The Pennsylvania State University
The Graduate School
Eberly College of Science

BOUNDARY MAPS AND THEIR NATURAL EXTENSIONS ASSOCIATED WITH FUCHSIAN AND KLEINIAN GROUPS

A Dissertation in
Mathematics
by
Adam J. Zydney

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2018

The dissertation of Adam J. Zydney was reviewed and approved* by the following:

Svetlana Katok
Professor of Mathematics
Dissertation Advisor, Chair of Committee

Anatole Katok
Raymond N. Shibley Professor of Mathematics

Federico Rodriguez Hertz
Professor of Mathematics

Kyusun Choi
Associate Professor of Computer Science and Engineering

Yuxi Zheng
Francis R. and Helen M. Pentz Professor of Science
Head of the Department of Mathematics

* Signatures are on file in the Graduate School.

Abstract

Geodesic flows on surfaces of constant negative curvature are a rich source of examples in ergodic theory, and geodesic flow on the modular surface in particular has deep connections to real continued fractions from number theory. This thesis deals with two extensions of this setting: either replacing the modular group $\mathrm{PSL}(2,\mathbb{Z})$ with a cocompact torsion-free Fuchsian group, or working with three-dimensional hyperbolic space and relating the boundary maps to continued fractions of complex numbers.

The Fuchsian results (Chapter III) are joint with Svetlana Katok and build on results of Katok and Ugarcovici, who studied a family of maps generalizing the Bowen-Series boundary map. When the parameters satisfy the short cycle property, i.e., the forward orbits at each discontinuity point coincide after one step, the natural extension map has a global attractor with finite rectangular structure. In this thesis, we generalize several results of Adler and Flatto, describing a conjugacy between the geometric and arithmetic maps and showing that the attractor parameterizes an associated arithmetic cross-section. This allows us to represent the geodesic flow as a special flow over a symbolic system. In cases where the cycle ends are discontinuity points, the resulting symbolic system is sofic.

In Chapter IV, we consider three-dimensional real hyperbolic space, in which the boundary is the Reimann sphere $\mathbb{C} \cup \{\infty\}$ and the boundary maps use generators of Kleinian groups. Here the endpoints of geodesics can be described by complex continued fractions, which have been studied from the number theoretic perspective by Doug Hensley, S. G. Dani, and Arnaldo Nogueira, among others. In this thesis, a new "partition property" is described, substituting for the cycle property seen in the modular and Fuchsian literature. We state some results that apply to a wide range of boundary maps satisfying this partition property, and we discuss several specific algorithms in more detail. In many cases, the attractor of the natural extension map can be expressed as a finite union of products in $\mathbb{C} \times \mathbb{C}$; this "finite product structure" is explicitly demonstrated for certain algorithms.

Table of Contents

List of Figures

Acknowledgements

I would like to express my thanks and gratitude to all those who helped and supported me during my time as a graduate student and throughout my life. Deserving of special recognition are the following:

- My parents, Laurel and Andrew Zydney. All of my successes both past and future would be impossible without their dedication to support and provide for me, and I will never be able to thank them enough for all that they have given me. My siblings, Sarah and Benjamin, are also a constant source of love and inspiration.

- My advisor, Svetlana Katok. I first worked with Svetlana as an undergraduate, taking several upper-level courses with her. Ultimately, she became my graduate advisor, and my research is heavily based on results and concepts from her work. Over the years, her efforts as a mentor, advisor, and advocate for me have been invaluable.

- My other committee members: Anatole Katok, Federico Rodriguez-Hertz, and Kyusun Choi. In addition to forming my committee, each of them have also taught undergraduate or graduate courses I have taken, and I greatly value all of their contributions to my education.

- The dynamical systems and broader faculty of the Math Department, including Yakov Pesin, Misha Guysinsky, Mark Levi, Sergei Tabachnikov, and others.

- The many dynamics students at Penn State, especially Dom Veconi, David Hughes, Alena Erchenko, Changguang Dong, and Daren Wei, as well as Shilpak Banerjee, Dong Chen, and Kurt Vinhage.

- Other friends and mentors, most notably David Zach, whose unwavering friendship I have particularly appreciated in recent years, and Jeff Bennett, whose conversations at an early age through today never fail to excite my academic curiosity.

Chapter I: Introduction

The generators

$$T(x) = x + 1 \qquad \text{and} \qquad S(x) = \frac{-1}{x}$$

of $\mathrm{PSL}(2, \mathbb{Z})$ can be thought of as Möbius transformations on the hyperbolic plane, and they naturally appear in the construction of minus continued fractions:

$$a_0 - \cfrac{1}{a_1 - \cfrac{1}{a_2 - \cfrac{1}{\ddots - \frac{1}{a_n}}}} = T^{a_0} S T^{a_1} S T^{a_2} \cdots S T^{a_n} S(\infty).$$

The connection between continued fractions and geodesic flow on the modular surface $\mathcal{H}^2 \backslash \mathrm{PSL}(2, \mathbb{Z})$ goes back to Artin [6], with further development by Caroline Series [36, 38] and Adler and Flatto [2, 3]. Katok and Ugarcovici [22, 23, 24] detailed a two-parameter family of continued fraction algorithms leading to the so-called (a, b)-continued fractions, which have applications in both number theory and dynamics.

Continued fractions for complex numbers have been studied by Adolf Hurwitz [16], Doug Hensley [15], and more recently by S. G. Dani and Arnaldo Nogueira [12].

The main result of [23, 25] is that almost any (a, b)-continued fraction algorithm can be used to code geodesics on the modular surface "arithmetically," and the associated arithmetic cross-section is parameterized by a set in $\mathbb{R}^2$ that has "finite rectangular structure."

In this thesis, we discuss some extensions of these ideas. In particular, we either replace $\mathrm{PSL}(2, \mathbb{Z})$ with certain other Fuchsian groups, or we work with 3-dimensional hyperbolic space, whose boundary is the Reimann sphere $\mathbb{C} \cup \{\infty\}$.

This document is organized as follows. Chapter I introduces the basic concepts and content. Chapter III describes results, joint with Svetlana Katok, allowing arithmetic coding of geodesics on factors of the hyperbolic plane by Fuchsian surface groups. Chapter IV describes several algorithms for complex continued fractions and demonstrates important properties of their associated dynamical systems. Each of Parts III and IV first presents a summary of previously known results (Sections 4 and 6) and then discusses new results (Sections 5 and 7).

1. Background

1.1. Symbolic dynamics

Let $\mathcal{A}$ be a finite or countable alphabet. Then the space $\mathcal{A}^{\mathbb{Z}} = \{\, s = \{a_i\}_{i \in \mathbb{Z}} \,:\, a_i \in \mathcal{A} \,\}$ consists of all bi-infinite sequences, and we endow this space with the Tikhonov

(product) topology. The map

$$\sigma : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}} \text{ defined by } (\sigma s)_i = a_{i+1}$$

is a left shift. A *symbolic dynamical system* is a pair (Λ, σ), where $\Lambda \subset \mathcal{A}^{\mathbb{Z}}$ is a closed σ-invariant subset. It is also possible to do symbolic dynamics with one-sided sequences in $\mathcal{A}^{\mathbb{N}}$, in which case the shift is not invertible.

There are some important classes of symbolic dynamical systems (definitions here are for two-sided shifts).

- The space $(\mathcal{A}^{\mathbb{Z}}, \sigma)$ is called the *full shift* or a *topological Bernoulli shift*.
- For any $a, b \in \mathcal{A}$, let $m_{a,b}$ be either zero or one. Then let

  $$\Sigma_M = \left\{ \{a_i\} \in \mathcal{A}^{\mathbb{Z}} \,:\, m_{a_i, a_{i+1}} = 1 \text{ for all } i \in \mathbb{Z} \right\}.$$

  The system (Σ_A, σ) is called a *one-step topological Markov chain* or simply a *topological Markov chain* or sometimes a *subshift of finite type*. When the alphabet is finite, it is common to think of $\{m_{i,j}\}$ as a matrix.
- A factor of a topological Markov chain is called a *sofic shift*. That is, (Λ, σ) is a sofic shift if there exists a surjective continuous map $h : \mathcal{A}^{\mathbb{Z}} \to \Lambda$ with $h \circ \sigma = \sigma \circ h$.

1.2. Continued fractions

A *continued fraction* is any expression of the form

$$a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{\ddots + \frac{b_n}{a_n}}}} \qquad \text{or} \qquad a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{\ddots}}}.$$

Often, a_i and b_i are real or complex numbers and are required to come from restricted sets of numbers (e.g., only positive integers). From a dynamics perspective, infinite continued fractions are much more interesting.

Given two sequences $\{a_n\}$ and $\{b_n\}$ in any field, one can define sequences $\{p_n\}$ and $\{q_n\}$ by

$$\begin{array}{lll} p_{-2} = 0 & p_{-1} = 1 & p_n = a_n p_{n-1} + b_n p_{n-2} \quad \text{for } n \geq 0; \\ q_{-2} = -1 & q_{-1} = 0 & q_n = a_n q_{n-1} + b_n q_{n-2} \quad \text{for } n \geq 0. \end{array}$$

Then algebraic manipulations show that for all $n \geq 0$,

$$\frac{p_n}{q_n} = a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{\ddots + \frac{b_n}{a_n}}}} \tag{1}$$

assuming $a_n \neq 0$. The fraction p_n/q_n is called the n^{th} *convergent* of the continued fraction.

A *plus continued fraction* is one in which $b_n = 1$ for all n, and a *minus continued fraction* is one in which $b_n = -1$ for all n. When one of these conventions is used throughout a work—for the remainder of this thesis, all c.f. are minus—we can denote the expression in (1) by $[a_0; a_1, a_2, \ldots a_n]$ and its limit, if it exists, by $[a_0; a_1, a_2, \ldots]$.

The a_i are called the *digits* of the continued fraction (in some works they are called "partial quotients"), and in most cases they are required to be in a restrictive set such as $\mathbb{N}$ (here a_0 is allowed to be in $\mathbb{Z}$), $\mathbb{Z}$, or $\mathbb{Z}[i]$.

Given a value $x \in \mathbb{R}$ or $x \in \mathbb{C}$, there are various algorithms that can be used to construct a finite or infinite sequence $(a_0, a_1, \ldots)$ such that $[a_0; a_1, \ldots]$ converges to x (or equals x at some point if the sequence is finite). See Section 6.1 for more details.

1.3. Fuchsian and Kleinian groups

There are many models of hyperbolic spaces and many different constructions of these models. See [35] or [29, Ch. 4] for general definitions involving Hermitian forms. In this thesis we will mainly consider the disk model of 2-dimensional hyperbolic space and the half-space model of 3-dimensional hyperbolic space.

Two-dimensional hyperbolic space

Let $\mathbb{D} = \{ z \in \mathbb{C} : |z| < 1 \}$, and endow this disk with the Poincaré metric

$$\mathrm{d}s^2 = \frac{4\,\mathrm{d}z\,\mathrm{d}\overline{z}}{(1 - |z|^2)^2}.$$

In this model, geodesics are diameters and arcs of circles orthogonal to the boundary $\partial\mathbb{D} = \mathbb{S}$.

A *Fuchsian group* is a discrete group of orientation-preserving isometries of hyperbolic 2-space. The most basic example of a Fuchsian group is $\mathrm{PSL}(2, \mathbb{Z})$, often called the *modular group*. For a finitely generated Fuchsian group, the factor space $M = \Gamma\backslash\mathcal{H}^2$ is a surface of constant negative curvature, possibly with some singularities (fixed points of elliptic elements) and punctures (cusps), and, in the case of infinite volume, funnels. For more detail, see [7] or [19]. If Γ acts freely on $\mathbb{D}$ with $\mathbb{D}\backslash\Gamma$ a compact domain, then Γ is called a *surface group*, and the quotient $\mathbb{D}\backslash\Gamma$ is a compact surface of constant negative curvature -1 of a certain genus $g > 1$. This is the main setting of Chapter III.

Three-dimensional hyperbolic space

The upper half-space $\{\,(x,y,h)\in\mathbb{R}^3\,:\,h>0\,\}$ can be endowed with the hyperbolic metric

$$\mathrm{d}s^2=\frac{\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}h^2}{h^2},$$

and we denote the resulting space by $\mathcal{H}^3$. In the half-plane model, geodesics are vertical rays and semicircles orthogonal to the xy-plane. The boundary of $\mathcal{H}^3$ is a Riemann sphere, and we naturally associate any $(x,y,0)$ with the complex number $x+iy$ (there is also one point at infinity).

An orientation-preserving isometry of $\mathcal{H}^3$ acting on the boundary $\mathbb{C}\cup\{\infty\}$ can be modeled as some

$$z\mapsto\frac{az+b}{cz+d},\qquad a,b,c,d\in\mathbb{C},\ ad-bc=1.$$

Using the trace $t=a+d$, these can be classified as elliptic ($t^2<4$), parabolic ($t^2=4$), purely hyperbolic ($t^2>4$), or loxodromic ($t^2\notin\mathbb{R}$, which clearly cannot occur in Fuchsian groups).

A *Kleinian group* is a discrete group of orientation-preserving isometries of $\mathcal{H}^3$. The most basic example of a Kleinian group is $\mathrm{PSL}(2,\mathbb{Z}[i])$, often called the Picard group. Standard generators for $\mathrm{PSL}(2,\mathbb{Z}[i])$ acting on the boundary $\mathbb{C}\cup\{\infty\}$ are the maps sending z to $z+1, z+i, -1/z$, and $-z$.

Two specific Kleinian groups will be used in this thesis. The first is

$$\Gamma_1:=\left\langle\, z\mapsto z+1,\ z\mapsto z+i,\ z\mapsto\tfrac{-1}{z}\,\right\rangle,$$

which is associated to minus continued fractions with digits in $\mathbb{Z}[i]$. The second,

$$\Gamma_\alpha:=\left\langle\, z\mapsto z+(1+i),\ z\mapsto z+(1-i),\ z\mapsto\tfrac{-1}{z}\,\right\rangle,$$

occurs when the digits are restricted to satisfying $\operatorname{Re}a+\operatorname{Im}a$ even. There do not appear to be any standard names or notations for either of these groups in literature.

1.4. Geodesic flow and cross-sections

The *geodesic flow* $\{\widetilde{\varphi}^t\}$ on a Riemannian manifold H is defined as an $\mathbb{R}$-action on the unit tangent bundle SH that moves a tangent vector along the geodesic defined by this vector with unit speed. If Γ is some discrete group of orientation-preserving isometries of H, then the geodesic flow $\{\widetilde{\varphi}^t\}$ on H descends to the geodesic flow $\{\varphi^t\}$ on the factor

$$M=\Gamma\backslash H$$

via the canonical projection $\pi:SH\to SM$ of the unit tangent bundles (again, geodesic flow "on M" is in fact an $\mathbb{R}$-action on SM). Geodesics on M are orbits of the geodesic flow $\{\varphi^t\}$.

Often the quotient $M = H\backslash\Gamma$ will have cusps (for example, with the half-plane model of $\mathcal{H}^2$, the modular surface $\mathcal{H}^2\backslash\mathrm{PSL}(2,\mathbb{R})$ has one cusp at ∞). We will consider only oriented geodesics which do not go to the cusp(s) of M in either direction, and the phrase "every geodesic" in this document will refer to every oriented geodesic from this set. The set of excluded geodesics is insignificant from the measure-theoretic point of view (see [21, p. 59]).

A *cross-section* C for the geodesic flow $\{\varphi^t\}$ is a subset of the unit tangent bundle SM visited by (almost) every geodesic infinitely often both in the future and in the past. It is well-known that the geodesic flow can be represented as a *special flow* on the space

$$C^h = \{\, (v,s) : v \in C, 0 \le s \le h(v) \,\}.$$

It is given by the formula $\varphi^t(v,s) = (v, s+t)$ with the identification $(v, h(v)) = (R(v), 0)$, where the ceiling function $h : C \to \mathbb{R}$ is the *time of the first return* of the geodesic defined by v to C, and $R : C \to C$ given by $R(v) = \varphi^{h(v)}(v)$ is the *first return map* of the geodesic to C.

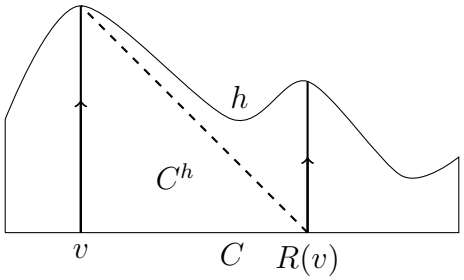


FIGURE 1: Geodesic flow is a special flow.

In order to represent the geodesic flow as a special flow over a symbolic dynamical system, one needs to choose an appropriate cross-section C and "code it," that is, find an appropriate symbolic dynamical system (Λ, σ) and a continuous surjective map $\mathrm{Cod} : \Lambda \to C$ (in some cases the actual domain of Cod is Λ except a finite or countable set of excluded sequences) defined such that $R \circ \mathrm{Cod} = \mathrm{Cod} \circ \sigma$. We can then talk about *coding sequences* for geodesics defined up to a shift that corresponds to a return of the geodesic to the cross-section C. Usually the coding map is not injective but only finite-to-one (see, e.g., Examples 3 and 4 in Section 5.4).

Chapter II: Modular setting

2. Cycle structure

Katok and Ugarcovici [23, 24] produced many results around a two-parameter family of continued fraction algorithms called (a, b)-continued fractions. For $a, b \in \mathbb{R}$ satisfying

$$a \leq 0 \leq b, \quad b - a \geq 1, \quad -ab \leq 1,$$

they define the "generalized integer part" function

$$\lfloor x \rceil_{a,b} = \begin{cases} \lfloor x - a \rfloor & \text{if } x < a \\ 0 & \text{if } a \leq x \leq b \\ \lceil x - b \rceil & \text{if } x \geq b \end{cases}$$

and the map

$$f_{a,b}(x) = \begin{cases} x + 1 & \text{if } x < a \\ -1/x & \text{if } a \leq x \leq b \\ x - 1 & \text{if } x \geq b. \end{cases}$$

Definition. The parameters (a, b) satisfy the *cycle property* if there exist $k_1, m_1, k_2, m_2 > 0$ such that

$$f_{a,b}^{m_1}(Sa) = f_{a,b}^{k_1}(Ta) \qquad \text{and} \qquad f_{a,b}^{m_2}(T^{-1}b) = f_{a,b}^{k_2}(Sb),$$

where $T(x) = x + 1$ and $S(x) = -1/x$.

Figure 2 shows the cycle structure for the example $(a, b) = (-\frac{4}{5}, \frac{2}{5})$.

In [23] it is proved that except for a set of zero measure on the boundary $b - a = 1$ of the parameter space, either the orbits of $Sa, Ta, T^{-1}b, Sb$ under $f_{a,b}$ are eventually periodic or the cycle property is satisfied.

