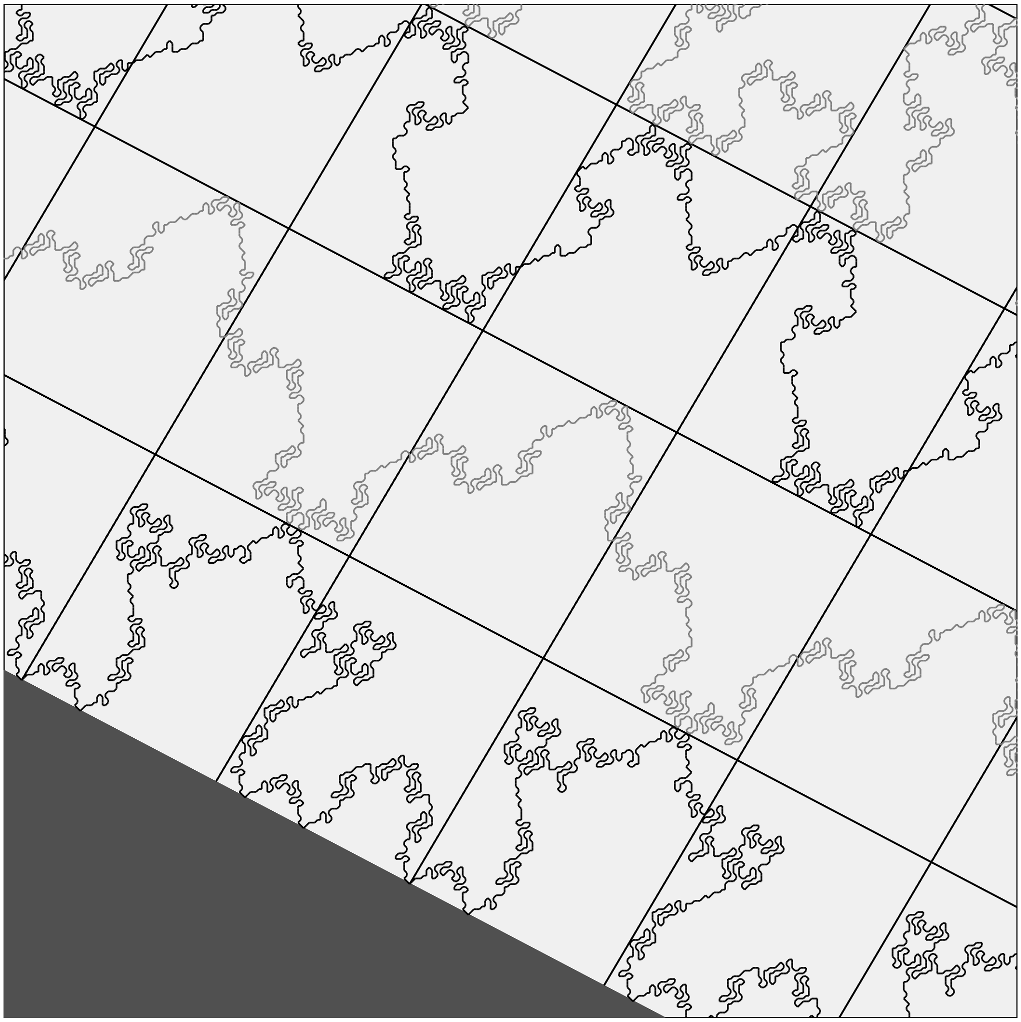

Figure 3.4: *RG*$(73/13\ )$ *and some of*$(73/13\ )$.

There are several things to notice about this picture. First of all, the polygonal arcs are stacked on top of each other, and separated by the floors of the room grid. Second of all, each polygonal arc only crosses each wall once. In other words, we can break each polygonal arcs into pieces, each of which lives in a single room. The pieces of then cross the walls to join each other. Two consecutive rooms comprise a single period for each polygonal arc. The next thing we would like to understand is where the arcs cross the walls. We hope that the reader is starting to see the reason behind our terminology.

