UNIVERSITY OF CALIFORNIA
RIVERSIDE

An Investigation of the Koch Snowflake Fractal Billiard: Experimental and Theoretical Results

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Mathematics

by

Robert Garrett Niemeyer

June 2012

Dissertation Committee:

Dr. Michel L. Lapidus, Chairperson
Dr. Fred Wilhelm
Dr. James Kelliher

The Dissertation of Robert Garrett Niemeyer is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

When one writes a dissertation, he or she never knows how bad it's going to get. After the writing is done, one somewhat forgets how bad it was. Completing a thesis is an arduous task. Moreover, one sometimes discovers what exactly his or her dissertation was about halfway through writing it. Such an epiphany typically comes when one is the most discouraged, and this is what fuels the rest of the writing process.

I have—like so many others—had my doubts and failures, and each time I got back up, there was someone standing there right beside me. It may not have been the same person each time, but to say I made it through alone would be a lie. Below I thank a few key people in my life; by no means is this list complete.

I must first thank Jack Bennett, who is now teaching at Odessa Junior College in Odessa, Texas. Jack is a good friend with an incredible work ethic. He and I studied an entire summer, at least ten hours per day and only stopping for bathroom breaks and lunch. He kept me on task for the algebra qualifier. This wasn't the only qualifier for which we prepared together; complex analysis and topology were the other two. Of course, I must also thank him for humoring me in my attempts at humor and allowing me to be a part of his wonderful family.

Professor Lapidus has been a great adviser; he has also been a great friend. From my first project as an undergraduate student at UCR to the thesis defense, Professor Lapidus has been an excellent mentor through it all; the lessons I have learned from him as an undergraduate and graduate student have shaped me as a mathematician. Never giving me a drink, but always leading me to the water, I will never forget what he said when I told him that I wanted to pursue a career in research: "I'm proud of you."

My brother, Rick, has always believed in me. He is my best friend and someone

for whom I will always be there. He has always served as an example to me on what to do (and, perhaps, *what not to do*), and he is someone I continue to look up to. As we finally leave California for our respective careers, we both know the other will be just a phone call away (or a seven-hour drive). Words cannot properly express how much I have appreciated his guidance over the years.

Finally, I must address the two people most responsible for the eventual writing of this thesis: my mother and father. I must thank my mother for letting me take the occasional 'sick day' in high school to get caught up on schoolwork, and I can still remember my father and me going for ice-cream on those tougher days. My parents, Rick and Renee, raised my brother and me to believe in the power of education. Though they never went to college, they knew how important it was for my brother and me to go further in our schooling. Ultimately, because of their countless sacrifices and their message to my brother and me, we both obtained our PhDs. Though my father never went to college, he had the perfect advice for learning any subject: "Did you reread the previous chapter?" This has always stuck with me and this is exactly the question I ask all my students when they are struggling. Their example has shaped me as a person and an academic. As I get older, my appreciation for them grows.

There are many more people I could acknowledge. The list is truly extensive, and even the words I've written above do not adequately express my sentiment. Please know that if you are someone who has impacted my life, your influence was greatly appreciated and I have benefited tremendously from having known you.

To my parents, Rick and Renee, for their constant support and motivation. And to my brother, Rick, for always being there.

ABSTRACT OF THE DISSERTATION

An Investigation of the Koch Snowflake Fractal Billiard:
Experimental and Theoretical Results

by

Robert Garrett Niemeyer

Doctor of Philosophy, Graduate Program in Mathematics
University of California, Riverside, June 2012
Dr. Michel L. Lapidus, Chairperson

In this dissertation, we will focus our attention on the limiting behavior of a sequence of compatible orbits of Koch snowflake prefractal billiards. We give a number of theoretical results. In one instance, we construct a well-defined billiard orbit of the Koch snowflake fractal billiard. In another instance, we construct what we call a nontrivial polygonal path. Such a path is constructed from a sequence of basepoints from orbits of prefractal billiards that converges to what we call an elusive limit point of the Koch snowflake. Many other results are concerned with particular properties of particular types of sequences of compatible orbits. One major result is a topological dichotomy for sequences of compatible orbits. A number of figures are used to illustrate the various definitions and concepts. We discuss experimental results indicating the possible existence of a well-defined law of reflection (or some appropriate analogue) for the Koch snowflake fractal billiard. Future directions for research are discussed and a number of conjectures and open questions are given.

Contents

List of Figures

Chapter 1

Introduction

This thesis constitutes the culmination of our research on the Koch snowflake fractal billiard ([LapNie1, LapNie2]). Specifically, we expand upon the content of [LapNie2] by further investigating what constitutes a billiard orbit of the Koch snowflake fractal billiard $\Omega(KS)$; see Figure 1.1. The Koch snowflake KS is everywhere nondifferentiable. The absence of a well-defined tangent at any point of KS is what, a priori, prevents one from determining a billiard flow on $\Omega(KS)$. While the focus of [LapNie1] was to provide experimental evidence of such an appropriately defined fractal law of reflection, we do not propose to determine a "*fractal law of reflection*" in this thesis. A more complete discussion of and conjectures on such such a law of reflection and the Koch snowflake billiard can be found in Chapter 6 and [LapNie2, §6]. However, we do determine a number of properties of what we are calling *sequences of compatible orbits*, and a family of sequences of compatible orbits whose suitable limits constitute likely candidates for orbits of the Koch snowflake fractal billiard $\Omega(KS)$.

The unique nature of the subject of fractal billiards requires that we provide certain background material so that those familiar with only some of the subjects find

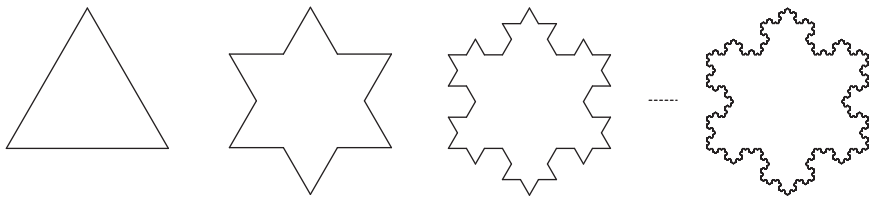

Figure 1.1: The Koch snowflake curve KS and its prefractal approximations KS_n, for $n = 0, 1, 2, ..., 6$.

immediately accessible the others. Moreover, we make an attempt at catering to the advanced undergraduate student who might find this material of interest. To such end, we provide a brief treatment of measure theory, key definitions and theorems from ergodic theory (e.g., Birkhoff and Maximal ergodic theorems), a detailed discussion of the subject of flat surfaces and the necessary concepts from fractal geometry. In addition to this, we touch on the notion of an inverse limit and a few necessary facts for establishing the fractality of the inverse limit of what we call *footprints of piecewise Fagnano orbits*. The interested reader will find references to [Ma] for further details on covering spaces, [McL] for category theory, [Bo] for general topology and [HoYo] for the specialized topic of inverse and direct limits in the context of the category of topological spaces.

The Koch snowflake curve, as depicted in Figure 1.1, is a fractal. In particular, it is the union of three *self-similar* Koch curves, with the Koch curve being a continuous, nowhere differentiable curve with infinite length (see Figures 1.2 and 1.3). Consequently, any attempt to construct a line tangent to the Koch snowflake curve may seem like an exercise in futility. This poses a unique problem for defining the trajectory of a *billiard ball* (i.e., a pointmass traversing the interior of the planar region bounded by the Koch snowflake KS). Specifically, when this pointmass collides with the boundary KS with unit speed, the absence of a well-defined tangent results in multiple choices for the angle of reflection, meaning there is, a priori, no well-defined angle of reflection.

We give a thorough description of the billiard flow associated with a billiard table $\Omega(B)$ with (piecewise) smooth boundary B. We also recall the notion of a *flat surface* and how one can construct a flat surface from a *rational polygonal billiard table* (that is, a planar billiard table whose boundary is a polygon with interior angles that are rational multiples of π); see Definitions 64 and 69. In this context, a flat surface is a mathematical device used to rigorously describe the billiard flow on $\Omega(B)$ in terms of the geodesic flow on the surface.

We explain how the Cantor set $\mathscr{C}$ can be viewed as the inverse limit of an *inverse limit sequence* of its prefractal approximations, denoted by $\mathscr{C}_n$, with the index n corresponding to the approximation with 2^n many points. Similarly, the snowflake curve KS can be viewed as the inverse limit of its prefractal approximations KS_n given in terms of finite addresses. Here, KS_n is the nth (inner) polygonal approximation to KS, and hence defines a rational polygonal billiard table $\Omega(KS_n)$; see Definition 64 for the definition of *rational billiard*. Since the theory of rational polygonal billiards is very well developed (see, e.g., [GaStVo,Gtk1,GtkJu1–2,HuSc,HaKa,KaZe,Mas,MasTa,Ve1–3,Vo,Zo]), it is then natural to define the dynamics on the fractal "billiard table" $\Omega(KS)$ in terms of the dynamics on its prefractal approximations $\Omega(KS_n)$. As a result, much of the focus of this thesis will be to first obtain a solid understanding of the periodic orbits of $\Omega(KS_n)$, and then discuss theoretical and experimental results indicating the existences of well-defined orbits of the Koch snowflake fractal billiard $\Omega(KS)$.

The main results of the thesis are presented throughout Chapters 3–5. Chapter 3 contains results on the prefractal flat surface $\mathcal{S}(KS_n)$ and consequences for the fact that such a surface can be tiled by hexagons of a particular scale. Primarily, we show that $\mathcal{S}(KS_n)$ is a branched cover of the hexagonal torus $\mathcal{S}(KS_0)$. Such a result is tantamount to the billiard flow on $\Omega(KS_0)$ being dynamically equivalent to the billiard flow

on $\Omega(KS_n)$, for all $n \geq 0$. This fact is used extensively in Chapter 4.

Using the main results of Chapter 3, we show in Chapter 4 how to form what we call sequences of compatible orbits. We show that directions for which an orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ of $\Omega(KS_n)$ is periodic are exactly the same for which an orbit $\mathscr{O}_0(x_0^0, \theta_0^0)$ of $\Omega(KS_0)$ is periodic. This aides us in constructing what we call a *sequence of compatible periodic orbits*. We initially focus our investigation on orbits with an initial direction of $\pi/3$. As such, we describe what we call a *sequence of compatible piecewise Fagnano orbits*, a *sequence of compatible Cantor orbits* and *sequence of compatible approximate piecewise Fagnano orbits*. The period and length of *piecewise Fagnano orbits*, *Cantor orbits* and *approximate piecewise Fagnano orbits* of $\Omega(KS_n)$ are given in terms of the ternary representation of the initial basepoint of the initial orbit of the respective sequence of compatible periodic orbits.[1] In directions not $\pi/3$, we define what we are calling hybrid orbits of $\Omega(KS_n)$ and provide a set of sufficient conditions for which a sequence of compatible orbits is a sequence of compatible hybrid orbits. Such orbits (and sequences of compatible orbits) have particular properties that indicate that suitable limits of sequences of hybrid orbits may indeed be orbits of the Koch snowflake fractal billiard $\Omega(KS)$. In general, it is our intention to insinuate that such sequences should have suitable limits that constitute orbits of the Koch snowflake billiard.

As we describe in Chapter 5, the only family of well-defined orbits of the Koch

[1] The initial direction of $\pi/3$ is special, but much can be said about orbits of prefractal approximations with such an initial direction. As such, much of [LapNie2] was devoted to understanding the nature of orbits of prefractal approximations with an initial direction of $\pi/3$ and specific properties of such orbits. Such properties include, but are not limited to, the period and length of orbits of prefractal approximations with an initial direction of $\pi/3$ and the fractality of an inverse limit of footprints of piecewise Fagnano orbits. We elaborate on such topics, this being the focus of a sizable portion of Chapter 4.

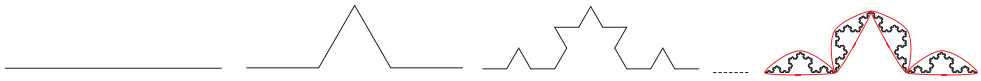

Figure 1.2: The Koch curve KC and its prefractal approximations KC_n, for $n = 0, 1, 2, ..., 6$. The Koch curve is *self-similar*, meaning that there are scaled (and rotated/reflected) copies of the Koch curve found as subsets of KC. In this figure, we circle four copies of KC scaled by $1/3$. In general, one can find 4^n copies of the Koch curve scaled by $1/3^n$ as subsets of KC such that the disjoint union (disjoint except at the endpoints of each copy) comprises the whole Koch curve. This is, in fact, the essence of self-similarity. By abuse of language, we say that the Koch snowflake curve itself is "self-similar"; see Figure 1.3.

snowflake billiard $\Omega(KS)$ is the family consisting of what we call *finitely stabilizing periodic orbits*. A footprint $\mathcal{F}_n(x_n^0, \pi/3)$ of an orbit $\mathcal{O}_n(x_n^0, \pi/3)$ of a prefractal approximation $\Omega(KS_n)$ amounts to the points of the Poincaré section constituting the collision points of the orbit. That which we propose to be a piecewise Fagnano orbit of $\Omega(KS)$ has a footprint $\mathcal{F}(x^0, \pi/3)$ that is the inverse limit of footprints of piecewise Fagnano orbits of the prefractal approximations. While we say "piecewise Fagnano orbit," we are making an abuse of language in that we do not mean to imply that such an orbit of $\Omega(KS)$ actually exists, but that whatever the orbit truly *is*, it has a footprint $\mathcal{F}(x^0, \pi/3)$. Furthermore, even less is known about what we have called the *approximate piecewise Fagnano orbits*. Again, there really is no orbit to speak of, nor is there, a priori, any footprint to speak of.

In general, we show that different notions of limit applied to different types of sets yield sets of interest that are analogous with similar sets found in the classical case. We show that 1) the inverse limit of a *particular* inverse limit sequence of footprints exists and 2) the standard notion of limit involving the Euclidean metric yields a particular collection of basepoints coming from a *sequence of compatible orbits* that converges to what we call an *elusive limit point* of KS.

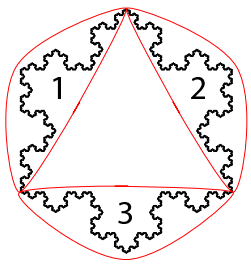

Figure 1.3: The self-similarity of the Koch snowflake curve KS. Shown here is the Koch snowflake curve KS, viewed as the union of three isometric, abutting copies of the Koch curve.

The more experienced reader may notice that we do not need much of the background material presented in order to understand the main results. Such an effort was made to make the thesis accessible to everyone ranging from the uninitiated undergraduate to the experienced billiards expert. On the other hand, presenting the background material in such a way is meant to indicate the direction in which we intend to go in the long-term. While we do not make any specific claims about the ergodicity of the billiard flow in "irrational" directions, we do make use of these concepts and discuss a number of open questions in Chapter 6 that draw upon the ideas from ergodic theory. Such a discussion will largely focus on the possible ergodic properties of the proposed flow on the conjectured fractal flat surface $\mathcal{S}(KS)$. In particular, we lead up to the conjecture on the existence of a meaningful group of affine automorphisms of the conjectured fractal flat surface $\mathcal{S}(KS)$.

Considering the fact that the field of "fractal billiards" is still in its infancy, we provide many open questions and conjectures in Chapter 6. We stress that $\Omega(KS)$ does not constitute a well-defined mathematical billiard, in the sense that we have not provided a well-defined phase space, let alone a geodesic flow on such a phase space. Such a mathematical object has yet to be precisely defined, but the work we have completed in Chapter 5 and the remarks made in Chapter 6 indicate a possible path for constructing

such a phase space and geodesic flow.

In addition to determining the nature of such a geodesic flow and whether it could be dynamically equivalent to the billiard flow, we ask questions regarding the *ergodic* nature of the conjectured geodesic and billiard flows on the hypothesized 'fractal flat surface' and the corresponding fractal billiard, respectively.

In the long-term, we hope that the present (preliminary) study of the Koch snowflake billiard will help lay the foundations for a general theory of fractal billiards. Other work in this direction has already begun in the form of a joint paper with Joe P. Chen in which we are attempting to understand the nature of periodic orbits of self-similar and non-self-similar Sierpinski carpets. An immediately relevant project consists of investigating the relationship between the Veech group $\Gamma(KS_n)$ and $\Gamma(KS_{n+1})$ and whether or not one can determine $\varprojlim \Gamma(KS_n)$.

Chapter 2

Background

2.1 Basic Measure theory

A set is a collection of elements satisfying particular properties. We know that additional structure can be placed on a set in the form of what is called a topology. Such structure determined the definition of *open* and *closed*. In this section, we will place additional structure on a set and describe a general notion of *volume*.

2.1.1 Measures from outer measures

Throughout single variable calculus, many students shed the idea that every function is differentiable and every function is integrable. The first example of this is given by the function

$$\chi_A(x) = \begin{cases} 0 & \text{if } x \in \mathbb{Q} \cap A \\ 1 & \text{if } x \in \mathbb{Q}^c \cap A. \end{cases} \tag{2.1}$$

One can show that such a function is not Riemann integrable. However, a generalization of the Riemann integral called the *Lebesgue integral* is defined in such a way so as to ignore sets like the rational numbers and examine the elements of a set $A \subseteq \mathbb{R}$ that make

the bulk of the contribution to the total volume of A.

It is not until a student reaches upper-division mathematics or even graduate-level mathematics, that he or she learns that there are different notions of "differentiable" and "integrable." This same fact is encountered at the same time he or she sees behind the curtain and learns that an interval is a very nice example of a set. A basic fact taught to upper-division students is that the rationals are dense in the real line. A more surprising fact is that no interval in the real line contains only rational numbers. We will show later that this implies that the rational numbers have length zero, and explain exactly what we mean by length.

We now describe the notion of the Lebesgue outer measure defined on $\mathbb{R}$, this being the generalization of length that we have alluded to above. Consider a closed, connected interval $A \subset \mathbb{R}$. We know that the length of $A = [a, b]$ is $b - a$. What if we want to describe the length of two sets A and B, both subsets of $\mathbb{R}$? In general, two such sets may intersect. So, how does one logically describe the length of $A \cup B$? What if one considers a different collection of subsets of the real line in which not every element is an interval? We will show that the Lebesgue measure is a generalization of the notion of *length* that provides a framework in which to answer all of these questions.

Let A be a set. For the sake of simplicity, let us suppose $A \subset \mathbb{R}$. We first consider what is called a *cover* of A. Consider a collection of intervals $U := \{U_j\}_{j=1}^{\infty}$ where $A \subset \bigcup_{j=1}^{\infty} U_j$. Then U is a cover of A. Now, let us suppose we wanted to then shrink the cover U so that it constricts and tightens around the set A. Doing so results in a new cover of A. One can continue this process until it is impossible to constrict any further. In the limit, we have determined what is called the *outer measure* of the set A. It is called *outer*, because one starts with covers that encapsulate A and then shrink down to fit exactly to A. It is called *outer measure*, because the word measure is a more

general notion of length, and we are using the lengths of the intervals U_j to determine the overall *measure* of the set A. We now give the formal definition of *Lebesgue outer measure*.

Definition 1 (Lebesgue outer measure)**.** The Lebesgue outer measure of a subset $A \subseteq \mathbb{R}$ is given as

$$\lambda^*(A) := \inf \left\{ \sum_{j=1}^{\infty} |U_j| \, : \, A \subseteq \bigcup_{j=1}^{\infty} U_j, \text{ where } U_j \text{ are bounded intervals in } \mathbb{R} \right\}.$$

Definition 2 (Set function)**.** Let X be a set and $\mathscr{B}$ a collection of subsets of X. Then $\mu : \mathscr{B} \to \mathbb{R}^*$ is a (extended, real-valued) set function defined on $\mathscr{B}$. Typically, $\mathscr{B}$ is taken to be the powerset $\mathscr{P}(X)$ or a σ-algebra determined from X.

Definition 3 (Outer measure)**.** Let $\mu^* : \mathscr{P}(X) \to \mathbb{R}^*$ be a set function defined on the powerset of X. Then μ^* is an *outer measure* if it satisfies the following properties.

1. $\mu^*(\emptyset) = 0$.
2. $\mu^*(U) \leq \mu^*(V)$ for every $U, V \in \mathscr{P}(X)$ and $U \subseteq V$.
3. $\mu^*(\cup_{j=0}^{\infty} U_j) \leq \sum_{j=0}^{\infty} \mu^*(U_j)$.

Definition 4 (An algebra)**.** Consider a set X. An algebra A of subsets of X is a collection of subsets of X with the following properties:

1. $\emptyset \in A$, $X \in A$.
2. A is closed under finite unions, intersections and complements.

Definition 5 (σ-algebra)**.** Let X be a set. A σ-Algebra $\mathcal{A}$ of subsets of X is a collection of subsets of a set X with the following properties:

1. $\emptyset \in \mathcal{A}$, $X \in \mathcal{A}$.

2. $\mathcal{A}$ is closed under *countable* unions, intersections and complements.

Definition 6 (Additive and countably additive set function)**.** Let X be a set and μ a set function defined on a collection of subsets $\mathscr{B}$ of X containing at least the empty set. If $\mu(\emptyset) = 0$, then we say μ is additive. If $\mu(\cup_{j=1}^{\infty} B_j) = \sum_{j=1}^{\infty} \mu(B_j)$ whenever $\{B_j\}_{j=1}^{\infty}$ is a collection of pairwise disjoint sets in $\mathscr{B}$, then we say μ is countably additive. (Note, some authors may refer to countably additive as *completely additive*. We do not do so in this thesis.)

Definition 7 (Measure)**.** A measure μ is a nonnegative, completely additive set function defined on a σ-algebra $\mathcal{A}$.

In practice, one constructs a measure from an outer measure by examining a particular collection of subsets of a set X.

Definition 8 (Measurable set)**.** Let μ^* be an outer measure defined on $\mathbb{R}$. Let $E \subseteq \mathbb{R}$. We say that E is measurable if, for any subset $A \subseteq \mathbb{R}$, we have that

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c). \tag{2.2}$$

Then, the collection of subsets satisfying Equation (2.2) in Definition 8 is a σ-algebra $\mathcal{A}$. Moreover, μ^* restricted to $\mathcal{A}$ is a set function that satisfies the criteria for being a measure. Hence, μ^* can be made into a measure when restricting to the proper σ-algebra.

Definition 9 (Lebesgue Measure)**.** Consider the Lebesgue outer measure λ^*. Then λ^* restricted to the σ-algebra of measurable sets is called the *Lebesgue measure* on $\mathbb{R}$.

Definition 10 (Measure space)**.** Let X be a set and $\mathscr{B}$ a σ-algebra on X. If $\mu : X \to \mathbb{R}$ is a measure, then a measure space is a triple $(X, \mathscr{B}, \mu)$. When there is no ambiguity in what σ-algebra we are considering or what measure μ we are specifying, we will refer to the measure space $(X, \mathscr{B}, \mu)$ simply as X.

Definition 11 (Measurable function)**.** Let $(X, \mathscr{B}, \mu)$ be a measure space. A function $f : X \to X$ is measurable if for every $B \in \mathscr{B}$, $f^{-1}(B) \in \mathscr{B}$. That is, if the preimage of a measurable set is measurable, then f is called a measurable function.

Definition 12 (Measure-preserving transformations)**.** Let $(X, \mathscr{B}, \mu)$ be a measure space. Let $\sigma : X \to X$ is a measurable transformation. Then f is a *measure-preserving* transformation if for every $B \in \mathscr{B}$, $\mu(f^{-1}B) = \mu(B)$. That is, the measure of the preimage of a measurable set is the measure of that set.

Remark 13. When f is a measure-preserving transformation of a measure space X and is also 1-1, we say f is invertible.

Example 14. Let z be a complex number such that $|z| = 1$. The transformation $f : S^1 \to S^1$ given by $f(z) = z^2$ is not measure-preserving. This follows from the fact that the arc length of an interval of S^1 doubles with each iteration of the transformation. Note, we were implicitly assuming the measure was the arc length measure.

Example 15. The transformation f given by $f(x) = x + 1$ is a measure-preserving transformation of the real line when we take our measure to be the Lebesgue measure on $\mathbb{R}$.

2.1.2 Integration

We now proceed to construct the Lebesgue integral. In order to do so we need to first consider a few preliminary ideas.

Definition 16 (Characteristic function)**.** Let E be a measurable set. Then

$$\chi_E(x) := \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases} \tag{2.3}$$

is called the *characteristic function* of E. Note, some authors may refer to such a function as the *indicator function* and use the notation $\mathbb{I}_E$ or I_E. We will not use this convention

in this thesis.

Definition 17 (Simple function)**.** A function f is called simple if f is expressed as the finite linear combination of characteristic functions. That is, if $E_1, E_2, ..., E_k$ are measurable sets and

$$f(x) := \sum_{j=1}^{k} a_i \chi_E(x), \tag{2.4}$$

where a_j are real numbers, then f is called a *simple function.*

When attempting to integrate a simple function f with respect to the Lebesgue measure λ, we suppose f is expressed as a linear combination of characteristic function $\chi_{E_j}(x)$ where $\{E_j\}_{j=1}^{k}$ is a pairwise disjoint collection of closed sets, perhaps only overlapping at the endpoints.

Definition 18 (The integral of a simple function)**.** If f is a simple function, then we define the integral of f to be

$$\int f d\lambda := \sum_{j=1}^{k} a_j \lambda(E_j). \tag{2.5}$$

This is certainly inline with our intuition. Moreover, when E_j is an interval, this is nothing more than the sum of the area of rectangles.

We now proceeed to construct the integral for more general functions.

Definition 19 (The integral of a nonnegative function)**.** Let f be a nonnegative, measurable function. It can be show that f is the pointwise limit of an increasing sequence of simple functions f_j. Define the integral of f to be

$$\int f d\lambda := \lim_{j \to \infty} \int f_j d\lambda. \tag{2.6}$$

When f is an arbitrary measurable function, we can decompose f into its positive, negative and zero parts. Since there is no need to write the zero portion, we can write $f = f^+ - f^-$.

Definition 20. Let f be a measurable function. If f^+ and f^- are integrable (that is, the integrals of each are finite), then we write the integral of f as

$$\int f d\lambda := \int f^+ d\lambda - \int f^- d\lambda. \tag{2.7}$$

Theorem 21 (Lebesgue Dominated Convergence)**.** Let $\{f_j\}_{j=1}^{\infty}$ be a collection of measurable functions defined on a set E of finite, positive measure. Suppose that there is a constant $M > 0$ such that $|f_j(x)| \leq M$ for all $j \geq 1$. If $\lim_{j=1}^{\infty} f_j(x)$ exists for almost every x, then

$$\lim_{j\to\infty} \int f_j d\lambda = \int \lim_{j=1}^{\infty} f_j d\lambda. \tag{2.8}$$

Theorem 22 (Lebesgue Monotone Convergence Theorem)**.** Let $\{f_j\}_{j=1}^{\infty}$ be a sequence of measurable functions such that $0 \leq f_1 \leq f_2 \leq ....$ Then

$$\lim_{j\to\infty} \int f_j d\lambda = \int \lim_{j\to\infty} f_j d\lambda. \tag{2.9}$$

Theorem 23 (Fatou's Lemma)**.** Let $\{f_j\}_{j=1}^{\infty}$ be a collection of nonnegative, measurable functions on $\mathbb{R}$. Then

$$\int \liminf_{j\to\infty} f_j d\lambda \leq \liminf_{j\to\infty} \int f_j d\lambda. \tag{2.10}$$

2.2 Basic group theory

Definition 24 (Equivalence relation)**.** Let X be a nonempty set. If R is a relation on X, then R is an equivalence relation if the following are satisfied.

1. (Reflexive) $x \in X$ implies xRx. That is, every element $x \in X$ is related to itself.
2. (Symmetric) xRy implies yRx. That is, x related to an element y implies y is related to x.

3. (Transitive) xRy and yRz, then xRz. That is, if x is related to y and y is related to z, then it must be the case that x is related to z.

Remark 25. When R is an equivalence relation on a set X, we often denote it by $\sim$.

Definition 26 (A group)**.** A group G is a collection of elements equipped with an operation denoted by $+$ such that

1. (closure) for every $g, h \in G$, $g + h \in G$.
2. (associativity) for every $g, h, k \in G$, $(g + h) + k = g + (h + k)$.
3. (identity) there exists $e \in G$ such that for every $g \in G$, $g + e = e + g = g$.
4. (inverse) for every $g \in G$ there exists $h \in G$ such that $g + h = h + g = e$.

Example 27 (The integers)**.** The integers constitute a group under the usual notion of addition. The (additive) identity e is indeed 0. The fact that $\mathbb{Z}$ contains negative numbers indicates that $\mathbb{Z}$ contains an additive inverse $-z$ for each z.

Example 28 (The integers mod n)**.** Let n be a positive integer. Consider the equivalence relation $\sim_n$ on $\mathbb{Z}$ given by $x \sim_n y$ if and only if $x = y \mod n$. Then, $\mathbb{Z}_n := \{[0], [1], ...[n-1]\}$ is referred to as the *integers mod n* where the operation $+$ is defined as $[a] + [b] = [a + b]$.

Example 29 (General linear group)**.** Consider the space of $n \times n$ matrices with real entries with nonzero determinant (it is not necessary for the entries to be real. They could be complex numbers or quarternions). This collection constitutes what is called the *general linear group* with real entries and is denoted by $GL_n(\mathbb{R})$.

Example 30 (Special linear group)**.** The elements $A \in GL_n(\mathbb{R})$ that have determinant $\det A = 1$ constitute a subgroup of $GL_n(\mathbb{R})$ called the *special linear group* and

is denoted by $SL_n(\mathbb{R})$. Elements of $SL_n(\mathbb{R})$ preserve the volume and orientation of n-dimension subsets of $\mathbb{R}^n$.

Example 31 (Orthogonal group)**.** The elements of $A \in GL_n(\mathbb{R})$ that preserve the length of vectors in $\mathbb{R}^n$ constitute a subgroup of $GL_n(\mathbb{R})$ called the *orthogonal group*.

Example 32 (Special orthogonal group)**.** The elements $A \in GL_n(\mathbb{R})$ that have determinant $\det A = 1$ and preserve the length of vectors in $\mathbb{R}^n$ constitute a subgroup of $GL_n(\mathbb{R})$ called the *special orthogonal group* and it is denoted as $SO_n(\mathbb{R})$.

Example 33 (Dihedral group)**.** Let P_n be a regular polygon with n sides. Then the collection of symmetries generated by reflection and rotation of P_n about the origin constitute a group. This group has $2n$ many elements.