3. Partitions

Katok and Ugarcovici describe the structure of the attractors for their maps using "levels:" the y- and x-values of horizontal and vertical segments, respectively, of the boundary of an attractor. Here we present an alternative description of an attractor as a finite union of rectangles, analogous to the methods in Chapter IV of this thesis. We start by constructing two partitions of $[a, b]$.

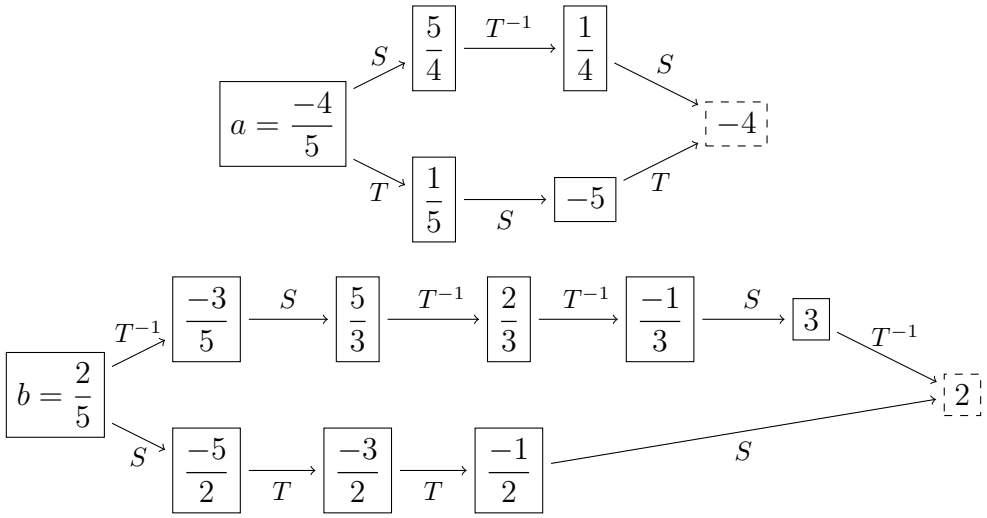


FIGURE 2: Cycles for $(a, b) = (\frac{-4}{5}, \frac{2}{5})$.

Let $\mathscr{C}_{a,b}$ be the set of all points in the cycles of a and b (not including the cycle ends), and let $\mathcal{P}_{a,b}$ be the partition of $[a, b]$ into intervals using the values in $\mathscr{C}_{a,b} \cap [a, b]$ as endpoints of the intervals. Setting

$$X_{a,b}(n) = \left\{ x \in [a, b] \,:\, \lfloor -1/x \rceil_{a,b} = n \right\},$$

we then define a countable partition $\mathcal{C}_{a,b}$ of $[a, b]$ by

$$\mathcal{C}_{a,b} = \{ \mathscr{X} \cap X_{a,b}(n) \,:\, \mathscr{X} \in \mathcal{P}_{a,b}, n \in \mathbb{Z} \}.$$

From the cycles in Figure 2, we see that for $(a, b) = (-\frac{4}{5}, \frac{2}{5})$ we have

$$\mathscr{C}_{\frac{-4}{5},\frac{2}{5}} \cap [-\tfrac{4}{5}, \tfrac{2}{5}] = \left\{ \frac{-4}{5}, \frac{-3}{5}, \frac{-1}{2}, \frac{-1}{3}, \frac{1}{5}, \frac{1}{4}, \frac{2}{5} \right\},$$

so we can construct $\mathcal{P}_{\frac{-4}{5},\frac{2}{5}}$ using these values as endpoints of intervals. This partition is shown in the top part of Figure 3 (the superscript indices 1 through 6 correspond to the ordering of the values in the intervals in this example, but the numeric values of the indices of $\mathscr{X}^{(j)}_{a,b} \in \mathcal{P}_{a,b}$ are irrelevant in general). The middle of Figure 3 shows the sets $X_{\frac{-4}{5},\frac{2}{5}}(n)$ (note that $X_{\frac{-4}{5},\frac{2}{5}}(-1) = X_{\frac{-4}{5},\frac{2}{5}}(0) = \emptyset$), and the bottom shows the partition $\mathcal{C}_{\frac{-4}{5},\frac{2}{5}}$.

All but finitely many elements of $\mathcal{C}_{a,b}$ will be some $X_{a,b}(n)$, with the only exceptions being that the $X_{a,b}(n)$ containing points in $\mathscr{C}_{a,b}$ must be split into multiple intervals of the form $X_{a,b}(n) \cap \mathscr{X}^{(j)}$. For $(a, b) = (\frac{-4}{5}, \frac{2}{5})$, these are $n = 2, 3, -4, -5$.

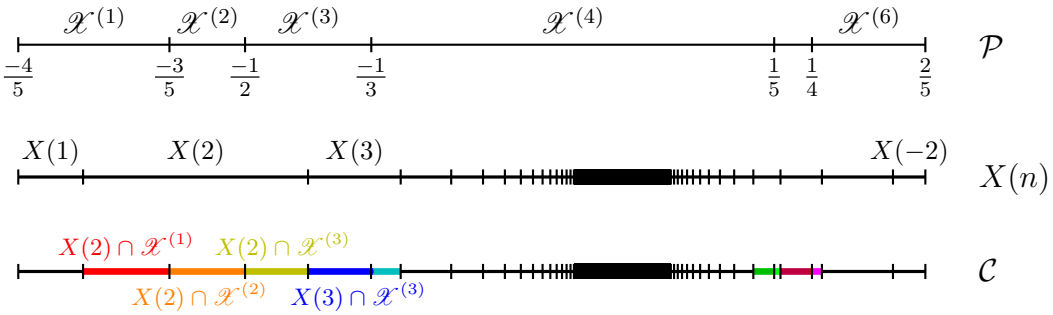


FIGURE 3: Partitions of $[-\frac{4}{5}, \frac{2}{5}]$ (subscripts omitted for readability).

Using the results of [24], the attractor $\hat{\Lambda}_{a,b}$ of

$$\hat{F}_{a,b}(x,y) = (T^{-n}Sx, ST^{-n}y), \qquad n = \lfloor Sx \rceil_{a,b},$$

has finite rectangular structure. We can describe the specific example $\hat{\Lambda}_{\frac{-4}{5},\frac{2}{5}}$ as

$$\hat{\Lambda}_{\frac{-4}{5},\frac{2}{5}} = \bigcup_{j=1}^{6} \mathscr{X}^{(j)}_{\frac{-4}{5},\frac{2}{5}} \times \mathscr{Y}^{(j)}_{\frac{-4}{5},\frac{2}{5}},$$

(see Figure 4) where the $\mathscr{X}^{(j)}_{\frac{-4}{5},\frac{2}{5}}$ are the elements of the finite partition $\mathcal{P}_{\frac{-4}{5},\frac{2}{5}}$ and the intervals $\mathscr{Y}^{(j)}_{\frac{-4}{5},\frac{2}{5}}$ are

$$\mathscr{Y}^{(1)}_{\frac{-4}{5},\frac{2}{5}} = [\tfrac{-1}{2}, 0] \qquad \mathscr{Y}^{(2)}_{\frac{-4}{5},\frac{2}{5}} = [\tfrac{-1}{2}, \tfrac{1}{2}] \qquad \mathscr{Y}^{(3)}_{\frac{-4}{5},\frac{2}{5}} = [\tfrac{-1}{3}, \tfrac{1}{2}]$$
$$\mathscr{Y}^{(4)}_{\frac{-4}{5},\frac{2}{5}} = [\tfrac{-1}{3}, \tfrac{2}{3}] \qquad \mathscr{Y}^{(5)}_{\frac{-4}{5},\frac{2}{5}} = [0, \tfrac{2}{3}] \qquad \mathscr{Y}^{(6)}_{\frac{-4}{5},\frac{2}{5}} = [0, 1].$$

Note that the $\mathscr{Y}^{(j)}_{\frac{-4}{5},\frac{2}{5}}$ intervals overlap (they do not form a partition). While the endpoints of $\mathscr{X}^{(j)}_{\frac{-4}{5},\frac{2}{5}}$ come from the the cycles, the endpoints of each $\mathscr{Y}^{(j)}_{\frac{-4}{5},\frac{2}{5}}$ come from an overdetermined system as described in [23, Sec. 5].

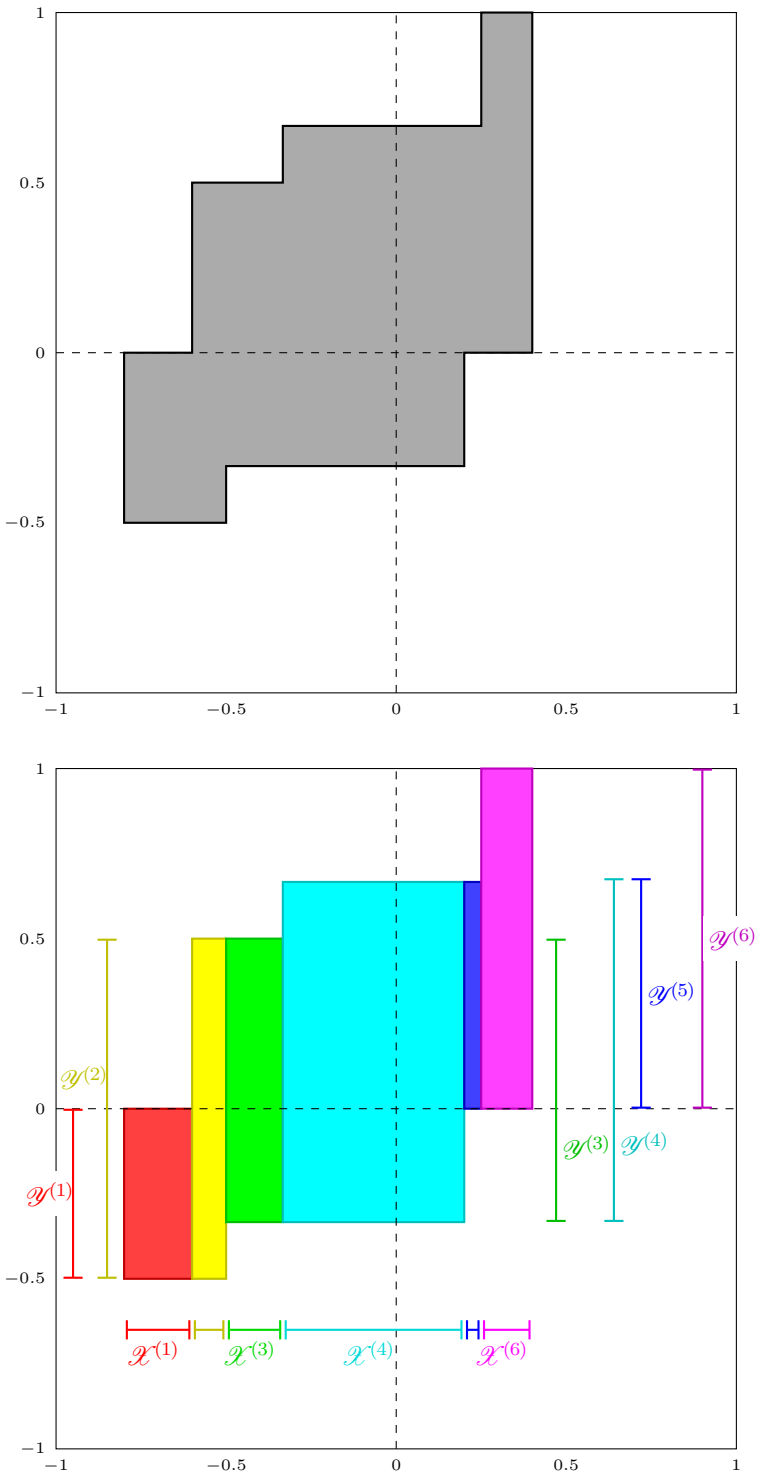


FIGURE 4: Top: the attractor $\hat{\Lambda}_{\frac{-4}{5},\frac{2}{5}}$ of $\hat{F}_{\frac{-4}{5},\frac{2}{5}}$.
Bottom: colored by rectangle and showing the intervals $\mathscr{X}^{(j)}_{\frac{-4}{5},\frac{2}{5}}$ and $\mathscr{Y}^{(j)}_{\frac{-4}{5},\frac{2}{5}}$.

Chapter III: Fuchsian setting

The contents of Chapter III are essentially reproduced from [27].

This chapter is organized as follows. Section 4 recalls the definitions of the relevant maps as well as previous results on attractors, while Section 5 presents the new results of [27]. Section 5.1 describes a conjugacy between curvilinear and rectilinear maps, extending Adler-Flatto's phenomenon of "bulges mapping to corners" to a broader class of boundary maps. In Section 5.2, we define a class of arithmetic cross-section for the geodesic flow on $M = \Gamma\backslash\mathbb{D}$, and Section 5.3 describes the arithmetic coding of geodesics using these cross-sections. Although the arithmetic cross-sections we consider actually coincide with the geometric one, the symbolic dynamical systems are different (some examples are discussed in Section 5.4). While the set of geometric coding sequences is never Markov, for some choice of parameters the arithmetic ones are. This is discussed in Sections 5.5 and 5.6. In Section 5.7 we discuss dual codes, and in Section 5.8 we apply results of [25] to obtain explicit formulas for invariant measures and calculate the measure-theoretic entropy.

4. Background

4.1. Adler and Flatto

Adler and Flatto have written three papers [2, 3, 4] devoted to the representation of the geodesic flow on surfaces of constant negative curvature as special flow over a topological Markov chain. The first two papers are devoted to the modular surface and the third to the compact surface $M = \Gamma\backslash\mathbb{D}$ of genus $g \geq 2$, where $\mathbb{D} = \{ z \in \mathbb{C} : |z| < 1 \}$ is the unit disk endowed with hyperbolic metric

$$\frac{2\,|dz|}{1 - |z|^2} \tag{2}$$

and Γ is a finitely generated Fuchsian group of the first kind acting freely on $\mathbb{D}$.

Recall that geodesics in this model are half-circles or diameters orthogonal to $\mathbb{S} = \partial\mathbb{D}$, the circle at infinity. The geodesic flow $\{\widetilde{\varphi}^t\}$ on $\mathbb{D}$ is defined as an $\mathbb{R}$-action on the unit tangent bundle $S\mathbb{D}$ that moves a tangent vector along the geodesic defined by this vector with unit speed. The geodesic flow $\{\widetilde{\varphi}^t\}$ on $\mathbb{D}$ descends to the geodesic flow $\{\varphi^t\}$ on the factor $M = \Gamma\backslash\mathbb{D}$ via the canonical projection

$$\pi : S\mathbb{D} \to SM$$

of the unit tangent bundles. The orbits of the geodesic flow $\{\varphi^t\}$ are oriented geodesics on M.

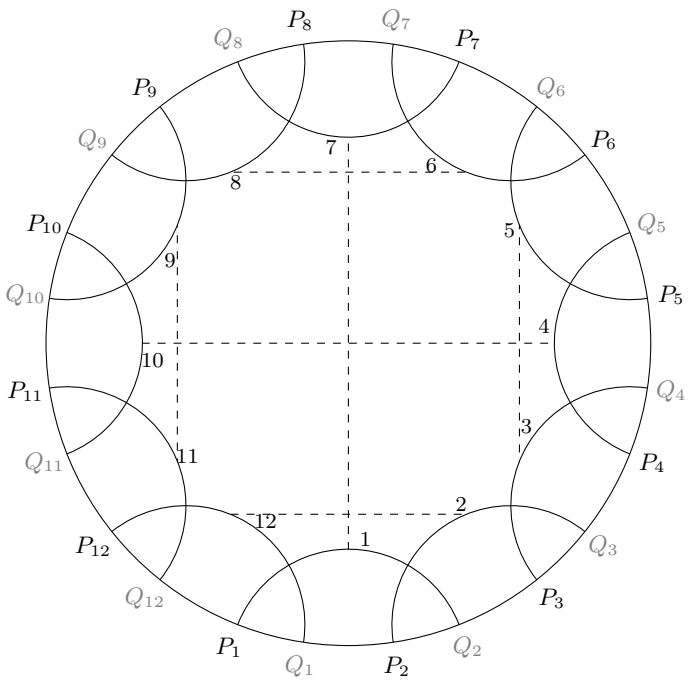


FIGURE 5: Side-pairing of the fundamental domain $\mathcal{F}$ for genus $g = 2$.

A classical (Ford) fundamental domain for Γ is a $4g$-sided regular polygon centered at the origin. In [4], Adler and Flatto used another fundamental domain—an $(8g-4)$-sided polygon $\mathcal{F}$—that was much more convenient for their purposes. Its sides are geodesic segments which satisfy the *extension condition*: the geodesic extensions of these segments never intersect the interior of the tiling sets $\gamma\mathcal{F}$, $\gamma \in \Gamma$.

The fundamental polygon $\mathcal{F}$ will play an important role in this paper, so let us describe it in detail. We label the sides of $\mathcal{F}$ in a counterclockwise order by numbers $1 \leq i \leq 8g-4$ and label the vertices of $\mathcal{F}$ by V_i so that side i connects V_i to V_{i+1} (mod $8g-4$).

We denote by P_i and Q_{i+1} the endpoints of the oriented infinite geodesic that extends side i to the circle at infinity $\mathbb{S}$.[1] The order of endpoints on $\mathbb{S}$ is the following:

$$P_1, Q_1, P_2, Q_2, \ldots, Q_{8g-4}.$$

The identification of the sides of $\mathcal{F}$ is given by the side pairing rule

$$\sigma(i) := \begin{cases} 4g - i \bmod (8g-4) & \text{if } i \text{ is odd} \\ 2 - i \bmod (8g-4) & \text{if } i \text{ is even} \end{cases}$$

[1] The points P_i and Q_i from this thesis and [9, 25, 27] are denoted by a_i and b_{i-1}, respectively, in [4].

The generators T_i of Γ associated to this fundamental domain are Möbius transformations satisfying the following properties:

$$T_{\sigma(i)}T_i = \mathrm{Id} \tag{3}$$
$$T_i(V_i) = V_{\sigma(i)+1} \tag{4}$$
$$T_{\rho^3(i)}T_{\rho^2(i)}T_{\rho(i)}T_i = \mathrm{Id}, \tag{5}$$

where $\rho(i) = \sigma(i) + 1$.

Additionally, the following useful property of T_i will be referenced throughout.

