Chapter One

Introduction

1.1 DEFINITIONS AND HISTORY

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

Outer billiards is a dynamical system defined (typically) in the Euclidean plane. Unlike the more familiar variant, which is simply called *billiards*, outer billiards involves a discrete sequence of moves outside a convex shape rather than inside it. To define an outer billiards system, one starts with a bounded convex set $K \subset \mathbf{R}^2$ and considers a point $x_0 \in \mathbf{R}^2 - K$. One defines x_1 to be the point such that the segment $\overline{x_0x_1}$ is tangent to K at its midpoint and K lies to the right of the ray $\overrightarrow{x_0x_1}$. The iteration $x_0 \to x_1 \to x_2 \to \cdots$ is called the *forward outer billiards orbit* of x_0. It is defined for almost every point of $\mathbf{R}^2 - K$. The backward orbit is defined similarly.

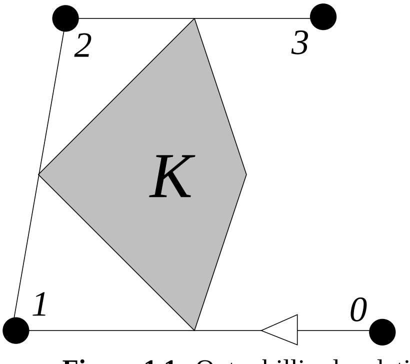

Figure 1.1: Outer billiards relative to K.

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

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

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

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

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

---

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

1.2 THE ERRATIC ORBITS THEOREM

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

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

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

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

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

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

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

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

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

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

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

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

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

1.3 COROLLARIES OF THE COMET THEOREM

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

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

The interval

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

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

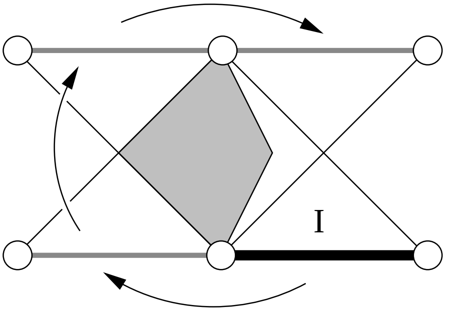

Figure 1.2: I and its first 3 iterates.

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

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

1. *U_A is minimal: Every orbit in U_A is dense in U_A and all but at most 2 orbits in U_A are both forward dense and backward dense in U_A.*

2. *U_A is locally homogeneous: Every two points in U_A have arbitrarily small neighborhoods that are isometric to each other.*

3. *$U_A \cap I = C_A^\#$ for some Cantor set C_A.*

Remarks:

(i) One endpoint of C_A is the kite vertex $(0, -1)$. Hence Statement 1 implies that all but at most 2 unbounded special orbits are erratic. The remaining special orbits, if any, are each erratic in one direction.

(ii) Statements 2 and 3 combine to say that every point in U_A lies in an interval that intersects U_A in a trimmed Cantor set. This gives us a good local picture of U_A. One thing we are missing is a good global picture of U_A.

(iii) The Comet Theorem describes C_A explicitly.

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

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

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

that conjugates f_1 to f_2.

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

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

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

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

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

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

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

Remarks:

(i) The Comet Theorem explicitly describes $\mathcal{Z}_A$ in terms of a sequence we call the *remormalization sequence*. This sequence is related to the continued fraction expansion of A. We will give a description of this sequence in the next section.
(ii) Theorem 1.4 is part of a larger result. There is a certain suspension flow over the odometer, which we call *geodesic flow on the cusped solenoid*. It turns out that the time-one map for this flow serves as a good model, in a certain sense, for the dynamics on U_A. §24.3.

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

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

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

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

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

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

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

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

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

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

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

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

---

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

1.4 THE COMET THEOREM

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

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

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

We define

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

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

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

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

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

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

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

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

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

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

The identification works like this.