Remark 34. We will later see in §2.4, §2.5.1 and Chapter 3 that the dihedral group plays a major role in our analysis of billiard trajectories. We will be using a heuristic device[1] referred to as *unfolding the billiard orbit* so as to determine an equivalent geodesic line on an associated flat surface (see Definition 68 and 69). For further discussion of how one unfolds an orbit, see §2.4.1, [HuSc] and [MasTa].

If G is a group and X is a set, then the map given by $\varphi : G \times X \to X$ is called a (left) group action on the set X.

Definition 35 (An orbit Gx)**.** Let X be a set and G a group acting on X. Consider $Gx := \{gx : g \in G\}$. This set is called the orbit of an element $x \in X$.

An orbit constitutes an equivalence class of elements. That is, the condition relating elements of an orbit constitutes and equivalence relation on X.

Definition 36 (Transitive group action)**.** Let X be a set and G a group acting on X.

[1]Such a device is actually quite rigorous and is an example of what is called a *developing map*; see [Thur] for a discussion of such an abstract concept. We will not need the full power of such a concept and refer the reader for any and all explanation of developing maps to [Thur]

If for every $x, y \in X$ there exists $g \in G$ such that $x = gy$, then we say G acts *transitively* on X or the group action $\sigma : G \times X \to X$ is a *transitive group action.*

Proposition 37. Let $\sigma : G \times X \to X$ be a transitive group action. Then, for any $x \in X$, the orbit Gx is exactly X.

2.3 Basic dynamics and ergodic theory

2.3.1 Definitions and terminology in dynamical systems

We initially assume the spaces we are considering are only topological spaces. As such, this section is primarily concerned with defining concepts in *topological dynamics.* In a way, the subject of ergodic theory can be thought of as *measure theoretic dynamics*, since the primary assumption is that one is dealing with a probability measure space (that is, a space X with $\mu(X) = 1$).

Definition 38 (Continuous dynamical system)**.** Consider a topological space X, a rule φ describing the evolution of the elements of X and t a real variable. Then (X, φ, t) is called a continuous dynamical system. The flow is then denoted by φ_t.

In the sections and chapters that follow, we will be primarily concerned with the situation in which the time-variable is a discrete variable.

Definition 39 (Discrete dynamical system)**.** Let X be a topological space, $f : X \to X$ a map describing the evolution of the system with respect to the discrete time variable n. Then (X, f, n) is called a *discrete dynamical system.*

There is a relationship between a continuous dynamical system and a particular discrete dynamical system. If the flow φ_t in a space $M \subset \mathbb{R}^m$ is continuous, then there is a way to reduce this continuous flow to a discrete flow f^n that still encodes enough information to discern certain properties of the flow φ_t on M. This is what is referred to

as the Poincaré section (or, simply the section). The discrete flow f^n then determines a discrete set, when φ_t happened to be closed.

Definition 40 (Poincaré section)**.** Let $M \subseteq \mathbb{R}^m$ be an m-dimensional subspace of $\mathbb{R}^m$. Let φ_t be a flow on M. Then a Poincaré section S of M is an $(m-1)$-dimensional subspace of M that is transversal to the flow.

Definition 41 (Return time)**.** If S is a Poincaré section of $M \subseteq \mathbb{R}^m$ and φ_t is a flow on M and $x \in S$, then the time t_0 for $\varphi_t(x)$ to return to the Poincaré section S of the flow is referred to as the *return time*.

Definition 42 (Poincaré map or first return map)**.** Let (X, φ, t) be a continuous dynamical system and S a section of X. If the return time t_0 of the flow $\varphi_t(x)$ is finite, then there is a map f such that $f(x) = x' \in S$. Such a map is called the *Poincaré map* or the *first return map*.

Definition 43 (Periodic orbit)**.** Consider a flow $\varphi_t(x)$. If there exists some time $t_0 > 0$ such that $\varphi_{t_0}(x) = x$, then we say the orbit $\{\varphi_t(x)\}_{t\in\mathbb{R}}$ is a *periodic orbit*. The least positive value $t_0 > 0$ for which $\varphi_{t_0}(x) = x$ is called the *period* of the orbit. If f is a Poincaré map, then there exists a least positive integer n such that $f^n(x) = x$. This is then the period of the orbit given by $\{f^j(x)\}_{j=0}^{\infty}$. We note that the notation f^j is to indicate j compositions of f with itself.

Definition 44 (Fixed point of a map)**.** Let (X, f, n) be a discrete dynamical system. A fixed point for the map f is a point x such that $f(x) = x$.

Example 45. If $\{f^j(x)\}_{j=0}^{\infty}$ is a periodic orbit with period n, then the initial condition x for the orbit is a fixed point of the map f^n.

Definition 46 (Transitive orbit)**.** Let (X, f, n) be a discrete dynamical system where f is a homeomorphism. Then the orbit $\{f^j(x)\}_{j\geq 0}$ is a transitive orbit if it is dense in X.

Remark 47. The notion of topological transitivity is heuristically linked with the notion of transitive orbit found in the subject of group theory (see Definition 36). By Proposition 37, the orbit Gx of a point $x \in X$, where G is a group, has an orbit that was all of X. Here, the orbit of the point is "nearly" all of X, where "nearly" means *dense*. hence, this is why the notion of transitivity defined in Definition 46 is of a topological nature.

2.3.2 Definitions and terminology in ergodic theory

The word 'ergodic' is an amalgamation of two Greek words ergon (work) and odos (path). Boltzmann was the first to coin the term 'ergodic' as it related to his ergodic hypothesis, which stated that each surface of constant energy consists of a single trajectory. Unfortunately, as discussed by Mark Policott and Yuri Michinko [PoYu], Boltzmann's Greek language skills may have not been as sharp as one initially believed. The more appropriate mashing together of two Greek words would be something along the lines of ergoidic or erchodic, all relating to 'work along a path'. The term ergodic is defined as follows.

Definition 48 (Ergodic measure)**.** Let (X, β, μ) be a probability measure space and T a measure-preserving transformation defined on X. Then μ is an *ergodic measure* if for every measurable $A \subset X$ such that $\mu(T^{-1}(A) \cap A) = 0$, we have that $\mu(A) = 0$ or $\mu(A) = 1$.

Remark 49. If μ is an ergodic measure, then we can equivalently refer to the measure-preserving transformation T as an *ergodic transformation*. Formally, the discrete dynamical system (X, β, μ, T, n) is referred to as an ergodic dynamical system when μ is an ergodic measure.

Definition 50 (Uniquely ergodic)**.** Let (X, β, μ, T) be an ergodic dynamical system.

If for any other ergodic dynamical system (X, β, ν, T) we have that $\nu(A) = \mu(A)$ for every $A \in \beta$, then (X, β, μ, T) is called a uniquely ergodic dynamical system. Equivalently, T is a uniquely ergodic measure-preserving transformation and μ is the unique ergodic measure.

An ergodic average is, in a way, an average of some dynamical phenomenon modeled by a measure-preserving transformation T.

Definition 51 (Ergodic average)**.** Let $(X, \mathscr{B}, \mu)$ be a measure space. Let $T : X \to X$ be a measure-preserving transformation and $f : X \to \mathbb{R}$ be a continuous function. Then

$$s_n(x) := \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$$

is called an *ergodic average*.

Before we state the Birkhoff Ergodic Theorem or the Maximal Ergodic Theorem, we give the Poincaré Recurrence Theorem. While the notion of an ergodic average was not known at the time of Poincaré (especially when Poincaré proved the theorem in 1899), such a result historically marks the genesis of the subject. In order to properly state the Poincaré Recurrence theorem, we first present a number of definitions and lemmas.

Definition 52 (Recurrent)**.** Let $(X, \mathscr{B}, \mu)$ be a measure space. If $T : X \to X$ is a measure-preserving transformation, then we say T is *recurrent* if for every measurable set $A \in \mathscr{B}$ with $0 < \mu(A) < \infty$, there exists a null set $N \subset A$ such that for all $x \in A \setminus N$, there exists $n > 0$ such that $T^n(x) \in A$.

We would like to point out that nothing about the definition of recurrent requires that $T^n(x)$ *not* be in N. We will see that there is a way of guaranteeing that $T^n(x)$ not be in *some* null set N. The definition of recurrent says in a rigorous fashion that a point should, more often than not, come back to the set it started in. This is

intuitively clear if X is a finite measure space. We will see that this is the content of the Poincaré Recurrence Theorem.

Definition 53 (Conservative)**.** Let $(X, \mathscr{B}, \mu)$ be a measure space. A measure preserving transformation is said to be *conservative* if for any set A of positive measure, there exists an integer $n > 0$ such that $\mu(A \cap T^{-n}(A)) > 0$.

Remark 54. $T^n(x) \in N$ or $T^n(x) \notin N$ does not make a difference when discussing the iterates of T. That is, N does not have to (and, in general, will not) be the same null set each time the notion of recurrent or conservative is being applied.

Lemma 55. Let $(X, \mathscr{B}, \mu)$ be a measure space and $T : X \to X$ a recurrent, measure-preserving transformation. By definition, this implies that there exists a null set N_1 such that for every $x \in A \setminus N_1$ there exists $n > 0$ such that $T^n(x) \in A$. Define $N := \bigcup_{k=0}^{\infty} T^{-k}(N_1)$. Then, for every $x \in A \setminus N$ we have that $T^n(x) \in A \setminus N$.

Proof. Let $x \in A \setminus N$. Then $x \notin N$. Therefore, $T^n(x) \notin N$. Otherwise, $T^n(x) \in N$ implies that $T^n(x) \in T^{-k}(N_1)$ for some $k > 0$. Hence, $x \in T^{-k-n}N_1$, which is impossible. ■

Lemma 56. Let $(X, \mathscr{B}, \mu)$ be a measure space and $T : X \to X$ a recurrent measure-preserving transformation. Then for all sets A of positive measure, there exists a null set N such that for all $x \in A \setminus N$, there is an increasing sequence $m_i > 0$ with $T^{m_i}(x) \in A \setminus N$ for all $i \geq 1$.

Proof. Let n and N be as they were in Lemma 55. Letting $z = T^n(x)$, there exists $n_1 = n(z) > 0$ such that $T^{n_1}(z) \in A \setminus N$. Let $m_2 = n + n_1$. Then

$$\begin{aligned} T^{m_2}(x) &= T^{n+n_1}(x) \\ &= T^{n_1}(z), \end{aligned} \tag{2.11}$$

and $T^{n_1}(z) \notin N$. Therefore, $T^{m_2}(x) \notin N$. Let $m_1 := n$. Continuing in this fashion, we

can construct a monotonically increasing sequence $\{m_i\}_{i=1}^{\infty}$ such that $T^{m_i}(x) \in A \setminus N$ for every $i \geq 1$. ■

Lemma 57 (Characterization of recurrence)**.** Let $(X, \mathscr{B}, \mu)$ be a measure space. A measure-preserving transformation T is recurrent if and only if T is conservative.

Theorem 58 (Poincaré Recurrence Theorem)**.** Let $(X, \mathscr{B}, \mu)$ be a finite measure space. If $T : X \to X$ is a measure-preserving transformation, then f is recurrent.

Proof. By Lemma 57, it suffices to show that for any set A of positive measure, there is an intenger $n > 0$ such that $\mu(A \cap T^{-n}(A)) > 0$. Suppose, by way of contradiction, that $\mu(A \cap T^{-n}(A)) = 0$ for every $n > 0$. Then, for every k and $l := n + k$, we have that

$$\begin{aligned} \mu(T^{-l}(A) \cap T^{-k}(A)) &= \mu(T^{-n-k}(A) \cap T^{-k}(A)) \\ &= \mu(T^{-k}(T^{-n}(A) \cap A)) \\ &= \mu(T^{-n}(A) \cap A) \\ &= 0. \end{aligned} \tag{2.12}$$

Hence, $\{T^{-n}(A)\}_{n=0}^{\infty}$ is an almost everywhere pairwise disjoint collection of sets. However, $\mu(T^{-n}(A)) = \mu(A)$ for every $n \geq 0$. So,

$$\begin{aligned} \mu(X) &\geq \mu(\bigcup_{n=0}^{\infty} T^{-n}(A)) \\ &= \sum_{n=0}^{\infty} \mu(T^{-n}(A)) \\ &= \sum_{n=0}^{\infty} \mu(A) = \infty. \end{aligned} \tag{2.13}$$

■

Generally speaking, ergodic theorems are concerned with relating a time-average to a volume average. In the context of the current setting, time is discrete and the notion of volume is dictated by a probability measure μ.

The Maximal Ergodic Theorem can be used to prove the Birkhoff Ergodic Theorem. Recall that the ergodic average $s_n(x)$ was defined in Definition 51.

Theorem 59 (Maximal Ergodic Theorem)**.** Let T be a measure-preserving transformation of the measure space $(X, \mathscr{B}, \mu)$ and let f be integrable. Let $A = \{x : \sup_{n\geq 0} s_n(x) > 0\}$. Then $\int_A f(x)dm \geq 0$.

Sketch of the proof (see [Si] for the complete proof). Let $p \geq 1$ and define $A_p := \{x \in X | \sup_{n\geq 0} s_n(x) > 0 \text{ for some } n,\ 1 \leq n \leq p\}$. It can be shown that

$$n \int f^+ d\mu \leq n \int_{A_p} f d\mu + \sum_{i=n-p}^{n-1} \int |f| d\mu. \tag{2.14}$$

Then, dividing by n and letting $n \to \infty$,

$$0 \leq \int_{A_p} f d\mu. \tag{2.15}$$

We observe that $A = \bigcup_{p=0}^{\infty} A_p$. By the Dominated Convergence Theorem, it follows that $\int_A f d\mu \geq 0$. ■

Theorem 60 (Birkhoff Ergodic Theorem)**.** Let T be a measure preserving transformation of the measure space $(X, \mathscr{B}, \mu)$ and let f be an integrable function. Then the ergodic averages $s_n(x)$ converge for almost every x to a limit function $f^*(x)$ which is again an integrable function. The function f^* is constant on orbits, i.e., $f^*(T^k x) = f^*(x)$ a.e. In the case $\mu(X) < \infty$, we also have $\int_X f dm = \int_X f^* d\mu$.

Sketch of the proof. We consider two cases: when T is an ergodic transformation and when T is not necessarily an ergodic transformation. The first case will not require the Maximal Ergodic Theorem, but the second case will.

We then let T be an ergodic, measure-preserving transformation of the probability measure space X. We will show that for every integrable function f,

$$\limsup_{n\to\infty} s_n(x) \leq \int f d\mu \leq \liminf_{n\to\infty} s_n(x) \tag{2.16}$$

to establish the result. To such end, we define $A := \{x : \liminf_{n\to\infty} s_n(x) < \int f d\mu\}$. If $\mu(A) = 0$, then we have established one of the inequalities in Equation (2.16). We can write A as follows.

$$A := \bigcup_{r\in\mathbb{Q}} \{x : \liminf_{n\to\infty} s_n(x) < r < \int f d\mu\}. \tag{2.17}$$

Then, define $C_r := \{x : \liminf_{n\to\infty} s_n(x) < r < \int f d\mu\}$. Suppose $\mu(A) > 0$. Then there exists $r \in \mathbb{Q}$ such that $\mu(C_R) > 0$. We claim that C_r is T-invariant $\mod \mu$. Since T is an ergodic transformation, $\mu(C_r) = 1$.

Consider the set $E_p^r := \{x : s_n(x) \geq r, \text{ for all } 1 \leq n \leq p\}$. The fact that $\mu(C_r) = 1$, we have that $\mu(\bigcap_{p=1}^{\infty} E_p^r) = 0$. Since $\{E_p^r\}_{p=1}^{\infty}$ is a decreasing, nested sequence of measurable sets, we have that $\lim_{p\to\infty}$ contradiction of our choice of r. Therefore, $\mu(A) = 0$. Replacing f by $-f$, the left-hand inequality in Equation (2.16)

We refer the reader to [Si, Chapter 5] for the proof of the Birkhoff Ergodic Theorem in the case when μ is not an ergodic measure. We only point out here that the Maximal Ergodic Theorem, the Dominated Convergence Theorem and a few key lemmas are used to prove the theorem when μ is not an ergodic measure. ■

We end this discussion of ergodic theory with an important application: uniformly distributed sequences.

Definition 61 (Uniform distribution)**.** Let A be a closed interval in $\mathbb{R}$ and λ the Lebesgue measure on $\mathbb{R}$. A sequence of points $\{a_i\}_{i=0}^{\infty}$ in A is said to be uniformly distributed in A if for every subinterval $I \subset A$ we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \#\{i : 0 \leq i < n,\ a_i \in I\} = \lambda(I), \tag{2.18}$$

where # denotes the cardinality of a set.

Remark 62. The left-hand side of Equation (2.18) can be written as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_I(a_i), \tag{2.19}$$

where χ_I is the characteristic function of I; see Definition 16.

Proposition 63. Let A be a closed interval in $\mathbb{R}$. If $(A, \beta, \lambda, T, n)$ is an ergodic dynamical system (i.e., T is an ergodic transformation), where λ is the Lebesgue measure, then, for almost every $x \in A$, $\{T^i(x)\}_{i=0}^{\infty}$ is a uniformly distributed sequence in A.

Proof. Let $I \subset A$. Since χ_I is a measurable function and T is an ergodic transformation, by Theorem 60, it follows that the ergodic averages converge to the integral of the characteristic function defined on I. Since we know that such an integral is in fact the measure of the set I, it follows that the ergodic averages converge to $\lambda(I)$. Since I was arbitrary, the sequence of points $\{T^i(x)\}_{i=0}^{\infty}$ is uniformly distributed in A for almost every $x \in A$. ■

2.4 Mathematical Billiards

Under ideal conditions, we know that a point mass making a perfectly elastic collision with a C^1 surface (or curve) will reflect at an angle which is equal to the incoming angle, both measured relative to the normal at the point of collision. That is, the angle of reflection is equal to the angle of incidence. Such a law is referred to as the *Law of Reflection.*

Consider a compact region $\Omega(B)$ in the plane with connected boundary B. Then, $\Omega(B)$ is called a *planar billiard* when B is smooth enough to allow the Law of Reflection to hold, off of a set of measure zero (where the measure is taken to be the Hausdorff measure or the arc length measure). Though the Law of Reflection implicitly states that the angles of incidence and reflection be determined with respect to the

normal to the line tangent at the basepoint, we adhere to the equivalent convention in the field of mathematical billiards that the vector describing the position and velocity of the billiard ball (which amounts to the position and angle, since we are assuming unit speed) be reflected in the tangent to the point of incidence. That is, employing such a law so as to determine the path on which the billiard ball departs after impact essentially amounts to identifying certain vectors.

Then we can rigorously reformulate the Law of Reflection as follows: the vector describing the motion is the reflection of the incoming vector through the tangent at the point of collision. One may express the Law of Reflection in terms of equivalence classes of vectors. Hence, identifying these two vectors to form an equivalence class of vectors in the unit tangent bundle corresponding to the billiard table $\Omega(B)$ (see Figure 2.1). (See [Sm] for a detailed discussion of this equivalence relation on the unit tangent bundle $\Omega(B) \times S^1$.) When B is a nontrivial, connected polygon in $\mathbb{R}^2$, $\Omega(B)$ is called a *polygonal billiard*. The collection of vertices of $\Omega(B)$ forms a set of zero measure (when we take our measure to be the Hausdorff measure or simply, the arc-length measure on B), since there are finitely many vertices. A *rational billiard* is defined below.

For the remainder of this thesis, we will almost exclusively be dealing with either rational billiard tables or billiard tables with self-similar fractal boundary. This is to imply that we do not at all discuss the current research on irrational billiard tables and their associated flat surfaces and extensions to Teichmüller theory. For a detailed discussion of such topics, as well as a source for additional references discussing flows on moduli spaces, see [Hoop].

To clearly understand how one forms equivalence classes from elements of $B \times S^1$, we let $(x, \theta), (y, \gamma) \in B \times S^1$ and say that $(x, \theta) \sim (y, \gamma)$ if and only if $x = y$ and one of the following is true:

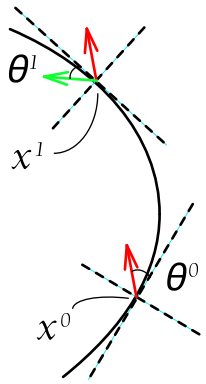

Figure 2.1: A billiard ball traverses the interior of a billiard and collides with the boundary. The velocity vector is pointed outward at the point of collision. The presence of a well-defined tangent at this point provides for the existence of a normal to the tangent and the recovery of the Law of Reflection (i.e., Snell's Law). The resulting direction of flow is found by either reflecting the vector through the tangent or by reflecting the incidence vector through the normal and reversing the direction of the vector. We use the former method throughout this paper. A rigorous discussion of the Law of Reflection in this context is given in [Sm].

1. $x = y$ is not a vertex of the boundary B and $\theta = \gamma$,

2. $x = y$ is not a vertex of the boundary B, but $x = y$ is a point on a segment s_i of the polygon B and $\theta = r_i(\gamma)$, where r_i denotes reflection in the segment s_i,

3. If $x = y$ is a vertex of B, then we identify (x, θ) with $(y, g(\gamma))$ for every g in the group generated by reflections in the two adjacent sides having x (or y) as a common vertex.

Definition 64 (Rational polygon and rational billiard)**.** If B is a nontrivial connected polygon such that for each interior angle θ_j of B there are relatively prime integers $p_j > 0$ and $q_j > 0$ such that $\theta_j = \frac{p_j}{q_j}\pi$, then we call B a *rational polygon* and $\Omega(B)$ a *rational billiard*.

Denote by S^1 the unit circle, which we let represent all the possible directions (or angles) in which a billiard ball may initially move. The phase space for the billiard

dynamics is $(\Omega(B)\times S^1)/\sim$. In practice, one restricts his or her attention to $(B\times S^1)/\sim$ when discussing the phase space of the billiard $\Omega(B)$.

Let x^0 be an initial basepoint on the boundary B of a billiard table $\Omega(B)$ and θ^0 an initial direction. The billiard map is determined from the continuous flow. Let $\varphi_t(x^0,\theta^0)$ be a flow line in the phase space $(\Omega(B)\times S^1)/\sim$. We consider a section of the phase space given by $(B\times S^1)/\sim$. The values t_j for which $\varphi_{t_j}(x^0,\theta^0)\in(B\times S^1)/\sim$ constitute the *return times* (i.e., times for which $\varphi_t(x^0,\theta^0)$ returns to the section, or intersects it in a non-tangential way). Then, the discrete map $f^{t_j}(x^0,\theta^0)$ constitutes the section map. In terms of the configuration space, $f^{t_j}(x^0,\theta^0)$ constitutes the point and angle of incidence in the boundary B. Since $\Omega(B)$ is the billiard and it is the collision points in which we are interested, it is only fitting that such a map be called the *billiard map*. More generally, f^{t_j} is denoted by f^j and is called the *Poincaré map* and the section is called the *Poincaré section*[2]. Furthermore, the obvious benefit of having a visual representation of $f^j(x^0,\theta^0)$ in the configuration space is exactly why one restricts their attention to the section $(B\times S^1)/\sim$. Specifically, all one really cares about in the end, from the perspective of one examining a planar billiard, are the collision points, which are clearly determined by the Poincaré map.

Remark 65. In the sequel, we will simply refer to an element $[(x^k,\theta^k)]$ by (x^k,θ^k), since the vector corresponding to θ^k is inward pointing at the basepoint x^k. So as not to introduce cumbersome notation, when we begin discussing the billiard map f_{KS_n} corresponding to the nth prefractal billiard $\Omega(KS_n)$, we will simply write f_{KS_n} as f_n. When discussing the discrete billiard flow on $(\Omega(KS_n)\times S^1)/\sim$, the kth point in an orbit $(x^k,\theta^k)\in(\Omega(KS_n)\times S^1)/\sim$ will instead be denoted by $(x_n^{k_n},\theta_n^{k_n})$, so as to be clear as to which space such a point belongs. Specifically, k_n refers to the number of iterates

[2]In the context of mathematical billiards, the billiard map is the Poincaré map.

of the billiard map f_n necessary to produce the pair $(x_n^{k_n}, \theta_n^{k_n})$. An initial condition of an orbit of $\Omega(KS_n)$ will always be referred to as (x_n^0, θ_n^0).

In the event that a basepoint x^j of $f_B^j(x^0, \theta^0)$ is a corner of $\Omega(B)$ (that is, a vertex of the polygonal boundary B) and θ^0 was a direction for which the billiard flow would be periodic, then the resulting closed orbit is said to be *singular*. In addition, since θ^0 is a direction for which the resulting orbit is periodic, there exists a positive integer k such that the basepoint x^{-k} of $f_B^{-k}(x^0, \theta^0)$ is a corner of $\Omega(B)$. (Here, f_B^{-k} denotes the kth inverse iterate of f_B.) The path then traced out by the billiard ball connecting x^j and x^{-k} is called a *saddle connection*.

Definition 66 (Footprint of an orbit)**.** Let $\mathscr{O}_n(x_n^0, \theta_n^0)$ be an orbit of $\Omega(KS_n)$. Then

$$\mathcal{F}_n(x_n^0, \theta_n^0) := \mathscr{O}_n(x_n^0, \theta_n^0) \cap KS_n \tag{2.20}$$

is called the footprint of the orbit $\Omega(KS_n)$.

Remark 67. Within the subject of mathematical billiards, there appears to be a slight abuse of language. One may refer to the orbit of a billiard as the path traced out by the billiard ball or as the collection of incidence points in the boundary. In the latter case, such a set of points is referred to as the images of iterates of the billiard map f_B. When we want to be clear as to which concept we are referring, we will specifically write 'the path corresponding to the orbit' or 'the footprint of the corresponding orbit,' respectively. When no ambiguity shall arise, we simply refer to the orbit as either the "path" or the "footprint" of the corresponding orbit.

2.4.1 Unfolding a billiard orbit

Consider a rational polygonal billiard $\Omega(B)$ and an orbit $\mathscr{O}(x^0, \theta^0)$. Reflecting the billiard $\Omega(B)$ and the orbit in a side of the billiard containing a basepoint of the

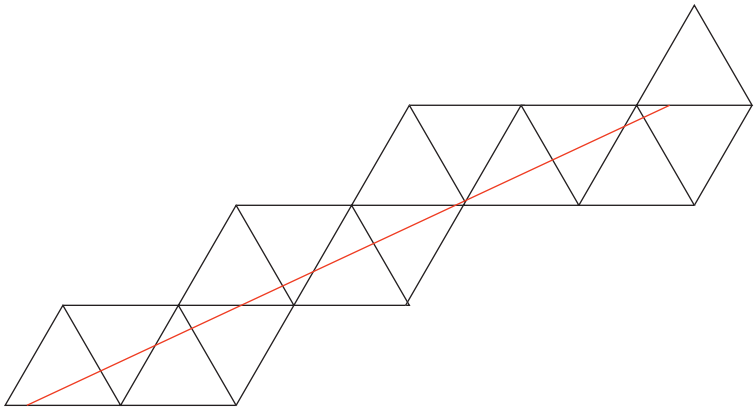

Figure 2.2: Unfolding an orbit of the equilateral triangle billiard $\Omega(KS_0)$.

orbit (or an element of the footprint of the orbit) partially unfolds the orbit $\mathscr{O}(x^0, \theta^0)$; see Figure 2.2. Reflection in the side constitutes an affine transformation of $\Omega(B)$. The linear portion of the affine transformation is an element of the group of symmetries of B. Continuing this process until the orbit is a straight line produces as many copies of the billiard table as there are elements of the footprint. That is, if the period of an orbit $\mathscr{O}(x^0, \theta^0)$ is some positive integer p, then the number of copies of the billiard table in the unfolding is also p. Therefore, we refer to such a straight line as the unfolding of the billiard orbit.

2.5 Flat surfaces and properties of the flow

2.5.1 A flat surface

In this section, we deal only with flat surfaces constructed from rational billiards; see Definition 64 in §2.4 above. Heuristically, a flat structure on a connected surface M is an atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in \mathscr{A}}$ with a finite number of singularities such that, away

from these singularities, each coordinate changing map

$$\varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(U_\alpha \cap U_\beta) \to \varphi_\alpha(U_\alpha \cap U_\beta) \tag{2.21}$$

is a translation in $\mathbb{R}^2$. Specifically, a flat structure satisfies the following definition.

Definition 68 (Flat structure)**.** Let M be a compact, connected, orientable surface. A *flat structure* on M is an atlas ω, consisting of charts of the form $(U_\alpha, \varphi_\alpha)_{\alpha \in \mathscr{A}}$, where U_α is a domain (i.e., a connected open set) in M and φ_α is a homeomorphism from U_α to a domain in $\mathbb{R}^2$, such that the following conditions hold:

1. the collection $\{U_\alpha\}_{\alpha \in \mathscr{A}}$ cover the whole surface M except for finitely many points $z_1, z_2, ..., z_k$, called *singular points*;

2. all coordinate changing functions are translations in $\mathbb{R}^2$;

3. the atlas ω is maximal with respect to properties (1) and (2);

4. for each singular point z_j, there is a positive integer m_j, a punctured neighborhood $\dot{U}_j$ not containing other singular points, and a map ψ_j from this neighborhood to a punctured neighborhood $\dot{V}_j$ of a point in $\mathbb{R}^2$ that is a shift in the local coordinates from ω, and is such that each point in $\dot{V}_j$ has exactly m_j preimages under ψ_j.

Definition 69 (Flat surface)**.** We say that a connected, compact surface equipped with a flat structure is a *flat surface*.

Calling a connected, compact (2-dimensional) manifold with flat structure a flat surface is somewhat of an abuse of language, but it enables us to be brief and refer to related notions with greater ease. Note that in the literature on billiards and dynamical systems, the terminology and definitions pertaining to this topic are not completely uniform; see, for example, [GaStVo, Gtk1, GtkJu1, GtkJu2, HaKa, HuSc, Mas, MasTa,

Ve1, Ve2, Ve3, Vo, Zo]. We have adopted the above definition for clarity and the reader's convenience.

The singular points of a flat surface $\mathcal{S}$ are called *singularities of the flow*. There are two types of singularities in a flat surface: *removable and nonremovable*. They are called such, because it may or may not be possible to define the flow at these points. We now turn to a discussion of why these singularities can be termed "removable" or "nonremovable" and how to discern between the two types. Consider z_j in the set of singularities of a flat structure. A neighborhood of this point is homeomorphic to a cone. Recall that the measure (in radians) of an angle is the length of the arc in the unit circle intersected by the two rays forming the angle. Consequently, a unit circle with a circumference of 4π is not a circle (with radius $r = 1$) living in the plane; see Figures 2.3 and 2.4 and Example 72. A *conic angle* is the number of radians required to form a closed circle about the cone point z_j.