Proposition 4.1 (Proposition 2.2 in [25] and Theorem 3.4 in [4])**.** *The map T_i sends the points P_{i-1}, P_i, Q_i, P_{i+1}, Q_{i+1}, Q_{i+2} respectively to $P_{\sigma(i)+1}$, $Q_{\sigma(i)+1}$, $Q_{\sigma(i)+2}$, $P_{\sigma(i)-1}$, $P_{\sigma(i)}$, $Q_{\sigma(i)}$.*

According to B. Weiss [42], the existence of such a fundamental polygon is an old result of Dehn, Fenchel, and Nielsen, while J. Birman and C. Series [8] attribute it to Koebe [28]. Adler and Flatto [4, Appendix A] give a careful proof of existence and properties of the fundamental $(8g-4)$-sided polygon for any surface group Γ such that $\Gamma\backslash\mathbb{D}$ is a compact surface of genus g. Notice that in general the polygon $\mathcal{F}$ need not be regular. If $\mathcal{F}$ is regular, it is the Ford fundamental domain, i.e., P_iQ_{i+1} is the isometric circle for T_i, and $T_i(P_iQ_{i+1}) = Q_{\sigma(i)+1}P_{\sigma(i)}$ is the isometric circle for $T_{\sigma(i)}$ so that the inside of the former isometric circle is mapped to the outside of the latter, and all internal angles of $\mathcal{F}$ are equal to $\frac{\pi}{2}$. Figure 5 shows such a construction for $g = 2$.

Remark 4.2. Although the results in [25] were obtained using the regular fundamental polygon, the general case is reduced to this specific situation by the Fenchel-Nielsen theorem without affecting the results of [25], as explained in [4, Appendix A] and in [26]. More precisely, given $\Gamma'\backslash\mathbb{D}$ a compact surface of genus $g > 1$, there exists a Fuchsian group Γ and (by the Fenchel-Nielsen theorem [41]) an orientation-preserving homeomorphism h from $\bar{\mathbb{D}}$ onto $\bar{\mathbb{D}}$ such that

1. $\Gamma\backslash\mathbb{D}$ is a compact surface of the same genus g;
2. Γ has a fundamental domain $\mathcal{F}$ given by a regular $(8g-4)$-sided polygon;
3. $\Gamma' = h \circ \Gamma \circ h^{-1}$;
4. Γ' has a fundamental domain $\mathcal{F}'$ given by an $(8g-4)$-sided polygon whose sides are produced by the geodesics $h(P_i)h(Q_{i+1})$ and identified by $T_i' = h \circ T_i \circ h^{-1}$, where $\{T_i\}$ are generators of Γ identifying the sides of $\mathcal{F}$.

We use the regular polygon in some proofs for convenience.

Slightly paraphrased, Adler and Flatto's method of representing the geodesic flow on M as a special flow is the following. Consider the set C_G of unit tangent vectors $(z, \zeta) \in S\mathbb{D}$ based on the boundary of $\mathcal{F}$ and pointed inside $\mathcal{F}$.[2] Every geodesic γ in $\mathbb{D}$ is equivalent to one that intersects $\mathcal{F}$ and thus is comprised of countably many segments in $\mathcal{F}$. More precisely, if we start with a segment $\gamma \cap \mathcal{F}$ which enters $\mathcal{F}$ through side j and exits $\mathcal{F}$ through side i, then $T_i\gamma \cap \mathcal{F}$ is the next segment in $\mathcal{F}$, and $T_j\gamma \cap \mathcal{F}$ is the previous segment. The canonical projection of γ to SM, $\pi(\gamma)$, visits C_G infinitely often, hence C_G is a cross-section, which we call *the geometric cross-section*. The geodesic $\pi(\gamma)$ can be coded by a bi-infinite sequence of generators of Γ identifying the sides of $\mathcal{F}$, as explained below. This method of coding goes back to Morse; for more details see [20].

The set Ω_G of oriented geodesics in $\mathbb{D}$ tangent to the vectors in C_G coincides with the set of geodesics in $\mathbb{D}$ intersecting $\mathcal{F}$; this is depicted in Figure 6 in coordinates $(u, w) \in \mathbb{S} \times \mathbb{S}, u \neq w$, where u is the repelling fixed point and w is the attracting fixed point of a geodesic. The coordinates of the "vertices" of Ω_G are (P_j, Q_{j+1}) (the upper part) and (Q_j, P_j) (the lower part). Let $\mathcal{G}_i$ be the "curvilinear horizontal slice" comprised of points (u, w) such that geodesics uw exit $\mathcal{F}$ through side i; this slice is in the horizontal strip between P_i and Q_{i+1}. The map $F_G(u, w)$ piecewise transforms variables by the same Möbius transformations that identify the sides of $\mathcal{F}$, that is,

$$F_G(u, w) = (T_i u, T_i w) \quad \text{if } (u, w) \in \mathcal{G}_i, \tag{6}$$

and $\mathcal{G}_i$ is mapped to the "curvilineal vertical slice" belonging to the vertical strip between $P_{\sigma(i)}$ and $Q_{\sigma(i)+1}$. F_G is a bijection of Ω_G. Due to the symmetry of the fundamental polygon $\mathcal{F}$, the "curvilinear horizontal (vertical) slices" are congruent to each other by a Euclidean translation.

For a geodesic γ, the *geometric coding sequence*

$$[\gamma]_G = [\ldots, n_{-2}, n_{-1}, n_0, n_1, n_2, \ldots]$$

is such that $F_G^k\gamma$ exits $\mathcal{F}$ through side $\sigma(n_k)$. By construction, the left shift in the space of geometric coding sequences corresponds to the map F_G, and the geodesic flow becomes a special flow over a symbolic system.

The set of geometric coding sequences is natural to consider, but it is never Markov. In order to obtain a special flow over a Markov chain, Adler and Flatto replaced the curvilinear boundary of the set Ω_G by polygonal lines, obtaining what they call a "rectilinear" set $\Omega_{\bar{P}}$ and defining an auxiliary map $F_{\bar{P}}$ mapping any horizontal line into a part of a horizontal line. We denote by $F_G : \Omega_G \to \Omega_G$ and $F_{\bar{P}} : \Omega_{\bar{P}} \to \Omega_{\bar{P}}$, respectively, their *curvilinear* and *rectilinear* transformations.[3] They prove that these maps are conjugate and that the rectilinear map is sofic.

[2] Adler and Flatto [4, p. 240] used unit tangent vectors pointed out of $\mathcal{F}$, but the results are easily transferable.

[3] In the remainder of this chapter we deal with a family of generalized rectilinear maps $F_{\bar{A}}$, and the notations $F_{\bar{P}}$ and $\Omega_{\bar{P}}$ are in conformance with those.

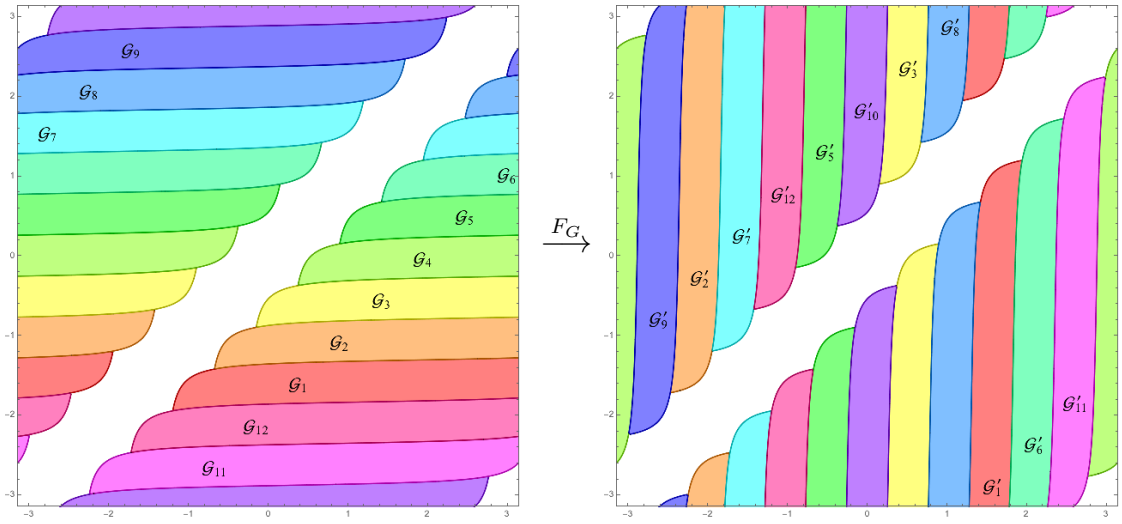


FIGURE 6: Curvilinear set Ω_G and its image for $g = 2$. Here $\mathcal{G}'_i = F_G(\mathcal{G}_i)$.

There is a relation between [4] and the series of papers by C. Series [36, 37, 39] where she also studies the problem of representing the geodesic flow by symbolic systems following her fundamental paper with Bowen [9]. Adler and Flatto acknowledge that they reach essentially the same mathematical conclusions. However, the use of the $(8g-4)$-sided fundamental polygon $\mathcal{F}$ in [4] makes their exposition much more accessible and allows for extensions of their results.

4.2. Katok and Ugarcovici

Let $f_{\bar{A}}$ be a generalized Bowen-Series boundary map studied in [25] and defined by the formula

$$f_{\bar{A}}(x) = T_i(x) \quad \text{if } x \in [A_i, A_{i+1}), \tag{7}$$

where

$$\bar{A} = \{A_1, \ldots, A_{8g-4}\}, \quad A_i \in (P_i, Q_i),$$

and let $F_{\bar{A}}$ be the corresponding two-dimensional map:

$$F_{\bar{A}}(x, y) = (T_i(x), T_i(y)) \quad \text{if } y \in [A_i, A_{i+1}). \tag{8}$$

The map $F_{\bar{A}}$ is a *natural extension* of the boundary map $f_{\bar{A}}$. Setting $\Delta = \{\,(x, y) \in \mathbb{S} \times \mathbb{S} : x = y\,\}$, we have $F_{\bar{A}}$ as a map on $\mathbb{S} \times \mathbb{S} \setminus \Delta$. If we identify a geodesic in $\mathbb{D}$ from u to w with a point in $\mathbb{S} \times \mathbb{S} \setminus \Delta$, $F_{\bar{A}}$ may also be considered as a map on geodesics.

Adler and Flatto [4] studied the partition

$$\bar{P} = \{P_1, \ldots, P_{8g-4}\},$$

which was previously considered by Bowen and Series [9], and their "rectilinear map T_R" is exactly our $F_{\bar{P}}$.

A key ingredient in analyzing the map $F_{\bar{A}}$ is what we call the *cycle property* of the partition points $A_1, \ldots, A_{8g-4}$. Such a property refers to the structure of the orbits of each A_i that one can construct by tracking the two images $T_i A_i$ and $T_{i-1} A_i$ of these points of discontinuity of the map $f_{\bar{A}}$. If a cycle closes up after one iteration, that is,

$$f_{\bar{A}}(T_i A_i) = f_{\bar{A}}(T_{i-1} A_i), \tag{9}$$

we say that the point A_i satisfies the *short cycle property*. We say a partition $\bar{A}$ has the *short cycle property* if each $A_i \in \bar{A}$ has the short cycle property.

Theorem 4.3 ([25, Thm. 1.3])**.** *If $\bar{A}$ satisfies the short cycle property, then there exists a set $\Omega_{\bar{A}} \subset \mathbb{S}\times\mathbb{S}\setminus\Delta$ with the following properties:*

1. *$\Omega_{\bar{A}}$ has a finite rectangular structure, i.e., it is bounded by non-decreasing step-functions with a finite number of steps.*
2. *The map $F_{\bar{A}}$ is essentially bijective on $\Omega_{\bar{A}}$.*
3. *Almost every point $(u, w) \in \mathbb{S}\times\mathbb{S}\setminus\Delta$ is mapped to $\Omega_{\bar{A}}$ after finitely many iterations of $F_{\bar{A}}$, and $\Omega_{\bar{A}}$ is a global attractor for the map $F_{\bar{A}}$, that is,*

$$\Omega_A = \bigcap_{n=0}^{\infty} F_{\bar{A}}^n(\mathbb{S}\times\mathbb{S}\setminus\Delta).$$

Explicitly, the y-levels of the upper and lower connected component of $\Omega_{\bar{A}}$ are, respectively,

$$B_i := T_{\sigma(i-1)} A_{\sigma(i-1)} \qquad \text{and} \qquad C_i := T_{\sigma(i+1)} A_{\sigma(i+1)+1}. \tag{10}$$

The x-levels in this case are the same as for the Bowen-Series map $F_{\bar{P}}$, and the set $\Omega_{\bar{A}}$ is determined by the corner points located in the strip $\{\, (x, y) \in \mathbb{S}\times\mathbb{S} : y \in [A_i, A_{i+1}) \,\}$ with coordinates

$$(P_i, B_i) \quad \text{(upper part)} \qquad \text{and} \qquad (Q_{i+1}, C_i) \quad \text{(lower part)}.$$

See Figure 7 for an example of such an attractor.

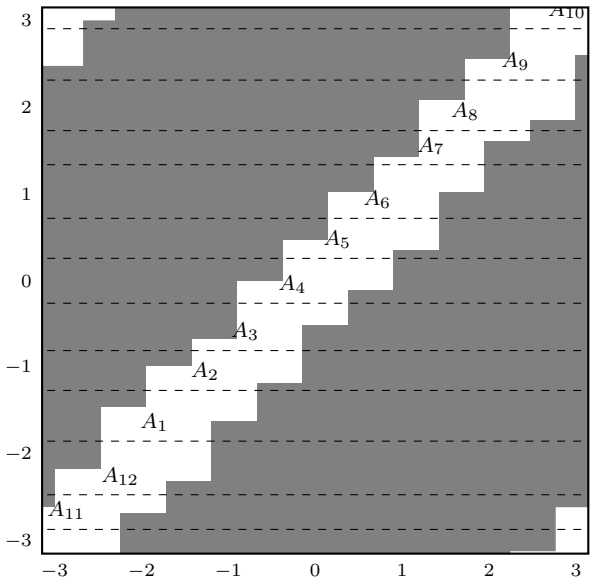


FIGURE 7: Attractor $\Omega_{\bar{A}}$ for generic $F_{\bar{A}}$ with $g = 2$.

5. New results

5.1. Conjugacy between $F_{\bar{A}}$ and F_G

We will make extensive use of the map that we call here U_i, which is used in both [4] and [25] without its own notation. The following are all equivalent definitions:

$$\begin{aligned} U_i &:= (T_{\sigma(i)+1}T_i)^{-1} = T_i^{-1}T_{\sigma(i)+1}^{-1} = T_{\sigma(i)}T_{\tau(i-1)} \\ &= (T_{\sigma(i-1)-1}T_{i-1})^{-1} = T_{i-1}^{-1}T_{\sigma(i-1)-1}^{-1} = T_{\sigma(i-1)}T_{\tau(i)}, \end{aligned} \tag{11}$$

where

$$\tau(i) := i + (4g-2) \mod (8g-4). \tag{12}$$

Most importantly, the map U_i sends $\mathcal{F}$ to the "corner image" $\mathcal{F}_i := U_i\mathcal{F}$ that touches $\mathcal{F}$ only at the vertex V_i (Figure 11 on page 26 shows $\mathcal{F}_2$ in gray). Note that $\{T_i\}$ and $\{U_i\}$ are both closed under inversion; specifically, $T_i^{-1} = T_{\sigma(i)}$ via (3) and

$$U_i^{-1} = U_{\tau(i)},$$

which can be shown using (3) and Lemma 5.1 below. Additionally, U_i^{-1} maps A_i to the end of the cycle (9), hence the two main expressions in (11).

Lastly, we will need the following properties of τ and σ:

Lemma 5.1. *If $\theta(i) := \sigma(i) - 1$ and $\rho(i) := \sigma(i) + 1$, then*

1. $\theta^2(i) = \rho^2(i) = \tau(i)$,
2. $\rho(i+1) = \theta(i-1) = \tau(\sigma(i))$,
3. σ *and* τ *commute.*

Proof. These can be verified by direct calculation. As one instance, we show that for even i, $\rho(i+1) = \tau(\sigma(i))$. In this case

$$\begin{aligned} \sigma(i) &= 2 - i \qquad (i \text{ is even}) \\ \tau(\sigma(i)) &= 2 - i + (4g-2) = 4g - i \\ \sigma(i+1) &= 4g - (i+1) \qquad (i+1 \text{ is odd}) \\ \rho(i+1) &= 4g - (i+1) + 1 = 4g - i \end{aligned}$$

The other equalities are shown similarly. □

Note that $\tau(i-1) = \tau(i) - 1$ because τ is a shift; these two forms are used interchangeably in several equations. In general, $\sigma(i-1) \neq \sigma(i) - 1$.

For the case $\bar{A} = \bar{P}$, Adler and Flatto [4, Sec. 5] explicitly describe a conjugacy between F_G and $F_{\bar{P}}$. They define the function $\Phi : \Omega_G \to \Omega_{\bar{P}}$ given by

$$\Phi = \begin{cases} \text{Id} & \text{on } \mathcal{O} \\ U_{\tau(i)+1} & \text{on } \mathcal{B}_i, \end{cases}$$

where $\mathcal{O} = \Omega_G \cap \Omega_{\bar{P}}$ and the set $\mathcal{B}_i$ is the "bulge" that is the part of $\Omega_G \setminus \Omega_{\bar{P}}$ with $w \in [P_i, P_{i+1}]$. They then show that the diagram

$$\begin{array}{ccc} \Omega_G & \xrightarrow{F_G} & \Omega_G \\ {\scriptstyle\Phi}\downarrow & & \downarrow{\scriptstyle\Phi} \\ \Omega_{\bar{P}} & \xrightarrow[F_{\bar{P}}]{} & \Omega_{\bar{P}} \end{array}$$

commutes. Note that for $\bar{A} = \bar{P}$, the bulges (comprising all points in $\Omega_G \setminus \Omega_{\bar{P}}$) are affixed only to the lower part of the rectilinear set $\Omega_{\bar{P}}$.

For the case of a generic $\bar{A}$ having the short cycle property, the set $\Omega_G \setminus \Omega_{\bar{A}}$ has pieces affixed to both the upper and lower parts of $\Omega_{\bar{A}}$. Thus we must define both upper and lower bulges:

$$\begin{aligned} &\textit{lower bulge } \mathcal{B}_i \text{ is the part of } \Omega_G \setminus \Omega_{\bar{A}} \text{ with } u \in [Q_{i+1}, Q_{i+2}]; \\ &\textit{upper bulge } \mathcal{B}^i \text{ is the part of } \Omega_G \setminus \Omega_{\bar{A}} \text{ with } u \in [P_{i-1}, P_i]. \end{aligned} \tag{13}$$

We also define corners, which comprise $\Omega_{\bar{A}} \setminus \Omega_G$:

$$\begin{aligned} &\textit{lower corner } \mathcal{C}_i \text{ is the part of } \Omega_{\bar{A}} \setminus \Omega_G \text{ with } u \in [Q_{i+1}, Q_{i+2}]; \\ &\textit{upper corner } \mathcal{C}^i \text{ is the part of } \Omega_{\bar{A}} \setminus \Omega_G \text{ with } u \in [P_{i-1}, P_i]. \end{aligned} \tag{14}$$

Figure 8 shows all four of these sets for a single i.