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

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

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

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

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

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

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

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

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

Define

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

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

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

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

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

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

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

Remarks:

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

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

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

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

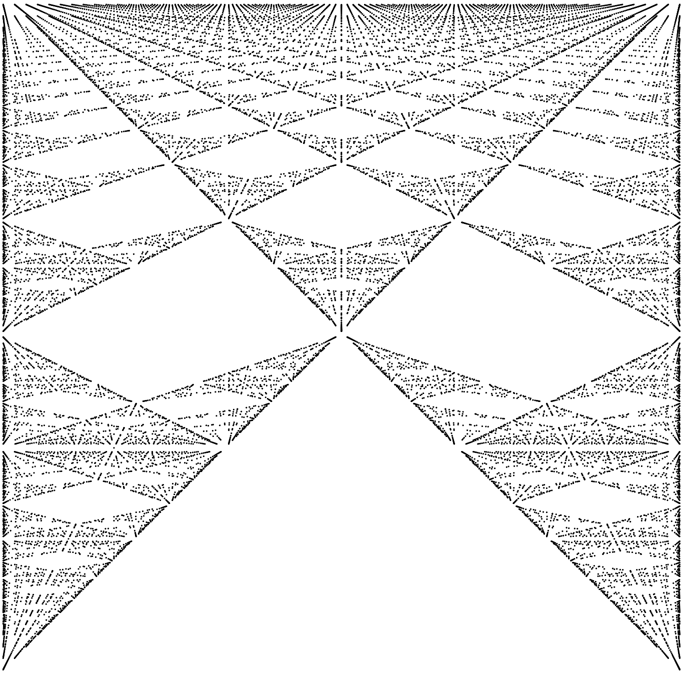

Figure 1.3: The set C. The bottom is $A=0$ and the top is $A=1$.

1.5 RATIONAL KITES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

---

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

1.6 THE ARITHMETIC GRAPH

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

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

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

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

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

Here $A = p/q$.