3.3 The Door Multigrid

The **door multigrid** $DG(A)$ *is obtained by extending the sides of* $Q(A)$ *and then taking their orbit under the lattice* $\mathbf{Z}\ V, W$ *. Figure 3.5 shows a picture for the parameter* $73/13$ *. Together* $RG(A)$ *and* $DG(A)$ *give us infinite families of lines. We call these two grids together the* **hexagrid***, and we denote it by* $G(A)$*.*

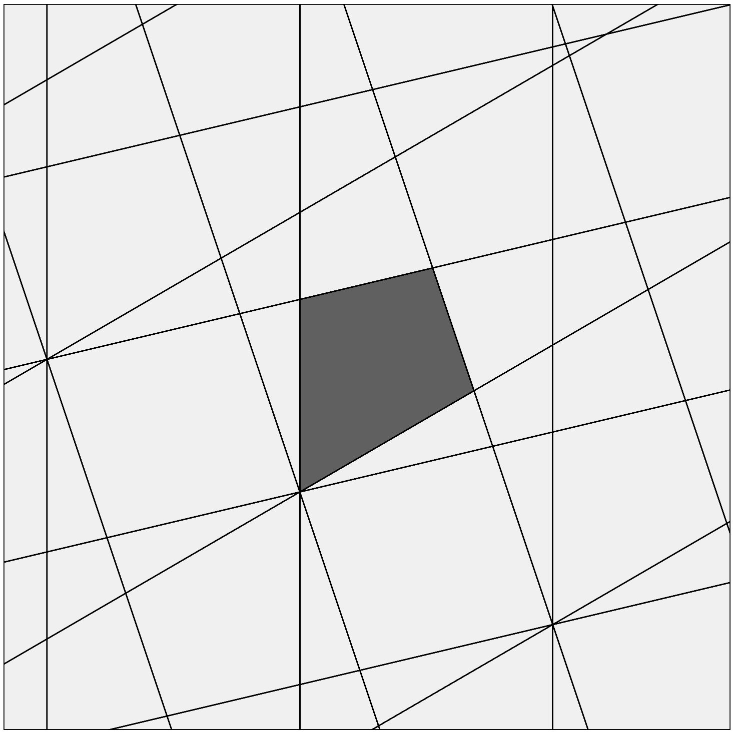

Figure 3.5: *$DG(73/13\ )$ and $Q(73/13\ )$.*

Say that a **room wall** *is the line segment, contained in a wall, that divides two adjacent rooms. So, in other words, each wall of the room grid is divided into infinitely many room walls. Say that a* **door** *is a point of intersection between a wall of* $RG(A)$ *and room wall. At every door at least 2 door lines cross a wall. In other words, every door is a triple point. Moreover, there is exactly one door on each room wall. Occasionally, a door occurs at the corner of a room, as for (,). In this case, we count the door twice, once for the room above and once for the room below. The first coordinate of a door is always an integer. See §2 for details about this.*

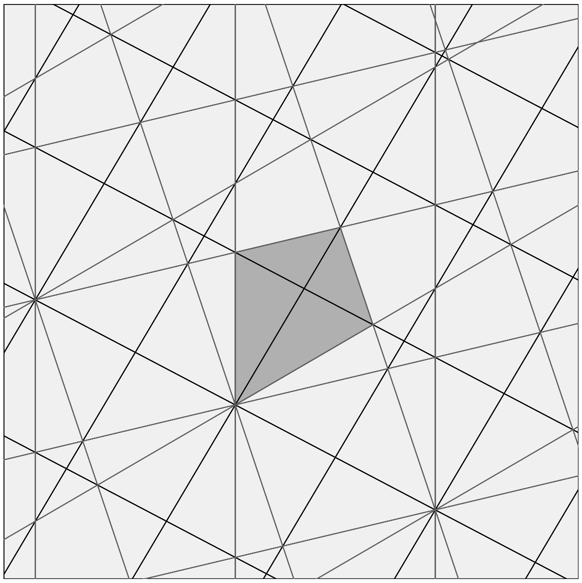

Figure 3.6: *$G(73/13\ )$. and $Q(73/13\ )$.*

3.4 The Hexagrid Theorem

Now we come to the main result of the Chapter. Say that an edge e of **crosses a wall** *if e intersects a wall L at an interior point. Say that a union of two incident edges of* **crosses a wall** *if the common vertex lies on a wall, and the two edges of incident point to opposite sides of the wall. The point (,) 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. We say that the* **crossing point** *is the point where the crossing cell intersects the wall.*

Theorem 3.2 (Hexagrid) Let A be an odd rational. Then the following two statements are true about the arithmetic graph (A)

1. **(A) never crosses a floor of $RG(A)$.**
2. **There is a bijection between the set of doors and the set of crossing cells. The bijection pairs a given door with the unique crossing cell that crosses the room wall containing the door. The crossing point occurs within 1 unit of the door.**

Proof: *See Part III of the monograph.* ♠

Figure 3.7 illustrates the Hexagrid Theorem. The shaded parallelogram, a region we denote by $R(73/13\ )$, isolates one period of $(73/13\ )$. We will discuss this region in detail below. The top right corner of $R(73/13\ )$ is the vector V, namely the point of $Q(73/13\ )$ where the diagonals cross.