We state the following definitions for completeness.

Definition 70 (Removable conic singularity)**.** A singularity of a flat structure given by $(U_\alpha, \varphi_\alpha)_{\alpha\in\mathscr{A}}$ is a *removable singularity* of the flow if the conic angle about such a point is 2π.

Definition 71 (Nonremovable conic singularity)**.** A singularity of a flat structure given by $(U_\alpha, \varphi_\alpha)_{\alpha\in\mathscr{A}}$ is a *nonremovable singularity* of the flow if the conic angle about such a point is $2\pi c$, for some integer $c > 1$.

With regards to the geodesic flow on a flat surface $\mathcal{S}$, removable singularities pose no problem for the flow and the flow lines through such points can continue unimpeded. However, when a singularity is nonremovable, the flow cannot continue unimpeded. Moreover, when z_j is a removable singularity of a flat structure ω, there exists a flat structure $\tilde{\omega}$ on $\mathcal{S}$ such that z_j is not a singular point of $\tilde{\omega}$. Technically speaking, when

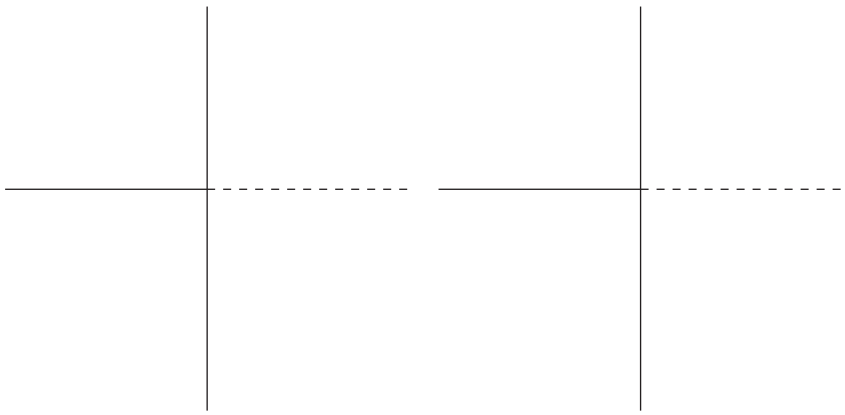

Figure 2.3: Two coordinate planes properly identified share the origin.

z_j is a removable singularity of the flow on a surface with a flat structure, there exist neighborhoods U_β and U_γ of z_j and maps φ_β and φ_γ such that the associated transition map $\varphi_\beta \circ \varphi_\gamma^{-1}$ is a shift in the local coordinates at z_j, where the number of preimages of a point $x \in \dot{V}_j$ is $m_j = 1$. As previously mentioned, the geodesic flow may then be continuously extended at removable singularities of the flat surface.

Example 72. Consider two coordinate planes as shown in Figure 2.3. If we identify the underside of the nonnegative x-axis in the left coordinate plane with the upperside of the nonnegative x-axis in the right coordinate plane and then identify the upperside of the nonnegative x-axis in the left coordinate plane with the underside of the nonnegative x-axis in the right coordinate plane, then one point these two planes will have in common is the origin. The origins are identified and any attempt to form a closed circle about this new point in the new space results in a cone point with a conic angle of 4π; see Figure 2.4.

We now discuss how to construct a flat surface from a rational billiard. Consider a rational polygon billiard $\Omega(P)$ with k sides and interior angles $\frac{p_j}{q_j}\pi$ at each vertex

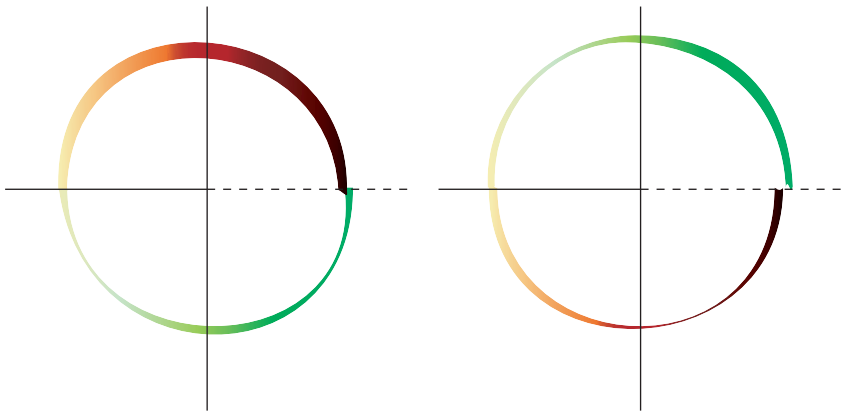

Figure 2.4: In this figure, we see that the origin has a conic angle of 4π.

z_j, for $1 \leq j \leq k$. Here, p_j and q_j are relatively prime positive integers. Then, a (laborious) calculation shows that for some $j \leq k$, $q_j = \text{lcm}(q_i)_{i=1,i\neq j}^{k}$. Consequently, the linear portions of the planar symmetries generated by reflection in the sides of the polygonal billiard $\Omega(P)$ generate the dihedral group D_N, where $N := \text{lcm}\{q_j\}_{j=1}^{k}$. Here, by definition, the *dihedral group* D_N denotes the group of symmetries of the regular N-gon. So as to be clear, we mention that the notation D_N does not refer to the (wrong) fact that such a finite group has N elements.[3] We next consider $\Omega(P) \times D_N$ (equipped with the product topology). We want to glue 'sides' of $\Omega(P) \times D_N$ together and construct a natural atlas on the resulting surface M so that M becomes a flat surface.

To such end, let $p_1 = p_2$ be a point on a side s_a of $\Omega(P)$, r_a be the linear portion of the reflection determined by reflecting $\Omega(P)$ in the side a, and $(p_1, r_1), (p_2, r_2) \in P \times D_N$. Then, by definition, $(p_1, r_1) \sim (p_2, r_2)$ if and only if

1. $(p_1, r_1) = (p_2, r_2)$, or

[3] Actually, D_N has $2N$ elements, and the standard group theory notation for the dihedral group is then D_{2N}, since the cardinality of the group is often given more importance, from the perspective of group theory.

2. $r_a = r_1^{-1} r_2$, or

3. p_1 and p_2 are the same vertex of $\Omega(P)$ having adjacent sides s_a and s_b, with $r_1^{-1} r_2$ belonging to the subgroup generated by r_a and r_b.

It is not difficult to check that $\sim$ is an equivalence relation on $\Omega(P) \times D_N$. More work is required to show that $M := (\Omega(P) \times D_N)/\sim$ is a compact, connected, orientable surface. As a result of the identification, the points of M that correspond to the vertices of $\Omega(P)$ constitute (removable or nonremovable) conic singularities of this surface. Heuristically, $\Omega(P) \times D_N$ can be represented as $\{r_j \Omega(P)\}_{j=1}^{2N}$, in which case it is easy to see what points are made equivalent under the action of $\sim$.

Denote by $\Omega(P)^\circ$ the interior of the billiard $\Omega(P)$. To construct a flat structure on M, let $U_j = \Omega(P)^\circ \times \{r_j\}$. Then $\{U_j, \varphi_j\}_{j=1}^{2N}$ can be naturally extended to constitute a flat structure on M, in the sense of Definition 68. Hence, M is a flat surface, in the sense of Definition 69. The map $\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$ is a translation in the local coordinates of a point $z \in U_i \cap U_j$; i.e., $\varphi_i \circ \varphi_j^{-1}(z) = z + d$, where the constant d is independent of the choice of i and j.[4]

Example 73. Consider the following example of a triangle with interior angles

$$(\pi/6, \pi/6, 2\pi/3). \tag{2.22}$$

If we fix the vertex corresponding to the angle $2\pi/3$ and make successive reflections in one of the adjacent sides to generate the picture in Figure 2.5, then we see that this vertex must have conic angle $2\pi c$, with $c > 1$. The reason for this is that after the three

[4] A priori, the choice of d describing the translation in the local coordinates *does* depend on i and j. However, given the fact that one constructs the flat surface by identifying parallel and opposite sides of the polygon B, for a fixed direction θ, one can describe a parallel line field in the direction $\theta := \arctan\frac{\delta}{\gamma}$, with $d =: \gamma + \sqrt{-1}\delta$.

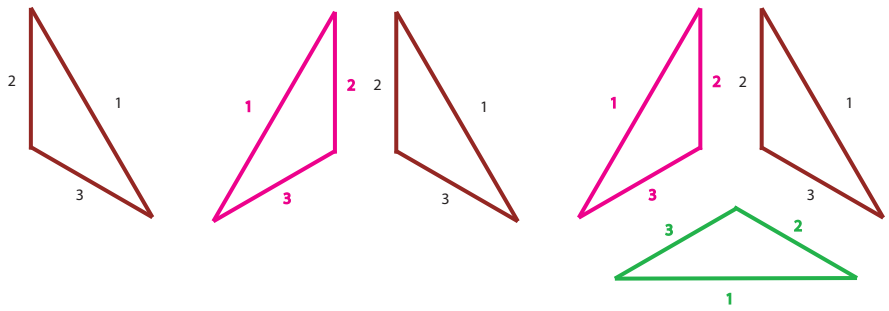

Figure 2.5: Successive reflections in the sides adjacent to the vertex and gluing such sides together produces a space for which the origin (the common point in all the triangles) has a conic angle of 4π.

reflections, the reflected copy of the triangle will not have the proper orientation. A total of six reflections will produce a copy of the triangle with the proper orientation and 4π radians are required to form a closed circle about this vertex, a point that all the triangles have in common.

Example 74 (The equilateral triangle billiard $\Omega(KS_0)$)**.** We consider the equilateral triangle $\Delta := KS_0$, an important example in the literature (see, e.g., [BaUm]) and an even more important example when considering the Koch snowflake prefractal billiard. Since the interior angles $\pi p_i/q_i$ are all the same and equal to $\pi/3$, we have

$$N = \operatorname{lcm}\{q_i\}_{i=1}^{3} = \operatorname{lcm}\{3,3,3\} = 3. \tag{2.23}$$

Consequently, the associated surface is given by $\mathcal{S}(KS_0) := (\Delta \times D_3)/\sim$. Moreover, there is a flat structure on this surface and since all the singularities are removable conic singularities, such a structure can be extended to include the singular points and the resulting surface is topologically equivalent to a torus. (Really, the associated flat surface $\mathcal{S}(KS_0)$ is the hexagonal torus, which is topologically equivalent to the torus determined from identifying the four sides of the unit square appropriately; see Figure 3.3 in §3.)

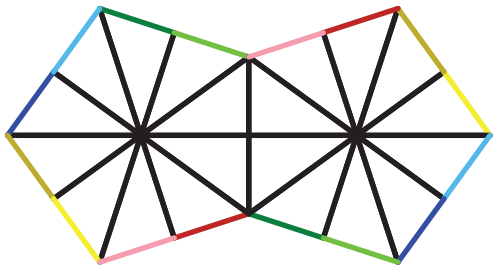

Figure 2.6: The associated flat surface for the triangle $(\pi/2, \pi/5, 3\pi/10)$.

Example 75. Consider the triangle given by $(\pi/5, \pi/2, 3\pi/10)$. If we fix the vertex with angle measuring $\pi/5$ and make successive reflections in one of the sides, the result is a pentagon with 10 copies of the triangle. However, a quick calculation shows that there should be 20 copies of the triangle. The group of symmetries generated by the linear portions of the planar symmetries given by the sides of the triangle corresponds to D_{10}. While 10 reflections produced a triangle with the appropriate orientation (meaning the conic angle about the point coming from the vertex with angle $\pi/5$ in the surface is 2π), it follows that the collection of ten triangles is not the whole surface (once we identify sides appropriately).

If one wanted to generate the surface by reflecting successively in an adjacent side of a fixed vertex, that vertex would have to be the corner with an angle measuring $3\pi/10$. However, the conic angle of that point is 6π and the successive reflections would not live in the plane, rather would have to lift out of the plane.

We close this discussion by recalling an important fact about the geodesic flow on a flat surface $\mathcal{S}(B)$ constructed from a rational billiard $\Omega(B)$ and the billiard flow on that rational billiard. These two flows can be shown to be (dynamically) equivalent under the action of the group D_N associated with the construction of the corresponding flat surface. Heuristically, one may view the corresponding equivalence as follows: a billiard

flow line may be straightened (i.e., unfolded) to a geodesic flow line and additionally, a geodesic flow line may be collapsed into a billiard flow line. In order to be more technically correct, we further explain the details of the equivalence of these two flows. If we consider the geodesic flow on the surface, the quotient of the flat surface by the group of symmetries associated with the construction of $\mathcal{S}(B)$ has the effect of collapsing the space to a space with a quotient flow that is isomorphic to the billiard flow. On the other hand, any given billiard orbit may be straightened by making successive reflections, via the action of D_N, in the identified sides of the flat surface, therefore producing a straight-line flow line on the flat surface $\mathcal{S}(B)$.

2.5.2 The Veech group and the Veech Dichotomy

Consider a flat surface $\mathcal{S}(B)$, where B is a rational polygon. The geodesic flow is dynamically equivalent to the billiard flow, as discussed at the end of §2.5.1. The stabilizer of the flat surface, when it has a particular structure, will be able to provide information about the nature of the geodesic flow on $\mathcal{S}(B)$.

Definition 76 (Affine automorphism)**.** Suppose $\mathcal{S}(B)$ is a flat surface (and B is a rational polygon). An *affine automorphism* of $\mathcal{S}(B)$ is a homeomorphism $\varphi : \mathcal{S}(B) \to \mathcal{S}(B)$ such that the following hold:

1. φ permutes the nonremovable conic singularities of $\mathcal{S}(B)$.

2. Every point $p \in \mathcal{S}(B)$ for which the cone angle is 2π, there is a neighborhood U_p and a neighborhood $\varphi(U_p) = U_q$, $q = \varphi(p)$, in which $\varphi(U_p)$ is isometric (via a map f_p) with a neighborhood of $\mathbb{R}^2$ and $f_q \circ \varphi \circ f_p^{-1} : f_p(U_p) \to f_q(U_q)$ is an affine transformation.

Definition 77 (The Veech group $\Gamma(B)$)**.** Let B be a connected rational polygon and

$\mathcal{S}(B)$ be a flat surface. The collection $\Gamma(B)$ of affine automorphisms of $\mathcal{S}(B)$ is called the *Veech group* of the flat surface.

Statement 78. The *Veech dichotomy* says that the flow is either closed or uniquely ergodic.

We appeal to the fact that G has a representation in $\mathrm{PSL}_2(\mathbb{R})$ to state the following theorem, due to A. Veech ([Ve4]).

Theorem 79 (Veech)**.** If the Veech group $\Gamma(B)$ is co-compact in $\mathrm{PSL}_2(\mathbb{R})$, then the Veech dichotomy holds for the flat surface $\mathcal{S}(B)$.

Example 80 (The Veech group of the flat torus $(\mathbb{R}/\mathbb{Z})^2$)**.** Consider the flat surface given by properly identifying sides of the square. Since the flat surface has no nonremovable singularities and is locally isometric with $\mathbb{R}^2$, it follows that the corresponding Veech group is the collection of affine automorphisms $Ax + By$, $A, B \in \mathbb{R}$.

2.6 Concepts in fractal geometry

We present here the concepts in the field of fractal geometry necessary for understanding the content of this thesis. For a more developed treatment of fractal geometry, see [Fc].

2.6.1 Iterated function systems

Definition 81 (Contraction mapping)**.** Let D be a closed subset of $\mathbb{R}^m$. A map $\psi : D \to D$ is called a *contraction map* on D if there is a constant c with $0 < c < 1$ such that $|\psi(x) - \psi(y)| \leq c|x - y|$ for every $x, y \in D$.

Definition 82 (Similarity map)**.** Let D be a closed subset of $\mathbb{R}^m$. A map $\psi : D \to D$ is called a *contracting similarity map* on D if there is a constant c with $0 < c < 1$ such

that $|\psi(x) - \psi(y)| = c|x - y|$ for every $x, y \in D$.

Remark 83. In the context of this thesis, every contracting similarity map will be referred to as a *similarity map*, for the sake of brevity. If we should be discussing a more general similarity map (i.e., $c > 0$), we will be very clear to point this out. One case of this occurs in Theorem 95.

Definition 84 (Iterated function system (or IFS))**.** Let D be a nonempty, compact subset of $\mathbb{R}^m$. For every $j \leq k$, let $\psi_j : D \to D$ be a contraction map. Then $\{\psi_j\}_{j=1}^k$ is called an *iterated function system* (or IFS) and is denoted by Ψ.

Definition 85 (Fixed-point attractor)**.** Let $\{\varphi_j\}_{j=1}^k$ be an IFS. A nonempty subset $F \subset D$ is called an *attractor* of the IFS if

$$F = \bigcup_{j=1}^{k} \psi_j(F) \tag{2.24}$$

This same set F is also a fixed-point of the map $\Lambda_\Psi(\cdot) := \bigcup_{j=1}^k \psi_j(\cdot)$ defined on the space of compact subsets of D. That is, $\Lambda_\Psi(F) = F$.

As we will see, the fact that F is called an attractor of an IFS is motivated by how we show there exists such a set. Moreover, the procedure we go through to do so will end up validating another statement and that is that such an attractor is unique.

Theorem 86 (Contraction Mapping Principle)**.** Let X be a nonempty, complete metric space. Let $\psi : D \to D$ be a contraction mapping on a closed subset $D \subseteq X$. Then the map ψ admits one and only one fixed-point x^* in D.

Theorem 87. Let $\varphi = \{\varphi_j\}_{j=1}^k$ be an IFS defined on a closed subset $D \subseteq \mathbb{R}^m$ and $\Lambda_\Psi(\cdot) := \bigcup_{j=1}^k \psi_j(\cdot)$. Then there exists a unique, nonempty, compact, fixed-point attractor F such that for any compact subset $A \subseteq D$ such that $\psi_j(A) \subseteq A$ for every

$j \leq k$, we have that

$$F = \bigcap_{i=0}^{\infty} \bigcup_{\alpha \in J_i} \psi_\alpha(A) \tag{2.25}$$

where $J_i = \{\alpha = (\alpha_1, \alpha_2, ..., \alpha_i) | \alpha_t \leq k, t \leq i\}$ and $\psi_\alpha(A) := \psi_{\alpha_i} \circ ... \circ \psi_{\alpha_1}(A)$.

Finally, we close this section with the notion of self-similarity and the similarity dimension.

Definition 88 (Self-similar set)**.** Let F be the unique, nonempty, fixed-point attractor of an IFS Ψ. If every contraction map ψ_j is also a similarity map, then we say F is a *self-similar set*.

Remark 89. Throughout the thesis, we will be referring to certain sets as *self-similar* even when it is clear they are not. Really, in these particular situations, we mean that either some obvious subset (which we would clearly indicate) of the set is self-similar or the set is the finite union of non-overlapping (or, rather, abutting) copies of a self-similar set.

Theorem 90 (The Moran Equation)**.** Let Ψ be an IFS and c_j the corresponding scaling ratio of the contraction mapping ψ_j. Then the following equation (known as the Moran Equation) has a unique, nontrivial (and nonnegative) solution.

$$\sum_{j=1}^{k} c_j^s = 1 \tag{2.26}$$

That is, there exists a unique real value s_0 such that the Moran Equation (2.26) is satisfied.

Definition 91 (Open set condition)**.** Let Ψ be an iterated function system. Then, for each $1 \leq j \leq k$, the contraction mapping ψ_j is said to satisfy the *open set condition* if there exists a non-empty bounded open set V such that

$$\bigcup_{j=1}^{k} \psi_j(V) \subseteq V. \tag{2.27}$$

2.6.2 Hausdorff dimension and measure

Definition 92 (Diameter of a set)**.** Let $U \subseteq \mathbb{R}^m$ be a nonempty subset. The *diameter* of the set U is given as follows.

$$|U| := \sup\{|x - y| : x, y \in U\}, \tag{2.28}$$

where $|x - y|$ is the n-dimensional Euclidean distance from x to y.

Definition 93 (δ-cover of a set A)**.** Let $\delta > 0$ and $\{U_j\}_{j=1}^{\infty}$ be a countable collection of sets covering a set A with $|U_j| < \delta$ for all $j \geq 1$. Then $\{U_j\}_{j=1}^{\infty}$ is called a *δ-cover* of the set A.

Let $A \subseteq \mathbb{R}^m$ and $s \geq 0$. For any $\delta > 0$, we define the following.

$$\mathscr{H}_{\delta}^{\sigma}(X) := \inf\left\{\sum_{j=1}^{\infty} |U_j|^{\sigma} \mid \{U_j\}_{j=1}^{\infty} \text{ is a } \delta\text{-cover of } X\right\}. \tag{2.29}$$

We want to note that $\mathscr{H}_{\delta}^{\sigma}$ is not an outer measure. Unrelated to this note, but worth remarking on is the fact that as $\delta \to 0$, the collection of permissible covers is reduced. Therefore, $\mathscr{H}_{\delta}^{\sigma}(A)$ increases as $\delta \to 0$.

Definition 94 (σ-dimensional Hausdorff measure of A)**.** Let A be a subset of $\mathbb{R}^n$. Then the σ-dimensional Hausdorff measure $\mathscr{H}^{\sigma}(A)$ is defined as follows.

$$\mathscr{H}^{\sigma}(X) := \lim_{\delta \to 0} \mathscr{H}_{\delta}^{\sigma}(X). \tag{2.30}$$

One can show that $\mathscr{H}^{\sigma}$ is, indeed, a measure in the sense defined in Definition 7. In fact, when $\sigma = m \geq 1$, the measure $\mathscr{H}^{\sigma}$ and Lebesgue measure λ on $\mathbb{R}^m$ agree on the σ-algebra of Lebesgue measurable sets, up to a constant. When $\sigma = 1$, then the two measures agree exactly on the space of Lebesgue measurable sets.

Theorem 95 (Scaling property of Hausdorf measure)**.** Let φ be a similarity transformation of scaling ratio $c > 0$. If $X \subset \mathbb{R}^d$, then

$$\mathscr{H}^{\sigma}(\varphi(X)) = c^{\sigma}\mathscr{H}^{\sigma}(X). \tag{2.31}$$

Proposition 96. If $t > \sigma$ and $\{U_j\}_{j=1}^{\infty}$ is a δ-cover of a set X, then we have

$$\sum_{j=1}^{\infty} |U_j|^t \leq \sum_{j=1}^{\infty} |U_j|^{t-\sigma} |U_j|^{\sigma} \tag{2.32}$$

$$\leq \delta^{t-\sigma} \sum_{j=1}^{\infty} |U_j|^{\sigma}. \tag{2.33}$$

Therefore, $\mathscr{H}_{\delta}^{t}(X) \leq \delta^{t-\sigma} \mathscr{H}_{\delta}^{\sigma}(X)$.

Proposition 97. If $\mathscr{H}^{\sigma}(X) < \infty$, then $\mathscr{H}^{t}(X) = 0$ for every $t > \sigma$.

Proposition 98. If $\mathscr{H}^{\sigma}(X) > 0$, then $\mathscr{H}^{t}(X) = \infty$ for every $t < \sigma$.

Theorem 99. Let Ψ be an IFS with each contraction map ψ_j satisfying the open set condition. If F is the attractor of the IFS, then the Hausdorff dimension σ is the unique value s_0 satisfying the Moran Equation. Moreover, for this value s_0, the s_0-dimensional Hausdorff measure is nonzero and finite.

2.7 The ternary Cantor set

The ternary Cantor set $\mathscr{C}$ (hereafter referred to as the Cantor set) is usually the first fractal one learns about and is the canonical example of a topological Cantor set, if not the canonical example of a self-similar fractal set.

2.7.1 Geometric construction

One may construct the Cantor set by starting with the unit interval $I := [0, 1]$ and removing the open middle third $(1/3, 2/3)$. Subsequent middle thirds are then removed from each remaining closed interval. The surprising fact is that one removes enough intervals so that none remain, but there are still uncountably many points remaining; see Figure 2.7.

The Cantor set is the unique, nonempty, compact, fixed-point attractor of a particular IFS. The particular contraction maps comprising the iterated function system

Figure 2.7: The construction of the ternary Cantor set.

Ψ are as follows.

$$\psi_1(x) = \frac{1}{3}x \tag{2.34}$$
$$\psi_2(x) = \frac{1}{3}x + \frac{2}{3}.$$

As one can see from Figure 2.7, the result of each iteration of the map $\Lambda_\Psi := \bigcup_{j=1}^{2} \psi_j$ is a collection of intervals that effectively amounts to removing middle thirds from the previous approximation. As one can (or, perhaps, cannot) deduce from the geometric construction, there are no intervals remaining in the limit. As we will discuss in §2.7.3, the ternary Cantor set is an example of a topological Cantor set.

2.7.2 Symbolic construction

The geometric construction given above in §2.7.1 can be described symbolically. Consider the alphabet $\{l, c, r\}$. What we are about to show amounts to describing a ternary *representation* of elements of the Cantor set. Let us begin by first describing a few interesting elements of $\mathscr{C}$. The rational value 1/3 has a ternary expansion given by 0.1 or, equivalently, $0.0\overline{2}$. The rational value 1/4 has a ternary expansion given by $0.\overline{02}$. In terms of our alphabet, we say 1/3 has a ternary *representation* given by $l\overline{r}$ and 1/4 has a ternary *representation* given by $\overline{lr}$. We want to stress that a ternary representation will always consist of infinitely many characters, while a ternary *expansion* of an element

of the unit interval I can be finite, which is illustrated by the example of the rational value $1/3$.

We want to be careful and note that a ternary representation of an element does not translate into a ternary expansion of the same value perfectly. That is, an l does not translate into a 0, a c does not translate into a 1 and an r does not translate into a 2; again, $x = 1/3$ serves as an example of this. Also, it may be possible for an elements of the Cantor set $\mathscr{C}$ to have a ternary representation consisting of infinitely many l's and finitely many c's and r's or infinitely many r's and finitely many l's and c's. For example, $1/3$ has a representation given by $l\overline{r}$ and, equivalently, $c\overline{l}$. For the sake of simplicity, we take every element of $\mathscr{C}$ to have a ternary representation that contains no c's.[5] The symbolic description of elements of $\mathscr{C}$ can be tied to the geometric construction via the IFS described in Equation (2.34). To see this, note that each subscript in a composition of length j, $\psi_{k_j} \circ \cdots \circ \psi_{k_1}$, constitutes a character in a word α_x entirely comprised of either l's and r's. This is not really entirely unexpected, especially since the labels for the characters of the alphabet $\mathscr{A}$ were chosen to represent the choices of *left*, *center* and *right*.

In the sequel, the *type* of ternary representation will provide us with important information. Particular qualities of the representation will, in part, dictate the type of orbit that results and the nature of the sequence of compatible orbits. For example, $1/4 = \overline{lr}$, but we are more interested in the fact that such a representation consists of infinitely many l's and r's and no $c's$.

Notation 100. The *type* of ternary representation can be summarized as follows. If $x \in I$, then the first coordinate of $[\cdot, \cdot]$ describes the characters that occur infinitely

[5] What we want to prevent is a point of $\mathscr{C}$ having a representation determined by approaching it from the complement of $\mathscr{C}$. In this way, every point of $\mathscr{C}$ has a unique ternary representation.

often and the second coordinate of $[\cdot, \cdot]$ describes the characters that occur finitely often. If we want to discuss many different types of ternary representations, then we use the exclusive or. That is, the notation $[\cdot, \cdot] \oplus [\cdot, \cdot] ... \oplus [\cdot, \cdot]$ is to be read as $[\cdot, \cdot]$ *or* $[\cdot, \cdot] ...$ *or* $[\cdot, \cdot]$. If the collection of characters occurring finitely often is empty, then we write the type of ternary representation as $[\cdot, \emptyset]$.

Example 101. If an element $x \in I$ has a ternary representation consisting of infinitely many c's and l's but finitely many r's, then we write this type of ternary representation as $[lc, r]$. If we have a collection of points in I such that each point has a ternary representation consisting of infinitely many c's and l's or infinitely many l's and r's then we write the types of ternary representations as $[lc, r] \oplus [lr, c]$. Concretely, the element $1/4 \in \mathscr{C}$ has an explicit ternary representation given by $\overline{lr}$ and the type of ternary representation is described as $[lr, \emptyset]$. Even though the ternary representation of $1/10 \in \mathscr{C}$ is not the same as $1/4$, both have the same type of ternary representation.

2.7.3 Properties

The Cantor set $\mathscr{C}$ is an example of a topological Cantor set, meaning $\mathscr{C}$ is a perfect set that contains no segments of positive length. In the context of a subset of a measure space equipped with the Lebesgue measure, $\mathscr{C}$ is a set of measure zero. However, the Cantor set is an uncountable set (in fact, it has the cardinality of the continuum). The fact that it is uncountable is proven by using Cantor's diagonalization argument. Constructing the bijection $f : \mathscr{C} \to [0, 1]$ is slightly more involved.

As alluded to in §2.7.1, the Cantor set $\mathscr{C}$ has no volume. One can see this by summing up the lengths of the open intervals removed as each stage of the geometric construction. At the first stage, one interval of length $1/3$ has been removed. At the second stage, two intervals of length $1/9$ have been removed. At the nth stage, 2^{n-1}

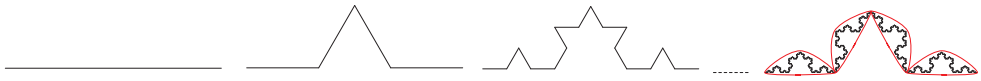

Figure 2.8: The Koch curve construction.

intervals of length $1/3^n$ have been removed. Summing up these lengths, we find that the total length removed is 1, which was the whole length of the unit interval I.

The Cantor set has a similarity dimension of $\log_3 2$. That is, the solution s_0 to the Moran Equation is $s_0 = \log_3 2$. Since it can be shown the ternary Cantor set satisfies the open set condition, the Hausdorff dimension σ of $\mathscr{C}$ is s_0.

2.8 The Koch snowflake

The Koch snowflake KS, the main fractal used throughout this dissertation, is intimately tied to $\mathscr{C}$.

2.8.1 Geometric construction

We first describe the construction of the Koch curve, and from that, construct the Koch snowflake KS.

The Koch curve is constructed as shown in Figure 2.8. One begins with the unit interval, removing the open middle third and placing at the endpoints 1/3 and 2/3 two uprights that would have formed the sides of an equilateral triangle. Heuristically, one can say that they are replacing open middle thirds of length $1/3^n$, for some positive integer n, with an equilateral triangle of the same scale but with no base.