Remark 5.2. The curvilinear and rectilinear sets of Adler and Flatto include one boundary but not the other (see [4, Fig. 4.7, 5.1]). Our convention is that Ω_G and $\Omega_{\bar{A}}$ are closed, so for us each bulge includes its curved boundary but not its straight boundaries and each corner includes its straight boundaries but not its curved boundary. This does not affect the overall dynamics in any significant way.

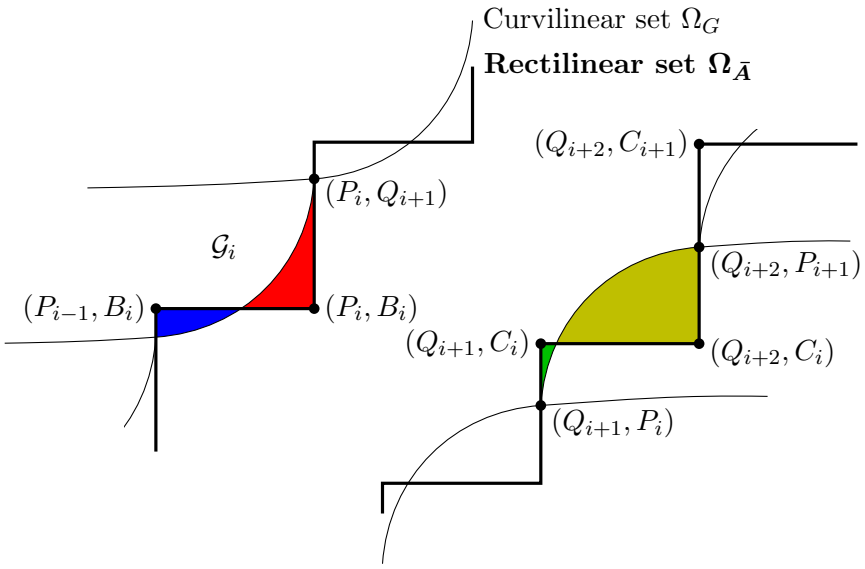


FIGURE 8: Bulges $\mathcal{B}^i$ in blue and $\mathcal{B}_i$ in gold; corners $\mathcal{C}^i$ in red and $\mathcal{C}_i$ in green.

Now let $\mathcal{O} := \Omega_G \cap \Omega_{\bar{A}}$ and define the map Φ with domain Ω_G as

$$\Phi = \begin{cases} \text{Id} & \text{on } \mathcal{O} \\ U_{\tau(i)+1} & \text{on } \mathcal{B}_i \\ U_{\tau(i)} & \text{on } \mathcal{B}^i. \end{cases} \tag{15}$$

Proposition 5.3. *Let $\bar{A}$ have the short cycle property. Then the map Φ is a bijection from Ω_G to $\Omega_{\bar{A}}$. Specifically, $\Phi(\mathcal{B}_i) = \mathcal{C}^{\tau(i)+1}$ and $\Phi(\mathcal{B}^i) = \mathcal{C}_{\tau(i)-1}$.*

Proof. All the sets $\mathcal{B}_i, \mathcal{B}^i, \mathcal{C}_i, \mathcal{C}^i$ are bounded by one horizontal line segment, one vertical line segment, and one curved segment that is part of $\partial\Omega_G$. Since each U_j is a Möbius transformation, we need only to show that these boundaries are mapped

accordingly. The boundaries of $\mathcal{B}_i$ are the vertical segment $\{Q_{i+2}\}\times[C_i, P_{i+1}]$, part of the horizontal segment $[Q_{i+1}, Q_{i+2}]\times\{C_i\}$, and the part of $\partial\Omega_G$ connecting (Q_{i+1}, P_i) to (Q_{i+2}, P_{i+1}). By definition, $\Phi(\mathcal{B}_i) = U_{\tau(i)+1}(\mathcal{B}_i)$. Recall that $U_{\tau(i)+1} = T_{\sigma(i)-1}T_i$. From Proposition 4.1, we have

$$T_iP_i = Q_{\sigma(i)+1}, \qquad T_iP_{i+1} = P_{\sigma(i)-1}, \qquad T_iQ_{i+1} = P_{\sigma(i)}, \qquad T_iQ_{i+2} = Q_{\sigma(i)}.$$

Applying $T_{\sigma(i)-1}$ to both sides and using Lemma 5.1, we get

$$\begin{aligned} U_{\tau(i)+1}P_i &= T_{\sigma(i)-1}Q_{\sigma(i)+1} = Q_{\tau(i)+1}, & U_{\tau(i)+1}P_{i+1} &= T_{\sigma(i)-1}P_{\sigma(i)-1} = Q_{\tau(i)+2}, \\ U_{\tau(i)+1}Q_{i+1} &= T_{\sigma(i)-1}P_{\sigma(i)} = P_{\tau(i)}, & U_{\tau(i)+1}Q_{i+2} &= T_{\sigma(i)-1}Q_{\sigma(i)} = P_{\tau(i)+1}. \end{aligned}$$

Additionally, by (10),

$$U_{\tau(i)+1}C_i = (T_{\sigma(i+1)+1}T_{i+1})(T_{\sigma(i+1)}A_{\tau\sigma(i)}) = T_{\sigma(\tau(i))}A_{\sigma(\tau(i))} = B_{\tau(i)+1}. \tag{16}$$

Therefore

$$\begin{aligned} U_{\tau(i)+1}\big(\{Q_{i+2}\}\times[C_i, P_{i+1}]\big) &= \{P_{\tau(i)+1}\}\times[B_{\tau(i)+1}, Q_{\tau(i)+2}], \\ U_{\tau(i)+1}\big([Q_{i+1}, Q_{i+2}]\times\{C_i\}\big) &= [P_{\tau(i)}, P_{\tau(i)+1}]\times\{B_{\tau(i)+1}\}, \\ U_{\tau(i)+1}(Q_{i+1}, P_i) &= (P_{\tau(i)}, Q_{\tau(i)+1}), \\ U_{\tau(i)+1}(Q_{i+2}, P_{i+1}) &= (P_{\tau(i)+1}, Q_{\tau(i)+2}). \end{aligned}$$

The corner $\mathcal{C}^{\tau(i)+1}$ is exactly the set bounded by the vertical segment $\{P_{\tau(i)+1}\}\times[B_{\tau(i)+1}, Q_{\tau(i)+2}]$, part of the horizontal segment $[P_{\tau(i)}, P_{\tau(i)+1}]\times\{B_{\tau(i)+1}\}$, and part of segment of $\partial\Omega_G$ connecting $(P_{\tau(i)}, Q_{\tau(i)+1})$ to $(P_{\tau(i)+1}, Q_{\tau(i)+2})$. Thus $U_{\tau(i)+1}(\mathcal{B}_i) = \mathcal{C}^{\tau(i)+1}$. A similar argument shows $U_{\tau(i)}(\mathcal{B}^i) = \mathcal{C}_{\tau(i)-1}$. Taking $\Phi(\mathcal{O}) = \mathrm{Id}(\mathcal{O}) = \mathcal{O}$ together with $\Phi(\mathcal{B}_i) = \mathcal{C}^{\tau(i)+1}$ and $\Phi(\mathcal{B}^i) = \mathcal{C}_{\tau(i)-1}$, we have that $\Phi(\Omega_G) = \Omega_{\bar{A}}$. □

Definition. Let $u, w \in \mathbb{S} = \partial\mathbb{D}$, $u \neq w$. An oriented geodesic in $\mathbb{D}$ from u to w is called $\bar{A}$-*reduced* if $(u, w) \in \Omega_{\bar{A}}$.

Corollary 5.4. *If a geodesic $\gamma = uw$ intersects $\mathcal{F}$, then either γ is $\bar{A}$-reduced or $U_j\gamma$ is $\bar{A}$-reduced, where $j = \tau(i)$ if $(u, w) \in \mathcal{B}^i$ and $j = \tau(i) + 1$ if $(u, w) \in \mathcal{B}_i$.*

Proof. The geodesic $\gamma = uw$ intersecting $\mathcal{F}$ is equivalent to (u, w) being in the set Ω_G. If $(u, w) \in \mathcal{O} = \Omega_G \cap \Omega_{\bar{A}}$, then γ is $\bar{A}$-reduced as well. If not, then (u, w) is in some upper or lower bulge since the bulges comprise all of $\Omega_G \setminus \Omega_{\bar{A}}$. If $(u, w) \in \mathcal{B}_i$, then $U_{\tau(i)+1}(u, w) \in \mathcal{C}^{\tau(i)+1} \subset \Omega_{\bar{A}}$ and so $U_{\tau(i)+1}\gamma$ is $\bar{A}$-reduced. If $(u, w) \in \mathcal{B}^i$, then $U_{\tau(i)}(u, w) \in \mathcal{C}_{\tau(i)-1} \subset \Omega_{\bar{A}}$ and so $U_{\tau(i)}\gamma$ is $\bar{A}$-reduced. It follows that γ and $U_j\gamma$ cannot be $\bar{A}$-reduced simultaneously. □

Remark 5.5. Note that although the geodesic γ might intersect two different corner images, the point (u, w) cannot be in two different bulges simultaneously. The index of the bulge $\mathcal{B}^i$ or $\mathcal{B}_i$ determines a specific j for which $U_j\gamma$ is $\bar{A}$-reduced, as stated in Corollary 5.4.

Corollary 5.6. *If a geodesic $\gamma = uw$ is $\bar{A}$-reduced, then either γ intersects $\mathcal{F}$, or γ intersects $\mathcal{F}_j = U_j(\mathcal{F})$, where $j = i$ if $(u, w) \in \mathcal{C}^i$ and $j = i + 1$ if $(u, w) \in \mathcal{C}_i$.*

Proof. The geodesic γ is $\bar{A}$-reduced means that $(u, w) \in \Omega_{\bar{A}}$. If $(u, w) \in \mathcal{O} = \Omega_G \cap \Omega_{\bar{A}}$, then γ intersects $\mathcal{F}$. If not, then (u, w) is in some upper or lower corner. If $(u, w) \in \mathcal{C}_i$, then $U_i^{-1}(u, w) \in \mathcal{B}_{\tau(i)-1} \subset \Omega_G$, so $U_i^{-1}(\gamma)$ intersects $\mathcal{F}$, or γ intersects $U_i(\mathcal{F}) = \mathcal{F}_i$.

If $(u, w) \in \mathcal{C}^i$, then $U_{i+1}^{-1}(u, w) \in \mathcal{B}^{\tau(i)+1}$, so $U_{i+1}^{-1}(\gamma)$ intersects $\mathcal{F}$, or γ intersects $U_{i+1}(\mathcal{F}) = \mathcal{F}_{j+1}$. □

Remark 5.7. Here again, although the geodesic γ might intersect two different corner images, the index of the corner $\mathcal{C}^i$ or $\mathcal{C}_i$ determines a specific j for which γ intersects $\mathcal{F}_j$, as stated in Corollary 5.6.

Theorem 5.8. *Let $\bar{A}$ have the short cycle property, and let $\Phi : \Omega_G \to \Omega_{\bar{A}}$ be as in (15). Then Φ is a conjugacy between F_G and $F_{\bar{A}}$. That is, the following diagram commutes:*

$$\begin{array}{ccc} \Omega_G & \xrightarrow{F_G} & \Omega_G \\ {\scriptstyle\Phi}\downarrow & & \downarrow{\scriptstyle\Phi} \\ \Omega_{\bar{A}} & \xrightarrow[F_{\bar{A}}]{} & \Omega_{\bar{A}} \end{array} . \tag{17}$$

Proof. The diagram commuting is equivalent to

$$F_{\bar{A}}^{-1} \circ \Phi \circ F_G \circ \Phi^{-1} = \mathrm{Id}. \tag{18}$$

Let (u, w) be any point in $\Omega_{\bar{A}}$; we aim to prove that (18) is true for (u, w) by building out the diagram

$$\begin{array}{ccc} & \xrightarrow{F_G} & \\ {\scriptstyle\Phi}\downarrow & & \downarrow{\scriptstyle\Phi} \\ (u, w) & \xrightarrow[F_{\bar{A}}]{} & \end{array} .$$

If $(u, w) \in \mathcal{O} = \Omega_G \cap \Omega_{\bar{A}}$, then $\Phi^{-1}(u, w) = (u, w)$. Let i be the side through which the geodesic $\gamma = uw$ exits $\mathcal{F}$; then $F_G(u, w) = T_i(u, w)$. At this point (18) is

$$F_{\bar{A}}^{-1} \circ \Phi \circ (T_i) \circ (\mathrm{Id})^{-1} = \mathrm{Id},$$

corresponding to

$$\begin{array}{ccc} (u, w) & \xrightarrow{F_G = T_i} & \\ {\scriptstyle\Phi = \mathrm{Id}}\downarrow & & \downarrow{\scriptstyle\Phi} \\ (u, w) & \xrightarrow[F_{\bar{A}}]{} & \end{array} .$$

Since $\gamma = uw$ exits $\mathcal{F}$ through side i, the point w must be in $[P_i, Q_{i+1}]$, which means that $F_{\bar{A}}$ must act by T_{i-1}, T_i, or T_{i+1} (see Figure 9).

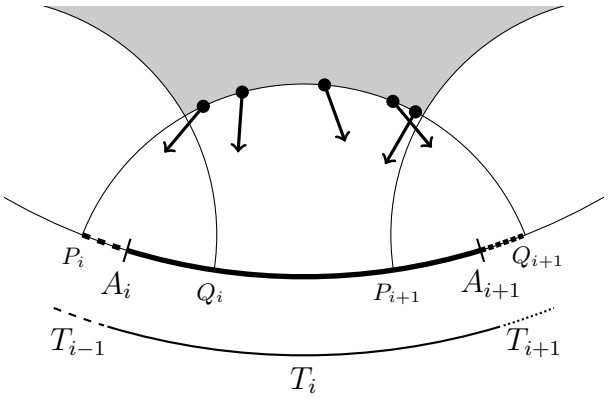


FIGURE 9: Action of $F_{\bar{A}}$ based on side exit.

- If $F_{\bar{A}} = T_i$, then $F_{\bar{A}}\gamma = T_i\gamma$ is both $\bar{A}$-reduced and intersects $\mathcal{F}$. Thus Φ acts on $T_i\gamma$ as the identity, and (18) becomes

$$(T_i)^{-1} \circ (\mathrm{Id}) \circ (T_i) \circ (\mathrm{Id})^{-1} = \mathrm{Id}.$$

- If $F_{\bar{A}} = T_{i-1}$, then (18) becomes

$$(T_{i-1})^{-1} \circ \Phi \circ (T_i) \circ (\mathrm{Id})^{-1} = \mathrm{Id},$$

  which is true if and only if $\Phi = T_{i-1}T_i^{-1}$, or equivalently, by (11), $\Phi = U_{\sigma(i-1)}$. Since we know $T_i\gamma$ intersects $\mathcal{F}$, the geodesic

$$F_{\bar{A}}\gamma = T_{i-1}\gamma = T_{i-1}T_i^{-1}(T_i\gamma) = U_{\sigma(i-1)}(T_i\gamma)$$

  must intersect the corner image $\mathcal{F}_{\sigma(i-1)}$. Therefore Φ does act by $U_{\sigma(i-1)}$, and (18) is true in this case.

- If $F_{\bar{A}} = T_{i+1}$, then (18) becomes

$$(T_{i+1})^{-1} \circ \Phi \circ (T_i) \circ (\mathrm{Id})^{-1} = \mathrm{Id},$$

  which is true if and only if $\Phi = T_{i+1}T_i^{-1}$. Similar to the previous case, the geodesic

$$F_{\bar{A}}\gamma = T_{i+1}\gamma = T_{i+1}T_i^{-1}(T_i\gamma) = U_{\sigma\tau(i)}(T_i\gamma)$$

  must intersect $F_{\sigma\tau(i)}$, so Φ acts by $U_{\sigma\tau(i)} = T_{i+1}T_i^{-1}$ and (18) is true.

If $(u, w) = \Omega_{\bar{A}} \backslash \mathcal{O}$, then it is in an upper or lower corner. Assume (u, w) is in an upper corner, and let i be such that $(u, w) \in \mathcal{C}^i$. Since, by Lemma 5.3, $U_i(\mathcal{B}_{\tau(i)-1}) = \mathcal{C}^i$, we

know that Φ^{-1} acts on $\mathcal{C}^i$ by U_i^{-1}. Defining $\gamma' = u'w' = U_i^{-1}\gamma$, we have the partial diagram

$$\begin{array}{ccc} (u', w') & \xrightarrow{F_G} & \\ {\scriptstyle \Phi = U_i}\downarrow & & \downarrow{\scriptstyle \Phi} \\ (u, w) & \xrightarrow[F_{\bar{A}}]{} & \end{array}$$

corresponding to

$$F_{\bar{A}}^{-1} \circ \Phi \circ F_G \circ (U_i)^{-1} = \mathrm{Id}$$

at this point.

Since $\mathcal{C}^i \subset [P_{i-1}, P_i] \times [B_i, Q_{i+1}]$, we have

$$\begin{aligned} u' &\in U_i^{-1}[P_{i-1}, P_i] = [Q_{\tau(i)}, Q_{\tau(i)+1}], \\ w' &\in U_i^{-1}[B_i, Q_{i+1}] = [C_{\tau(i)-1}, P_{\tau(i)}] \end{aligned} \tag{19}$$

based on Proposition 4.1 and equation (16). In most instances γ' will exit $\mathcal{F}$ through side $\tau(i) - 1$, but if $C_{\tau(i)-1}$ is very close to $P_{\tau(i)-1}$ it is possible for γ' to exit through side $\tau(i) - 2$ instead. Thus F_G might act on (u', w') by $T_{\tau(i)-1}$ or $T_{\tau(i)-2}$.

The remainder of the proof is broken down into four cases:

Case 1: $F_{\bar{A}}\gamma = T_i\gamma$ and $F_G\gamma' = T_{\tau(i)-1}\gamma'$.

Case 2: $F_{\bar{A}}\gamma = T_i\gamma$ and $F_G\gamma' = T_{\tau(i)-2}\gamma'$.

Case 3: $F_{\bar{A}}\gamma = T_{i+1}\gamma$ and $F_G\gamma' = T_{\tau(i)-1}\gamma'$.

Case 4: $F_{\bar{A}}\gamma = T_{i+1}\gamma$ and $F_G\gamma' = T_{\tau(i)-2}\gamma'$.