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

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

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

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

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

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

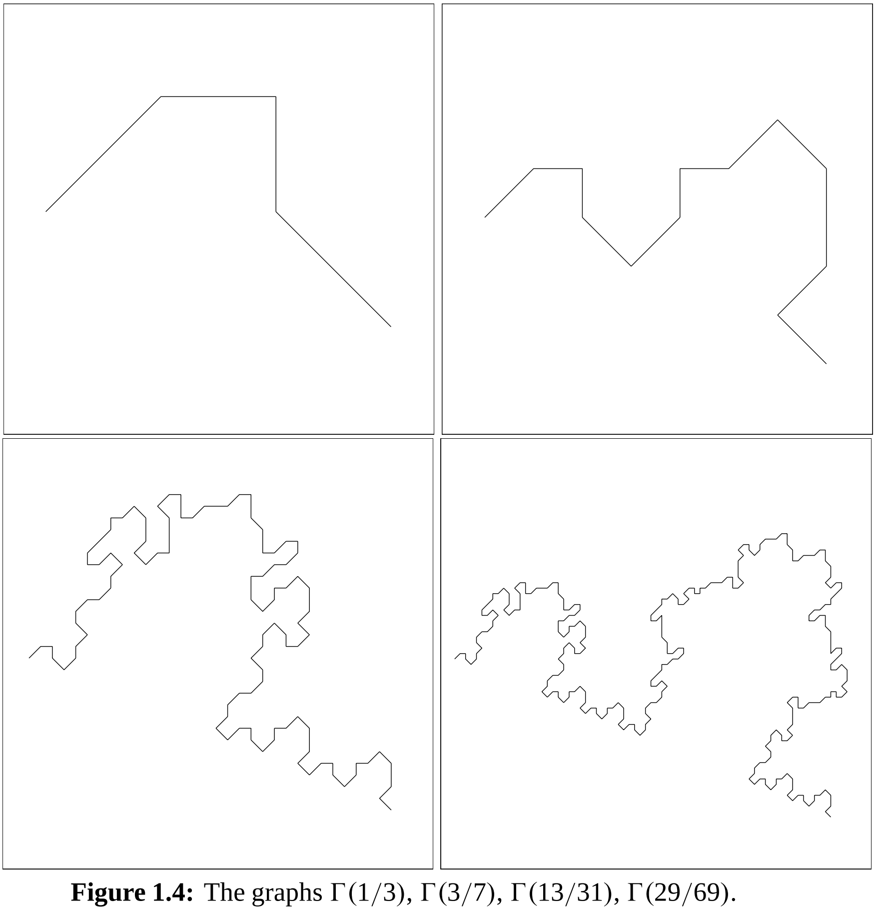

Figure 1.4: The graphs $\Gamma(1/3)$, $\Gamma(3/7)$, $\Gamma(13/31)$, $\Gamma(29/69)$.

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

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

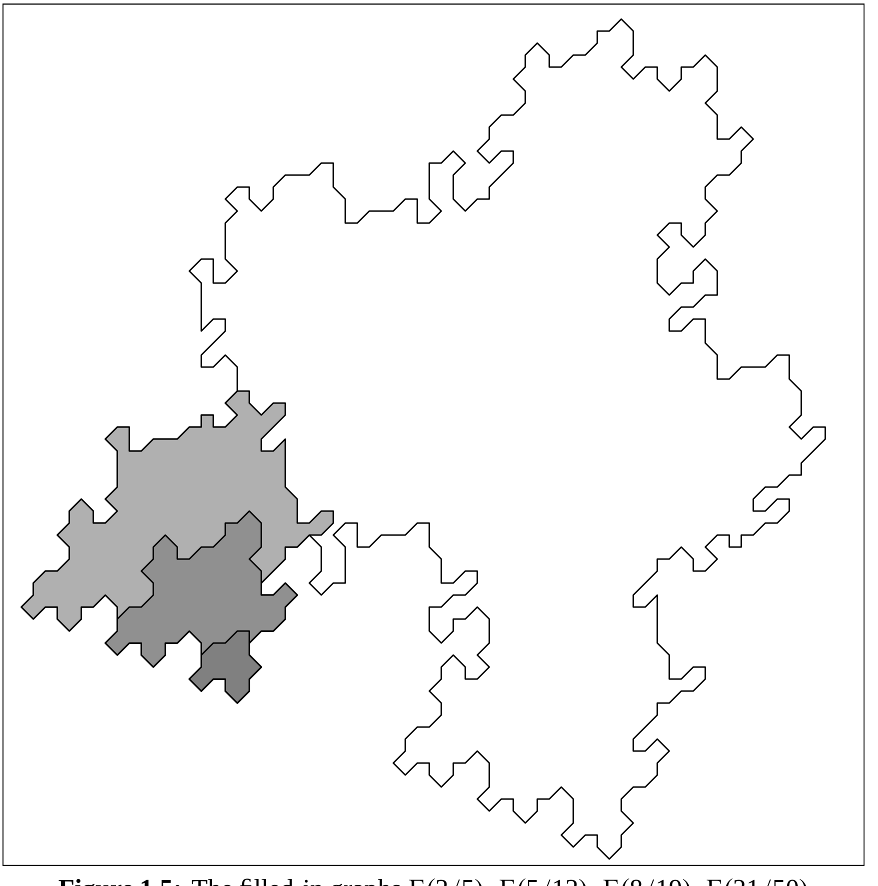

Figure 1.5: The filled-in graphs $\Gamma(2/5)$, $\Gamma(5/12)$, $\Gamma(8/19)$, $\Gamma(21/50)$.

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

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

1.7 THE MASTER PICTURE THEOREM

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

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

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

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

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

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

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

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

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

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

1.8 REMARKS ON COMPUTATION

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

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

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

1.9 ORGANIZATION OF THE BOOK

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

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

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

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