One may also show that the Koch curve is the unique, nonempty, compact, fixed-point attractor of a particular IFS. The IFS is comprised of the contraction maps

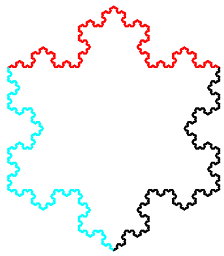

Figure 2.9: The Koch snowflake fractal.

given as:

$$\begin{aligned}
\varphi_1(x) &= \frac{1}{3}x \\
\varphi_2(x) &= \frac{1}{3}x + \left(\frac{2}{3}, 0\right) \\
\varphi_3(x) &= \frac{1}{3}\rho x + \left(\frac{1}{3}, 0\right) \\
\varphi_4(x) &= \frac{1}{3}R\rho x + \left(\frac{2}{3}, 0\right),
\end{aligned} \tag{2.35}$$

where R is reflection through the vertical axis and ρ is rotation by $\pi/3$. The generating set, however, in this case is not the unit interval, but a rectangle with unit length base and height measuring $h = \sqrt{3}/6$.

Then, the Koch snowflake KS is constructed by appending three copies of the Koch curve in such a way that the result is a closed and connected curve; see Figure 2.9. Alternately, one could view the Koch snowflake fractal KS as the limit of particular polygonal approximations. By starting with the equilateral triangle and replacing middle thirds of each successive side with a smaller triangle missing its base, the set that results in the limit is also the Koch snowflake; see Figure 2.10.

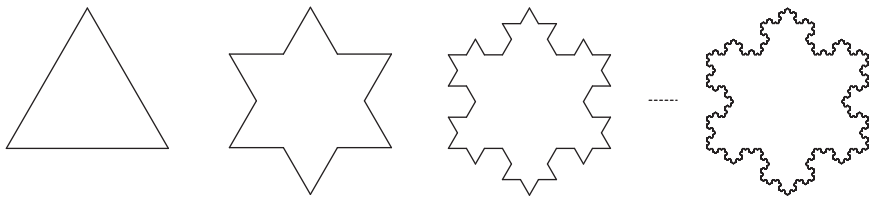

Figure 2.10: The Koch snowflake fractal as a limit of polygonal approximations.

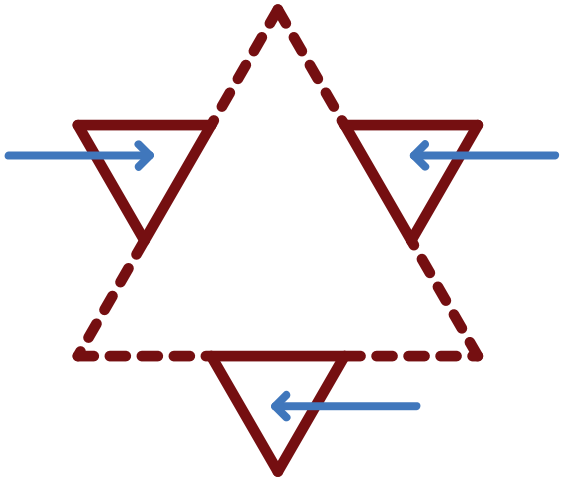

Figure 2.11: An illustration of Definition 102 in terms of $KS_0 = \Delta$ and KS_1.

2.8.2 Symbolic construction

The symbolic description of points of the Koch snowflake is motivated—as one would expect—by the geometric construction described in §2.8.1. In order to describe how one goes about assigning addresses to points of the snowflake KS, we first define what a *cell* of the Koch snowflake billiard is.

Definition 102 (A cell $C_{n,k}$ of $\Omega(KS_n)$)**.** Consider (the 'set-theoretic difference') $\Omega(KS_n) \setminus \Omega(KS_{n-1})$. Each resulting triangular region is then called a *cell* of $\Omega(KS_n)$. We denote a cell of $\Omega(KS_n)$ by $C_{n,k}$, where $k \leq 3 \cdot 4^{n-1}$ refers to the side $s_{n-1,k}$ of $\Omega(KS_{n-1})$ to which the cell was glued; see Figure 2.11.[6]

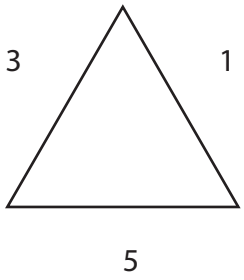

Figure 2.12: The only occurrence of a 5 in an address of a point of KS will be at the very beginning of the address. Essentially, this labeling indicates, heuristically speaking, in which copy of the Koch curve is the point being addressed contained.

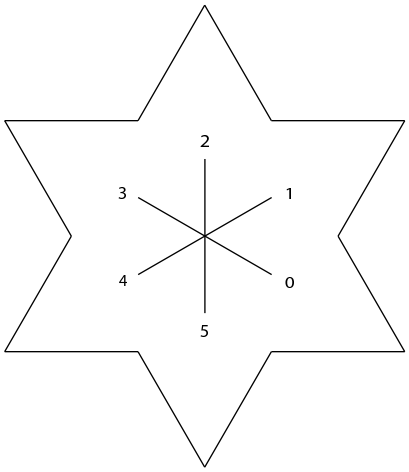

Figure 2.13: A directional in $\Omega(KS_1)$ indicating six possible directions of travel. It is this directional that will essentially provide a 'path' to any point on the snowflake.

It is here that we want to describe an addressing system for the Koch snowflake KS. We note that such an addressing system was first introduced in [LapPa] and used again in [LapNie2]. Consider the alphabet given by $\mathscr{A} = \{0, 1, 2, 3, 4, 5\}$. We first suppose that the sides of the equilateral triangle are labeled as shown in Figure 2.12. A directional indicating six possible directions is shown in Figure 2.13.

Notation 103. If α_x is an address of an element $x \in KS$, then $\alpha_x|_n$ is the first n characters of the address, and $(\alpha_x)_n$ is the nth character of the word.

[6]Hence, there are $3 \cdot 4^{n-1}$ cells $C_{n,k}$ of $\Omega(KS_n)$ and so $1 \leq k \leq 3 \cdot 4^{n-1}$, in Definition 102.

Let $x \in KS$ have an address given by the infinite word α_x given in terms of the characters of $\mathscr{A}$. Then, the characters 0, 1, 2, 3 and 4 have the following geometric meaning. We first deal with the case where $(\alpha_x)_1 = 1, 3$. Suppose a segment of KS_n has a finite address $\alpha_x|_{n+1}$ ending in:

- the character 1. If $(\alpha_x)_{n+2}$ is:
  - the character 2, then this indicates one chooses the left third of the side.
  - the character 0, then this indicates one chooses the right third of the side.
  - the character 1, then this indicates a rotation of the directional by $-\pi/3$, to move forward into a new cell and choose the right side of the cell.
  - the character 3, then this indicates a rotation of the directional by $\pi/3$, to move forward into a new cell and choose the left side of the cell.
- the character 3. If $(\alpha_x)_{n+2}$ is
  - the character 4, then this indicates one chooses the left third of the side.
  - the character 2, then this indicates one chooses the right third of the side.
  - the character 1, then this indicates a rotation of the directional by $-\pi/3$, to move forward into a new cell and choose the right side of the cell.
  - the character 3, then this indicates a rotation of the directional by $\pi/3$, to move forward into a new cell and choose the left side of the cell.

Suppose the address begins with a 5. Until a 1 or 3 appear in the address, the characters 0 and 4 indicate to choose the right and left third of the base of $\Omega(KS_0) = \Omega(\Delta)$, respectively. Then,

- the character 1 occurs in the word beginning with a 5 indicates a rotation of the directional by π, to move forward into a new cell and choose the right side of the cell.

- the character 3 occurs in the word beginning with a 5 indicates a rotation of the directional by π, to move forward into a new cell and choose the left side of the cell.

The rules previously described in the case of when $(\alpha_x)_1 = 1, 3$ then apply once a 1 or 3 has occurred in the address α_x. Examples of addresses of sides of an approximation of the Koch snowflake are given in Figure 2.14.

Notation 104. We may use the same notation given in Notation 100 when describing types of addresses of points of the snowflake KS.

2.8.2.1 Representing elements of KS_n

The addressing system used to identify points of KS will become particularly useful in Chapter 5. In Chapter 4, we will be more interested in the information provided by the ternary representation of an element x_n^0 of a side $s_{n,k}$, $k \leq 3 \cdot 4^n$, of $\Omega(KS_n)$. In order to make that material more palatable, we discuss how exactly we will represent elements of KS_n below.

We have already seen how to represent elements of the unit interval (in particular, elements of $\mathscr{C}$) and elements of the fractal KS. We want to use those two addressing systems to represent elements of the sides of a prefractal KS_n. Consider a side $s_{n,k}$ of KS_n. This side can be identified using the addressing system discussed in §2.8.2. Since the addressing system implicitly codes for a notion of '*left*, 'center' and '*right*,' we can give the ternary representation of an element x_n^0 of the side $s_{n,k}$. Based on this method,

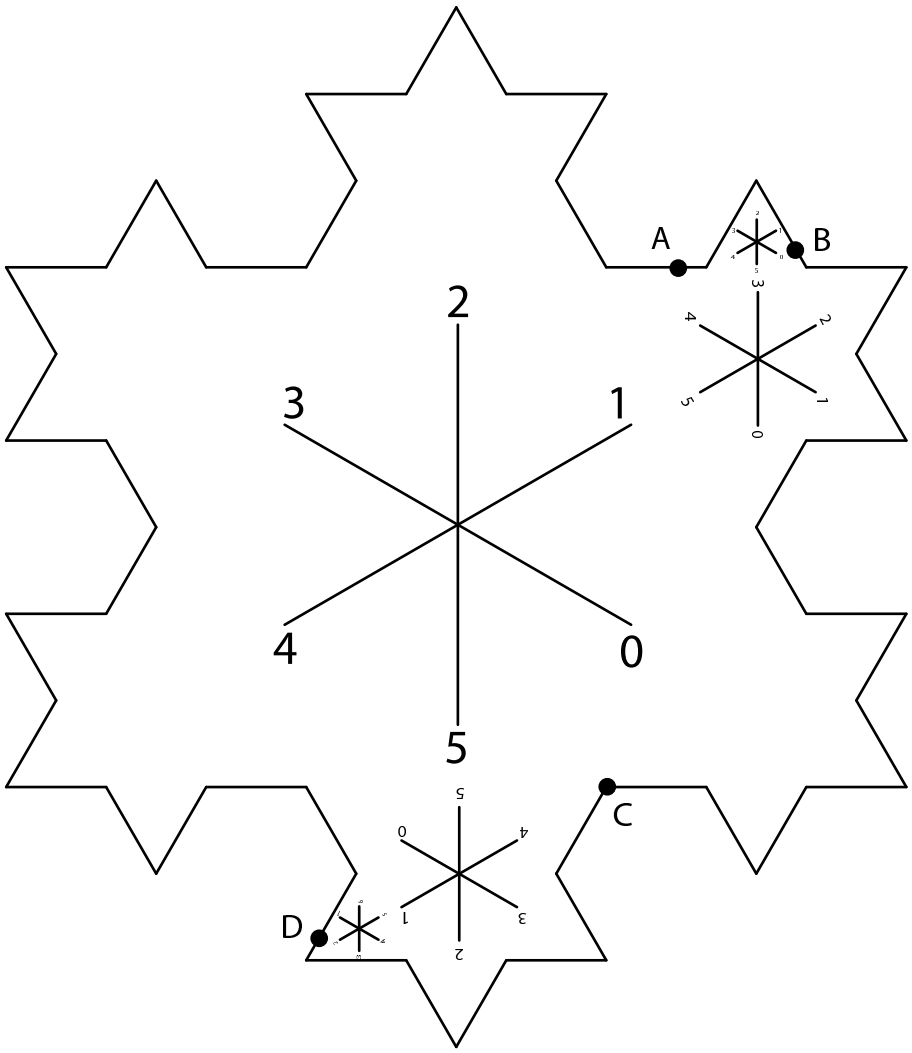

Figure 2.14: Examples of exactly how the rules for addressing points are implemented with the use of the directional. In the context of a finite approximation KS_n, the addressing system serves to label sides of the prefractal approximation KS_n. Points A, B and D then lie on sides which are labeled by the addresses 134, 131, 511, respectively. The point C, however, does not lie exactly on a side, but at a vertex. Hence, one could say that C lies on the side with address 504 or 534. Moreover, in the limit, C may be represented by the equivalent addresses $50\overline{4}$ and $53\overline{4}$, where (as before) the overbar indicates that 4 is repeated ad infinitum. Since a billiard orbit is a collection of dynamically ordered points, it follows that each basepoint of an element of an orbit can be given an address in terms of this addressing system and still be distinguished from other basepoints of elements of the orbit with the same address.

we can see that even though two points on two different sides may have the same ternary representation, we can distinguish between the two points by associating with each point the finite address of the sides on which each point lies.

For example, the representation $\overline{lr}$ may be used to represent 1/4 in the unit-interval base of the equilateral triangle or 1/12 on the side $s_{1,k}$ corresponding to the left third of the unit interval I. Moreover, in $\Omega(KS_1)$, the ternary representation of 1/4 is $\overline{rl}$ and not $\overline{lr}$ as it was in $\Omega(KS_0)$. One can see, however, that the representation of 1/4 in $\Omega(KS_1)$ is the left shift of the representation of $\Omega(KS_0)$, which is true in general.

2.8.3 Properties

The Koch snowflake is a compact, connected and infinitely long curve in the plane. At no point of the snowflake KS can one form a well-defined tangent. This last property is what makes the Koch snowflake billiard so impenetrable. In terms of the so-called *fractal properties*, KS has a Hausdorff dimension of $\log_3 4$. Since it is the union of three self-similar curves, each of which satisfies the open set condition, the similarity dimension and Minkowski dimension are both equal to the Hausdorff dimension.

2.8.4 The geography of the Koch snowflake

At first glance, the title of this section may seem a bit out of place. However, the Koch snowflake is a terribly difficult set to deal with. Having some language in which to describe where it is one may be could be useful. Hence, we describe various points and regions of the Koch snowflake and its prefractal approximations.

We have already defined *a cell* of $\Omega(KS_n)$. We have shown how to describe the addresses of points of the snowflake and the fact that a finite address can be used to address a side of a prefractal approximation $\Omega(KS_n)$. In fact, the addresses of sides of a

prefractal approximation $\Omega(KS_n)$ will always consist of $n+1$ many characters.

In the Koch snowflake KS, there are three types of points: *corners*, *Cantor-points* and *elusive limit points*. We first describe these points in terms of their qualities and then describe them in terms of their addresses.

A *corner* of the Koch snowflake KS is an element that is a corner of a finite approximation KS_n, for some $n \geq 0$. A *Cantor-point* of KS is an element of a finite approximation KS_n, for some $n \geq 0$, such that the ternary representation of this point (with respect to the side on which it lies) consists of infinitely many l's *and* r's. Therefore, a Cantor-point is an element of KS that is not a corner, but definitely a tangible point. An *elusive limit point* of KS is an element of KS that is never an element of any finite approximation KS_n. This may seem like a quick way of brushing most of the elements of KS under the carpet, to be forgotten or ignored, but we will see when we discuss the addresses of these types of points that they have a clearly defined representation. Moreover, it is the Cantor-points and elusive limit points that will be of the greatest interest in the later chapters.

In terms of a ternary representation, a corner of KS has a ternary representation $[l, cr] \oplus [r, lc]$. As one would then expect, the address of a point consists of infinitely many of some character, but never infinitely many of more than one character. For example, consider the address $\alpha_x = 13\overline{2}$. This is an element of a side with address 13 that corresponds with a point having a ternary representation given by (relative to the side 13) $\overline{r}$. We then define a corner of KS in the following way.

Definition 105 (A corner of KS)**.** If $x \in KS$ has an address α_x $[2, 01345] \oplus [0, 12345] \oplus [4, 01235]$, then x is a *corner* of KS.

In terms of a ternary representation, a Cantor-point of KS has a ternary representation $[lr, c]$. As one would then expect, the address of such a point consists of

infinitely many of exactly two characters. For example, $\alpha_x = 13\overline{42}$ is an element of KS with a ternary representation $[lr, c]$. In fact, relative to the side 13 of $\Omega(KS_1)$, the ternary representation of x is $\overline{lr}$. We then define a Cantor-point of KS in the following way.

Definition 106 (A Cantor-point of KS)**.** If $x \in KS$ has an address α_x $[02, 1345] \oplus [24, 0135] \oplus [04, 1235]$, then we say x is a *Cantor-point* of KS.

Remark 107. The name *Cantor-point* is slightly misleading. It is true that every corner of KS is in the Cantor set of some side of some approximation. However, we make the distinction between corners and so-called Cantor-points even though both types of points are certainly points of a Cantor set.

Remark 108. We note that the method for identifying Cantor-points of a side $s_{n,k}$ described in §2.8.2.1 produces a representation that is equivalent to the representation given in terms of the addressing system.

An *elusive limit point* of KS is quite elusive. As previously mentioned, it is a point that is never an element in any finite approximation KS_n of KS. We have seen that the effect of a 1, 3 or 5 in an address is to 'move forward' into a cell of a finer approximation. It is this quality that makes an elusive limit point so elusive. We define an *elusive limit point* in terms of the type of address as follows.

Definition 109 (Elusive limit point)**.** Let $x \in KS$ have an address α_x $[1, 02345] \oplus [3, 01245] \oplus [13, 0245] \oplus [12, 0345] \oplus [01, 2345] \oplus [012, 345] \oplus [23, 0145] \oplus [34, 0125] \oplus [234, 015] \oplus [013, 245] \oplus [134, 025] \oplus [0123, 45] \oplus [0134, 25] \oplus [01234, 5]$. Then we say x is an *elusive limit point* of the Koch snowflake KS.

When we begin discussing billiard orbits of $\Omega(KS_n)$, we will find it more convenient to measure angles of incidence and reflection relative to a fixed coordinate system. As such, we suppose the left corner of the equilateral triangle with side length measuring

$l = 1$ constitutes the origin. However, on occasion, we will find it useful to refer to the angle of reflection measured relative to a particular side. So that no confusion arises, when we are discussing such a situation, if ϖ is an angle measured relative to a side of $\Omega(KS_n)$, then $\theta(\varpi)$ is the same angle measured relative to the fixed coordinate system.

2.8.4.1 Straightening addresses of the Koch snowflake

Let x be an elusive limit point of the Koch snowflake. Then x has an address as described in Definition 109. Without loss of generality, we may assume the address begins with a 1. Since the only characters that can follow a 1 before encountering the next 3 (should there be a 3 in the address) are 0 and 2, there is a least n_1 such that the n_1 character of the word is a 1 or 3 and not a 0 nor a 2. If it is a 3, we continue on to the next character of the address. If it is a 1, then we may switch this 1 for a 3 and until the next 1 or 3 (occurring at the n_2 position of the word), switch every 4 for a 0 while keeping the 2's unchanged. Continuing this process, making sure that every subsequent 1 is followed eventually by a 3 and not a 1, we end up with what we call the *straightening of* x, which we denote by $s(x)$. Since $s(x)$ has an address consisting of 1's and 3's interspersed with 0's and 2's and/or 2's and 4's such that the 1's and 3's alternate, we see that this point is collinear with a point of I along the base of the equilateral triangle and such a line connecting them contains no other points of the snowflake.

Example 110. Let n and k be positive integers. Consider the address of a side $s_{n,k}$ of $\Omega(KS_n)$ given by $\alpha = 13123232113133100324$. The straightening of this address is 13123212313131344120. Next, suppose the address of a particular point is given by $\overline{\alpha} = \overline{13123232113133100324}$. Then $s(x) = \overline{13123212313131344120}$.

Remark 111. Many of the *elusive limit points* of the Koch snowflake may be reached via a piecewise linear logarithmic spiral. As the name suggests, such spirals may be

straightened to straight line segments that begin in the interior of the snowflake and end on the boundary of the snowflake, thus identifying an initial basepoint x_0^0 of an orbit of the billiard $\Omega(KS_0)$ and the basepoint x_0^1 of $f_0(x_0^0, \pi/3)$.

2.9 Additional topics

2.9.1 Covering spaces

Definition 112 (Covering map)**.** Let E and B be topological spaces. A *covering map* $p : E \to B$ is a continuous surjective map such that each point $b \in B$ admits an open neighborhood U of b for which the preimage $p^{-1}(U)$ is a disjoint collection of open sets in E, each of which is mapped homeomorphically onto U via p. One then says that U is *evenly covered* by p, and that E is a *covering space* for B; see, for example, [Ma, Chap. 5].

Definition 113 (Branched (or ramified) cover)**.** Let E and B be topological spaces. A continuous map $p : E \to B$ is a *branched cover* of B if for all but a finite number of points of B, p is a covering map of E onto B. The set of points of B that are not evenly covered by p is called the *branch locus* (or *set of ramification points*).

Example 114 (The map $p : \mathbb{C} \to \mathbb{C}$)**.** The map $p : \mathbb{C} \to \mathbb{C}$, given by $p(z) = z^2$, is a branched covering of $\mathbb{C}$, with branch locus $\{0\}$. Hence, it is certainly not a covering map. On the other hand, $p : \mathbb{C} - \{0\} \to \mathbb{C} - \{0\}$, given by the same expression $p(z) = z^2$, is a covering map, since it is *locally trivial*: indeed, each nonzero complex number z in the target space has an open neighborhood U such that p, restricted to $p^{-1}(U)$, is equivalent to the projection onto $U \times \{+, -\}$.

2.9.2 Inverse limit sequence and inverse limit

Inverse limits of various topological (algebraic, or geometric) objects will play an important role in this paper. Hence, since such a notion may not be familiar to all readers, it may be helpful to recall some basic facts pertaining to this subject. Further information can be found, for example, in [HoYo] and, in a more general context, in [Bo, McL].

We discuss the inverse limit in the context of the category Set. The objects are sets and the morphisms are set maps.[7] Consider a partially ordered collection of sets $\{X_j\}_{j=1}^{\infty}$, with each X_j equipped (for each integer $n \geq 1$) with a map $F_n : \prod_{j=1}^{\infty} X_j \to X_n$ defined by $F_n((x_j)_{j=1}^{\infty}) = x_n$. If for every n and every $m \leq n$, where $m, n \in \mathbb{N}^* := \{1, 2, ...\}$, we define the map $F_{nm} : X_n \to X_m$ by $F_{nm}(x_n) = x_m$, then the set

$$\varprojlim X_j := \left\{ (x_j)_{j=1}^{\infty} \in \prod_{j=1}^{\infty} X_j | F_{nm}(x_n) = x_m, \text{ for all } m \leq n \right\} \tag{2.36}$$

exists and is unique; it is called the *inverse limit* of the *inverse limit sequence* $\{X_j, F_j\}_{j=1}^{\infty}$. Furthermore, the maps F_{nm} are called the *transition maps* of the inverse limit system.

Naturally, if we work instead within the category of Topological Spaces, then all the maps involved should be morphisms of that same category (and hence, here, continuous maps).

We next give an example of the inverse limit of a particular collection of sets (or rather, of topological spaces).

Example 115 (The ternary Cantor set $\mathscr{C}$)**.** Recall that a ternary number is a number expressed in terms of a base-3 number system, say, in terms of the characters (or symbols)

[7] Discussing the inverse limit in the context of Set is purely a formality that allows us to speak in more concrete terms and utilize important existences and uniqueness properties of the inverse limit in the category Set. See [Bo, McL] for a general discussion.

$\{0, 1, 2\}$.[8] For example, 1102 in base-3 is actually the number 38 in the base-10 number system. When one constructs the Cantor set $\mathscr{C}$, one may do so by removing middle thirds (open intervals) of successive approximations.[9] Let $n \geq 1$. In removing the middle third from an interval of length $1/3^{n-1}$, we are essentially producing a left third and a right third interval of length $1/3^n$. As such, we can label the left third as 0 and the right third as 2. One quickly sees that the Cantor set $\mathscr{C}$ is a collection of infinite words written entirely in terms of 0's and 2's. There are no ternary numbers in $\mathscr{C}$ whose address contains 1, because that would mean that we did not remove a middle third from some interval in the construction process.[10] More importantly, our labeling system described above enables us to determine particular ternary numbers.

Let $\mathscr{C}_n$ be the nth prefractal approximation of the Cantor set. In the context of the inverse limit construction and addressing system above, $\mathscr{C}_n$ is the collection of 2^n points, each having an address given by a finite ternary expansion of length n and each never containing the character 1. On the other hand, in the context of the geometric construction detailed in the previous paragraph, $\mathscr{C}_n$ consists of 2^n compact intervals of

[8] Any three symbols suffice. One can, and we do so in §2.7.2, represent elements of the Cantor set in terms of the characters $\{l, c, r\}$, where such characters stand for *left*, *center* and *right*, respectively. In general, any element in the unit interval I can be represented by a finite or an infinite sequence expressed in terms of an alphabet consisting of the characters l, c, r.

[9] This is not the only way to construct the ternary Cantor set, but most, if not all, methods amount essentially to the same process.

[10] For the endpoints of the deleted intervals (also called 'ternary points' and necessarily of the form $p/3^q$, for p, q nonnegative integers with $p \leq 3^q$ and p not divisible by 3, when we restrict our attention to the unit interval I), the resulting address is not unique and may contain 1's, although only finitely many 1's. For example, one may represent 1/3 by the finite ternary expansion 0.1 or by the infinite ternary expansion $0.0\overline{2}$, where the overbar indicates that 2 is repeated ad infinitum. We adhere to the convention that such ternary numbers are always represented by infinite expansions given in terms of only 0's and 2's.

length $1/3^n$ (i.e., of 'scale n'). Furthermore, each address in $\mathscr{C}_n$ may be thought of as an address of a particular segment that remains after removing 2^{n-1} segments from the unit interval I; see Figure 2.7.[11] We take the more geometric interpretation as the definition of $\mathscr{C}_n$; in that case, the sequence $\{\mathscr{C}_n\}_{n=1}^{\infty}$ of prefractal approximations converges to the Cantor set $\mathscr{C}$. That is, $\mathscr{C}_n \to \mathscr{C}$ as $n \to \infty$, in the sense of the Hausdorff metric. (Also, more simply, $\mathscr{C}_n$ is monotonically decreasing and $\mathscr{C} = \bigcap_{n=1}^{\infty} \mathscr{C}_n$.) We will next discuss another way in which $\mathscr{C}$ can be viewed as the 'limit' of $\{\mathscr{C}_n\}_{n=1}^{\infty}$.

If for positive integers $m \leq n$, we now define the transition map $\tau_{nm} : \mathscr{C}_n \to \mathscr{C}_m$ as the truncation map that truncates addresses of segments in the prefractal approximations $\mathscr{C}_n$ to addresses of segments in the prefractal approximation $\mathscr{C}_m$ by simply removing the last $n-m$ characters from the address, then we form an inverse limit of the prefractal approximations $\mathscr{C}_n$ that is exactly the Cantor set $\mathscr{C}$:

$$\mathscr{C} = \varprojlim \mathscr{C}_j = \left\{ (c_j)_{j=1}^{\infty} \in \prod_{j=1}^{\infty} \mathscr{C}_j | \tau_{nm}(c_n) = c_m \text{ for all } m \leq n \right\}. \tag{2.37}$$

More background information about inverse limits is provided, for example, in [HoYo, §2–14 and §2–15], where the following well-known theorems can be found; see [HoYo], Theorems 2–95 and 2–97, along with Corollaries 2–98 and 2–99. (All the topological spaces considered in Theorems 116, 117 and 118 below are implicitly assumed to be metrizable. Moreover, by the "Cantor set", we mean the classic ternary Cantor set discussed in Example 115 just above.)

Theorem 116. The inverse limit of finite sets is a compact and totally disconnected set. Conversely, any such topological space is homeomorhpic to an inverse limit of finite sets.

[11] Such a construction is called *construction by tremas*, where one removes segments ad infinitum, thereby producing the fractal set. At each stage, middle thirds are removed from the remaining segments, thereby producing the Cantor set in the limit.

Theorem 117. Any compact and totally disconnected space is homeomorphic to a (closed) subset of the Cantor set.

Theorem 118. Any two totally disconnected and perfect[12] compact spaces are homeomorphic to one another (and hence also to the Cantor set).

Remark 119. In the literature on dynamical systems, it is common to use the term *topological Cantor set* to refer to a totally disconnected and perfect compact space (i.e., to a metrizable space that is homeomorphic to the Cantor set).

Remark 120. In §5, we will show that what we will be referring to as the *footprints* of the *primary piecewise Fagnano orbit* and, in general, *piecewise Fagnano orbits* of the Koch snowflake billiard $\Omega(KS)$ are, in fact, topological Cantor sets (see Theorem 181). Moreover, the footprint of the primary piecewise Fagnano orbit will be the analog of the classic ternary Cantor set, and as a subset of the Koch snowflake KS, the union of the footprints of every piecewise Fagnano orbit will constitute a subset of what we will call the *elusive limit points* of the Koch snowflake KS.

[12]Recall that a subset of a topological space is called *perfect* if it is closed and contains no isolated points. Hence, a compact space is perfect if it has no isolated points or equivalently, if each of its points is a limit point.

Chapter 3

The Koch snowflake prefractal flat surface $\mathcal{S}(KS_n)$

We denote the flat surface M as constructed from a particular rational billiard $\Omega(P)$ by $\mathcal{S}(P)$. In particular, $\mathcal{S}(KS_n)$ is the flat surface associated with the prefractal billiard $\Omega(KS_n)$. The flat surfaces $\mathcal{S}(KS_n)$, $n = 1, 2, 3$, are given in Figure 3.1. For each billiard $\Omega(KS_n)$, the group of symmetries D_N, where $N = \text{lcm}\{q_j\}_{j=1}^{3\cdot 4^n}$ (that is, the second component in the product $\Omega(KS_n) \times D_N$) is the dihedral group D_3, and thus is independent of n. From this, we deduce that for any $n \geq 0$, there are six copies of the prefractal billiard table $\Omega(KS_n)$ (with sides appropriately identified) used in the construction of the associated flat surface $\mathcal{S}(KS_n) := (\Omega(KS_n) \times D_3)/\sim$; see Figure 3.1. We refer the reader back to §2.5.1 for a general discussion of flat surfaces and the associated conical singularities.

Remark 121. For the remainder of the paper, when we say that a regular polygon is of *scale* n, we mean that the side length of the regular polygon is $1/3^n$. For example, an equilateral triangle of scale n is one for which the side length is $1/3^n$.