For cases 1 and 2, $w \in [B_i, A_{i+1})$, while for 3 and 4 we have $w \in [A_{i+1}, Q_{i+1}]$. For cases 1 and 3, γ' exits through side $\tau(i) - 1$, so the geometrically next segment of $\bar{\gamma}$ in $\mathcal{F}$ starts at $z_1 := T_{\tau(i)-1}z_1'$ and is part of the geodesic

$$T_{\tau(i)-1}\gamma' = T_{\tau(i)-1}U_i^{-1}\gamma = T_{\tau(i)-1}T_{\sigma(i)+1}T_i\gamma = T_{\tau(i)-1}T_{\sigma(\tau(i)-1)}T_i\gamma = T_i\gamma. \tag{20}$$

Case 1. In this case, $F_{\bar{A}}\gamma = T_i\gamma$ and, as shown in (20) above, $F_G\gamma' = T_i\gamma$ as well.

Case 2. Here γ' exits $\mathcal{F}$ through side $\tau(i) - 2$, so the geometrically next segment of $\bar{\gamma}$ in $\mathcal{F}$ is part of $T_{\tau(i)-2}\gamma'$. The expression $T_{\tau(i)-2}U_i^{-1}$ does not simplify as easily as $T_{\tau(i)-1}U_i^{-1} = T_i$ did. However, we do have

$$T_{\tau(i)-2}U_i^{-1} = T_{\tau(i)-2}T_{\rho(i)}T_i = T_{\sigma(\rho(i))-1}T_{\rho(i)}T_i = U_{\rho(j)+1}^{-1}T_j. \tag{21}$$

The geodesic $T_{\tau(i)-2}\gamma'$ must intersect $\mathcal{F}$. This is equivalent to $U_{\rho(i)+1}T_{\tau(i)-2}\gamma'$ intersecting $U_{\rho(i)+1}\mathcal{F} = \mathcal{F}_{\rho(i)+1}$. Since

$$U_{\rho(i)+1}T_{\tau(i)-2}\gamma' = U_{\rho(i)+1}(T_{\tau(i)-2}U_i^{-1})\gamma = U_{\rho(i)+1}(U_{\rho(i)+1}^{-1}T_i)\gamma = T_i\gamma,$$

the geodesic $T_i\gamma = F_{\bar{A}}\gamma$ must intersect $\mathcal{F}_{\rho(i)+1}$. Thus when we "pull back" $F_{\bar{A}}\gamma = T_i\gamma$ to $\mathcal{F}$, we get $U_{\rho(i)+1}^{-1}T_i\gamma$, which by (21) is exactly $T_{\tau(i)-2}U_i^{-1}\gamma = F_G\gamma'$.

Case 3. Here γ' exits $\mathcal{F}$ through side $\tau(i)-1$, so $F_G\gamma' = T_i\gamma$ by (20). However, $F_{\bar{A}}\gamma = T_{i+1}\gamma$ instead of $T_i\gamma$.

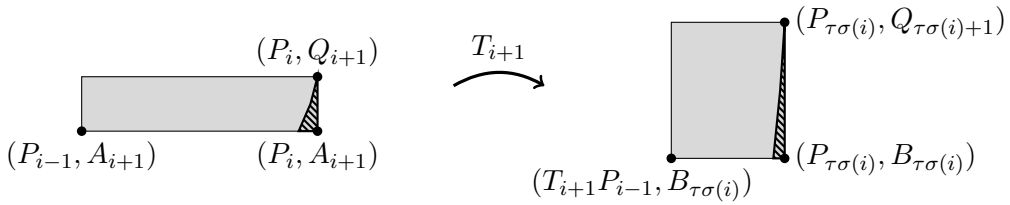


FIGURE 10: The "triangular" region (the "hypotenuse" is part of $\partial\Omega_G$ so is curved) is the set of clockwise $\bar{A}$-reduced geodesics that intersect $\mathcal{F}_i$ but not $\mathcal{F}$ and for which $F_{\bar{A}} = T_{i+1}$.

Figure 10 shows the upper tip of $\mathcal{C}^i$ and its image under T_{i+1}. The right wall $\{P_i\} \times [A_{i+1}, Q_{i+1}]$ is mapped to $\{P_{\tau\sigma(i)}\} \times [B_{\tau\sigma(i)}, Q_{\tau\sigma(i)}]$. As was mentioned in Section 4, the striped region at the vertex $(P_{\tau\sigma(i)}, Q_{\tau\sigma(i)+1})$ is bounded by the curve obtained by a Euclidean translation of the boundary of the striped region at the vertex (P_i, Q_{i+1}), and since the map T_{i+1} is hyperbolic (it expands in the y-direction and contracts in the x-direction), the image of the upper tip of the striped region at the vertex (P_i, Q_{i+1}) is a narrow curvilinear triangular region contained completely inside the striped region at the vertex $(P_{\tau\sigma(i)}, Q_{\tau\sigma(i)+1})$. Thus for all geodesics γ in this case, $F_{\bar{A}}\gamma$ intersects $\mathcal{F}_{\tau\sigma(i)}$.

When we "pull back" $F_{\bar{A}}\gamma = T_{i+1}\gamma$ to $\mathcal{F}$, we get

$$U_{\tau\sigma(i)}^{-1}T_{i+1}\gamma = T_{\sigma(\tau\sigma(i)-1)-1}T_{\tau\sigma(i)-1}T_{i+1}\gamma = T_iT_{\sigma(i+1)}T_{i+1}\gamma = T_i\gamma,$$

which by (20) is exactly $F_G\gamma'$.

Case 4. This case, in which $F_{\bar{A}}\gamma = T_{i+1}\gamma$ and $\gamma' = U_i^{-1}\gamma$ exits $\mathcal{F}$ through side $\tau(i)-2$, will be shown to be impossible.

Since the curvilinear horizontal slice $\mathcal{G}_k$, that is, the set of all geodesics exiting $\mathcal{F}$ through side k, is contained in the horizontal strip $\mathbb{S} \times [P_k, Q_{k+1}]$, the endpoint w' of γ' must be in $[P_{\tau(i)-2}, Q_{\tau(i)-1}]$. Since $w' \in [C_{\tau(i)-1}, P_{\tau(i)}]$ by (19), we have

$$w' \in [P_{\tau(i)-2}, Q_{\tau(i)-1}] \cap [C_{\tau(i)-1}, P_{\tau(i)}] = [C_{\tau(i)-1}, Q_{\tau(i)-1}]$$

(in fact, w' will be quite close to $C_{\tau(i)-1}$). This means that the endpoint w of $\gamma = U_i\gamma'$ will be in the interval

$$I = U_i[C_{\tau(i)-1}, Q_{\tau(i)-1}] = [B_i, U_iQ_{\tau(i)-1}].$$

We have

$$U_iQ_{\tau(i)-1} = T_{\sigma(i)}T_{\tau(i)-1}Q_{\tau(i)-1} = T_{\sigma(i)}Q_{\sigma(i)+3}.$$

In order to locate $T_{\sigma(i)}Q_{\sigma(i)+3}$, we use Proposition 4.1. $T_{\sigma(i)}$ maps the geodesic $P_{\sigma(i)}Q_{\sigma(i)+1}$, which is its isometric circle, to the geodesic $Q_{i+1}P_i$. We also see that $T_{\sigma(i)}Q_{\sigma(i)+2} = Q_i$ and $T_{\sigma(i)}P_{\sigma(i)-1} = P_{i+1}$, and since $T_{\sigma(i)}$ maps the outside of $P_{\sigma(i)}Q_{\sigma(i)+1}$ to the inside of $Q_{i+1}P_i$ and preserves the order of points, we conclude that $T_{\sigma(i)}Q_{\sigma(i)+3} \in (Q_i, P_{i+1})$. Therefore $I \subset [B_i, P_{i+1}] \subset (A_i, A_{i+1})$. Thus $w \in I$ implies $F_{\bar{A}}\gamma = T_i\gamma$, which contradicts the assumption $F_{\bar{A}}\gamma = T_{i+1}\gamma$ of this case.

Cases 1–3 cover all potential ways that

$$\begin{array}{ccc} (u', w') & \xrightarrow{F_G} & \\ {\scriptstyle \Phi = U_i}\downarrow & & \downarrow{\scriptstyle \Phi} \\ (u, w) & \xrightarrow[F_{\bar{A}}]{} & \end{array}$$

can be completed, showing that (18) is true if $(u, w) \in \mathcal{C}^i$. Similar arguments show that (18) is true for $(u, w) \in \mathcal{C}_i$ as well.

Thus we have proven (18) for all $(u, w) \in \Omega_{\bar{A}}$, meaning that

$$\begin{array}{ccc} \Omega_G & \xrightarrow{F_G} & \Omega_G \\ {\scriptstyle \Phi}\downarrow & & \downarrow{\scriptstyle \Phi} \\ \Omega_{\bar{A}} & \xrightarrow[F_{\bar{A}}]{} & \Omega_{\bar{A}} \end{array}$$

is indeed a commutative diagram. □

5.2. Cross-sections

Based on Corollary 5.6, we introduce the notion of the $\bar{A}$-*cross-section point*. It is the entrance point of an $\bar{A}$-reduced geodesic γ to $\mathcal{F}$, or, if γ does not intersect $\mathcal{F}$, the first entrance point to $\mathcal{F}_j$, where j is as in Corollary 5.6.

Now we define a map

$$\varphi : \Omega_{\bar{A}} \to S\mathbb{D}, \quad \varphi(u, w) = (z, \zeta),$$

where z is the $\bar{A}$-cross-section point on the geodesic γ from u to w and ζ is the unit tangent vector to γ at z. This map is clearly injective. Composed with the canonical projection $\pi : S\mathbb{D} \to SM$, we obtain a map

$$\pi \circ \varphi : \Omega_{\bar{A}} \to SM. \tag{22}$$

The set $C_{\bar{A}} := \pi \circ \varphi(\Omega_{\bar{A}})$ can be described as follows: $\pi(z, \zeta) \in C_{\bar{A}}$ if the geodesic γ in $\mathbb{D}$ through (z, ζ) is $\bar{A}$-reduced or if $U_j\gamma$ is $\bar{A}$-reduced for $j \in \mathbb{A}$ determined in Corollary 5.4. It follows from Corollary 5.4 that the map $\pi \circ \varphi$ is injective and continuous, and hence $C_{\bar{A}}$ is parametrized by $\Omega_{\bar{A}}$. Since $\Omega_{\bar{A}}$ is an attractor for $F_{\bar{A}}$, $C_{\bar{A}}$ is a cross-section for the geodesic flow $\{\varphi^t\}$; we call $C_{\bar{A}}$ an *arithmetic cross-section.*

The *geometric cross-section* C_G can be described in similar terms. We define a (clearly injective) map

$$\psi : \Omega_G \to S\mathbb{D}, \quad \varphi(u, w) = (z, \zeta),$$

where z is the entrance point on the geodesic γ from u to w to $\mathcal{F}$, and ζ is the unit tangent vector to γ at z. Then

$$\pi \circ \psi : \Omega_G \to SM$$

is injective, and $C_G := \pi \circ \psi(\Omega_G)$ consists of *all* $\pi(z, \zeta)$ for which $z \in \partial\mathcal{F}$ and ζ points inward.

A priori, we only know that $C_{\bar{A}} \subset C_G$; the first return to C_G is not necessarily the first return to $C_{\bar{A}}$. In the $\mathrm{PSL}(2, \mathbb{Z})$ case, a geodesic can return to the geometric cross-section multiple times before reaching the arithmetic one [24], but here the situation is simpler.

Corollary 5.9. $C_{\bar{A}} = C_G$.

Proof. Since $\pi \circ \varphi$, $\pi \circ \psi$, and Φ are bijections and Φ acts by elements of Γ, the diagram

$$\begin{array}{ccc} \Omega_G & \xrightarrow{\pi \circ \psi} & C_G \\ {\scriptstyle \Phi}\downarrow & & \downarrow{\scriptstyle \mathrm{Id}} \\ \Omega_{\bar{A}} & \xrightarrow[\pi \circ \varphi]{} & C_{\bar{A}} \end{array} \tag{23}$$

commutes. Indeed, let $\gamma = uw$ with $(u, w) \in \Omega_G$. If γ is $\bar{A}$-reduced, then $\Phi = \mathrm{Id}$, $\varphi = \psi$, and we are done. If not, then for $j \in \mathbb{A}$ determined in Corollary 5.4, $\gamma' := U_j\gamma$ is $\bar{A}$-reduced. In this case $\Phi = U_j$, and again we have $\varphi \circ \Phi = \psi$. □

Corollary 5.10. *Given any tangent vector in the geometric cross-section C_G, its first return to C_G is also its first return to the arithmetic cross-section $C_{\bar{A}}$.*

Proof. This is essentially equivalent to Theorem 5.8 via Corollary 5.9.

Combining commutative diagrams (17) and (23), we obtain a diagram

$$\begin{array}{ccccccc} C_G & \xrightarrow{(\pi\circ\psi)^{-1}} & \Omega_G & \xrightarrow{F_G} & \Omega_G & \xrightarrow{\pi\circ\psi} & C_G \\ \downarrow{\scriptstyle \mathrm{Id}} & & \downarrow{\scriptstyle \Phi} & & \downarrow{\scriptstyle \Phi} & & \downarrow{\scriptstyle \mathrm{Id}}\,. \\ C_{\bar{A}} & \xrightarrow[(\pi\circ\varphi)^{-1}]{} & \Omega_{\bar{A}} & \xrightarrow[F_{\bar{A}}]{} & \Omega_{\bar{A}} & \xrightarrow[\pi\circ\varphi]{} & C_{\bar{A}} \end{array}$$

The composition of the maps in the upper row is the first return map to C_G, and the composition in the lower row is the first return to $C_{\bar{A}}$. The result follows from the commutativity of the diagram. □

Figure 11 depicts the first return to the cross-section $C_{\bar{A}}$. Here γ is $\bar{A}$-reduced but does not intersect $\mathcal{F}$, while $\gamma' := \Phi^{-1}\gamma = U_2^{-1}\gamma$ intersects $\mathcal{F}$ but is not $\bar{A}$-reduced. The unit tangent vector at the point of entrance of γ' to $\mathcal{F}$ belongs to both $C_{\bar{A}}$ and C_G. Its first return to $C_{\bar{A}}$ is the unit tangent vector at the point of entrance of $F_{\bar{A}}\gamma$ to $\mathcal{F}$, and its first return to C_G is the unit tangent vector at the point of entrance of $F_G\gamma'$ to $\mathcal{F}$. Since $F_G\gamma' = F_{\bar{A}}\gamma$, the first returns coincide.

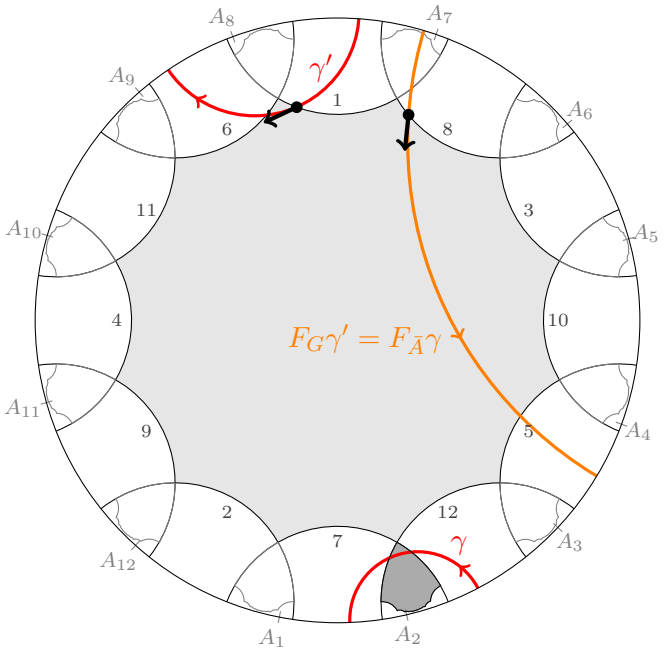


FIGURE 11: The first return map to the cross-section $C_{\bar{A}}$.

5.3. Symbolic coding of geodesics

In this section we describe how to code geodesics for a partition $\bar{A}$ for which $F_{\bar{A}}$ possesses an attractor $\Omega_{\bar{A}}$ with finite rectangular structure. A large class of examples was given in [25, Theorems 1.3 and 2.1]. A reduction algorithm described there for (almost) every geodesic γ from u to w in $\mathbb{D}$ produces in finitely many steps an $\bar{A}$-reduced geodesic Γ-equivalent to γ, and an application of this algorithm to an $\bar{A}$-reduced geodesic produces another $\bar{A}$-reduced geodesic.

We associate to any $w_0 \in \mathbb{S}$ a sequence of symbols from the alphabet $\mathbb{A}$

$$[w_0]_{\bar{A}} = [n_0, n_1, n_2, \dots], \tag{24}$$

where $n_k = \sigma(i)$ if $f_{\bar{A}}^k(w_0) \in [A_i, A_{i+1})$ for $k \geq 0$. We call this the *(forward)* $\bar{A}$-*expansion* of w_0. If $\gamma = u_0 w_0$ is an $\bar{A}$-reduced geodesic, we also call this sequence the *future* of γ.

By successive application of the map $F_{\bar{A}}$, we obtain a sequence of pairs $(u_k, w_k) = F_{\bar{A}}^k(u_0, w_0)$, $k \geq 0$, such that each geodesic γ_k from u_k to w_k is $\bar{A}$-reduced and

$$[w_k]_{\bar{A}} = [n_k, n_{k+1}, \dots]. \tag{25}$$

Using the bijectivity of the map $F_{\bar{A}}$, we extend the sequence (24) to the past to obtain a bi-infinite sequence

$$[\gamma]_{\bar{A}} = [\dots, n_{-2}, n_{-1}, n_0, n_1, n_2, \dots], \quad n_i \in \mathbb{A}, \tag{26}$$

called the *arithmetic coding sequence* or *arithmetic code* of γ (or the *coding sequence* or *code* when the context is clear) as follows: from the bijectivity of the map $F_{\bar{A}}$ on $\Omega_{\bar{A}}$, there exists a pair $(u_{-1}, w_{-1}) \in \Omega_{\bar{A}}$ such that $F_{\bar{A}}(u_{-1}, w_{-1}) = (u_0, w_0)$, i.e., $f_{\bar{A}} w_{-1} = w_0$. Then $w_{-1} \in [A_i, A_{i+1})$ for some $i \in \mathbb{A}$, and for $n_{-1} = \sigma(i)$ we have

$$[w_{-1}]_{\bar{A}} = [n_{-1}, n_0, n_1, \dots].$$

Continuing inductively, we define the sequence $n_{-k} \in \mathbb{A}$ and the pairs $(u_{-k}, w_{-k}) \in \Omega_{\bar{A}}$ $(k \geq 2)$, where

$$[w_{-k}]_{\bar{A}} = [n_{-k}, n_{-k+1}, n_{-k+2}, \dots],$$

by $F_{\bar{A}}(u_{-k}, w_{-k}) = (u_{-(k-1)}, w_{-(k-1)})$. We call the sequence

$$[n_{-1}, n_{-2}, \dots, n_{-k} \dots] \tag{27}$$

the *past* of γ.

Notice that the future of γ depends only on w_0 while, in general, the past of γ depends on both, w_0 and u_0. In some rare cases the past only depends on u_0, and, in fact, the sequence (27) is an expansion of u_0 with respect to a different (dual) partition. This will be discussed in Section 5.7.

We also associate to $\gamma = \gamma_0$ a bi-infinite sequence $\{\gamma_k = u_k w_k\}_{k\in\mathbb{Z}}$ of $\bar{A}$-reduced geodesics Γ-equivalent to γ. The left shift of the this sequence corresponds to an application of the map $F_{\bar{A}}$ to the corresponding geodesic: $F_{\bar{A}}\gamma_k = \gamma_{k+1}$.

Combining results of Subsections 5.1 and 5.2, we obtain the following result:

Proposition 5.11. *Let γ be an $\bar{A}$-reduced geodesic on $\mathbb{D}$ and $\bar{\gamma}$ its projection to M. Then*

1. *each geodesic segment of $\bar{\gamma}$ between successive returns to the cross-section $C_{\bar{A}}$ produces an $\bar{A}$-reduced geodesic on $\mathbb{D}$, and each reduced geodesic Γ-equivalent to γ is obtained this way;*

2. *the first return of $\bar{\gamma}$ to the cross-section $C_{\bar{A}}$ corresponds to a left shift of the coding sequence of γ.*

Additionally, (1) and (2) hold for the case $\bar{A} = \bar{P}$.

Proof. Let γ be an $\bar{A}$-reduced geodesic on $\mathbb{D}$. Then its projection $\bar{\gamma}$ to M can be represented as a (countable) sequence of geodesic segments in $\mathcal{F}$. By Corollary 5.4, each such segment either extends to an $\bar{A}$-reduced geodesic on $\mathbb{D}$ or its image under U_j is $\bar{A}$-reduced (j is specified in Corollary 5.4). If γ' is an $\bar{A}$-reduced geodesic Γ-equivalent to γ, then both project to the same geodesic in M. By Corollary 5.6, either γ' intersects $\mathcal{F}$ or its image under U_j^{-1} intersects $\mathcal{F}$ (j is specified in Corollary 5.6). In either case the intersection of the corresponding geodesic on $\mathbb{D}$ with $\mathcal{F}$ is another segment of the same geodesic in $\mathcal{F}$. This completes the proof of (1).

Since $F_{\bar{A}}\gamma = \gamma_1$, $f_{\bar{A}}w_1 = [n_1, n_2, \dots]$, the first digit of the past of γ_1 is n_0, and the remaining digits are the same as in the past of γ. Now (2) follows from Corollary 5.10.

Since Corollaries 5.4 and 5.6 are true for $\bar{A} = \bar{P}$, the arguments in this proof hold for $\bar{A} = \bar{P}$ as well. □

The following corollary is immediate.

Corollary 5.12. *If γ' is Γ-equivalent to γ, and both geodesics can be reduced in finitely many steps, then the coding sequences of γ and γ' differ by a shift.*

Thus we can talk about coding sequences of geodesics on M. To any geodesic γ that can be reduced in finitely many steps we associate the coding sequence (26) of a reduced geodesic Γ-equivalent to it. Corollary 5.12 implies that this definition does not depend on the choice of a particular representative.

An *admissible* sequence $s \in \mathbb{A}^{\mathbb{Z}}$ is one obtained by the coding procedure (26) from some reduced geodesic uw with $(u, w) \in \Omega_{\bar{A}}$. Given an admissible sequence, then, we can associate to it the vector $v \in C_{\bar{A}}$ such that $v = \pi\circ\phi(u, w)$, where $\pi\circ\varphi : \Omega_{\bar{A}} \to C_{\bar{A}}$ is defined in (22). We denote this vector v by $\mathrm{Cod}(s)$. The map Cod is essentially bijective (finite-to one: see Example 4 in Section 5.4). By Proposition 5.13 below,

Cod is uniformly continuous on the set of admissible coding sequences, and thus we can extend it to the closure $X_{\bar{A}} \subset \mathbb{A}^{\mathbb{Z}}$ of all admissible sequences.

The symbolic system $(X_{\bar{A}}, \sigma) \subset (\mathbb{A}^{\mathbb{Z}}, \sigma)$ is defined on the alphabet $\mathbb{A} = \{1, 2, \ldots, 8g-4\}$. The product topology on $\mathbb{A}^{\mathbb{Z}}$ is induced by the distance function

$$d(s, s') = \frac{1}{\max\{\, k : s_i = s'_i \text{ for } |i| \le k \,\}}.$$

Proposition 5.13. *The map* Cod *is uniformly continuous.*

Proof. Let s and s' be two admissible coding sequences obtained from reduced geodesics uw and $u'w'$, respectively. If $d(s, s') < \frac{1}{m}$, then the $\bar{A}$-expansions of the attracting end points w and w' of the corresponding geodesics given by (24) have the same first m symbols in their (forward) $\bar{A}$-expansions, and the same is true for any $y \in [w, w']$, hence $f^k_{\bar{A}}(y) = y_k \in [A_{\sigma(n_k)}, A_{\sigma(n_k)+1})$ for $0 \le k \le m$, and $y = T_{n_0} T_{n_1} \cdots T_{n_m}(y_m)$.

Since $[A_i, A_{i+1}) \subset (P_i, Q_{i+1})$, any point $y \in [A_i, A_{i+1})$ is inside the isometric circle for T_i, and thus $T'_i(y) > \mu_i > 1$ for some $\mu_i > 1$. Using this for $i = \sigma(n_0), \sigma(n_1), \ldots, \sigma(n_m)$ and the Chain Rule, we conclude that the derivative of the composite function

$$f^{m+1}_{\bar{A}}(y) = T_{\sigma(n_m)} \cdots T_{\sigma(n_1)} T_{\sigma(n_0)}(y)$$

is $> \mu^{m+1}$, and hence the derivative of the inverse function

$$f^{-(m+1)}_{\bar{A}}(y_m) = T_{n_0} T_{n_1} \cdots T_{n_m}(y_m)$$

is $< \lambda^{m+1}$ for some $\lambda < 1$. Then, by the Mean Value Theorem, the arc length distance

$$\ell(w, w') = \left(f^{-(m+1)}_{\bar{A}}\right)'(c) \cdot \ell(f^{-(m+1)}_{\bar{A}} w, f^{-(m+1)}_{\bar{A}} w')$$

for some $c \in [w, w']$. Since $(f^{-(m+1)}_{\bar{A}})'(c) < \lambda^{m+1}$ and the arclength distance between any two points is at most 2π, we have that

$$\ell(w, w') \le 2\pi \lambda^{m+1}.$$

Now we choose m large enough so that w and w' are so close that for any $x \in [u, u']$ $(x, y) \in \Omega_{\bar{A}}$ (here we use the rectangular structure of $\Omega_{\bar{A}}$ and are avoiding the corners). We recall that we used bijectivity of $F_{\bar{A}}$ to define the sequence of pairs $(x_{-i}, y_{-i}) \in \Omega_{\bar{A}}$ to the past, $1 \le i \le m$. Since the first m symbols with negative indices in the coding sequences s and s' are the same, for any $y \in [w, w']$ we have $y_{-i} \in [A_{\sigma(n_{-i})}, A_{\sigma(n_{-i}+1)})$ and hence

$$f^m_{\bar{A}}(y_{-m}) = T_{\sigma(n_{-1})} T_{\sigma(n_{-2})} \cdots T_{\sigma(n_{-m})}(y_{-m}) = y,$$

and at the same time we have

$$T_{\sigma(n_{-1})}T_{\sigma(n_{-2})}\cdots T_{\sigma(n_{-m})}(x_{-m}) = x$$

for $x \in [u, u']$.

But $T_{\sigma(n_{-i})}$ were determined by the second coordinate y_{-i}, and since $(x_{-i}, y_{-i}) \in \Omega_{\bar{A}}$, and y_{-i} is inside the isometric circle for $T_{\sigma(n_{-i})}$, x_{-i} is outside that isometric circle, so the derivative of the composite function $T_{\sigma(n_{-1})}T_{\sigma(n_{-2})}\cdots T_{\sigma(n_{-m})}(x_{-m})$ is $< \lambda^m$. Therefore

$$\ell(u, u') = \ell\big(T_{\sigma(n_{-1})}T_{\sigma(n_{-2})}\cdots T_{\sigma(n_{-m})}(u_{-m}), T_{\sigma(n_{-1})}T_{\sigma(n_{-2})}\cdots T_{\sigma(n_{-k})}(u'_{-m})\big) \leq 2\pi\lambda^m$$

for some $\lambda < 1$.

Therefore the geodesics are uniformly $2\pi\lambda^m$-close. But the tangent vectors $v = \mathrm{Cod}(s)$ and $v' = \mathrm{Cod}(s')$ in $C_{\bar{A}}$ are determined by the first intersection of the corresponding geodesic with the boundary of $\mathcal{F}$ or $\mathcal{F}_j$ for a particular j determined in Corollary 5.4. Hence, by making m large enough we can make v' as close to v as we wish. Thus we proved that the map Cod is uniformly continuous on the set of admissible coding sequences and therefore extends to the closure $X_{\bar{A}}$ as a uniformly continuous map. □

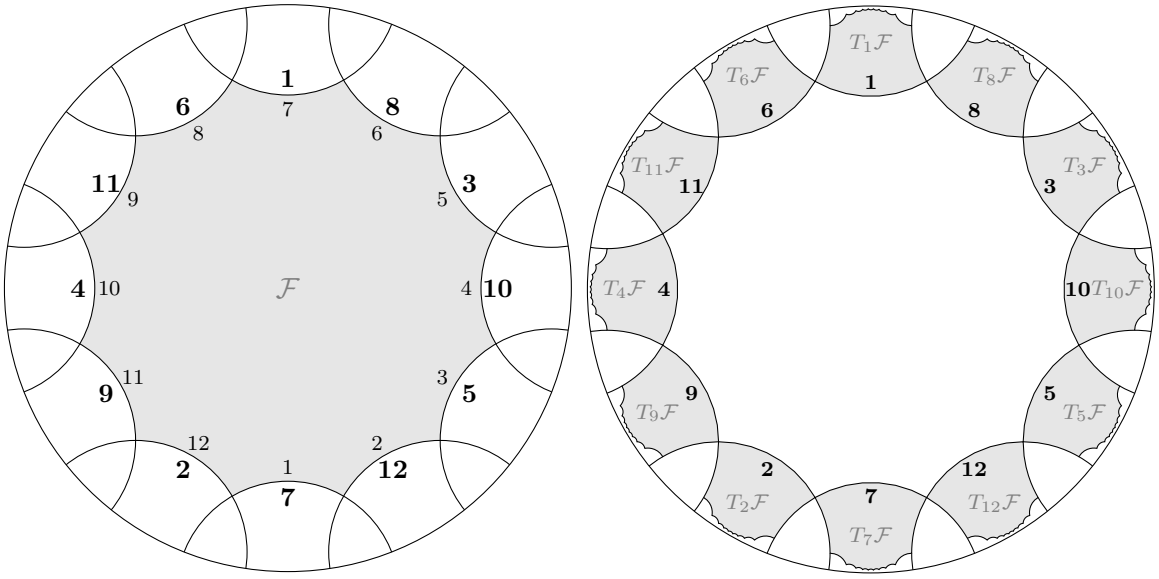
In conclusion, the geodesic flow becomes a special flow over a symbolic dynamical system $(X_{\bar{A}}, \sigma)$ on the finite alphabet $\mathbb{A}$. The ceiling function $g_{\bar{A}}(s)$ on $X_{\bar{A}}$ is the time of the first return to the cross-section $C_{\bar{A}}$ of the geodesic associated to s.

5.4. Examples of coding

Note that when a geodesic segment exits $\mathcal{F}$ through side i it is $\sigma(i)$ that is used in the geometric code, as described in the Introduction. Similarly, if the endpoint of a geodesic lies in $[A_i, A_{i+1})$, it is $\sigma(i)$ that is used in the arithmetic code, as described in Section 5.3. We could call i the "inside" numbering of the sides of $\mathcal{F}$ (this is the numbering used in the Introduction), while $\sigma(i)$ is the "outside" numbering, as shown in Figure 12. See [20, Introduction] for more detail on geometric codes and numbering conventions.

As described in [22, Sec. 2], an axis of a transformation in Γ (that is, a geodesic whose start and end points are the repelling and attracting fixed points of the transformation) becomes a closed geodesic in $\Gamma\backslash\mathbb{D}$ and has a periodic geometric code. For both arithmetic and geometric codes, we denote repeating codes by only showing the repetend, that is,

$$[n_0, \ldots, n_k] \quad \text{means} \quad [\ldots, n_k, n_0, n_1, n_2, \ldots, n_k, n_0, \ldots].$$



The "outside" labels in bold, corresponding to $T_k^{-1} = T_{\sigma(k)}$

The same numbers are "inside" labels of $T_k\mathcal{F}$.

FIGURE 12: Inside and outside numbering of $T_k\mathcal{F}$.

Example 1: codes agree.

Let γ_0 be the axis of $T_2T_8T_5$. Then

$$\begin{aligned}
\gamma_1 &= T_{12}\gamma_0 && \text{because } w_0 \in [A_{12}, A_1)\\
\gamma_2 &= T_6T_{12}\gamma_0 && \text{because } w_1 \in [A_6, A_7)\\
\gamma_3 &= T_3T_6T_{12}\gamma_0 && \text{because } w_2 \in [A_3, A_4)
\end{aligned}$$

and $\gamma_3 = \gamma_0$. Thus the arithmetic code is

$$[\gamma]_{\bar{A}} = [\sigma(12), \sigma(6), \sigma(3)] = [2, 8, 5].$$

The geometric code for this example is the same because for each geodesic γ_k the index i for which $w_k \in [A_i, A_{i+1})$ is the same i as the side through which γ_k exits $\mathcal{F}$; see Figure 13(a).

Example 2: codes disagree.

Let γ_0 be the axis of $T_5T_4T_7T_6$. Then

$$\begin{aligned}
\gamma_1 &= T_3\gamma_0 && \text{because } w_0 \in [A_3, A_4)\\
\gamma_2 &= T_{10}T_3\gamma_0 && \text{because } w_1 \in [A_{10}, A_{11})\\
\gamma_3 &= T_{12}T_{10}T_3\gamma_0 && \text{because } w_2 \in [A_{12}, A_1)\\
\gamma_4 &= T_1T_{12}T_{10}T_3\gamma_0 && \text{because } w_3 \in [A_1, A_2)
\end{aligned}$$

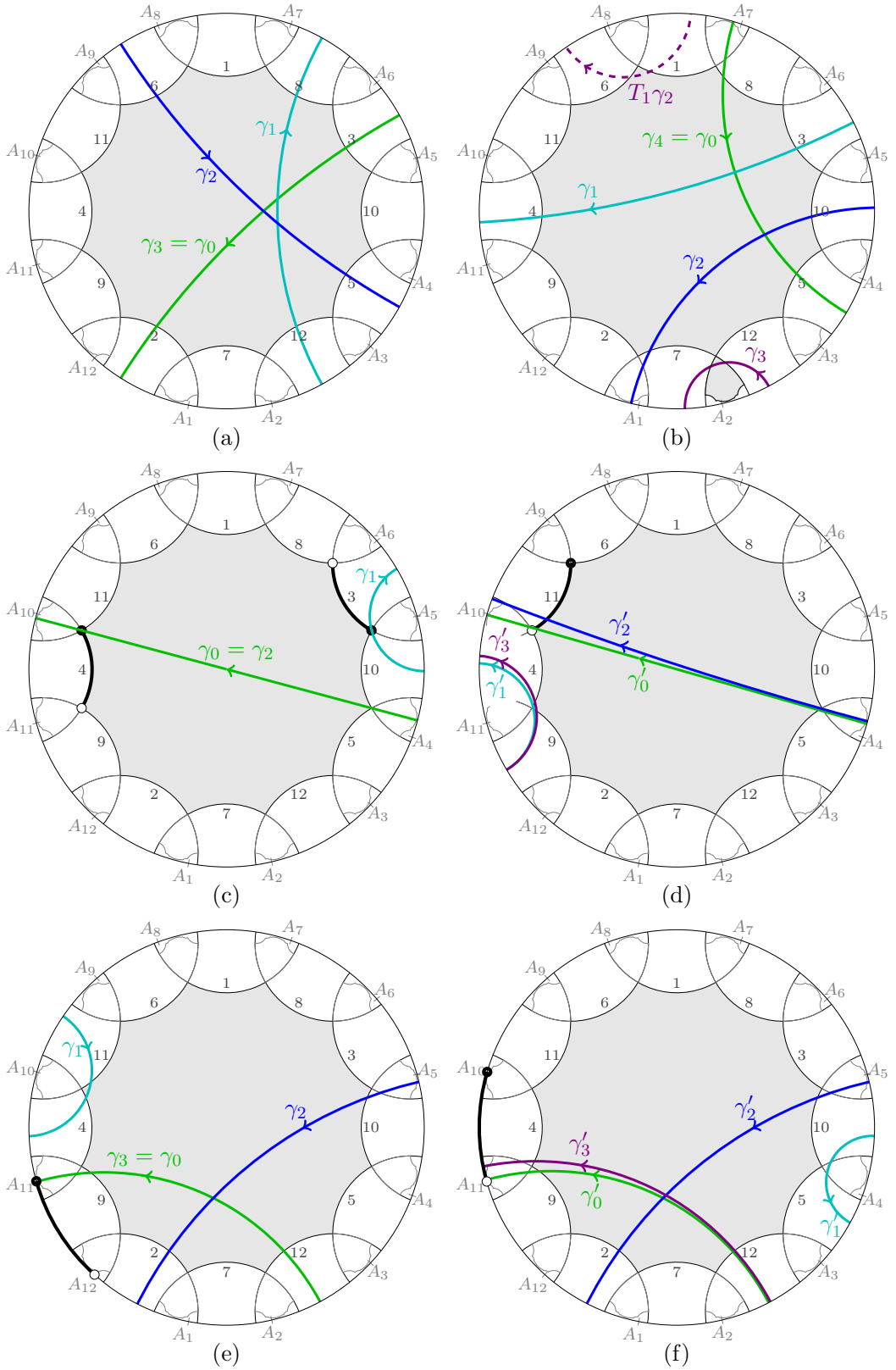


FIGURE 13: (a) Example 1. (b) Example 2. (c-d) Example 3. (e-f) Example 4.

and $\gamma_4 = \gamma_0$. The arithmetic code is thus

$$[\gamma]_{\bar{A}} = [\sigma(3), \sigma(10), \sigma(12), \sigma(1)] = [5, 4, 2, 7].$$

To find the geometric code of γ, we look at the sides through which segments exit $\mathcal{F}$. γ_0 exits through side 3 ("inside" numbering), so $T_3\gamma = \gamma_1$ is the next segment. Then γ_1 exits through side 10 and $T_{10}T_3\gamma = \gamma_2$ is the next segment. At this point, we see that γ_2 exits $\mathcal{F}$ through side 1, so the geometrically next segment is $T_1T_{10}T_3\gamma = T_1\gamma_2$ (this is *not* γ_3). The geodesic $T_1\gamma_2$ is shown as a dashed curve at the top of Figure 13(b). It exits $\mathcal{F}$ through side 8, and then $T_6(T_1\gamma_2) = \gamma_0$, so the geometric code for this example is

$$[\gamma]_G = [\sigma(3), \sigma(10), \sigma(1), \sigma(8)] = [5, 4, 7, 6].$$

The geometric code $[5, 4, 7, 6]$ differs from the arithmetic code $[5, 4, 2, 7]$ for this example precisely because γ_2 exits through side $1 = \sigma(7)$ but ends at $w_2 \in [A_{12}, A_1)$, not in the interval $[A_1, A_2)$. Note that γ_3 intersects the corner image $\mathcal{F}_2$ and that $T_1\gamma_2$ is exactly $U_2^{-1}\gamma_3$.