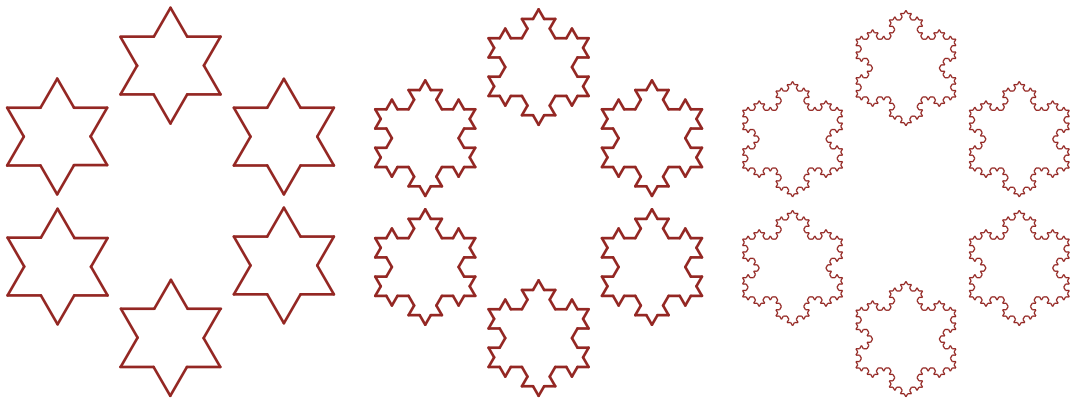

Figure 3.1: The flat surfaces $\mathcal{S}(KS_i)$, $i = 1, 2, 3$. Note that the proper identification is not shown in the figures above. Given the arrangement of the six copies of KS_n, one then identifies opposite and parallel sides to make the proper identification that results in a geodesic flow that is dynamically equivalent with the billiard flow on the associated billiards $\Omega(KS_1), \Omega(KS_2), \Omega(KS_3)$.

3.1 The geometry and topology of $\mathcal{S}(KS_n)$

The flat surface $\mathcal{S}(KS_n)$ is a surface with both types of conic singularities: removable and nonremovable. In constructing the flat surface $\mathcal{S}(KS_n)$ via $\Omega(KS_n)$, we see that the nonremovable conic singularities correspond to corners with obtuse angles of $\Omega(KS_n)$ and removable singularities correspond to corners with acute angles (both measured relative to the interior of $\Omega(KS_n)$). Since for every $n \geq 1$, the measure of every obtuse corner is the same (specifically, $4\pi/3$ radians), it follows that the conic angle of a nonremovable singularity is 8π. For the same reason, every corner with an acute angle gives rise to a removable singularity with conic angle 2π (which, in fact, is a defining characteristic of a removable singularity).

The genus of the surface $\mathcal{S}(KS_n)$ can be calculated using the standard formula for the Euler characteristic of polyhedra, namely

$$\chi = V - E + F, \tag{3.1}$$

where V is the number of vertices, E is the number of edges and F the number of faces

of the polyhedra. Such a formula holds for planar graphs.

Let B be a regular k-gon and $\Omega(B)$ be the rational billiard. If the measure of the angle formed by vertex V_j is $p_j/q_j\pi$ and $N = \text{lcm}\{q_j\}$, then the characteristic of a flat surface constructed from the rational billiard $\Omega(B)$ is

$$\chi = N\sum_{j=1}^{k}\frac{1}{q_j} - Nk + 2N. \tag{3.2}$$

Since the $\chi = 2 - 2g$, g being the genus of the surface $\mathcal{S}(B)$, we see that the genus g of the flat surface $\mathcal{S}(B)$ is[1]

$$g = 1 + \frac{N}{2}\left(k - 2 - \sum_{j=1}^{k}\frac{1}{q_j}\right). \tag{3.3}$$

The prefractal billiard $\Omega(KS_n)$ has $3 \cdot 4^n$ many sides and as many vertices. Therefore, the genus g_n of $\mathcal{S}(KS_n)$ is given by

$$g_n = 3 \cdot 4^n - 2. \tag{3.4}$$

One easily sees that the sequence of genuses $\{g_n\}_{n=0}^{\infty}$ increases without bound.

3.2 $\mathcal{S}(KS_n)$ is a branched cover of $\mathcal{S}(KS_0)$

Taking as inspiration the results and methods of Gutkin and Judge in [GtkJu1] and [GtkJu2], we now show that for each $n \geq 1$, the flat surface $\mathcal{S}(KS_n)$ is a branched cover of the hexagonal torus $\mathcal{S}(KS_0)$; see Figure 3.3. To such end, we establish several results culminating in this fact.

Lemma 122. Let $n \in \mathbb{N}$. Then, for any positive integer $k \geq n$, $\mathcal{S}(KS_n)$ can be tiled by equilateral triangles of scale k.

[1] See [HuSc] for a detailed description of how to calculate the genus of a surface that arises from a rational billiard table.

Proof. This follows from the construction of the Koch snowflake. We note that each triangle of scale n, denoted Δ_n, can be tiled by 9^{k-n} triangles of scale $k \geq n$; see Figure 3.2 for the case when $k = n + 1$. Note that $\mathcal{S}(KS_n) = (\Omega(KS_n) \times D_3)/\sim$ and that $\Omega(KS_n)$ is constructed from $\Omega(KS_{n-1})$ by gluing a copy of Δ_n to every side $s_{n-1,k}$ at the middle third of $s_{n-1,k}$. Since every triangle Δ_{n-1} tiling $\Omega(KS_{n-1})$ can also be tiled by Δ_n, it follows that $\Omega(KS_n)$ is tiled by Δ_n. So, $\mathcal{S}(KS_n)$ can be tiled by equilateral triangles of scale k, for every $k \geq n$. ■

In the sequel, given a bounded set $A \subseteq \mathbb{R}^2$, we will write that "A can be tiled by H_n" in order to indicate that A can be tiled by finitely many copies of hexagonal tiles H_n of scale n.

Proposition 123. Let $n \in \mathbb{N}$. Then the flat surface $\mathcal{S}(KS_n)$ can be tiled by H_k, for all $k \geq n + 1$, in such a way that each conic singularity is at the center of some tile H_k.

Proof. We see in Figure 3.4 that $\mathcal{S}(KS_1)$ can be tiled by H_2 such that each conic singularity is at the center of some tile H_2. Each H_2 is tiled by six equilateral triangles Δ_2. As seen in the proof of Lemma 122, each Δ_2 is tiled by nine Δ_3 such that six of these triangles form a hexagonal tile H_3; see Figure 3.2. At the center of H_2 is a copy of H_3. Hence, each conic singularity remains at the center of some tile H_3; see Figure 3.5. Continuing in this fashion, we see that for each $k \geq 2$, H_k tiles $\mathcal{S}(KS_2)$ in such a way that each conic singularity is at the center of some H_k.

Suppose there exists $N \in \mathbb{N}$ such that, for every $n \leq N$, $\mathcal{S}(KS_n)$ can be tiled by H_k, for every $k \geq n+1$. In particular, $\mathcal{S}(KS_N)$ can be tiled by H_k, for every $k \geq N+1$. We then have that, for every $k \geq N + 2$, $\mathcal{S}(KS_N)$ can be tiled by H_k. By Lemma 122, Δ_{N+1} tiles $\mathcal{S}(KS_{N+1})$. Each triangular region Δ_{N+1} in $\mathcal{S}(KS_{N+1})$ not in $\mathcal{S}(KS_N)$ is tiled by nine triangles Δ_{N+2} in such a way that six Δ_{N+2} comprise a tile H_{N+2}. Continuing in this fashion, we see that each Δ_k contributes to a hexagonal tile H_k (as

part of the embedded tiling) in such a way that each conic singularity is at the center of some hexagonal tile H_k.

■

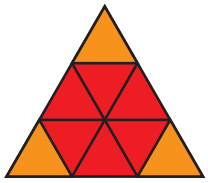

Figure 3.2: We see that Δ_n is tiled by nine copies of Δ_{n+1}, six of which form a hexagonal tile H_{n+1} in the center.

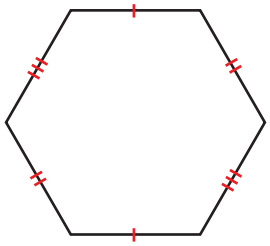

Figure 3.3: The hexagonal torus $\mathcal{S}(KS_0)$. (As usual, similarly marked sides are identified.) It should be noted that $\mathcal{S}(KS_0)$ is topologically (but not metrically) equivalent to the flat square torus.

Theorem 124. For every $n \in \mathbb{N}$, the prefractal Koch snowflake flat surface $\mathcal{S}(KS_n)$ is a branched cover of the prefractal Koch snowflake flat surface $\mathcal{S}(KS_0)$, which is the hexagonal torus. Such a covering map $p_n : \mathcal{S}(KS_n) \to \mathcal{S}(KS_0)$ is given by suitably defined translations on $\mathcal{S}(KS_n)$.

Proof. The center point x_0 of the flat hexagonal torus $\mathcal{S}(KS_0)$ is a branched locus of the cover $\mathcal{S}(KS_n)$ when $\mathcal{S}(KS_n)$ is tiled by H_{n+1} as described in Proposition 123. This follows from the fact that every nonremovable conic singularity of $\mathcal{S}(KS_n)$ is at the center of four hexagonal tiles. Specifically, this means that this center point x_0 is

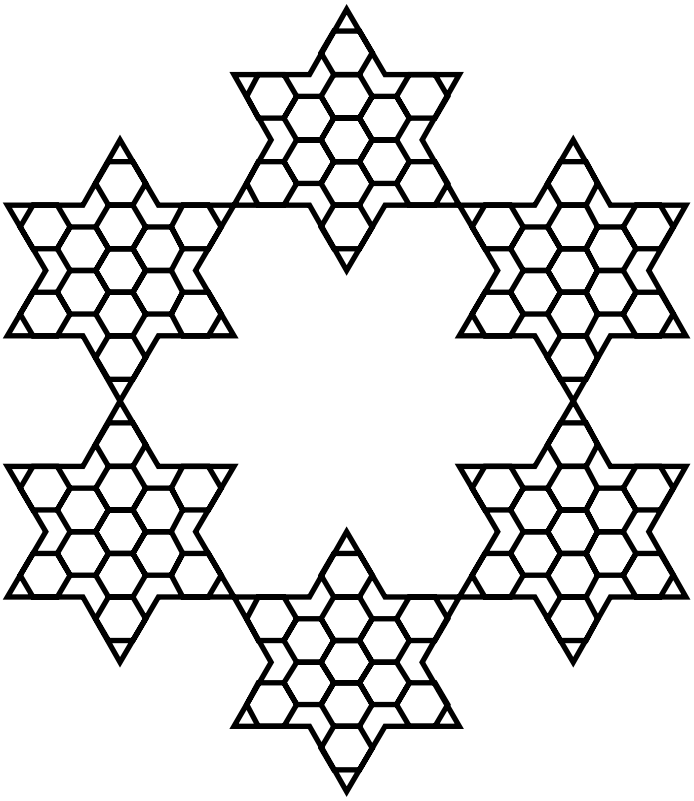

Figure 3.4: Tiling the flat surface $\mathcal{S}(KS_1)$ by hexagonal tiles H_2. We note that the conic singularities (both removable and nonremovable) are at the center of hexagonal tiles.

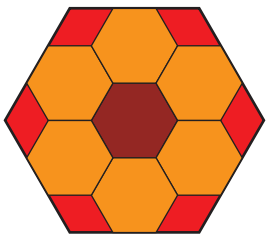

Figure 3.5: Six triangles Δ_n tile H_n. The hexagonal tile H_n is tiled by seven tiles H_{n+1} with six rhombic tiles.

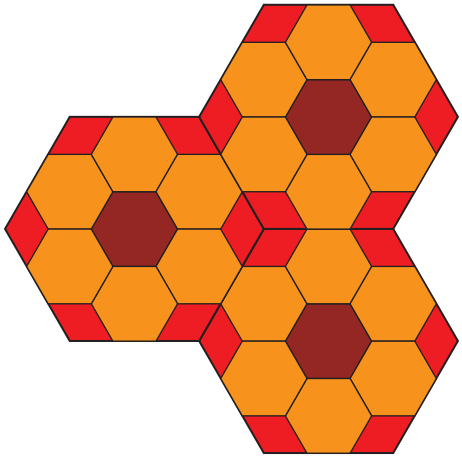

Figure 3.6: We see that three hexagonal tiles H_n tiled as in Figure 3.5 can be arranged so that a rhombic tile from each all combine to form another hexagonal tile H_{n+1}.

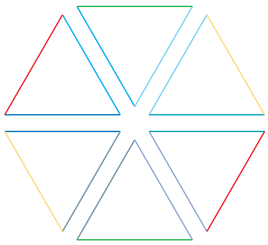

Figure 3.7: Six equilateral triangles glued together correctly constitute a hexagonal torus. We show the surface in an exploded view so as to emphasize the fact that there are six copies of the equilateral triangle embedded in the surface. Prior to identifying properly, this figure constitutes $\Omega(KS_0) \times D_3$.

not evenly covered by the covering map $p_n : \mathcal{S}(KS_n) \to \mathcal{S}(KS_0)$ determined by suitable translations of hexagonal tiles H_{n+1} on $\mathcal{S}(KS_n)$. Any other point in $\mathcal{S}(KS_0)$ is evenly covered since every element in the fiber $p_n^{-1}(z)$, $z \neq x_0$, has a conic angle of 2π. ■

3.3 The Veech group $\Gamma(KS_n)$

We now discuss the Veech group $\Gamma(KS_n)$ of the flat surface $\mathcal{S}(KS_n)$. The Veech group $\Gamma(KS_0)$ corresponding to the flat hexagonal torus is isomorphic with the Veech group for the flat square torus described in Example 80. The following result, due to E.

Gutkin and C. Judge ([GtkJu2]), is used in showing the Veech dichotomy holds for the prefractal flat surfaces $\mathcal{S}(KS_n)$.

Definition 125 (Arithmetic)**.** A flat surface $\mathcal{S}(B)$ is called *arithmetic* if it is the cover of a singly punctured torus.

Theorem 126 (E. Gutkin and C. Judge)**.** Let B be a rational polygon. The Veech dichotomy holds for the flat surface $\mathcal{S}(B)$ if and only if the flat surface is arithmetic.

We have shown that, for every $n \geq 0$, the flat surface $\mathcal{S}(KS_n)$ is indeed a branched cover of the torus (or a cover of the singly punctured torus). Therefore, the Veech dichotomy holds for $\mathcal{S}(KS_n)$, for every $n \geq 0$. Therefore, we have that the footprint $\mathcal{F}_n(x_n^0, \theta_n^0)$ is either finite or uniformly distributed in the boundary KS_n; or, we have that the path traversed by the billiard is closed or uniformly distributed in the billiard $\Omega(KS_n)$.

A result quoted in [Swz] and originally due to W. Veech [Ve4] states that a flat surface $\mathcal{S}(B)$, where B is a rational polygon, has a Veech group that is at most countably infinite if it contains at least one nonremovable singularity. We state the result below and sketch a proof of it for completeness.

Proposition 127. Let B be a rational polygon. If the corresponding flat surface $\mathcal{S}(B)$ has at least one (non-removable) conic singularity, then the corresponding Veech group is a discrete group.

Sketch of the proof of Proposition 127. Consider a rational polygon B and the associated flat surface $\mathcal{S}(B)$. We suppose $\mathcal{S}(B)$ has at least one non-removable conic singularity. Each singularity is contained in a neighborhood not containing any other singularities (there are finitely many nonremovable conic singularities, so it is possible to find a smallest ϵ-neighborhood that works for every cone point). Consider a saddle connection $\overline{PP'}$ connecting cone points P and P' (which may be the same point

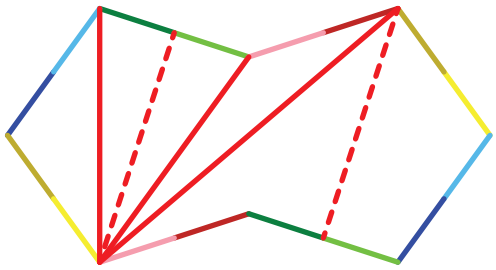

Figure 3.8: An illustration of what constitutes a saddle connection of a flat surface with at least one nonremovable conic singularity.

if there is only one singularity); Figure 3.8 illustrates a few saddle connections of the flat surface previously seen in Figure 2.6. We claim that there is a δ-tubular neighborhood about $\overline{PP'}$ not containing any other saddle connections. If this were not the case, then there would be arbitrarily close saddle connections. This would then imply that there are infinitely many singularities accumulating about P', which is a contradiction of our assumption. ■

For every $n \geq 1$, the flat surface $\mathcal{S}(KS_n)$ contains at least one nonremovable singularity. Hence, the following result.

Proposition 128. For every $n \geq 1$, The Veech group $\Gamma(KS_n)$ is at most countably infinite.

Chapter 4

The Koch snowflake prefractal billiard $\Omega(KS_n)$

4.1 Equivalence of geodesic flow and billiard flow

An orbit $\mathcal{O}_0(x_0^0, \theta_0^0)$ is closed if and only if the vector determined from the initial angle θ_0^0 is rational with respect to the basis $\{(1,0),(1/2,\sqrt{3}/2)\}$. In §3.2, we saw that $\mathcal{S}(KS_n)$ is a branched cover of $\mathcal{S}(KS_0)$. This implies that for every $n \in \mathbb{N}$, a closed geodesic on $\mathcal{S}(KS_n)$ will project down to a closed geodesic on the hexagonal torus $\mathcal{S}(KS_0)$ under the action of the covering map p_n defined in the proof of Theorem 124. Also, a closed geodesic γ on $\mathcal{S}(KS_0)$ lifts to a segment on $\mathcal{S}(KS_n)$ (that is not necessarily closed). However, according to the discussion in §3.2, there exists a positive integer k such that the lift of γ^k is a closed geodesic on $\mathcal{S}(KS_n)$.[1] Since the geodesic flow on $\mathcal{S}(KS_n)$ is dynamically equivalent[2] to the billiard flow on $(\Omega(KS_n) \times S^1)/\sim$,

[1] Recall that the notation γ^k is meant to represent $k-1$ many concatenations of γ with itself: $\gamma * \gamma * ... * \gamma = \gamma^k$

[2] The dynamical equivalence is due to the fact that $\mathcal{S}(KS_n)$ is a branched cover of $\mathcal{S}(KS_0)$, as stated in Theorem 124.

it follows that a direction giving rise to a closed orbit in $\Omega(KS_0)$ is a direction giving rise to a closed orbit in $\Omega(KS_n)$ for every $n \geq 0$, and vice-versa. We may put this more succinctly as periodic directions in $\mathcal{S}(KS_0)$ are exactly the periodic directions in $\mathcal{S}(KS_n)$, and vice-versa. We summarize the above discussion in the following theorem.

Theorem 129. The geodesic flow on $\mathcal{S}(KS_0)$ is closed if and only if for every $n \geq 0$, the geodesic flow on $\mathcal{S}(KS_n)$ is closed. Equivalently, the set of directions for which the discrete billiard flow f_n on $\Omega(KS_n)$ is closed[3] (i.e., regardless of the initial basepoint, a direction for which a geodesic will be closed) is exactly the set of directions for which the billiard flow f_0 is closed on $\Omega(KS_0)$.

Remark 130. If $\{e_1, e_2\}$ is a basis for $\mathbb{R}^2$, then a vector $z \in \mathbb{R}^2$ is called *rational with respect to* $\{e_1, e_2\}$ if $z = ne_1 + me_2$, $n/m \in \mathbb{Q}$ (that is, $n, m \in \mathbb{Z}$, $m \neq 0$). The plane can be tiled by $\Omega(KS_0)$. A more general result appearing in [Gtk2] can be used to show that the collection of directions that give rise to closed orbits of the equilateral triangle billiard $\Omega(KS_0)$ is exactly the set of directions that are rational with respect to the basis $\{e_1, e_2\} = \{(1, 0), (1/2, \sqrt{3}/2)\}$. By Theorem 129, for every $n \geq 0$, the same collection of rational directions (with respect to $\{e_1, e_2\}$) describes the directions for which the billiard flow on $\Omega(KS_n)$ is closed. Likewise, a vector that is irrational with respect to this basis will determine an angle that gives rise to a non-periodic orbit. In the case of $\Omega(KS_n)$, this will be a dense orbit.

Theorem 131. The geodesic flow on $\mathcal{S}(KS_0)$ in a fixed direction is uniquely ergodic if and only if for every $n \geq 0$, the geodesic flow in the same direction on $\mathcal{S}(KS_n)$ is uniquely ergodic. Moreover, the footprint $\mathcal{F}_n(x_n^0, \theta_n^0)$ corresponding to an orbit $\mathcal{O}_n(x_n^0, \theta_n^0)$ is uni-

[3]It should be noted that we are making a slight abuse of notation and language. The flow ϕ_t on the phase space $(\Omega(KS_n) \times S^1)/\sim$ is reduced to the billiard flow f_n on the Poincaré section. Iterates of the billiard map then yield elements of the Poincaré section.

formly distributed[4] in the boundary KS_n if and only if $\mathcal{F}_0(x_0^0, \theta_0^0)$ is uniformly distributed in KS_0.

Proof. Again, $\mathcal{S}(KS_n)$ is a branched cover of $\mathcal{S}(KS_0)$. We can clearly see that the projection of a uniquely ergodic flow on $\mathcal{S}(KS_n)$ will be a uniquely ergodic flow in $\mathcal{S}(KS_0)$, since the Veech dichotomy holds for every prefractal flat surface. If for the same direction θ, the geodesic $\phi_t(x, \theta)$ on $\mathcal{S}(KS_0)$ is dense, then lifting this path to the covering space results in a (not necessarily closed) path. If the lifting of the path results in a saddle connection on $\mathcal{S}(KS_n)$ (by possibly having to continue the lifted path), then the projection of the lift must also be a saddle connection of $\mathcal{S}(KS_0)$, which was not the case to begin with. ■

4.2 Orbits with an initial direction $\pi/3$

The equilateral triangle billiard $\Omega(\Delta)$ is an especially nice rational polygonal billiard. In the direction of $\pi/3$, there are exactly two types of orbits: the Fagnano orbit and everything else. The orbit $\mathcal{O}_0(\overline{c}, \pi/3)$ is called the *Fagnano orbit* of $\Omega(KS_0)$. Fagnano polygons are the shortest inscribed polygons for a given polygon. In the case of the equilateral triangle, the shortest inscribed polygon is a billiard orbit.[5] Deviating to the left or the right of $x_0^0 = \overline{c}$ results in an orbit of the equilateral triangle billiard that is twice the length of the Fagnano orbit. (See [BaUm] for a detailed classification of periodic orbits of $\Omega(KS_0)$, along with Remark 130 above for a more conceptual characterization of those orbits.)

In the context of the Koch snowflake prefractal billiard $\Omega(KS_n)$, we determine

[4] Recall that the definition of uniformly distributed sequences was given in Definition 61.

[5] In fact, given any triangle billiard, the Fagnano polygon is a billiard orbit of that particular triangle billiard. See [Ta2] for a more detailed discussion of this fact.

three types of periodic orbits, each with an initial direction of $\pi/3$: *piecewise Fagnano orbits*, *Cantor orbits* and *approximate piecewise Fagnano orbits*. Each are so named for the nature of the ternary representation of their respective initial basepoints x_n^0.

We can show that every orbit $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ passes through the interior of $\Omega(KS_0)$. By means of the local and global symmetry of the prefractal billiard table $\Omega(KS_n)$, as shown in Figure 4.3 and discussed in the corresponding caption, one can deduce that there must be a basepoint $y_n^{k_n}$ of the orbit $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ that, after colliding with the boundary KS_n at $y_n^{k_n}$, the billiard ball next traverses the interior of $\Omega(KS_n) \cap \Omega(KS_0)$ in the direction of $\pi/3$. Then, if we define x_n^0 by $x_n^0 := y_n^{k_n}$, the orbit $\mathscr{O}_n(x_n^0, \pi/3)$ contains $(y_n^0, \theta(\pi/3))$ and $\mathscr{O}_n(x_n^0, \pi/3) = \mathscr{O}_n(y_n^0, \theta(\pi/3))$.

Remark 132. Much of what follows relies heavily on the fact that an initial direction of $\theta(\pi/3)$ is one of the global axes of symmetry. This is particularly true in Propositions 133 and 163.

These facts and Remark 132 are used in the following proposition and throughout the remainder of this chapter.

Proposition 133. Let $n \geq 0$ and $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ a periodic orbit. Then there exists x_n^0 such that

$$\mathscr{O}_n(y_n^0, \theta(\pi/3)) = \mathscr{O}_n(x_n^0, \pi/3) \tag{4.1}$$

and a segment joining x_n^0 and x_n^1 intersects the unit interval I at a point we label as x_0^0.

Proof. A billiard ball with an orbit $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ will make at most two collisions in a cell $C_{n,k}$ before exiting that cell; see Figure 4.5. Then, the billiard ball traverses the interior of some previous approximation before entering into another cell $C_{n,k'}$. For every cell $C_{m,k}$ visited by the billiard ball, at most two cells $C_{n,k}$ and $C_{n,k'}$,

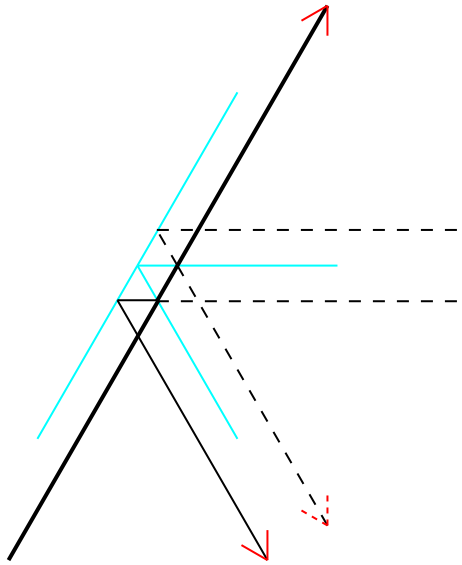

Figure 4.1: Depicted in this figure is a classical tool for analyzing billiard orbits: *unfolding the orbit*. We consider a billiard orbit's trajectory entering a cell $C_{n,k}$ at an angle of $\pi/3$. As expected, the billiard ball must exit the cell after two reflections and at an angle of $-\pi/3$. We verify this by considering the unfolded trajectory in the plane and noticing that the trajectory continues on unimpeded as it passes through a reflected copy of the opening. (See Figure 4.2 for a generalization of this discussion to the case of two collisions with the boundary.)

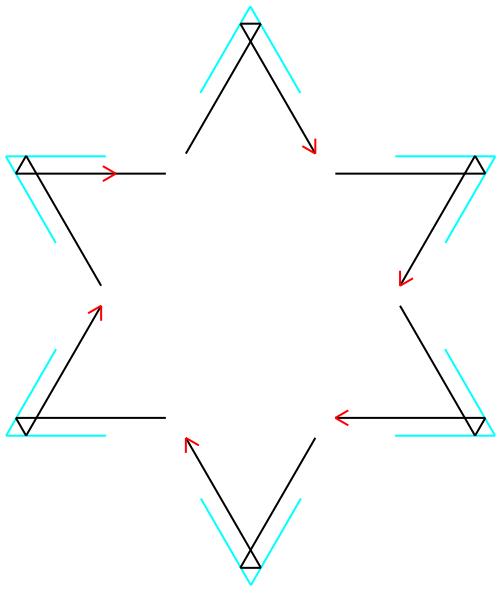

Figure 4.2: This figure generalizes what we have seen in Figure 4.1. An orbit's trajectory that enters a cell $C_{n,k}$ parallel to a side of $C_{n,k}$ will exit after two collisions.

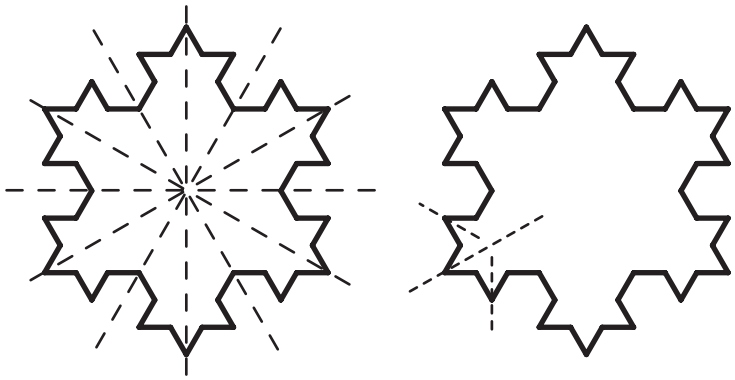

Figure 4.3: For $n \geq 1$, the symmetry group of KS_n is the dihedral group D_6 (see §2.5.1). Local symmetry is seen at the level of a cell $C_{n,k}$ of $\Omega(KS_n)$. The symmetry of interest is symmetry with respect to the angle bisector of the acute angle of the cell $C_{n,k}$.

$n \geq m$, can be visited by the billiard ball; see Figure 4.1. Exhausting all possibilities, a global symmetry is used to determine the next position of the billiard ball in the boundary KS_n. As such, the billiard ball must pass through the interior of $\Omega(KS_n) \cap \Omega(KS_0)$ to reach the point (which is the mirror image of some point through the global symmetry used in emulating the Law of Reflection). ■

4.2.1 Piecewise Fagnano orbits

We now discuss a generalization of the Fagnano orbit for $\Omega(KS_n)$.

Definition 134 (Piecewise Fagnano orbit)**.** Let $C_{n,k}$ be a cell of $\Omega(KS_n)$. Let x_n^0 be an element of one of the sides of $C_{n,k}$. If x_n^0 has a ternary representation consisting solely of infinitely many c's (i.e., a ternary representation $[c, \emptyset]$), then we call the orbit $\mathscr{O}_n(x_n^0, \theta(\pi/3))$ a *piecewise Fagnano orbit* of $\Omega(KS_n)$.