Example 3: multiple geometric codes.

The axis of U_{10} goes from M_4 to M_{10} and passes directly through the vertex V_{10} as it exits $\mathcal{F}$.

By convention, each vertex V_i is considered to be part of side i. Using this convention, the geometric code of γ should start with $\sigma(10) = 4$ (see Figure 13(c), in which side 10 is shown with a thick black arc). Then $\gamma_1 = T_{10}\gamma_0$ is seen to pass through vertex V_5, so we code with $\sigma(5) = 3$ (side 5 also has a thick black arc in Figure 13(c)). The next geodesic is $\gamma_2 = T_5\gamma_1 = T_5T_{10}\gamma_0 = U_4\gamma_0$, which is exactly γ_0 because U_4 fixes M_4 and M_{10} by Lemma 5.16. The conventional geometric code of γ is therefore $[4, 3]$.

Now consider a geodesic γ' very close to $\gamma = M_4M_{10}$ that exits $\mathcal{F}$ through side 9 instead of side 10, see Figure 13(d). Its geometric code starts with $\sigma(9) = 11$, and then $\gamma'_1 = T_9\gamma'$ will exit through side 10, so the code continues with $\sigma(10) = 4$. The geodesic $\gamma'_2 = T_{10}T_9\gamma'$ will be somewhat close to γ, but later iterates γ'_{2k} will eventually stop exiting through side 9. However, taking γ' sufficiently close to γ will give geodesics whose forward geometric codes begin with arbitrarily many repetitions of 11 and 4. For example, there might be a geodesic with code

$$[\ldots, 5, 7, \quad 11, 4, 11, 4, 11, 4, 11, 4, \quad 8, 3, 1, \ldots].$$

The repeating code $[11, 4]$ will not be the admissible geometric code of this geodesic (or in fact any geodesic), but it is in the closure of the space of admissible geometric codes. Note that the transformation $T_{11}T_4$ is exactly equal to T_4T_3 by (11), as both represent U_{10}.

The codes $[4, 3]$ and $[11, 4]$ both code the geodesic M_4M_{10}. Using one of $[4, 3]$ or $[11, 4]$ in the future and the other in the past gives two non-periodic codes as well.

Example 4: multiple arithmetic codes.

The repelling and attracting fixed points of $T_4T_5T_2(z)$ are, respectively, $e^{-1.07822\,i}$ and $e^{-2.86313\,i}$. The repelling point w_0 is in (P_{11}, Q_{11}), and for our particular choice of $\bar{A}$ in these examples it is exactly A_{11}.

In Figure 13(e), the interval $[A_{11}, A_{12})$ is shown with a thick black arc. Using the convention that T_i be applied to A_i, we have

$$\begin{aligned} \gamma_1 &= T_{11}\gamma_0 && \text{because } w_0 \in [A_{11}, A_{12}) \\ \gamma_2 &= T_{10}T_{11}\gamma_0 && \text{because } w_1 \in [A_{10}, A_{11}) \\ \gamma_3 &= T_{12}T_{10}T_{11}\gamma_0 && \text{because } w_2 \in [A_{12}, A_1) \end{aligned}$$

and $\gamma_3 = \gamma_0$. The arithmetic code is thus

$$[\gamma]_{\bar{A}} = [\sigma(11), \sigma(10), \sigma(12)] = [9, 4, 2].$$

Now consider a geodesic γ' very close to γ but with attracting endpoint slightly clockwise of A_{11}, see Figure 13(f). Let us denote the sequence of its iterates under $F_{\bar{A}}$ by γ'_1, γ'_2, etc.

$$\begin{aligned} \gamma'_1 &= T_{10}\gamma'_0 && \text{because } w'_0 \in [A_{10}, A_{11}) \\ \gamma'_2 &= T_3T_{10}\gamma'_0 && \text{because } w'_1 \in [A_3, A_4) \\ \gamma'_3 &= T_{12}T_3T_{10}\gamma'_0 && \text{because } w'_2 \in [A_{12}, A_1) \end{aligned}$$

The geodesic γ'_3 will be somewhat close to γ, but later iterates γ'_{3k} will eventually not end in $[A_{10}, A_{11})$. However, taking γ' sufficiently close to γ will give geodesics whose forward arithmetic codes have arbitrarily many repetitions of

$$[\sigma(10), \sigma(3), \sigma(12)] = [4, 5, 2].$$

The repeating code $[4, 5, 2]$ will not be the admissible arithmetic code of this (or any) geodesic, but it is in the closure $X_{\bar{A}}$ of the space of admissible arithmetic codes.

The codes $[9, 4, 2]$ and $[4, 5, 2]$ both code the geodesic γ. Using one of $[9, 4, 2]$ or $[4, 5, 2]$ in the future and the other in the past gives two non-periodic codes as well.

It is worth noting that the transformations $T_4T_5T_2$ and $T_9T_4T_2$ are identical: restating (5) as $T_{\rho(i)}T_i = (T_{\rho^2(i)})^{-1}(T_{\rho^3(i)})^{-1} = T_{\sigma\rho^2(i)}T_{\sigma\rho^3(i)}$, we have that

$$(T_4T_5)T_2 = (T_{\rho(5)}T_5)T_2 = (T_{\sigma\rho^2(5)}T_{\sigma\rho^3(5)})T_2 = (T_{\sigma(11)}T_{\sigma(10)})T_2 = (T_9T_4)T_2.$$

5.5. Markov and sofic partitions

Adler and Flatto [4] show that the symbolic system associated to the map we call $F_{\bar{P}} : \Omega_{\bar{P}} \to \Omega_{\bar{P}}$ is sofic with respect to the alphabet $\mathbb{A} = \{1, 2, \ldots, 8g-4\}$. This is not always possible for $F_{\bar{A}} : \Omega_{\bar{A}} \to \Omega_{\bar{A}}$, but there are several examples of $\bar{A}$ for which

the shift is sofic. In this subsection and the next, we first give a sufficient condition for sofic shifts to occur, and then we give some examples.

As described in [4, Appendix C], sofic systems are obtained from Markov ones by amalgamation of the alphabet. Conversely, Markov is obtained from sofic by refinement of the alphabet, which in this case is realized by splitting each symbol $i \in \mathbb{A}$ into three symbols i_1, i_2, i_3. A partition of $\Omega_{\bar{A}}$ whose shift is Markov with respect to the *extended alphabet* $\{ i_k : i \in \mathbb{A}, k = 1, 2, 3 \}$ will be sofic with respect to the alphabet $\mathbb{A}$.

Remark 5.14. The value $U_i^{-1} A_i$ is a "cycle end" (see [25, Sec. 3]) because

$$T_k B_k = T_k T_{\sigma(k-1)} A_{\sigma(k-1)} = U_{\sigma(k-1)}^{-1} A_{\sigma(k-1)}.$$

Proposition 5.15. *Let $\bar{A}$ have the short cycle property. A Markov partition (with respect to the extended alphabet) exists for $F_{\bar{A}} : \Omega_{\bar{A}} \to \Omega_{\bar{A}}$ if for all $i \in \bar{A}$ there exists a $j \in \bar{A}$ such that $U_i^{-1} A_i \in \{A_j, B_j, C_j\}$.*

Proof. Following the notion of the "fine partition" in [4], we split each strip of $\Omega_{\bar{A}}$ with $w \in [A_i, A_{i+1})$ into three rectangles R_{i_1}, R_{i_2}, and R_{i_3}. Assuming that $C_i \in (A_i, B_i]$, define

$$\begin{aligned} R_{i_1} &:= [Q_{i+1}, P_{i-1}] \times [A_i, C_i] \\ R_{i_2} &:= [Q_{i+2}, P_{i-1}] \times [C_i, B_i] \\ R_{i_3} &:= [Q_{i+2}, P_i] \times [B_i, A_{i+1}] \end{aligned}$$

(if $B_i \in (A_i, C_i)$ instead, the process is very similar except that $[A_i, A_{i+1}]$ is partitioned into $[A_i, B_i]$, $[B_i, C_i]$, and $[C_i, A_{i+1}]$). The left side of Figure 14 shows the partition $\{ R_{i_k} : i \in \mathbb{A}, k = 1, 2, 3 \}$ when each A_i is the midpoint of $[P_i, Q_i]$. Note that each R_{i_2} is extremely thin.

In accordance with [1, Theorem 7.9], it is sufficient to prove that for any pair of distinct symbols i_k and j_ℓ, the rectangles $F_{\bar{A}}(R_{i_k})$ and R_{j_ℓ} are either disjoint (that is, their interiors are disjoint) or intersect "transversally," i.e., their intersection is a rectangle with two horizontal sides belonging to the horizontal boundary of R_{j_ℓ} and two vertical sides belonging to the vertical boundary of $F_{\bar{A}}(R_{i_k})$. The images $F_{\bar{A}}(R_{i_k})$ are

$$\begin{aligned} F_{\bar{A}}(R_{i_1}) &= [T_i Q_{i+1}, T_i P_{i-1}] \times [T_i A_i, T_i C_i] \\ &= [P_{\sigma(i)}, P_{\sigma(i)+1}] \times [B_{\sigma(i)+1}, U_{\tau(\sigma(i))}^{-1} A_{\tau(\sigma(i))}] \\ F_{\bar{A}}(R_{i_2}) &= [T_i Q_{i+2}, T_i P_{i-1}] \times [T_i C_i, T_i B_i] \\ &= [Q_{\sigma(i)}, P_{\sigma(i)+1}] \times [U_{\tau(\sigma(i))}^{-1} A_{\tau(\sigma(i))}, U_{\tau(\sigma(i))+1}^{-1} A_{\tau(\sigma(i))+1}] \\ F_{\bar{A}}(R_{i_3}) &= [T_i Q_{i+2}, T_i P_i] \times [T_i B_i, T_i M_{i+1}] \\ &= [Q_{\sigma(i)}, Q_{\sigma(i)+1}] \times [U_{\tau(\sigma(i))+1}^{-1} A_{\tau(\sigma(i))+1}, C_{\sigma(i)-1}]. \end{aligned}$$

From this, it can be seen that $F_{\bar{A}}(R_{i_k}) \cap R_{j_\ell}$, if it is non-empty, will be a rectangle with vertical sides from $\bar{P} \cup \bar{Q}$ as required since the rectangles $R_{i_1}, R_{i_2}, R_{i_3}$ also use $\bar{P} \cup \bar{Q}$ for the values of their vertical sides. Let $m = \tau(\sigma(i))$. The horizontal sides may be $B_{\sigma(i)+1}$ or $C_{\sigma(i)-1}$ or may be one of the values $U_m^{-1}A_m$ or $U_{m+1}^{-1}A_{m+1}$, which in general are not values of the horizontal sides of the original partition elements. Thus the partition $\{ R_{i_k} : i \in \mathbb{A}, k = 1, 2, 3 \}$ is Markov exactly when the values $U_m^{-1}A_m$ and $U_{m+1}^{-1}A_{m+1}$ are already horizontal sides of $R_{i_1}, R_{i_2}, R_{i_3}$. This happens exactly when $U_m^{-1}A_m$ and $U_{m+1}^{-1}A_{m+1}$ are equal to some A_j, B_j, or C_j. □

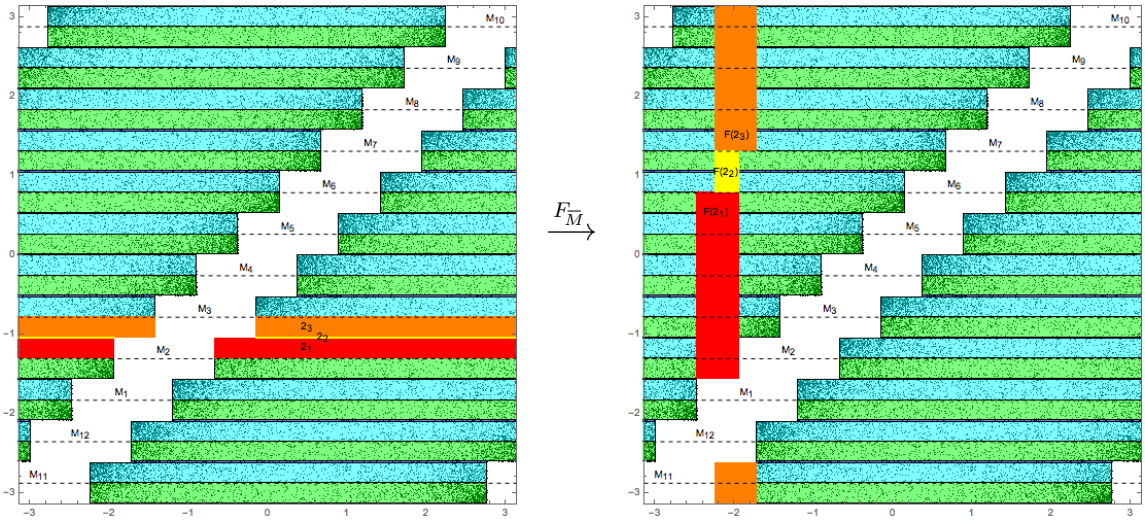


FIGURE 14: Markov partition of $\Omega_{\overline{M}}$ for genus $g = 2$.

The most notable example of a partition $\bar{A}$ satisfying the condition of Proposition 5.15 is the "midpoint" setup, in which each A_i is exactly halfway between P_i and Q_i. We label these midpoints M_i and write $F_{\overline{M}}$ for this specific case of $F_{\bar{A}}$. Note that each M_i does have the short cycle property, so all the previous results on short cycles hold for $\bar{A} = \overline{M}$.

To show that $\bar{A} = \overline{M}$ satisfies the condition of Proposition 5.15, we need that $U_i^{-1}M_i$ is equal to some M_j, B_j, or C_j. In fact, $U_i^{-1}M_i = M_i$ by the following lemma.

Lemma 5.16. *The fixed points of U_i are M_i and $M_{\tau(i)}$.*

Proof. Consider the image under $U_i = T_{\sigma(i)}T_{\tau(i)-1}$ of the vertices V_i and $V_{\tau(i)}$. To shorten notation, let C_j be the isometric circle P_jQ_{j+1}, which is the extension of side j of $\mathcal{F}$. Then each vertex V_j is the intersection of C_j and C_{j-1}. The map $T_{\tau(i)-1}$ sends $V_{\tau(i)}$ to $V_{\sigma(i)+1}$ because it maps the isometric circles $C_{\tau(i)-1}$ and $C_{\tau(i)}$ to $C_{\sigma(i)+1}$ and $C_{\sigma(i)}$, respectively. Then $T_{\sigma(i)}$ sends $V_{\sigma(i)+1}$ to V_i because it maps $C_{\sigma(i)+1}$ and $C_{\sigma(i)}$ to C_{i-1} and C_i, respectively. Thus $U_i(V_{\tau(i)}) = V_i$.

The geodesic connecting V_i and $V_{\tau(i)}$ in the disk model is a Euclidean line segment through the origin. Call this segment L. The isometric circles $C_{\tau(i)-1}$ and $C_{\tau(i)}$ intersect perpendicularly at $V_{\tau(i)}$, and L bisects the right angle $V_{\tau(i)-1}V_{\tau(i)}V_{\tau(i)+1}$. Thus the image U_iL is a geodesic in $U_i\mathcal{F}$ that bisects the right angle $(U_iV_{\tau(i)-1})V_i(U_iV_{\tau(i)+1})$. This means that U_iL is the extension of the line L into $U_i\mathcal{F}$.

The full geodesic containing L is precisely the line connecting M_i and $M_{\tau(i)}$, and this line must (as a set) be fixed by $U_i : \mathbb{D} \to \mathbb{D}$. Since U_i maps $\partial\mathbb{D} = \mathbb{S}$ to itself, the fixed points of U_i are precisely the two points M_i and $M_{\tau(i)}$. □

5.6. Examples of Markov/sofic setups

We now give a few examples of choices for $\bar{A}$ that satisfy the condition of Proposition 5.15 and therefore have a symbolic system that is sofic with respect to the alphabet $\mathbb{A} = \{1, 2, \ldots, 8g-4\}$.

Example 1: midpoints.

Suppose $U_i^{-1}A_i = A_i$ for all i. Then by Lemma 5.16, $A_i = M_i$ (since $M_{\tau(i)}$ is not in (P_i, Q_i), it cannot be A_i). See Figure 14 for the actual attractor and Markov partition when $g = 2$.

Example 2.

Suppose $U_i^{-1}A_i = A_{i+1}$ for all i. Then we can build a sequence of equations

$$\begin{aligned} A_2 &= U_1^{-1}A_1 \\ A_3 &= U_2^{-1}A_2 = U_2^{-1}U_1^{-1}A_1 \\ A_4 &= U_3^{-1}A_3 = U_3^{-1}U_2^{-1}U_1^{-1}A_1 \\ &\vdots \\ A_1 &= U_{8g-4}^{-1}A_{8g-4} = U_{8g-4}^{-1}U_{8g-5}^{-1}\cdots U_2^{-1}U_1^{-1}A_1 \end{aligned}$$

giving that A_1 has to be a fixed point of $U_1U_2\cdots U_{8g-4}$. By Lemma 5.17 below, this product does have an attracting fixed point in $[b_1, a_1] \subset (P_1, Q_1)$, so we can choose A_1 to be this fixed point. The remaining A_i, $i \neq 1$, will be the attracting fixed points of the cyclic product $U_iU_{i+1}\cdots U_{i-2}U_{i-1}$, where the indices are mod $8g-4$. By the cyclic permutation of indices in Lemma 5.17, $A_i \in [b_i, a_i]$. Notice that A_i, $i \neq 1$, will be shifts of A_1 by $(i-1)\frac{\pi}{4g-2}$ since each U_i differs from U_1 by conjugation with a rotation.