A piecewise Fagnano orbit $\mathscr{O}_n(x_n^0, \theta(\pi/3))$, where $x_n^0 = \overline{c}$ relative to the side $s_{n,k}$ containing x_n^0, is named as such for the fact that one can view this orbit as the result of appending scaled copies of the Fagnano orbit of $\Omega(KS_0)$ to every basepoint of an orbit $\mathscr{O}_{n-1}(y_{n-1}^0, \theta(\pi/3))$, where y_{n-1}^0 is the midpoint of a side $s_{n-1,k}$ of $\Omega(KS_{n-1})$;

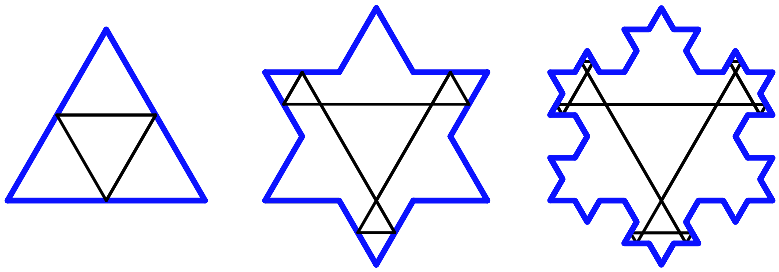

Figure 4.4: The first orbit is the Fagnano orbit of the equilateral triangle billiard $\Omega(KS_0)$. The second orbit is s *piecewise Fagnano* orbit of $\Omega(KS_1)$ and the third orbit is a piecewise Fagnano orbit of $\Omega(KS_2)$.

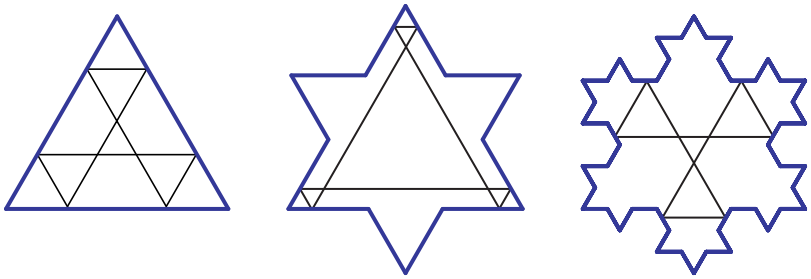

Figure 4.5: Other examples of orbits in the direction $\pi/3$.

see Figure 4.4.

By Proposition 133, we know that for every piecewise Fagnano orbit $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ of $\Omega(KS_n)$, there exists x_n^0 such that $\mathscr{O}_n(x_n^0, \pi/3) = \mathscr{O}_n(y_n^0, \theta(\pi/3))$. Later, we will see that for every orbit $\mathscr{O}_n(y_n^0, \theta(\pi/3))$, there exists a unique element $x_0^0 \in I$ (the unit interval $[0, 1]$ viewed as the base of $\Delta = KS_0$) that effectively determines a special sequence of orbits to which $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ belongs. This is the same value x_0^0 determined by Proposition 133. The relationship between the orbit $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ and a particular orbit $\mathscr{O}_0(x_0^0, \pi/3)$ will be referred to as *compatibility*. We defer a full treatment of this concept until §4.4.

4.2.2 Cantor orbits

It is possible that an orbit $\mathscr{O}_n(x_n^0, \pi/3)$ of $\Omega(KS_n)$, after a finite amount of time (that is, finitely many iterates of the prefractal construction of KS), will cease to be an orbit of subsequent approximations. The particular prefractal approximation at which an orbit $\mathscr{O}_n(x_n^0, \pi/3)$ ceases to be an orbit of $\Omega(KS_m)$, $m \geq n$, is determined from the ternary representation of the point x_n^0. Indeed, if the next occurrence of the character c in the ternary representation is in the $(k+n)$th position, then $\mathscr{O}_n(x_n^0, \pi/3)$ ceases to be an orbit of $\Omega(KS_{k+n})$.

A Cantor orbit is an orbit with exactly the opposite behavior; it is always an orbit of every subsequent prefractal approximation.

Definition 135 (Cantor orbit)**.** Let $s_{n,k}$ be a side of $\Omega(KS_n)$. Let x_n^0 be an element of $s_{n,k}$. If x_n^0 has a ternary representation $[lr, \emptyset]$ (i.e., x_n^0 is a Cantor-point of some side $s_{n,k}$ of $\Omega(KS_n)$), then we call the orbit $\mathscr{O}_n(x_n^0, \theta(\pi/3))$ a *Cantor orbit* of $\Omega(KS_n)$.

In terms of the addressing system on KS, the elements of the footprint of a Cantor orbit have a particular representation. As discussed in §2.8.4, a Cantor-point of a side $s_{n,k}$ is a point with an address $[02, 1345] \oplus [24, 0135] \oplus [04, 1235]$, where each type of address is determined by examining the character that preceded the infinite string of characters not containing 1, 3 or 5. More importantly, a Cantor orbit of $\Omega(KS_n)$ is a Cantor orbit for every subsequent prefractal approximation. That is, such an orbit remains fixed and well defined for every subsequent approximation. This follows from the fact that for every side $s_{n,k}$ of $\Omega(KS_n)$ and for every Cantor-point x of $s_{n,k}$ there exists an open, connected, 1-dimensional neighborhood about x. That is, the Law of Reflection holds at x for every subsequent finite approximation; see Figure 4.6.

We close this section by remarking on the fact that for every Cantor orbit

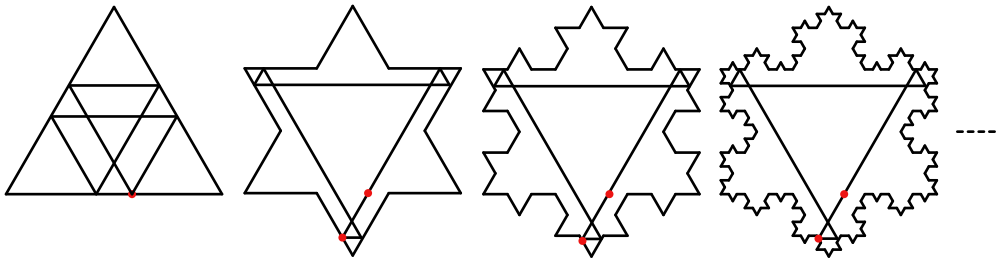

Figure 4.6: While the first orbit is not a Cantor orbit, the second orbit is. We note that each orbit is related to the next in a very particular way that will be further elaborated upon in the subsequent sections. However, we only mention that after the first level approximation, an orbit remains fixed for every subsequent prefractal approximation.

$\mathscr{O}_n(y_n^0, \theta(\pi/3))$ there exists $x_0^0 \in I$ such that $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ is in a particular sequence of Cantor orbits determined by $(x_0^0, \pi/3)$. Such a sequence is discussed in §4.4. In addition, there exists $y_n^{k_n} =: x_n^0$ such that $\mathscr{O}_n(x_n^0, \pi/3) = \mathscr{O}_n(y_n^0, \theta(\pi/3))$.

4.2.3 Approximate piecewise Fagnano orbits

Definition 136. Let $s_{n,k}$ be a side of $\Omega(KS_n)$. Let y_n^0 be an element of $s_{n,k}$. If y_n^0 has a ternary representation $[lc, r] \oplus [cr, l] \oplus [lcr, \emptyset]$, then we call the orbit $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ an *approximate piecewise Fagnano orbit* or $\Omega(KS_n)$.

Remark 137. We will later see that an approximate piecewise Fagnano orbit is an example of what we will call a periodic hybrid orbit.

Such an orbit is so named for the fact that it is *approximately* a piecewise Fagnano orbit. The nature of the ternary representation of the elements of the corresponding footprint would indicate that such points do not correspond to a midpoint of any side of any future approximation, but also never correspond to any point that would be considered a Cantor-point nor a corner of any future finite approximation. Hence, such an orbit would not remain fixed and no longer remain a valid orbit for some future approximation. Moreover, one would not be able to construct some orbit of some future approximation by simply appending scaled copies of a Fagnano orbit to each element of

the corresponding footprint as we did in the case of a piecewise Fagnano orbit. Although, visually, it would appear that scaled copies of Fagnano orbits are being appended to the elements of the footprint of the orbit in some previous approximation.

Finally, if $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ is an approximate piecewise Fagnano orbit, we remark that there exists $x_0^0 \in I$ such that $(x_0^0, \pi/3)$ determines a particular sequence of approximate piecewise Fagnano orbits to which $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ belongs. In addition, there exists $y_n^{k_n} =: x_n^0$ such that $\mathscr{O}_n(x_n^0, \pi/3) = \mathscr{O}_n(y_n^0, \theta(\pi/3))$.

4.3 Orbits with an initial direction not $\pi/3$

4.3.1 Hybrid orbits

The particular description of orbits with an initial direction of $\pi/3$ relied heavily on the fact that the ternary representations of the basepoints were all of the same type. The fact that every element $x_n^{k_n}$ of the footprint of an orbit $\mathscr{O}_n(x_n^0, \pi/3)$ had a ternary representation of the same nature as the initial basepoint x_n^0 was largely due to the fact that $\pi/3$ was one of the axes of symmetry of the prefractal billiard.

A hybrid orbit of a prefractal billiard is an orbit of $\Omega(KS_n)$ for which the ternary representation of the elements of the corresponding footprint do not have to all be of the same type. For example, elements of the footprint of a periodic hybrid orbit do not all have to consist of just infinitely many c's, but perhaps a mixture of types of points. We formally define a *hybrid orbit* below.

Definition 138 (Hybrid orbit)**.** Let $\mathscr{O}_n(x_n^0, \theta_n^0)$ be an orbit of $\Omega(KS_n)$. If all but at most two basepoints $x_n^{k_n} \in \mathcal{F}_n(x_n^0, \theta_n^0)$ have ternary representations (determined with respect to the side $s_{n,k}$ on which each point resides) $[c, lr] \oplus [cl, r] \oplus [cr, l] \oplus [lcr, \emptyset] \oplus [lr, \emptyset]$, then we call $\mathscr{O}_n(x_n^0, \theta_n^0)$ a *hybrid orbit* of $\Omega(KS_n)$.

Remark 139. It is clear from the definition of hybrid orbit that a Cantor orbit is a hybrid orbit. However, we focus our attention on the case where the basepoints of a hybrid orbit have at least two of the types of representations listed in Definition 138

Definition 140 (Closed hybrid orbit)**.** If $\mathscr{O}_n(x_n^0, \theta_n^0)$ is a hybrid orbit with exactly two basepoints or no basepoints corresponding to corners of $\Omega(KS_n)$, then we call $\mathscr{O}_n(x_n^0, \theta_n^0)$ a *closed hybrid orbit.*

Definition 141 (Periodic hybrid orbit)**.** If $\mathscr{O}_n(x_n^0, \theta_n^0)$ is a hybrid orbit with no basepoints corresponding to corners of $\Omega(KS_n)$, then we call $\mathscr{O}_n(x_n^0, \theta_n^0)$ a *periodic hybrid orbit.*

Definition 142 (Dense hybrid orbit)**.** A hybrid orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ that is dense in $\Omega(KS_n)$ is called a *dense hybrid orbit.*

Proposition 143. If $\mathscr{O}_n(x_n^0, \theta_n^0)$ is a hybrid orbit with at most one basepoint corresponding to a corner of $\Omega(KS_n)$ and $\mathscr{O}_n(x_n^0, \theta_n^0)$ is not closed, then $\mathscr{O}_n(x_n^0, \theta_n^0)$ is a dense hybrid orbit.

Proof. Let $\mathscr{O}_n(x_n^0, \theta_n^0)$ be a hybrid orbit with at most one basepoint corresponding to a corner of $\Omega(KS_n)$ that is also not closed. By the Veech dichotomy (see §2.5.2), such an orbit must then be dense in the billiard table $\Omega(KS_n)$. ■

Theorem 144. If $\mathscr{O}_n(x_n^0, \theta_n^0)$ is a dense orbit of $\Omega(KS_n)$, then $\mathscr{O}_n(x_n^0, \theta_n^0)$ is a dense hybrid orbit.

Proof. Suppose there were two basepoints $x_n^{k_n}$ and $x_n^{k'_n}$ of a dense orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ with ternary representations $[l, cr] \oplus [r, lc]$. Then, there exists $N \geq n$ such that the orbit connects two vertices of two equilateral triangles of scale N tiling $\Omega(KS_n)$. Since this orbit can be unfolded into the corresponding flat surface and then projected down onto the hexagonal torus, such an orbit (or flow line on the flat surface) must be at least a saddle connection of the equilateral triangle billiard. However, such a direction

θ_n^0 should yield a dense billiard flow in $\Omega(KS_0)$, which is not the case. Hence, $x_n^{k_n}$ and $x_n^{k'_n}$ do not both have a ternary representation $[l, cr] \oplus [r, lc]$. Moreover, if any basepoint of $\mathscr{O}_n(x_n^0, \theta_n^0)$ already corresponds to a corner of $\Omega(KS_n)$, then a similar argument shows that no other basepoint may have a ternary representation $[l, cr] \oplus [r, lc]$. Therefore, $\mathscr{O}_n(x_n^0, \theta_n^0)$ is a dense hybrid orbit. ■

Example 145 (Approximate piecewise Fagnano orbits)**.** By definition, an approximate piecewise Fagnano orbit $\mathscr{O}_n(x_n^0, \pi/3)$ of $\Omega(KS_n)$ is a periodic hybrid orbit. In fact, every periodic orbit of $\Omega(KS_n)$ with an initial direction of $\pi/3$ is an example of a periodic hybrid orbit.

Example 146 (Hook orbits)**.** Consider an orbit $\mathscr{O}_0(x_0^0, \theta_0^0)$ where $\theta_0^0 = \pi/6$. This orbit, by itself, is not terribly interesting. However, let us consider an orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ where, for every $n \geq 0$, $x_n^0 = x_0^0$ *and* $\theta_0^0 = \theta_n^0$, as shown in Figure 4.7. Then, the resulting degenerate[6] orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ appears to be hooking into the snowflake. Hence, the name *hook* for this orbit. As we will discuss later, $\mathscr{O}_0(x_0^0, \pi/3)$ and $\mathscr{O}_n(x_n^0, \pi/3)$ are related in a sense *and* such a sequence of related orbits will converge to what we will call a nontrivial polygonal path that has as "end points" two elusive limit points of KS. The element x_n^0 is a Cantor-point and every element of the corresponding footprint has either a ternary representation $[lr, \emptyset] \oplus [c, lr]$. In particular, only the endpoints of segments of this orbit that form right angles with the boundary have ternary representations $[c, lr]$.

Example 147. In Figure 4.8, three periodic hybrid orbits are displayed. What is significant about these three orbits is that they are related in such a way that the initial basepoints of each orbit (with x_0^0 being $\overline{c}$ in the unit interval base of $\Omega(KS_0)$) are collinear in the same initial direction relative to the fixed coordinate system, but each orbit looks drastically different from the next. We will later see in §5.2.1 that there may be reason

[6] A degenerate orbit is one that doubles back on itself.

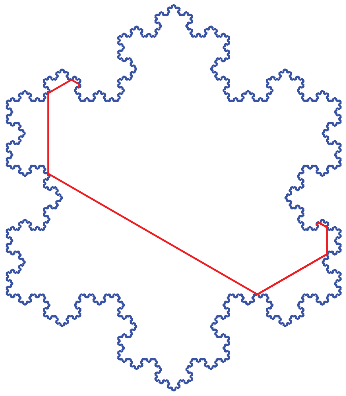

Figure 4.7: An example of a hook orbit.

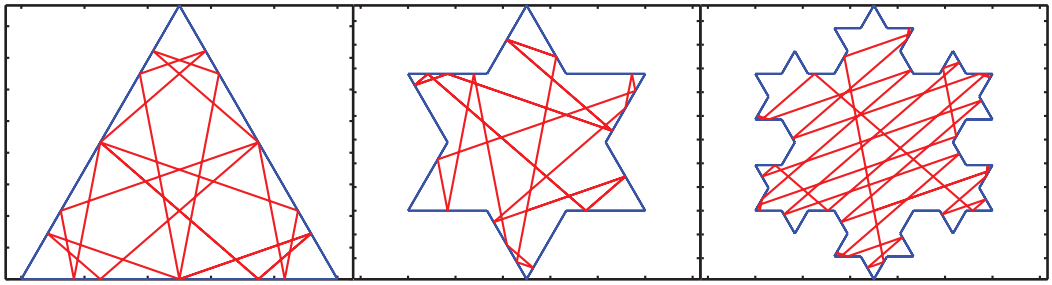

Figure 4.8: Examples of periodic hybrid orbits.

to believe that such orbits converge to a set that we may be able to call a periodic orbit of the Koch snowflake.

Example 148 (A dense hybrid orbit)**.** Consider an orbit of $\Omega(KS_0)$ with an initial condition (x_0^0, θ_0^0) where $x_0^0 = \overline{c}$ and $\tan\theta_0^0 = 1/9$. Then the resulting orbit $\mathscr{O}_0(x_0^0, \theta_0^0)$ is dense in $\Omega(KS_0)$. Figure 4.9 shows an orbit $\mathscr{O}_2(x_2^0, \theta_2^0)$ with $\theta_2^0 = \theta_0^0$ and x_2^0 collinear with x_0^0; the first orbit in the figure consists of 100 iterates, the second orbit consists of 500 iterates, the third orbit consists of 1,000 iterates and the fourth orbit consists of 2,000 iterates. We see in this figure that the orbit will most likely be dense, which is exactly what we expect since the slope of the initial segment is irrational with respect

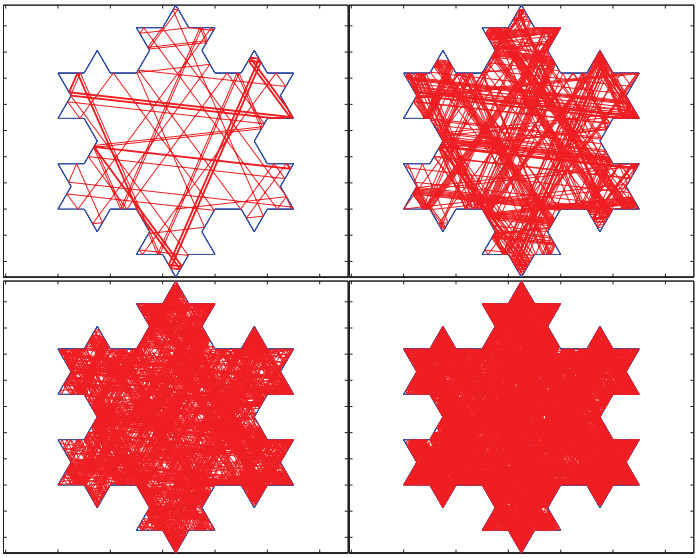

Figure 4.9: The first orbit consists of 100 iterates, the second consists of 500 iterates, the third consists of 1,000 iterates and the fourth consists of 2,000 iterates. It is clear from the figures that the billiard table becomes ever more filled as the number of iterates increases.

to the basis $\{\mathbf{u}_1, \mathbf{u}_2\} = \{(1,0), (1/2, \sqrt{3}/2)\}$.

4.4 Sequences of compatible orbits

We now discuss the notion of 'related' or 'compatible' that we have been referring to in the previous sections. It is intuitively clear that many of the examples of orbits we have been discussing are related by their initial conditions. We define a *sequence of compatible initial conditions* below.

Definition 149 (Compatible initial conditions)**.** Without loss of generality, suppose n and m are nonnegative integers such that $n > m$. Let $(x_n^0, \theta_n^0) \in (KS_n \times S^1)/\sim$ and $(x_m^0, \theta_m^0) \in (KS_m \times S^1)/\sim$ be two initial conditions of orbits $\mathscr{O}_n(x_n^0, \theta_n^0)$ and $\mathscr{O}_m(x_m^0, \theta_m^0)$, respectively, where we are assuming θ_n^0 and θ_m^0 are both inward pointing. If $\theta_n^0 = \theta_m^0$ and if x_n^0 and x_m^0 lie on a segment determined from θ_n^0 (or θ_m^0) that intersects KS_n only at x_n^0, then we say (x_n^0, θ_n^0) and (x_m^0, θ_m^0) are *compatible initial conditions.*

Remark 150. When two initial conditions (x_n^0, θ_n^0) and (x_m^0, θ_m^0) are compatible, then we simply write each as (x_n^0, θ^0) and (x_m^0, θ^0).

In §4.2, we discussed how it was that a $\pi/3$ orbit, regardless of where it begins on the boundary KS_n, must pass through the region of $\Omega(KS_n)$ corresponding to the interior of $\Omega(KS_0)$ in the way described by Proposition 133. Not every orbit must pass through the interior in this way, let alone pass through the interior of $\Omega(KS_m)$, for any $m < n$. Because of this, it may be the case that an initial condition (x_n^0, θ^0) is not compatible with (x_m^0, θ^0), for any $m < n$.

Definition 151 (Sequence of compatible initial conditions)**.** Let $\{(x_i^0, \theta_i^0)\}_{i=N}^{\infty}$ be a sequence of initial conditions, for some integer $N \geq 0$. We say that this sequence is a *sequence of compatible initial conditions* if for every $m \geq N$ and for every $n > m$, we have that (x_n^0, θ_n^0) and (x_m^0, θ_m^0) are compatible initial conditions. In such a case, we then write the sequence as $\{(x_i^0, \theta^0)\}_{i=N}^{\infty}$.

Definition 152 (Sequence of compatible orbits)**.** Consider a sequence of compatible initial conditions $\{(x_n^0, \theta^0)\}_{n=N}^{\infty}$. Then the corresponding sequence of orbits $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=N}^{\infty}$ is called *a sequence of compatible orbits.*

Building upon Proposition 133, we see that an orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ determines a sequence of compatible orbits, and, thus determines a *first orbit* $\mathscr{O}_m(x_m^0, \theta_m^0)$, $m \leq n$, of that sequence, and vice-versa.

Proposition 153. If $\mathscr{O}_m(y_m^0, \theta(\varpi_m^0))$ is an orbit of $\Omega(KS_m)$, then $\mathscr{O}_m(y_m^0, \theta(\varpi_m^0))$ is a member of a sequence of compatible orbits $\{\mathscr{O}_n(x_n^0, \varpi_0^n)\}_{n=N}^{\infty}$ for some $N \geq 0$.

Definition 154 (A sequence of compatible closed orbits)**.** If at least one orbit in a sequence of compatible orbits is a closed orbit, then we call such a sequence *a sequence of compatible closed orbits*

Definition 155 (A sequence of compatible periodic orbits)**.** If each orbit in a se-

quence of compatible closed orbits is a periodic orbit (that is, closed and nonsingular), then we call such a sequence *a sequence of compatible periodic orbits*.

It is clear from the definition of a sequence of compatible orbits that such a sequence is determined by the first orbit $\mathscr{O}_N(x_N^0, \theta^0)$, $N \geq 0$. Since the initial condition of an orbit determines the orbit, we can say without any ambiguity that a sequence of compatible orbits is determined by an initial condition (x_N^0, θ^0).

Definition 156 (A sequence of compatible dense orbits)**.** If at least one orbit in a sequence of compatible orbits is dense in its respective billiard table, then we call such a sequence a sequence of compatible dense orbits.

We know that for each fixed billiard table $\Omega(KS_n)$ and fixed direction θ_n^0, an orbit is either closed or dense, regardless of the initial basepoint x_n^0. Applying various results, we have the following.

Theorem 157 (A topological dichotomy for sequences of compatible orbits)**.** Let $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=N}^{\infty}$ be a sequence of compatible orbits. Then $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=N}^{\infty}$ is entirely comprised of either closed orbits or dense hybrid orbits.

Proof. This follows from Theorems 129 and 131. Indeed, since we have established that, for every $n \geq 1$, $\mathcal{S}(KS_n)$ is a branched cover of $\mathcal{S}(KS_0)$ and the fact that each orbit in a sequence of compatible orbits has the same initial direction, the qualitative behavior of each orbit must be the same, independently of n. ■

Corollary 158. A sequence of compatible orbits is either a sequence of compatible closed orbits or a sequence of compatible uniformly distributed orbits.

Proof. Since the Veech dichotomy holds for every prefractal billiard $\Omega(KS_n)$, $n \geq 0$, it follows that a sequence of compatible dense orbits must be a sequence of compatible orbits for which every orbit in the sequence is in fact uniformly distributed in the respective prefractal billiard $\Omega(KS_n)$. ■

4.4.1 Sequences of compatible piecewise Fagnano orbits

We have defined what it means for an orbit with an initial direction of $\pi/3$ to be a *piecewise Fagnano orbit* in Definition 134. We now define a *sequence of compatible piecewise Fagnano orbits*.

Definition 159. Let $x_0^0 \in [0,1]$ have a ternary representation $[c, lr]$ and $\mathscr{O}_0(x_0^0, \pi/3)$ be an orbit of $\Omega(KS_0)$. The sequence of compatible orbits determined by $(x_0^0, \pi/3)$ is called a *sequence of compatible piecewise Fagnano orbits*.

Remark 160. It is clear that finitely many orbits of a sequence of compatible piecewise Fagnano orbits are not piecewise Fagnano orbits of their respective approximations. Proposition 163 provides a criterion for determining (from the ternary representation) exactly when a sequence of compatible piecewise Fagnano orbits consists solely of piecewise Fagnano orbits.

Lemma 161. If $(x_0^0, \pi/3)$ and $(x_n^0, \pi/3)$ are compatible initial conditions, then the ternary representation of x_n^0 is the left shift of the ternary representation of x_0^0 by n characters.

Proof. Consider a sequence of compatible initial conditions $\{(x_n^0, \pi/3)\}_{n=0}^{\infty}$. The ternary representation of x_1^0 is the same as the left shift of the ternary representation of x_0^0. Let $N \geq 0$ and suppose for all $m \leq N$, the ternary representation of x_m^0 is the m left-shift of the ternary representation of x_0^0. Consider the compatible initial condition $(x_{N+1}^0, \pi/3)$. Suppose first that $x_{N+1}^0 = x_N^0$. Then x_N^0 lies on either the left or right third of a side $s_{N,k}$ of $\Omega(KS_N)$. The ternary representation of x_{N+1}^0 is then the same as the single left shift of the ternary representation of x_N^0. Now suppose $x_N^0 \neq x_{N+1}^0$. Then the segment connecting x_N^0 and x_{N+1}^0 is parallel to the side of a cell $C_{N+1,k'}$. The ternary representation of x_N^0, when described with respect to the deleted middle third

of the side $s_{N,k}$, is the left shift of the ternary representation of x_N^0 given with respect to the side $s_{N,k}$. Hence, the ternary representation of x_{N+1}^0 is the left shift of the ternary representation of x_N^0. Therefore, the ternary representation of x_{N+1}^0 is the $N+1$ left shift of the ternary representation of x_0^0. ■

Notation 162. Let $i \geq 0$. Then $(x_0^0)_i$ denotes the ith character in the ternary representation of x_0^0 (in terms of l, c, r).

Proposition 163. Let $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ be a sequence of compatible piecewise Fagnano orbits. Let N be the least value such that $(x_0^0)_m = c$ for every $m \geq N$. Then the first piecewise Fagnano orbit in $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is $\mathscr{O}_N(x_N^0, \pi/3)$.

Proof. Using Lemma 161 and the definition of piecewise Fagnano orbit, the result follows. In other words, the nth character such that the ternary representation consists of only the character c for every character thereafter is exactly when one appends a scaled copy of the Fagnano orbit to every basepoint of the preceding compatible piecewise Fagnano orbit, as dictated by the Definition 134 ■

4.4.2 Sequences of compatible Cantor orbits

Let $\mathscr{O}_N(x_N^0, \theta_N^0)$ be a Cantor orbit of $\Omega(KS_N)$. Then, the sequence of compatible orbits $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=N}^{\infty}$ is a constant sequence of compatible Cantor orbits. Since the corresponding path of the orbit $\mathscr{O}_N(x_N^0, \theta_N^0)$ traced out by the billiard ball in $\Omega(KS_N)$ must pass through the interior of $\Omega(KS_0)$, there must exist an orbit $\mathscr{O}_0(x_0^0, \theta_0^0)$ that is compatible with $\mathscr{O}_N(x_N^0, \theta_N^0)$. Therefore, we refer to the sequence of compatible orbits given by $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=0}^{\infty}$ as a *sequence of compatible Cantor orbits*. That is, we include in the sequence of compatible Cantor orbits those orbits which are compatible, but that are not, strictly speaking, Cantor orbits.

Definition 164 (Sequence of compatible Cantor orbits)**.** Let x_0^0 have a ternary rep-

resentation $[lr, c]$. Then, the sequence of compatible periodic orbits $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible Cantor orbits.

Proposition 165. Let $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ be a sequence of compatible Cantor orbits. Let N be the least value such that $(x_0^0)_m \neq c$ for every $m \geq N$. Then the first Cantor orbit in $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is $\mathscr{O}_{N-1}(x_{N-1}^0, \pi/3)$.

Proof. The result follows from the same exact reasoning as that found in the proof of Proposition 163. That is, there exists $N \geq 1$ such that $(x_0^0)_N \neq c$ and for every $n \geq N$, $(x_0^0)_n \neq c$. Since we have determined that x_N^0 has a representation determined from N left- shifts of the representation of x_0^0, it follows that for every $m \geq N - 1$, $\mathscr{O}_m(x_m^0, \pi/3)$ is a Cantor orbit. ■

4.4.3 Sequences of compatible approximate piecewise Fagnano orbits

Definition 166 (Sequence of compatible approximate piecewise Fagnano orbits)**.** Let $x_0^0 \in I$ have a ternary representation $[lc, r] \oplus [cr, l] \oplus [lcr, \emptyset]$. Then the sequence of compatible periodic orbits determined from $(x_0^0, \pi/3)$ is called a *sequence of compatible approximate piecewise Fagnano orbits.*

Proposition 167. Let $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ be a sequence of approximate piecewise Fagnano orbits. Then $N = 0$ is the least value for which $\mathscr{O}_N(x_N^0, \pi/3)$ is an approximate piecewise Fagnano orbit.

Proof. The result follows from the same exact reasoning as that found in the proof of Proposition 163 and immediately from the definition of an approximate piecewise Fagnano orbit. An approximate piecewise Fagnano orbit is defined in such a way that the first orbit of a sequence of compatible approximate piecewise Fagnano orbits is an approximate piecewise Fagnano orbit. ■

4.4.4 Properties of sequences of compatible $\pi/3$ orbits

By Proposition 133 and Proposition 153, we can deduce the following.

Proposition 168. Let $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ be an orbit. Then there exists $x_0^0 \in I$ such that $\mathscr{O}_n(y_n^0, \theta(\pi/3)) \in \{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ and $\mathscr{O}_n(x_n^0, \theta_n^0) = \mathscr{O}_n(y_n^0, \theta(\pi/3))$. In other words, there exists $x_0^0 \in I$ such that $(x_0^0, \pi/3)$ determines the sequence of compatible orbits to which $\mathscr{O}_n(y_n^0, \theta(\pi/3))$ belongs.

Notation 169. Let $n \geq 1$. We define $\omega_n(x_0^0)$ to be the cardinality of the set

$$\{(x_0^0)_i \,|\, (x_0^0)_i = 1,\, 1 \leq i \leq n\}, \tag{4.2}$$

where, for each $1 \leq i \leq n$, $(x_0^0)_i$ is as defined in Notation 162.

We recall that a corner has a ternary representation $[l, r] \oplus [r, l]$. If x_0^0 has a ternary representation $[l, cr] \oplus [r, cl]$, then there exists $n \geq 1$ such that $(x_n^0, \pi/3)$ is compatible with $(x_0^0, \pi/3)$ and x_n^0 is a corner of $\Omega(KS_n)$. Hence, the following.

Theorem 170 (Computation of the period of $\mathscr{O}_n(x_n^0, \pi/3)$)**.** Let $x_0^0 \in I$ be an initial basepoint of an orbit $\mathscr{O}_0(x_0^0, \pi/3)$ of $\Omega(KS_0)$ such that x_0^0 does not have a ternary representation $[l, cr] \oplus [r, lc]$ and $\mathscr{O}_n(x_n^0, \pi/3)$ is an orbit of $\Omega(KS_n)$ that is compatible with $\mathscr{O}_0(x_0^0, \pi/3)$. Then the period of $\mathscr{O}_n(x_n^0, \pi/3)$ (denoted by $\#\mathscr{O}_n(x_n^0, \theta_n^0)$) is given by

$$\#\mathscr{O}_n(x_n^0, \pi/3) = 3 \cdot 2^{\omega_n(x_0^0)}. \tag{4.3}$$

Proof. Let $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ be a sequence of compatible periodic orbits. Without loss of generality, we may assume $x_0^0 \in (1/3, 2/3)$. For the purpose of readibility, we consider the following three cases:

1. $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible piecewise Fagnano orbits.

2. $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible Cantor orbits.

3. $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible approximate piecewise Fagnano orbits.

Case 1: Suppose $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible piecewise Fagnano orbits. Recall from Proposition 163 that there exists $N \geq 0$ such that for every $m \geq N$, $\mathscr{O}_m(x_m^0, \pi/3)$ is a piecewise Fagnano orbit and for every $m < N$, $\mathscr{O}_m(x_m^0, \pi/3)$ is not a piecewise Fagnano orbit. We first consider the orbit $\mathscr{O}_0(x_0^0, \pi/3)$: the period is $\#\mathscr{O}_0(x_0^0, \pi/3) = 3$ if the orbit is the Fagnano orbit of $\Omega(KS_0)$, and $\#\mathscr{O}_0(x_0^0, \pi/3) = 6$ otherwise.

Now consider the orbit $\mathscr{O}_1(x_1^0, \pi/3)$. By assumption, $\omega_1(x_0^0) = 1$, and we have that $\#\mathscr{O}_1(x_1^0, \pi/3) = 6 = 3 \cdot 2^{\omega_1(x_0^0)}$.

Next, we proceed by induction. Let $M \geq 0$. Suppose that for every $n \leq M$, we have $\#\mathscr{O}_n(x_n^0, \pi/3) = 3 \cdot 2^{\omega_n(x_0^0)}$. We may as well assume that $N < M$. Therefore, $\mathscr{O}_M(x_M^0, \pi/3)$ is a piecewise Fagnano orbit of $\Omega(KS_M)$. By definition, $\mathscr{O}_{M+1}(x_{M+1}^0, \pi/3)$ is constructed from $\mathscr{O}_M(x_M^0, \pi/3)$ by appending $3 \cdot 2^{\omega_M(x_0^0)}$ many scale n copies of $\mathscr{O}_0(\overline{c}, \pi/3)$ (the Fagnano orbit of the equilateral triangle billiard $\Omega(KS_0)$) to each basepoint of $\mathscr{O}_M(x_M^0, \pi/3)$. Hence, we have that

$$\#\mathscr{O}_{M+1}(x_{M+1}^0, \pi/3) = 2 \cdot \#\mathscr{O}_M(x_M^0, \pi/3) = 2 \cdot 3 \cdot 2^{\omega_M(x_0^0)} = 3 \cdot 2^{\omega_M(x_0^0)+1}. \tag{4.4}$$

Since $(x_0^0)_n = 1$ for every $n \geq N$, we deduce that $\omega_{M+1}(x_0^0) = \omega_M(x_0^0) + 1$, and the result follows for a sequence of compatible piecewise Fagnano orbits.

Case 2: Suppose $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible Cantor orbits. Recall that there exists $N \geq 0$ such that for every $n \geq N$, $\mathscr{O}_n(x_n^0, \pi/3)$ is identical to $\mathscr{O}_N(x_N^0, \pi/3)$. Moreover, $\omega_n(x_0^0) = \omega_N(x_0^0)$ for every $n \geq N$. Suppose $N = 0$. Then $\#\mathscr{O}_N(x_N^0, \pi/3) = 6$. Suppose $N > 0$. Examining the ternary representation of x_N^0, we can find $y_0^0 \in I$ such that, for some $M \leq N$, $\mathscr{O}_M(y_M^0, \pi/3)$ forms a piecewise Fag-

nano orbit of $\Omega(KS_M)$ that is compatible with $\mathscr{O}_0(y_0^0, \pi/3)$, $\omega_N(x_0^0) = \omega_M(y_0^0)$ and $\#\mathscr{O}_N(x_N^0, \pi/3) = \#\mathscr{O}_M(y_M^0, \pi/3)$. In Case 1, we saw that the period of a piecewise Fagnano orbit was determined by the formula $\#\mathscr{O}_M(y_M^0, \pi/3) = 3 \cdot 2^{\omega_M(y_0^0)}$. Therefore,

$$\#\mathscr{O}_N(x_N^0, \pi/3) = \#\mathscr{O}_M(y_M^0, \pi/3) = 3 \cdot 2^{\omega_M(y_0^0)} = 3 \cdot 2^{\omega_N(x_0^0)}. \tag{4.5}$$

Since, for every $n \geq N$, $\mathscr{O}_n(x_n^0, \pi/3)$ is identical to $\mathscr{O}_N(x_N^0, \pi/3)$, the result follows for a sequence of compatible Cantor orbits.

Case 3: Suppose $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible approximate piecewise Fagnano orbits. For each $n \geq 1$, there exists $y_0^0 \in I$ and $M \leq n$ such that $\mathscr{O}_M(y_M^0, \pi/3)$ is a piecewise Fagnano orbit of $\Omega(KS_M)$ that is compatible with $\mathscr{O}_0(y_0^0, \pi/3)$ and $\omega_n(x_0^0) = \omega_M(y_0^0)$.

Let $N \geq 1$ and suppose that for every $n \leq N$, $\#\mathscr{O}_n(x_n^0, \pi/3) = 3 \cdot 2^{\omega_n(x_0^0)}$. Then, there exist $y_0^0 \in I$ and $M \leq N+1$ such that $\mathscr{O}_M(y_M^0, \pi/3)$ is a piecewise Fagnano orbit of $\Omega(KS_M)$ and $\omega_{N+1}(x_0^0) = \omega_M(y_0^0)$. Therefore,

$$\#\mathscr{O}_{N+1}(x_{N+1}^0, \pi/3) = \#\mathscr{O}_M(y_M^0, \pi/3) = 3 \cdot 2^{\omega_M(y_0^0)} = 3 \cdot 2^{\omega_{N+1}(x_0^0)}. \tag{4.6}$$

This concludes the proof of Theorem 170. ■

Notation 171. We now explain the notation that is about to be used in the following theorem and proof. The characteristic function χ is defined on the space of characters $\{l, c, r\}$ and is given by

$$\chi[\alpha] := \begin{cases} 0 & \text{if } \alpha = l, r \\ 1 & \text{if } \alpha = c. \end{cases} \tag{4.7}$$

In other words, it is the characteristic function of $\{c\}$.

Theorem 172 (Computation of the length of $\mathscr{O}_n(x_n^0, \pi/3)$)**.** Let $x_0^0 \in I$ be an initial basepoint of an orbit $\mathscr{O}_0(x_0^0, \pi/3)$ of $\Omega(KS_0)$ such that x_0^0 does not have a finite ternary

representation and $\mathscr{O}_n(x_n^0, \pi/3)$ an orbit that is compatible with $\mathscr{O}_0(x_0^0, \pi/3)$. Then the length of the orbit (denoted by $|\mathscr{O}_n(x_n^0, \theta_n^0)|$) is given by

$$|\mathscr{O}_n(x_n^0, \pi/3)| = 2\mathscr{L} + \sum_{i=2}^{n} \chi\left[(x_0^0)_i\right] \#\mathscr{O}_{i-1}(x_{i-1}^0, \pi/3)\frac{\mathscr{L}}{3^i}, \tag{4.8}$$

where $\mathscr{L}$ is the length of the Fagnano orbit $\mathscr{O}_0(\overline{c}, \pi/3)$ of $\Omega(KS_0)$.

Proof. Let $n = 0$. If $\overline{c}$ is the ternary representation (in terms of the characters l, c, r) of an element $x_0^0 \in I$, where I is viewed as the base of $\Omega(KS_0)$, then $\mathscr{O}_0(x_0^0, \pi/3)$ is the Fagnano orbit of $\Omega(KS_0)$ and has length $|\mathscr{O}_0(x_0^0, \pi/3)| = \mathscr{L}$. If x_0^0 has a ternary representation different from $\overline{c}$, but beginning in the character c, then $|\mathscr{O}_0(x_0^0, \pi/3)| = 2\mathscr{L}$. In either case, if x_1^0 is collinear with x_0^0, then $|\mathscr{O}_1(x_1^0, \pi/3)| = 2\mathscr{L}$.

Consider the basic case $n = 2$. Let $x_2^0 \in KS_2$ be collinear with $x_0^0 \in I$ (viewed as the base of the equilateral triangle KS_0). Then, x_2^1 is the basepoint in the compatible orbit $\mathscr{O}_2(x_2^0, \pi/3)$. As before, let $(x_0^0)_i$ be the ith character in the ternary expansion of x_0^0. We want to show that

$$|\mathscr{O}_2(x_2^0, \pi/3)| = 2\mathscr{L} + \sum_{i=2}^{2} \chi[(x_0^0)_i] \#\mathscr{O}_{i-1}(x_{i-1}^0, \pi/3)\frac{\mathscr{L}}{3^i}. \tag{4.9}$$

If $(x_0^0)_2 = c$, then, examining the ternary representation of $(x_0^0)_2$, we see that there exists $y_0^0 \in I$ such that $|\mathscr{O}_2(x_2^0, \pi/3)| = |\mathscr{O}_2(y_2^0, \pi/3)|$ and $(x_0^0)_i = (y_0^0)_i$ for every $i \leq n = 2$; y_0^0 will have infinitely many c's and finitely many l's and r's. There exists $y_1^0 \in KS_1$ that is collinear with y_0^0 and such that $|\mathscr{O}_2(y_2^0, \pi/3)| = |\mathscr{O}_1(y_1^0, \pi/3)| + \#\mathscr{O}_1(y_1^0, \pi/3)\frac{\mathscr{L}}{3^2}$, since $\mathscr{O}_2(y_2^0, \pi/3)$ is a piecewise Fagnano orbit of KS_2. Since y_0^0 and x_0^0 have the same first two characters in their respective ternary expansions, we have that

$$|\mathscr{O}_1(y_1^0, \pi/3)| = |\mathscr{O}_1(x_1^0, \pi/3)|$$

and

$$\#\mathscr{O}_1(y_1^0, \pi/3) = \#\mathscr{O}_1(x_1^0, \pi/3).$$

Since $\mathscr{O}_1(y_1^0, \pi/3) = 2\mathscr{L}$ (which is independent of the choice of basepoint, so long as such a choice is not a corner of the billiard table $\Omega(KS_1)$), it follows that

$$|\mathscr{O}_2(x_2^0, \pi/3)| = 2\mathscr{L} + \sum_{i=2}^{2} \#\mathscr{O}_{i-1}(x_{i-1}^0, \pi/3)\frac{\mathscr{L}}{3^i}. \tag{4.10}$$

If $(x_0^0)_2 \neq c$, then $\chi[(x_0^0)_2] = 0$ and

$$|\mathscr{O}_2(x_2^0, \pi/3)| = |\mathscr{O}_1(x_1^0, \pi/3)| = 2\mathscr{L}. \tag{4.11}$$

In either case, we have shown that Equation (4.9) holds.

Let us now proceed by induction and fix $N \geq 2$. Suppose that for every $n \leq N$,

$$|\mathscr{O}_n(x_n^0, \pi/3)| = 2\mathscr{L} + \sum_{i=2}^{n} \chi[(x_0^0)_i]\#\mathscr{O}_{i-1}(x_{i-1}^0, \pi/3)\frac{\mathscr{L}}{3^i}. \tag{4.12}$$

Then there exists $y_0^0 \in I$ with a ternary representation $[c, lr]$, $M \leq N+1$ such that for all $i \leq M$, $(x_0^0)_i = (y_0^0)_i$ and

$$|\mathscr{O}_{N+1}(x_{N+1}^0, \pi/3)| = |\mathscr{O}_M(y_M^0, \pi/3)|. \tag{4.13}$$

If $M = N+1$, then the nature of $\mathscr{O}_N(y_M^0, \pi/3)$ dictates that

$$|\mathscr{O}_{N+1}(y_{N+1}^0, \pi/3)| = |\mathscr{O}_N(y_N^0, \pi/3)| + \#\mathscr{O}_N(y_N^0, \pi/3)\frac{\mathscr{L}}{3^{N+1}}. \tag{4.14}$$

Applying this fact, along with the induction hypothesis (4.12), we have that

$$\begin{aligned}
|\mathscr{O}_{N+1}(x_{N+1}^0, \pi/3)| &= |\mathscr{O}_{N+1}(y_{N+1}^0, \pi/3)| \\
&= |\mathscr{O}_N(y_N^0, \pi/3)| + \#\mathscr{O}_N(y_N^0, \pi/3)\frac{\mathscr{L}}{3^{N+1}} \\
&= |\mathscr{O}_N(x_N^0, \pi/3)| + \#\mathscr{O}_N(x_N^0, \pi/3)\frac{\mathscr{L}}{3^{N+1}} \\
&= 2\mathscr{L} + \sum_{i=2}^{N} \chi[(x_0^0)_i]\#\mathscr{O}_{i-1}(x_{i-1}^0, \pi/3)\frac{\mathscr{L}}{3^i} + \\
&\quad \#\mathscr{O}_N(x_N^0, \pi/3)\frac{\mathscr{L}}{3^{N+1}} \\
&= 2\mathscr{L} + \sum_{i=2}^{N+1} \chi[(x_0^0)_i]\#\mathscr{O}_{i-1}(x_{i-1}^0, \pi/3)\frac{\mathscr{L}}{3^i}.
\end{aligned} \tag{4.15}$$

If $M < N+1$, then, again applying the previously mentioned fact, the induction hypothesis, we deduce that

$$\begin{aligned}
|\mathscr{O}_{N+1}(x^0_{N+1}, \pi/3)| &= |\mathscr{O}_M(y^0_M, \pi/3)| \\
&= |\mathscr{O}_{M-1}(y^0_{M-1}, \pi/3)| + \#\mathscr{O}_{M-1}(y^0_{M-1}, \pi/3)\frac{\mathscr{L}}{3^M} \\
&= |\mathscr{O}_{M-1}(x^0_{M-1}, \pi/3)| + \#\mathscr{O}_{M-1}(x^0_{M-1}, \pi/3)\frac{\mathscr{L}}{3^M} \\
&= 2\mathscr{L} + \sum_{i=2}^{M-1} \chi[(x^0_0)_i]\#\mathscr{O}_{i-1}(x^0_{i-1}, \pi/3)\frac{\mathscr{L}}{3^i} + \qquad (4.16) \\
&\quad \#\mathscr{O}_{M-1}(x^0_{M-1}, \pi/3)\frac{\mathscr{L}}{3^M} \\
&= 2\mathscr{L} + \sum_{i=2}^{M} \chi[(x^0_0)_i]\#\mathscr{O}_{i-1}(x^0_{i-1}, \pi/3)\frac{\mathscr{L}}{3^i} \\
&= 2\mathscr{L} + \sum_{i=2}^{N+1} \chi[(x^0_0)_i]\#\mathscr{O}_{i-1}(x^0_{i-1}, \pi/3)\frac{\mathscr{L}}{3^i},
\end{aligned}$$

where the last lines of the calculation in Equation (4.16) follow from the fact that the characters $(x^0_0)_i$, with $M < i \leq N+1$, are necessarily never equal to c, meaning that $\chi(x^0_0)_i = 0$ for $M < i \leq N+1$. ■

4.4.5 Sequences of compatible hybrid orbits

Definition 173 (Sequence of compatible hybrid orbits)**.** Let $\{\mathscr{O}_n(x^0_n, \theta^0)\}_{n=N}^{\infty}$ be a sequence of compatible orbits. If every orbit is a hybrid orbit, then we call the sequence of compatible orbits a *sequence of compatible hybrid orbits.*

Definition 174 (Sequence of compatible closed hybrid orbits)**.** Let $\{\mathscr{O}_n(x^0_n, \theta^0)\}_{n=N}^{\infty}$ be a sequence of compatible hybrid orbits. If at least one orbit is a closed hybrid orbit, then we call the sequence of compatible hybrid orbits a *sequence of compatible closed hybrid orbits.*

Definition 175 (Sequence of compatible periodic hybrid orbits)**.** Let $\{\mathscr{O}_n(x^0_n, \theta^0)\}_{n=N}^{\infty}$ be a sequence of compatible hybrid orbits. If every orbit is a periodic orbit, then we call

the sequence of compatible hybrid orbits a *sequence of compatible periodic hybrid orbits.*

Definition 176 (Sequence of compatible dense hybrid orbits)**.** Let $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=N}^{\infty}$ be a sequence of compatible hybrid orbits. If at least one orbit is a dense hybrid orbit, then we call the sequence of compatible hybrid orbits a *sequence of compatible dense hybrid orbits.*

Example 177. The orbits discussed in Example 148 constitute the first three elements of a sequence of compatible dense hybrid orbits.

Theorem 178. Consider a vector (a, b) that is rational with respect to the basis $\{u_1, u_2\} := \{(1, 0), (\sqrt{3}, 2)\}$. Let $x_0^0 = \frac{r}{4^s}$, $r, s \in \mathbb{Z}$ with $s \geq 1$ and $r < 4^s$ being an odd integer. If b is an odd integer and $\theta = \arctan \frac{b\sqrt{3}}{2a+b}$, then the sequence of compatible closed orbits $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=0}^{\infty}$ is a sequence of compatible periodic hybrid orbits.

Proof. Let $r, s \in \mathbb{Z}$, $s \geq 1$ and $r \leq 4^s$. Suppose a line segment starting at $(r/4^s, 0)$ with slope $\frac{b\sqrt{3}}{2a+b}$ intersects a point in $\mathbb{R}^2$ that would correspond to a lattice point of a lattice comprised of equilateral triangles at scale k. If $m, n, p, q, k \in \mathbb{Z}$, with $k \geq 1$ and $q \leq 3^k$, then such a point has the form $(m + p/3^k)u_1 + (n + q/3^k)u_2$. Then, using the equation for a line in the plane, we find that

$$\left(n + \frac{q}{3^k}\right)\frac{\sqrt{3}}{2} = \frac{b\sqrt{3}}{2a+b}\left(m + \frac{p}{3^k} + \frac{n}{2} + \frac{q}{2 \cdot 3^k} - \frac{r}{4^s}\right) \tag{4.17}$$

$$\left(\frac{3^k n + q}{3^k}\right)\frac{1}{2} = \frac{b}{2a+b}\left(\frac{4^s 3^k m + 4^s p + 2 \cdot 4^{s-1} 3^k n + 2 \cdot 4^{s-1} q - 3^k r}{3^k 4^s}\right) \tag{4.18}$$

$$2 \cdot 4^{s-1}(3^k n + q)(2a + b) = b(4^s 3^k m + 4^s p + 2 \cdot 4^{s-1} 3^k n + 2 \cdot 4^{s-1} q - 3^k r). \tag{4.19}$$

If b is odd, then the right-hand side of Equation (4.19) is not even, but the left-hand side is. Therefore, our assumption that such a point laid on the line with $\frac{b\sqrt{3}}{2a+b}$ was incorrect. It follows that such a line emanating from $x = q/4^s$, q and odd integer and $s \geq 1$ avoids all corners of all approximations $\Omega(KS_n)$, $n \geq 0$, giving us a sequence of compatible periodic hybrid orbits. ■

Theorem 179. Consider a vector $(a, 1)$ that is rational with respect to the basis $\{\mathbf{u}_1, \mathbf{u}_2\} = \{(1, 0), (\sqrt{3}, 2)\}$. Let $x_0^0 = \frac{r}{2^s}$, $r, s \in \mathbb{Z}$ with $s \geq 1$ and $r < 2^s$ being an odd integer. If $\theta = \arctan \frac{\sqrt{3}}{2a+1}$, then the sequence of compatible closed orbits $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=0}^{\infty}$ is a sequence of compatible periodic hybrid orbits.

Proof. The proof follows the same exact line of reasoning as the proof of Theorem 178. ■

Example 180 (A sequence of compatible hook orbits)**.** Let $x_0^0 \in I$ have a ternary representation given by $\overline{rl}$. Such a point has a value of 3/4. Considering an orbit of $\Omega(KS_0)$ with an initial direction of $\pi/6$, the ternary representation of the basepoints at which the billiard ball path forms right angles with the sides of $\Omega(KS_0)$ $[c, lr]$. This hook orbit (see Example 146) is in fact a periodic hybrid orbit. The next orbit in the sequence of compatible hook orbits has the initial condition $(x_1^0, \pi/6) = (x_0^0, \pi/6)$. Since the ternary representation of the basepoint of $f_0(x_0^0, \pi/6)$ is $r\overline{c}$, it follows that the basepoint of $f_1(x_1^0, \pi/6)$ is a Cantor-point. Then the basepoint of $f^2(x_1^0, \pi/6)$ has a ternary representation $[c, lr]$. This same pattern is repeated for every subsequent orbit in the sequence of compatible orbits. As a result, the sequence of compatible orbits forms a sequence of orbits that appears to be converging to a set that is well-defined. That is, such a set will be some path with finite length that is effectively determined by the law of reflection. We elaborate on this idea in §5.1.3.1.

Chapter 5

The Koch snowflake fractal billiard $\Omega(KS)$

5.1 Theoretical results

5.1.1 Piecewise Fagnano orbits

From §2.8.2, we now understand how to identify points of the boundary KS. From §4.2, if $\theta^0 = \pi/3$, then we know that we can determine the type of compatible sequence of periodic orbits $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ by examining the ternary representation of x_0^0. That is, a sequence of compatible $\pi/3$ orbits was either a sequence of compatible piecewise Fagnano orbits, Cantor orbits or approximate piecewise Fagnano orbits, depending on the nature of the ternary representation of x_0^0.

Let $X = \{0, 1, 2, 3, 4, 5\}$ and $X_n := \prod_{i=1}^{n} X = X^n$, the space of all words of length $n \geq 1$ with characters (or symbols) in the alphabet X; by default, we let $X^0 := \emptyset$. Define $X_\infty := \prod_{i=1}^{\infty} X = X^{\mathbb{N}}$ to be the space of infinite words expressed in terms of the elements of X. Let $x \in X_\infty$ and $n \in \mathbb{N}$. We define $\tau_n : X_\infty \to X_n$ by $\tau_n(x) := x|_n$,

where $x|_n \in X_n$ is the finite word of length n consisting of the first n characters of the infinite word x. That is, $x|_n \in X_n$ is the *truncation at level* n of $x \in X_\infty$.

Restricting τ_n to KS results in a map that is well defined on the points of KS which are not corners of KS. For example, $1\overline{2}$ and $3\overline{2}$ identify the same point in KS, but the words of length n given by 12...2 and 32...2 do not identify the same segments of KS_n.

So that we avoid any ambiguity, recall that the orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ is the billiard ball path and the footprint $\mathcal{F}_n(x_n^0, \theta_n^0)$ of the orbit $\mathscr{O}_n(x_n^0, \theta_n^0)$ is the intersection of the orbit with the boundary KS_n. If $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^\infty$ is a sequence of compatible piecewise Fagnano orbits, then we show that the inverse limit of the footprints $\mathcal{F}_n(x_n^0, \pi/3)$ exists. Such a task will require us to define the proper transition maps so that the definition of inverse limit is satisfied. (See §2.9.2 above for a brief discussion of inverse limits, and [Bo, HoYo] for further information.) In other words, we recast the sequence of footprints $\{\mathcal{F}_j(x_j^0, \pi/3)\}_{j=0}^\infty$ corresponding to a compatible sequence of orbits as an inverse limit sequence of footprints of periodic orbits, where each basepoint is given in terms of its address; see §2.8.2 (including Figures 2.14–2.12) for a description of the addressing system. Once (and if) an appropriate billiard flow can be defined, we will seek in future work to show that the inverse limit of footprints is a subset of some analogue of a Poincaré section of the suitably defined billiard flow.

Given $m, n \in \mathbb{N}^*$ with $m \leq n$, consider the map $\tau_{mn} : X_n \to X_m$, where $\tau_{mn}(x_n) = x_n|_m$ is the truncation of the finite word x_n (of length n) by $n - m$ characters (thereby producing a word of length m). Then $\tau_{mn}|_{KS_{n-1}}$ is a map that truncates the addresses of segments of KS_{n-1} of length n to produce addresses of the segments of KS_{m-1} of length m. Let the map $\iota = \iota_{mn}$ be the 'identity map' acting on the unit circle S^1. The map $\iota = \iota_{mn}$ will serve to preserve the compatibility of the two initial conditions

(x_m^0, θ_m^0) and (x_n^0, θ_n^0). In actuality, $\iota(\theta_n^0) = \theta_m^0$, meaning that (x_m^0, θ_m^0) is the same as writing (x_m^0, θ_n^0).[1]

Consider $x_0^0 \in I$ with a ternary representation $[c, lr]$. Recall from §4.2.1 that this is a necessary and sufficient condition for $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=0}^{\infty}$ forming a sequence of compatible piecewise Fagnano orbits. Specifically, as stated in Proposition 133, there exists a greatest integer $N \geq 1$ such that $\mathscr{O}_N(x_N^0, \pi/3)$ is a piecewise Fagnano orbit of $\Omega(KS_N)$, yet for every $m < N$, $\mathscr{O}_m(x_m^0, \pi/3)$ is not. Moreover, for every $n \geq N$, $\mathscr{O}_n(x_n^0, \pi/3)$ is a piecewise Fagnano orbit. It follows that the first element x_N^1 in the footprint of the orbit $\mathscr{O}_N(x_N^0, \theta_N^0)$ has a finite address ending in either 13 or 31. Likewise, for each $n \geq N$, x_n^1 has a finite address ending in either 13 or 31. Specifically, for every i such that $N \leq i \leq n$, $(x_n^0)_i \neq 0, 2, 4$.

By utilizing the local and global symmetry of the snowflake (see Figure 4.3), one may determine the footprint of $\mathscr{O}_n(x_n^0, \pi/3)$ as a dynamically ordered set of points. One may then consider the inverse limit of the footprints of $\mathcal{F}_n(x_n^0, \pi/3)$, where one forms the inverse limit by considering as our transition maps τ_{mn} (for $m \leq n$) the truncation of finite addresses, as defined above in this section. Letting $\mathcal{F}_n(x_n^0, \theta_n^0)$ be the footprint of the orbit $\mathscr{O}_n(x_n^0, \pi/3)$, we write the inverse limit of the footprints as

$$\varprojlim \mathcal{F}_n(x_n^0, \pi/3) = \left\{ (x_i^{k_i})_{i=N}^{\infty} \in \prod_{i=N}^{\infty} \mathcal{F}_n(x_n^0, pi/3) | \tau_{mn}(x_n^{k_n}) = x_m^{k_m} \text{ for all } N \leq m \leq n \right\}. \tag{5.1}$$

Due to the fact that this set lacks a dynamical ordering, a priori, we cannot say that it represents the iterates of a first return map defined on $\Omega(KS)$ (if it is even possible to define such a map). Therefore, we next examine how to formulate the inverse limit $\varprojlim \mathcal{F}_n(x_n^0, \pi/3)$ using transition maps that recapture the dynamical ordering on each

[1] The map ι preserves the compatibility of the initial basepoints x_n^0 and x_m^0, by relying on the fact that angles are measured with respect to a fixed coordinate system.

prefractal approximation.

Consider ζ_n, the number of pairs of points preceding $x_n^{k_n}$ in the footprint $\mathcal{F}_n(x_n^0, \pi/3)$ of the orbit $\mathscr{O}_n(x_n^0, \pi/3)$. One can easily see that $2\zeta_n + \beta_n = k_n$, for some $\beta_n \in \{1, 2\}$. Then, for $(x_i^{k_i})_{i=N}^{\infty} \in \varprojlim \mathcal{F}_i(x_i^0, \theta_i^0)$, we have that there exists k_{n-1} such that

$$x_{n-1}^{k_{n-1}} = \tau_{n-1,n}(x_n^{k_n}) = \tau_{n-1,n}(x_n^{2\zeta_n+\beta_n}) = x_{n-1}^{\zeta_n+1}, \tag{5.2}$$

where $x_{n-1}^{\zeta_n+1}$ is the basepoint of $f_{n-1}^{\zeta_n+1}(x_{n-1}^0, \pi/3)$. By the definition of a piecewise Fagnano orbit (see Definition 134), we see that $x_{n-1}^{k_{n-1}} = x_{n-1}^{\zeta_n+1}$, and hence that $k_{n-1} = \zeta_n + 1$. Consider the map $F_{n-1,n}$ given by

$$F_{n-1,n}(x_n^{k_n}, \theta_n^{k_n}) = f_{n-1}^{\zeta_n+1} \circ \tau_{n-1,n} \times \iota_{n-1,n} \circ f_n^{-k_n}(x_n^{k_n}, \theta_n^{k_n}). \tag{5.3}$$

The map $F_{n-1,n}$ is well defined. In general, if $m < n$, then the map

$$F_{m,n}(x_n^{k_n}, \theta_n^{k_n}) = f_m^{\zeta_{m+1}+1} \circ \tau_{m,n} \times \iota_{m,n} \circ f_n^{-k_n}(x_n^{k_n}, \theta_n^{k_n}) \tag{5.4}$$

constitutes the proper transition map needed to construct the inverse limit of footprints of a compatible sequence of piecewise Fagnano orbits $\{\mathscr{O}_i(x_i^0, \theta_i^0)\}_{i=0}^{\infty}$. We denote the resulting inverse limit as follows:

$$\mathcal{F}(x^0, \theta^0) := \varprojlim \mathcal{F}_n(x_n^0, \theta_n^0) = \tag{5.5}$$
$$\left\{(x_i^{k_i}, \theta_i^{k_i})_{i=N}^{\infty} \in \prod_{i=N}^{\infty} \mathscr{O}_i(x_i^0, \theta_i^0) | F_{nm}(x_n^{k_n}, \theta_n^{k_n}) = (x_m^{k_m}, \theta_m^{k_m}), \text{ for all } N \leq m \leq n\right\}.$$

We define $x^0 \in KS$ to be $\lim_{i\to\infty} x_i^0$, which is the limit (in the plane) of the compatible sequence of initial basepoints $\{x_i^0\}_{i=N}^{\infty}$.

As will be further discussed later on, we would like to think of $\mathcal{F}(x^0, \theta^0)$ as the '*footprint*' (or the analog of iterates of the billiard map) of a '*piecewise Fagnano periodic orbit*' of $\Omega(KS)$, namely, the piecewise Fagnano periodic orbit with initial basepoint $x^0 := \lim_{i\to\infty} x_i^0$ in the direction of $\pi/3$.

5.1.1.1 Topological properties of $\mathcal{F}(x^0, \theta^0)$

Theorem 181. If $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ is a sequence of compatible piecewise Fagnano orbits, then the 'footprint' $\mathcal{F}(x^0, \theta^0)$, given by the inverse limit $\varprojlim \mathscr{O}_i(x_i^0, \theta_i^0)$, is a topological Cantor set (i.e., a totally disconnected and perfect compact set). Moreover, the above inverse limit of the footprints is a self-similar Cantor set.

Proof. Let $\{\mathscr{O}_n(x_n^0, \pi/3)\}_{n=0}^{\infty}$ be a sequence of compatible piecewise Fagnano orbits. By Theorem 116, $\mathcal{F}(x^0, \theta^0)$ is totally disconnected and compact, being the inverse limit of finite sets. What remains to be shown is that $\mathcal{F}(x^0, \theta^0)$ is a perfect set.

Consider an element $(x_i^{k_i}, \theta_i^{k_i})_{i=0}^{\infty} \in \mathcal{F}(x^0, \theta^0)$. Fix $n \geq 0$. Then there exists $(y_i^{j_i}, \phi_i^{j_i})_{i=0}^{\infty} \in \mathcal{F}(x^0, \theta^0)$ such that 1) $(x_i^{k_i}, \theta_i^{k_i}) = (y_i^{j_i}, \phi_i^{j_i})$ for all $i \leq n$ and 2) $x_{n+1}^{k_{n+1}} \neq y_{n+1}^{j_{n+1}}$. Concretely, this element $(y_i^{j_i}, \phi_i^{j_i})_{i=0}^{\infty}$ is determined from $(x_i^{k_i}, \theta_i^{k_i})_{i=0}^{\infty}$ by way of the local symmetry. As such, we can continue to construct a sequence of elements in $\mathcal{F}(x^0, \theta^0)$ that converges (with respect to either the Euclidean metric or a metric defined on the space of addresses) to $(x_i^{k_i}, \theta_i^{k_i})_{i=0}^{\infty} \in \mathcal{F}(x^0, \theta^0)$.

To see that such a topological Cantor set is a self-similar Cantor set, one simply recognizes the fact that for every footprint $\mathcal{F}(x^0, \theta^0)$, there is a finite collection of IFS's $\{\Phi_{x_0^0,i}\}_{i=1}^{\Xi}$, each IFS $\Phi_{x_0^0,i}$ consisting of two contraction mappings, and each giving rise to a unique fixed point attractor that is a self-similar set in its own right. The union of these fixed point attractors is then the self-similar footprint. ■

Proposition 182. If $\{\mathcal{F}(x^0, \theta^0) | x_0^0$ has a representation $[c, lr]\}$ is the collection of all footprints of piecewise Fagnano orbits of the Koch snowflake billiard $\Omega(KS)$, then $x_0^0 \neq y_0^0$ if and only if $\mathcal{F}(x^0, \theta^0)$ and $\mathcal{F}(y^0, \theta^0)$ have no elements in common.

Proof. Suppose $\mathcal{F}(x^0, \theta^0) \neq \mathcal{F}(y^0, \theta^0)$. Recall from §2.8.4.1 that an address beginning with 5 of a point can be straightened so as to identify a point $x^0 \in KS$ that

is compatible with an element x_0^0 (in the direction of $\pi/3$ in the interior of KS) that is also an element of the unit interval base of KS_0, which we have been denoting by I throughout the paper. This element may then be taken to be $\lim_{i\to\infty} x_i^0$, with $\{x_i^0\}_{i=0}^{\infty}$ being a compatible sequence of initial basepoints (again, compatible in the direction of $\pi/3$). That is, $x^0 = \lim_{i\to\infty} x_i^0$ and $y^0 = \lim_{i\to\infty} y_i^0$. Suppose $x^0 \neq y^0$. Then there exists $\nu > 0$ such that $||x^0 - y^0|| > \nu$ (where we have taken $||\cdot||$ to be the Euclidean norm in $\mathbb{R}^2$). Therefore, there exists $n \geq 0$ such that

$$||x_n^0 - y_n^0|| > \nu \tag{5.6}$$

or

$$x_n^0 \neq y_n^0. \tag{5.7}$$

Since x_n^0 and x_0^0 are compatible in the direction of $\pi/3$ and y_n^0 and y_0^0 are compatible in the direction of $\pi/3$, it follows that $x_0^0 \neq y_0^0$. The converse holds since $x_0^0 \neq y_0^0$ implies $x^0 \neq y^0$. ■

5.1.2 Finitely stabilizing periodic orbits

We now are in a position to properly elaborate on the "stabilizing" nature of a Cantor orbit. We have alluded to the fact that a Cantor orbit is indeed a periodic orbit of the Koch snowflake, but have not said in what sense. Though we have shown that a Cantor orbit remains constant for all subsequent approximations, we have not said what exactly the consequence of this should be.

We say that such a sequence of compatible orbits has *stabilized after finite time* if the sequence of compatible orbits is constant after finitely many elements. A sequence of compatible Cantor orbits is a canonical example of such a phenomenon. As a result,

we call the trivial limit of a sequence of compatible orbits that has stabilized after finite time a *finitely stabilizing periodic orbit* of $\Omega(KS)$.

5.1.3 Periodic hybrid orbits

As of now, we cannot say whether or not there exists a 'periodic hybrid orbit' of the Koch snowflake fractal billiard. Moreover, the general notion of an orbit of the Koch snowflake billiard is not well defined. However, we have been able to determine a particular collection of sequences of compatible periodic hybrid orbits, each having a particular quality that allows us to determine from each sequence of compatible orbits a nontrivial polygonal path that is converging to an elusive limit point of the Koch snowflake fractal.

In Example 146, we gave an example of a hook orbit that was in fact a periodic hybrid orbit. Such an orbit, when the correct initial basepoint was chosen, was an element in a sequence of compatible periodic hybrid orbits $\{\mathscr{O}_n(x_n^0, \theta^0)\}_{n=N}^{\infty}$ (see Theorem 178. From each orbit in such a sequence, there exists a basepoint $x_n^{k_n}$ that is an element of a sequence of basepoints converging to an elusive limit point. Such a sequence of compatible periodic hybrid orbits is part of a larger collection of sequences of compatible periodic hybrid orbits with this same quality. A necessary condition for this quality (the quality that there exists a sequence of basepoints $\{x_n^{k_n}\}_{n=N}^{\infty}$ converging to an elusive limit point) is that for each n, the basepoint $x_n^{k_n+1}$ be on a side $s_{n,k}$ of $\Omega(KS_n)$, $x_n^{k_n+1}$ having a ternary representation $[c, lr]$ and that $x_{n+1}^{k_{n+1}}$ and $x_{n+1}^{k_{n+1}+1}$ be points on the sides of a cell $C_{n+1,k}$ of $\Omega(KS_{n+1})$ with $x_{n+1}^{k_{n+1}}$ being a Cantor-set point and $x_{n+1}^{k_{n+1}+1}$ having a ternary representation $[c, lr] \oplus [lc, r] \oplus [cr, l]$. In addition to hook orbits, the sequence of compatible orbits shown in Figure 4.8 is another example of a sequence of compatible periodic hybrid orbits with such a quality.

5.1.3.1 Nontrival Polygonal Paths

When one can determine a sequence of basepoints from a sequence of compatible periodic hybrid orbits that is converging to an elusive limit point of KS, then we say that such a sequence of points can be connected by what we call a *nontrivial polygonal path of* $\Omega(KS)$. More to the point, such a path is conjectured to be part of what would hopefully be a periodic orbit of the Koch snowflake. Evidence in support of this conjecture is provided later in §5.2.1.

5.2 Experimental results

5.2.1 Convergence to an eventually stabilizing periodic orbit

We have demonstrated the existence of a sequence of compatible orbits that stabilizes after finite time, namely a sequence of compatible Cantor orbits. We conjecture that a sequence of compatible periodic hybrid orbits will stabilize *eventually*, that is, after *infinite* time. In Figure 5.1 we demonstrate the fact that the compatible periodic hybrid orbits of the 2nd through 5th prefractal approximations appear to be converging to a particular path that would then constitute an orbit in the limit.

While there seems to be a trend demonstrated in Figure 5.1, such a sequence of simulated orbits does not constitute a proof. Though no simulation can provide a proof of anything (in the current context, that is), we must be careful not to deduce too much from these experiments. In order to prove that a sequence of compatible periodic hybrid orbits is indeed converging to some set (i.e., in order to show that the geometric limit of a sequence of compatible periodic hybrid orbits exists), we must consider the following question:

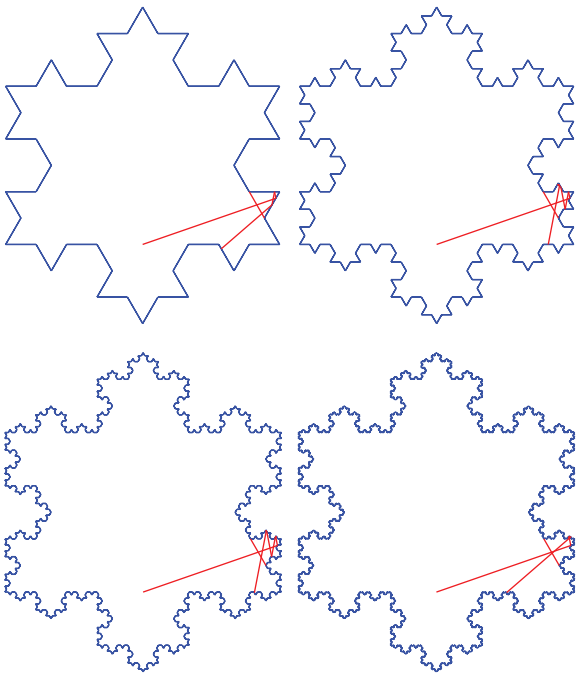

Figure 5.1: Beginning in the top left corner and moving left to right, top to bottom, we see that the point at which the path of the billiard ball exits the particular cell of $\Omega(KS_2)$ is converging to the point at which the particular path crossed into the cell. We replace the deleted segment in each figure to provide a visual frame of reference.

- Given a sequence of compatible periodic hybrid orbits, does there always exist a nontrivial polygonal path converging to an elusive limit point?

In the following chapter, we discuss a number of open questions and possible solutions to some of the questions posed thus far.

Chapter 6

Concluding remarks and future research

6.1 A well-defined billiard flow on $\Omega(KS)$

The goal of this thesis was to make progress in answering the question of whether or not there existed a well-defined billiard flow on the Koch snowflake. As was discussed in Chapter 5, there is a particular family of sequences of compatible hybrid periodic orbits, each of which can be used to determine a sequence of basepoints converging to an elusive limit point of the Koch snowflake KS. From such a sequence, one then can construct a nontrivial polygonal path. We ask the question whether or not such a nontrivial polygonal path does in fact constitute a portion of an orbit of the snowflake.

The difficulty in answering this question is the problem with determining an appropriate analog of a Poincaré section. It is intuitively clear that the Poincaré section of the billiard flow should be the Koch snowflake boundary KS. In the case of an initial direction of $\pi/3$, the appropriate analog of a piecewise Fagnano orbit would then have an

uncountable footprint, this then being an uncountable subset of the Poincaré section. As we have seen, a footprint of an orbit of a rational billiard table (or a mathematical billiard, in general) is at most countable. The closure of a dense footprint is the whole boundary B of a billiard table. Taking a literal interpretation of the footprint $\varprojlim \mathcal{F}_n(x_n^0, \pi/3)$ of a piecewise Fagnano orbit as a footprint in the classical sense would seem to pose a problem, given the current framework of mathematical billiards. More concretely, we must somehow resolve the question of how it is a periodic orbit can have an uncountable period. One cannot simply consider iterates of some Poincaré map as this will lead to at most a countably infinite footprint. Instead, we can consider the closure of a set of points determined from a suitable limit of a sequence of compatible orbits as the appropriate analogue of the classical footprint. This would be in line with our intuition and provide the necessary framework for incorporating our results on piecewise Fagnano orbits.

Another viable solution to this problem is to focus on determining finitely stabilizing periodic orbits. That is, can we demonstrate the existence of a finitely stabilizing periodic orbit with an initial condition (x^0, θ^0) with $\theta^0 \neq \theta(\pi/3)$? Determining a family of such orbits would certainly provide support for the argument that $\Omega(KS)$ can be treated as a billiard. As of yet, besides the trivial limit of a sequence of compatible Cantor orbits, no other finitely stabilizing periodic orbits have been found.

6.2 A well-defined fractal flat surface on $\mathcal{S}(KS)$

In §3.1, we calculated the genus g_n of $\mathcal{S}(KS_n)$. It is clear that the genus increases without bound. It is not clear that a suitable limit of flat surfaces $\mathcal{S}(KS_n)$ exists. For the sake of argument, let us suppose such a surface $\mathcal{S}(KS) := \lim \mathcal{S}(KS_n)$ exists. Then, the question is whether or not on $\mathcal{S}(KS) := \lim \mathcal{S}(KS_n)$ there exists a well-defined flow

in any direction.

The first example (and only example) we have seen of a billiard orbit being fixed and defined for every subsequent approximation is a Cantor orbit. Moreover, we then defined the trivial limit of a sequence of compatible Cantor orbits to be a *finitely stabilizing periodic orbit* of the Koch snowflake billiard $\Omega(KS)$. We then can easily see how such a flow line in the proposed surface $\mathcal{S}(KS)$ remains well-defined and fixed for every subsequent approximation. We can also see that a nontrivial polygonal path is determined by utilizing the law of reflection in each approximation and gives us a well-defined path in the limit. This also serves as an example of a geodesic of the proposed flat surface that *should* be well-defined in the limit.

The real problem with trying to force the existence of a flow on some limiting surface $\mathcal{S}(KS)$ is the fact that there are at least infinitely many non-removable conic singularities. Moreover, the process one goes through to construct $\mathcal{S}(KS_n)$ (i.e., by appropriately identifying opposite and parallel sides) certainly does not hold in the limit. Though it is reasonable to believe that a limiting fractal flat surface $\mathcal{S}(KS)$ will consist of six appropriately identified copies of the Koch snowflake billiard $\Omega(KS)$, it is not apparent how one will deal with the presence of infinitely many nonremovable conic singularities. In a much simpler (but not at all simple) case of a single convergent sequence of nonremovable conic singularities, Josh Bowman and Feran Valdez have made progress in understanding the nature of what they are calling *wild singularities* of a flat surface. The case of the Koch snowflake may or may not be dealt with in their current framework, but their work will most likely serve as a motivation for particular plans of attack in answering the question of how to define the flow at elusive limit points of the Koch snowflake fractal flat surface.

6.3 A Veech-like dichotomy

We know from [GtkJu1, GtkJu2] that for every $n \geq 0$, the Veech dichotomy holds for the prefractal flat surface $\mathcal{S}(KS_n)$. What we do not know is the explicit structure of the Veech group $\Gamma(KS_n)$ corresponding to $\mathcal{S}(KS_n)$. More importantly, we do not know how (or if) $\Gamma(KS_n)$ is somehow determined from $\Gamma(KS_0)$ or $\Gamma(KS_1)$. Specifically, $\Gamma(KS_0)$ is uncountable and $\Gamma(KS_n)$ is countable for every $n \geq 1$. This prompts the question: Does the inverse limit $\Gamma(KS) := \varprojlim \Gamma(KS_n)$ exist? If the answer to this question is *yes*, then what is this group *exactly*? Since the inverse limit of an inverse limit sequence of groups is a group, this question is well-posed, but the answer may not be easily determined. Assuming there is a well-defined flat surface $\mathcal{S}(KS)$, we ask whether or not $\Gamma(KS)$ acts on $\mathcal{S}(KS)$ in a well-defined way. The group $\Gamma(KS_n)$ is the group of affine automorphisms that preserves the geometric nature of the conical singularities of the flat surface $\mathcal{S}(KS_n)$. It has been suggested (in a personal communication) by J. Athreya of the University of Illinois, Urbana-Champaign, that a relationship between $\Gamma(KS_n)$ and $\Gamma(KS_m)$, $1 \leq m \leq n$ exists. Since a Veech group has a representation in $\mathrm{PSL}_2(\mathbb{R})$, it is reasonable to ask the following question: Does the group $\Gamma(KS)$ have a representation in $\mathrm{PSL}_2(\mathbb{R})$? If the answer is yes, what structure does this group have? This then begs the question: is there a Veech dichotomy for the fractal flat surface? That is, can we say the flow in a particular direction is either closed or uniquely ergodic?

Bibliography

[At] Athreya, J. S. Quantitative recurrence and large deviations for Teichmuller geodesic flow, *Geometriae Dedicata*, **119** (2006), 121–140.

[AtFo] Athreya, J. S., Forni, G.: Deviation for rational-angled billiards, *Duke Math. J.* No. 2, **144** (2008), 285–319.

[AtEsZo] Athreya, J. S., Eskin, A., Zorich, A.: Rectangular billiards and volumes of spaces of quadratic differentials, in preparation, (2011).

[AtBuEsMi] Athreya, J. S., Bufetov, A., Eskin, A., Mirzakhani, M.: Lattice Point Asymptotics and Volume Growth on Teichmuller space. [E-print: arXiv:math.DS/0610715v1, 2010.] (submitted to Duke Math. J.)

[BaUm] Baxter, A., Umble, R.: Periodic orbits of billiards on an equilateral triangle, *Amer. Math. Monthly* No. 8, **115** (2008), 479–491.

[Bo] Boshernitzan, M.: Billiards and rational periodic directions in polygons, *Amer. Math. Monthly*, **99** (1992), 522–529.

[BoGaKrTr] Boshernitzan, M., Galperin, G., Kruger, T., Troubetzkoy, S.: Periodic billiard orbits are dense in rational polygons, *Trans. Amer. Math. Soc.* **350** (1998), 3523–3535.

[BowVa] Bowman, J., Valdez, F.: Wild singularities of translation surfaces, in preparation, (2011).

[Ch] J. Chazarain, Formule de Poisson pour les variétés riemanniennes, *Invent. Math.* **24** (1974), 65–82.

[CheNie] Chen, J. P., Niemeyer, R. G.: Stabilizing periodic orbits of Sierpinski carpets, 8 pages, in progress, 2011.

[CiHaKo] Cipra, B., Hanson, R., Kolan, R.: Periodic trajectories in right triangle billiards, *Phys. Rev.*, **E52** (1995), 2066–2071.

[Ci] Cipra, B.: 2009 Joint Mathematics Meeting: Taking a cue from infinite kinkiness, *Science* (Feb. 13, 2009) No. 5916, **323** (2009), 874b–875b. [One of two articles written by Barry Cipra about the 2009 Joint Mathematics (AMS-MAA-SIAM) Annual Meeting in Washington, D.C. The present article describes the PI's joint work with Michel Lapidus in [LapNie1].]

[Co1] Y. Colin de Verdière, Spectre du laplacien et longueur des géodésiques périodiques, I et II, *Compositio Math.* **27** (1973), 83–106 and 159–184.

[Co2] Y. Colin de Verdière, Spectrum of the Laplace operator and periodic geodesics: thirty years after, *Ann. Inst. Fourier* No. 7, **57** (2008), 2429–2463.

[Du-CaTy] Durane-Cartagena, E., Tyson, J. T.: Rectifiable curves in Sierpiński carpets, to appear in *Indiana Univ. Math. J.*, 2011.

[DuGn] J. J. Duistermaat and V. Guillemin, The spectrum of positive elliptic operators and periodic bicharacteristics, *Invent. Math.* **29** (1975), 39–79.

[Ed] Edgar, G. A.: "*Measure, Topology, and Fractal Geometry*", 2nd. ed., Springer, New York.

[Fc] Falconer, K. J.: "*Fractal Geometry: Mathematical Foundations and Applications*", 2nd. ed., Wiley, Chichester, 2003.

[Ga] Galperin, G.: Non-periodic and not everywhere dense billiard trajectories in convex polygons and polyhedrons, *Comm. Math. Phys.*, 91 (1983), 187–211.

[GaStVo] Galperin, G., Vorobets, Ya. B., Stepin, A. M.: Periodic billiard trajectories in polygons, *Russian Math. Surveys* No. 3, **47** (1992), 5–80.

[GaZv] Galperin, G., Zvonkine, D.: Periodic billiard trajectories in right triangles and right-angled tetrahedra, *Regul. Chaotic Dyn.*, **8** (2003), 29–44.

[Gro] Gromov, M.: "*Metric Structures for Riemannian and Non-Riemannian Spaces*", Modern Birkhäuser Classics, Birkhäuser, Basel and Boston, 2001.

[Gtk1] Gutkin, E.: Billiards in polygons: Survey of recent results, *J. Stat. Phys.* **83** (1996), 7–26.

[Gtk2] Gutkin, E.: Billiards on almost integrable polyhedral surfaces, *Erg. Th. and Dyn. Syst.* **4** (1984), 569–584.

[GtkJu1] Gutkin, E., Judge, C.: The geometry and arithmetic of translation surfaces with applications to polygonal billiards, *Math. Res. Lett.* **3** (1996), 391–403.

[GtkJu2] Gutkin, E., Judge, C.: Affine mappings of translation surfaces: Geometry and arithmetic, *Duke Math. J.* **103** (2000), 191–213.

[GtkTr] Gutkin, E., Troubetzkoy, S.: Directional flows and strong recurrence for polygonal billiards, in "*Proceedings of the International Congress of Dynamical Systems*", (Montevideo, Uruguay), (F. Ledrappier et al., eds.), (1996), p. 21–45.

[Gz1] M. C. Gutzwiller, Periodic orbits and classical quantization conditions, *J. Math. Phys.* **12** (1971), 343–358.

[Gz2] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics*, Interdisciplinary Applied Mathematics, vol. 1, Springer-Verlag, New York, 1990.

[Hoop] Hooper, W. P.: On the stability of periodic billiard paths in triangles, Ph. D. Dissertation. Mathematics. Stonybrook University, Long Island, NY, 123+xii pages, 2006.

[HoYo] Hocking, J. G., Young, G. S.: "*Topology*", Dover Publ., Mineola, 1988.

[HuSc] Hubert, P., Schmidt, T.: An introduction to Veech surfaces, in: "*Handbook of Dynamical Systems*", vol. 1B (A. Katok and B. Hasselblatt, Eds.), Elsevier, Amsterdam, 2006, 501–526.

[Hut] Hutchinson, J. E.: Fractals and self-similarity, *Indiana Univ. Math. J.* **30** (1981), 713–747.

[HaKa] Katok, A., Hasselblatt, B: "*A First Course in Dynamics*: *With a panorama of recent developments*", Cambridge Univ. Press, Cambridge, 2003.

[KaZe] Katok, A., Zemlyakov, A.: Topological transitivity of billiards in polygons, *Math. Notes* **18** (1975), 760–764.

[Lap1] M. L. Lapidus, Fractal drum, inverse spectral problems for elliptic operators and a partial resolution of the Weyl–Berry conjecture, *Trans. Amer. Math. Soc.* **325** (1991), 465–529.

[Lap2] M. L. Lapidus, Vibrations of fractal drums, the Riemann hypothesis, waves in fractal media, and the Weyl–Berry conjecture, in: *Ordinary and Partial Differential Equations* (B. D. Sleeman and R. J. Jarvis, eds.), vol. IV, Proc. Twelfth Internat. Conf. (Dundee, Scotland, UK, June 1992), Pitman Research Notes in Math. Series, vol. 289, Longman, Scientific and Technical, London, 1993, pp. 126–209.

[LapNie1] Lapidus, M. L., Niemeyer, R. G.: Towards the Koch snowflake fractal billiard—Computer experiments and mathematical conjectures, in "*Gems in Experimental Mathematics*" (T. Amdeberhan, L. A. Medina and V. H. Moll, Eds.), Contemporary Mathematics, Amer. Math. Soc., Providence, R. I., **517** (2010), 231–263. [E-print: arXiv:math.DS.0912.3948v1, 2009.]

[LapNie2] Lapidus, M. L., Niemeyer, R. G.: Families of periodic orbits of the Koch snowflake fractal billiard, 63 pages, submitted for publication, 2011. [E-print: arXiv:1105.0737v1, 2011.]

[LapNie3] Lapidus, M. L., Niemeyer, R. G.: Hybrid periodic orbits of the Koch snowflake fractal billiard, 15 pages, in progress, 2011.

[LapNie4] Lapidus, M. L., Niemeyer, R. G.: Veech groups Γ_n of the Koch snowflake prefractal flat surfaces $\mathcal{S}(KS_n)$, in progress, 2011.

[LapNieRo1] Lapidus, M. L., Niemeyer, R. G., Rock, J. A.: "*An Invitation to Fractal Geometry Dimension Theory, Zeta Functions, and Applications*", book in preparation, 2011. [The book should discuss, in particular, some of the PI and his collaborators' research, as well as the research of MLL and JAR, but in a manner that can be understood by students and other readers with a diverse background (which is quite a challenge in itself). Approx. 120 page have been completed to date, jointly with M. Lapidus and J. Rock, much of which will eventually be significantly modified.]

[LapNieRo2] Lapidus, M. L., Niemeyer, R. G., Rock, J. A.: The Minkowski dimension of an attractor of a particular contraction system (tentative title) (with J. A. Rock and M. L. Lapidus), 2011.

[LapNRG] M. L. Lapidus, J. W. Neuberger, R. J. Renka and C. A. Griffith, Snowflake harmonics and computer graphics: Numerical computation of spectra on fractal domains, *Internat. J. Bifurcation & Chaos* **6** (1996), 1185–1210.

[LapPa] Lapidus, M. L., Pang, M. M. H.: Eigenfunctions of the Koch snowflake domain, Commun. Math. Phys. **172** (1995), 359–376.

[Lapv-Fra] Lapidus, M. L., van Frankenhuijsen, M.: Fractal Geometry, Complex Dimensions and Zeta Functions, (Subtitle: Geometry and Spectra of Fractal Strings.), Springer Monographs in Mathematics, Springer-Verlag, New York, approx. 490 pages (precisely, 460 + (xxiv) pages & 54 illustrations), August 2006.

[McL] S. Mac Lane, *Categories for the Working Mathematician*, 2nd ed., Graduate Text in Mathematics, vol. 5, Springer-Verlag, New York, 1989.

[Ma] W. S. Massey, *Algebraic Topology: An Introduction*, Springer-Verlag, New York, 1977.

[Mas] Masur, H.: Closed trajectories for quadratic differentials with an applications to billiards, *Duke Math. J.* **53** (1986), 307–314.

[MasTa] Masur, H., Tabachnikov, S.: Rational billiards and flat structures, in "*Handbook of Dynamical Systems*", vol. 1A (A. Katok and B. Hasselblatt, Eds.), Elsevier, Amsterdam, 2002, 1015–1090.

[Pe] Petersen, K.: "*Ergodic Theory*", Camb. Univ. Press, New York, 1983.

[PoYu] Pollicot, M., Yuri, M.: "*Dynamical Systems and Ergodic Theory*", Camb. Univ. Press, Cambridge, 1998.

[Ru] Ruijgrok, T.: Periodic orbits in triangular billiards, *Acta Physica Polonica*, **B22** (1991), 955–981.

[ScTr] Schmeling, J., Troubetzkoy, S.: Inhomogeneous Diophantine approximation and angular recurrence for polygonal billiards, *Math. Sb.*, **194** (2003), 295–309.

[Swz] Schwartz, R. E.: "*Mostly Surfaces*", Amer. Math. Soc., Rhode Island, 2011.

[Si] Silva, C. E.: "*Invitation to Ergodic Theory*", Amer. Math. Soc., Rhode Island, 2008.

[Sm] Smillie, J.: Dynamics of billiard flow in rational polygons, in "*Dynamical Systems*", Encyclopedia of Math. Sciences, vol. 100, Math. Physics 1 (Ya. G. Sinai, Ed.), Springer-Verlag, New York, 2000, pp. 360–382.

[Ta1] Tabachnikov, S.: "*Billiards*", Panoramas et Synthèses, Soc. Math. France, 1995.

[Ta2] Tabachnikov, S.: "*Geometry and Billiards*", Amer. Math. Soc., Rhode Island, 2005.

[Thur] Thurston, W.: "*Three-dimensional Geometry and Topology, Vol 1*", Princeton Univ. Press., Princeton, 1997.

[Tr] Troubetzkoy, S.: Recurrence and periodic billiard orbits in polygons, *Regul. Chaotic Dyn.* No. 1, **9** (2004), 1–12.

[Ve1] Veech, W.: Teichmüller geodesic flow, *Annals of Math.* **124** (1986), 441–530.

[Ve2] Veech, W.: Teichmüller curves in modular space, Eisenstein series, and an application to triangular billiards, *Invent. Math.* **97** (1989), 553–583.

[Ve3] Veech, W. A.: The billiard in a regular polygon, Geom. Fmwt. Anal. 2:341-379 (1992).

[Ve4] Veech, W. A.: Flat surfaces, *Amer. J. Math.* **115** (1993), 589–689.

[Vo] Vorobets, Ya. B.: Plane structures and billiards in rational polygons: The Veech alternative, *Russian Math. Surveys* **51** (1996), 779–817.

[We-Sc] Weitze-Schmithüsen, G.: An algorithm for finding the Veech group of an origami, *Experimental Mathematics* No.4, **13** (2004), 459–472.

[Zo] Zorich, A.: Flat surfaces, in "*Frontiers in Number Theory, Physics and Geometry* I" (P. Cartier, *et al.*, eds.), Springer-Verlag, Berlin, 2002, pp. 439–585.