Lemma 5.17. *The product $U_1U_2\cdots U_{8g-4}$ has an attracting fixed point in $[b_1, a_1]$.*

Proof. Recall that $a_i = T_{\sigma(i)}P_{\rho(i)+1}$ and $b_i = T_{\sigma(i-1)}Q_{\theta(i-1)}$, and any $A_i \in [b_i, a_i]$ satisfies the short cycle property (see [25, Corollary 8.2(i)]). Using Proposition 5.1 we obtain

$$a_i = T_{\sigma(i)}P_{\rho(i)+1} = T_{\sigma(i)}P_{\sigma(\tau(i-1))+1} = T_{\sigma(i)}T_{\tau(i-1)}P_{\tau(i)-2} = U_iP_{\tau(i)-2}$$
$$b_i = T_{\sigma(i-1)}Q_{\theta(i-1)} = T_{\sigma(i-1)}Q_{\sigma(\tau(i))} = T_{\sigma(i-1)}T_{\tau(i)}Q_{\tau(i)+2} = U_iQ_{\tau(i)+2}$$

Then $U_i([Q_{\tau(i)+2}, P_{\tau(i)-2}]) \subset [b_i, a_i]$. Take $x \in [b_1, a_1]$ and look at its orbit under a descending product of maps U_i:

$$\begin{aligned} U_{8g-4}(x) &\in [b_{8g-4}, a_{8g-4}] \\ U_{8g-5}U_{8g-4}(x) &\in [b_{8g-5}, a_{8g-5}] \\ U_{8g-6}U_{8g-5}U_{8g-4}(x) &\in [b_{8g-6}, a_{8g-6}] \\ &\vdots \\ U_1U_2\cdots U_{8g-4}(x) &\in [b_1, a_1]. \end{aligned}$$

Continuing, we obtain a sequence of points $x_n = (U_1U_2\cdots U_{8g-4})^n(x) \in [b_1, a_1]$, which by compactness has a limit point in the set, and this point is the attracting fixed point since the attracting fixed point is unique. □

Examples 1 and 2 are "equally spaced" in the sense that the distance $d(A_i, A_{i+1})$ on S^1 is the same for each i. The following example does not have this property.

Example 3.

Suppose $U_i^{-1}A_i = A_i$ for odd i and $U_i^{-1}A_i = A_{i+1}$ for even i. Then by Lemma 5.16, $A_i = M_i$ for odd i. For even i, we get that $A_i = U_iM_{i+1}$ simply by applying U_i to both sides of $U_i^{-1}A_i = M_{i+1}$. Since M_{i+1} is outside of the isometric circle for U_i, we see that indeed $A_i = U_iM_{i+1} \in (P_i, Q_i)$.

5.7. Dual codes

In [24, Sec. 5], Katok and Ugarcovici discuss cases in which expansions using two different parameters are "dual" and show that the existence of a dual code implies a sofic shift. The corresponding definition in current setup would be the following:

Definition. Let $\bar{A} = \{A_i\}$ and $\bar{A}' = \{A_i'\}$ be two partitions (not necessary having the short cycle property) such that $F_{\bar{A}}$ and $F_{\bar{A}'}$ have attractors $\Omega_{\bar{A}}$ and $\Omega_{\bar{A}'}$, respectively, with finite rectangular structure, and let $\phi(x, y) = (y, x)$ be the reflection of the plane about the line $y = x$. We say that $\bar{A}$ and $\bar{A}'$ are *dual* (equivalently, each is the

dual of the other) if $\Omega_{\bar{A}'} = \phi(\Omega_{\bar{A}})$ and the following diagram is commutative:

$$\begin{array}{ccc} \Omega_{\bar{A}} & \xrightarrow{\phi} & \Omega_{\bar{A}'} \\ {\scriptstyle F_{\bar{A}}^{-1}}\downarrow & & \downarrow{\scriptstyle F_{\bar{A}'}} \\ \Omega_{\bar{A}} & \xrightarrow[\phi]{} & \Omega_{\bar{A}'} \end{array} . \tag{28}$$

In order to state the next result, we must first introduce the *backward* $\bar{A}'$*-expansion* of u_0, defined as

$$[u_0]^-_{\bar{A}'} := [m_{-1}, m_{-2}, m_{-3}, \dots], \tag{29}$$

where, for $k < 0$, $m_k = i \in \mathbb{A}$ if $f_{\bar{A}'}^{-k+1}(u_0) \in [A'_i, A'_{i+1})$.

Theorem 5.18. *Suppose $\bar{A}$ admits a dual $\bar{A}'$, both $F_{\bar{A}}$ and $F_{\bar{A}'}$ have attractors $\Omega_{\bar{A}}$ and $\Omega_{\bar{A}'}$, respectively, with finite rectangular structure, and let γ be an $\bar{A}$-reduced geodesic from u to w. Then the coding sequence*

$$[\gamma]_{\bar{A}} = [\dots, n_{-2}, n_{-1}, n_0, n_1, n_2, \dots], \quad n_i \in \mathbb{A}, \tag{30}$$

is obtained by juxtaposing the forward $\bar{A}$-expansion of w, $[w]^+_{\bar{A}}$, and the backward $\bar{A}'$-expansion of u, $[u]^-_{\bar{A}'}$. This property is preserved under the left shift of the sequence.

Proof. Let $(u, w) \in \Omega_{\bar{A}}$. Then $\phi(u, w) = (w, u) \in \Omega_{\bar{A}'}$. Using (28) we obtain

$$F_{\bar{A}'}(w, u) = \phi \circ F_{\bar{A}}^{-1}(u, w) = (w_{-1}, u_{-1}).$$

Let $w_{-1} \in (A_i, A_{i+1})$. Then, by definition, the forward $\bar{A}$-expansion of w_{-1} begins with $n_{-1} = \sigma(i)$, $f_{\bar{A}} = T_i$, and

$$F_{\bar{A}}(u_{-1}, w_{-1}) = (T_i u_{-1}, T_i w_{-1}) = (u, w).$$

Therefore $w_{-1} = T_i^{-1} w = T_{\sigma(i)} w$.

On the other hand, let

$$u = [m_{-1}, m_{-2}, \dots].$$

This means that $u \in [A'_{m_{-1}}, A'_{m_{-1}+1})$ and

$$F_{\bar{A}'}(w, u) = T_{m_{-1}}(w, u) = (T_{m_{-1}} w, T_{m_{-1}} u) = (w_{-1}, u_{-1}).$$

Therefore $T_{m_{-1}} = T_{\sigma(i)}$ which implies $m_{-1} = n_{-1}$. Continuing by induction, one proves that all digits of the "past" of the sequence (30) are the digits of the backward $\bar{A}'$-expansion of u.

In order to see what happens under a left shift, we reverse the diagram to obtain

$$\begin{array}{ccc} \Omega_{\bar{A}} & \xrightarrow{\phi} & \Omega_{\bar{A}'} \\ {\scriptstyle F_{\bar{A}}}\downarrow & & \downarrow{\scriptstyle F^{-1}_{\bar{A}'}} \\ \Omega_{\bar{A}} & \xrightarrow[\phi]{} & \Omega_{\bar{A}'} \end{array} . \tag{31}$$

Let $(u, w) \in \Omega_{\bar{A}}$. Using (31) we obtain

$$F^{-1}_{\bar{A}'}(w, u) = \phi \circ F_{\bar{A}}(u, w) = (w_1, u_1).$$

Let $w \in (A_i, A_{i+1})$. Then, by definition, the forward $\bar{A}$-expansion of w begins with $n_0 = \sigma(i)$, $f_{\bar{A}} = T_i$, and

$$F_{\bar{A}}(u, w) = (T_i u, T_i w) = (u_1, w_1).$$

Therefore $w_1 = T_i w$.

On the other hand, let

$$u_1 = [m_0, m_{-1}, m_{-2}, \ldots].$$

This means that $u_1 \in [A'_{m_0}, A'_{m_0+1})$ and

$$F_{\bar{A}'}(w_1, u_1) = T_{m_0}(w_1, u_1) = (T_{m_0} w_1, T_{m_0} u_1) = (w, u).$$

Therefore $w_1 = T^{-1}_{m_0} w = T_{\sigma(m_0)} w$. Thus $T_{\sigma(m_0)} = T_i$ which implies $m_0 = n_0 = \sigma(i)$. □

By inspection, one can see that the partitions $\bar{P}$ and $\bar{Q}$ are dual. Future work will investigate other instances of duality among maps for which each A_i is either P_i or Q_i. Without giving the proofs here, it is interesting to note that the partitions

$$\overline{PQ} = \{P_1, Q_2, P_3, Q_3, \ldots, P_{8g-5}, Q_{8g-4}\}$$
$$\overline{QP} = \{Q_1, P_2, Q_3, P_4, \ldots, Q_{8g-5}, P_{8g-4}\}$$

are dual to each other and that each of

$$\overline{PPQQ} = \{P_1, P_2, Q_3, Q_4, P_5, P_6, \ldots\}$$

and

$$\overline{QQPP} = \{Q_1, Q_2, P_3, P_4, Q_5, Q_6, \ldots\}$$

is self-dual ($\bar{A}' = \bar{A}$). Proofs of these dualities, and indeed of the structure of attractors for these specific partitions are known at this time (see Figure 15), but the general case of $A_i \in \{P_i, Q_i\}$ is not completely understood.

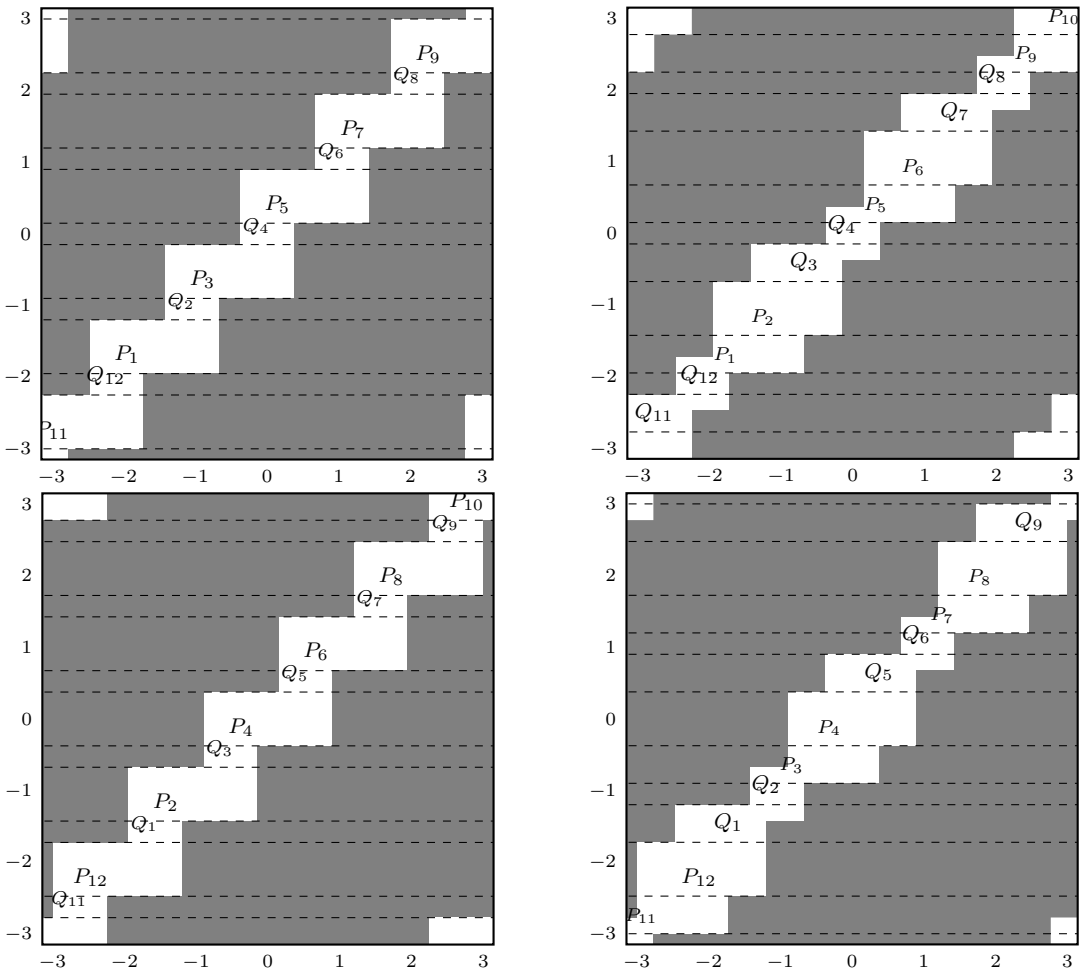


FIGURE 15: Attractors for different partitions, all with $g = 2$.

Restricting to short cycles with $A_i \in (P_i, Q_i)$ in fact yields no examples of duality:

Proposition 5.19. *There do not exist $\bar{A}$ and $\bar{A}'$, where $A_i, A_i' \in (P_i, Q_i)$, such that $\bar{A}$ and $\bar{A}'$ are dual and both satisfy the short cycle property.*

Proof. Recall that for short cycles the upper part of $\Omega_{\bar{A}}$ has corner points

$$(P_i, B_i) \quad \text{(upper part)} \qquad \text{and} \qquad (Q_{k+1}, C_i) \quad \text{(lower part)}.$$

where B_i and C_i are defined in (10); we define B_i' and C_i' similarly based on $\bar{A}'$.

Suppose we have $\bar{A}$ and $\bar{A}'$ dual with short cycles. Then $\psi(P_i, B_i) = (B_i, P_i)$ is a lower corner point for $\Omega_{\bar{A}'}$, meaning that $(B_i, P_i) = (Q_{j+1}, C_j')$ for some j. Specifically, $B_i = Q_{j+1}$. But since $B_i \in (Q_i, A_{i+1})$ by [25, Def. 3.10], B_i cannot be Q_{j+1} for any j, which leads to a contradiction.

This contradiction technically completes the proof, but since it relies on the lack of Q_{j+1} in the *open* interval (Q_i, A_{i+1}), it is interesting to consider the closed interval as well. There are two potential solutions to $B_i = Q_{j+1}$ in the closed interval $[Q_i, A_{i+1}]$, namely, $A_{i+1} = Q_{j+1}$ and $Q_i = Q_{j+1}$. For the former, we must use a value from $\bar{Q}$ in $\bar{A}$. For the latter, we have $j = i-1$, and from (10) and Proposition 4.1, we get that

$$\begin{aligned} B_i &= Q_i \\ T_{\sigma(i-1)} A_{\sigma(i-1)} &= Q_i \\ A_{\sigma(i-1)} &= T^{-1}_{\sigma(i-1)} Q_i = T_{i-1} Q_i = P_{\sigma(i-1)}, \end{aligned}$$

which is using a value from $\bar{P}$ in $\bar{A}$. Thus we see that combining the structure of $\Omega_{\bar{A}}$ given by short cycles with the existence of a dual forces $\bar{A}$ to use values from only $\bar{P} \cup \bar{Q}$. □

5.8. Application to the entropy calculation

Following [24], we use the representation of the geodesic flow on the compact surface $M = \Gamma\backslash\mathbb{D}$ as the special flow over the arithmetic cross-section $C_{\bar{A}}$ and Abramov's formula to compute the measure-theoretic entropy of the two maps $F_{\bar{A}}$ and $f_{\bar{A}}$.

Let (x, y, ψ), $z = x + iy \in \mathbb{D}$, $0 \le \psi < 2\pi$, be the standard coordinate system on the unit tangent bundle $S\mathbb{D}$. The hyperbolic measures on $\mathbb{D}$ and $S\mathbb{D}$, corresponding to the metric (2), are given by

$$dA = \frac{4dxdy}{(1 - |z|^2)^2}$$

and $dAd\psi$, respectively. They are preserved by Möbius transformations.

There is another coordinate system on $S\mathbb{D}$, introduced in [4] and used in [24], which proved to be more convenient than (x, y, ψ) in study of the geodesic flow, especially in the context of the cross-sections. Namely, to each $v \in S\mathbb{D}$ we assign the triple (u, w, s). where u and w are unit circle variables and s is real. The pair (u, w) designate points of intersection of the geodesic determined by v with the boundary $\partial\mathbb{D} = \{\, z : |z| = 1 \,\}$; u is the backward end and w is the forward end of this geodesic. The real parameter s is the hyperbolic length along the geodesic measured from its "midpoint" to the base point of v. As was pointed out in [25], it is a standard computation that the measure

$$d\nu = \frac{|du|\ |dw|}{|u - w|^2}$$

and the measure $dm = d\nu ds$ on $S\mathbb{D}$ are preserved by Möbius transformations, in the first case applied to unit circle variables u and w, and in the second case to the

variables (u, w, s). Therefore, $F_{\bar{A}}$ preserves the smooth probability measure

$$d\nu_{\bar{A}} = \frac{1}{K_{\bar{A}}} d\nu, \text{ where } K_{\bar{A}} = \int_{\Omega_{\bar{A}}} d\nu. \tag{32}$$

This can be also be derived from the representation of the geodesic flow $\{\varphi^t\}$ on $\Gamma\backslash S\mathbb{D}$ as the special flow over $\Omega_{\bar{A}}$ which parametrizes the arithmetic cross-section $C_{\bar{A}}$, as explained in Section 5.2, with $F_{\bar{A}}$ being the first return map to $\Omega_{\bar{A}}$ and the ceiling function $g_{\bar{A}} : \Omega_{\bar{A}} \to \mathbb{R}$ being the time of the first return to the cross-section $C_{\bar{A}}$ parametrized by $\Omega_{\bar{A}}$ (see Figure 11).

The circle map $f_{\bar{A}}$ is a factor of $F_{\bar{A}}$ (projecting on the w-coordinate), so one can obtain its smooth invariant measure $d\mu_{\bar{A}}$ by integrating $d\nu_{\bar{A}}$ over $\Omega_{\bar{A}}$ with respect to the u-coordinate.

We can immediately conclude that the systems $(F_{\bar{A}}, \nu_{\bar{A}})$ and $(f_{\bar{A}}, \mu_{\bar{A}})$ are ergodic from the fact that the geodesic flow $\{\varphi^t\}$ is ergodic with respect to dm.

The next result gives a formula for the measure-theoretic entropy of $(F_{\bar{A}}, \nu_{\bar{A}})$. Since $(F_{\bar{A}}, \nu_{\bar{A}})$ is a natural extension of $(f_{\bar{A}}, \mu_{\bar{A}})$, the measure-theoretic entropies of the two systems coincide, and we have the following result.

Proposition 5.20. $h_{\mu_{\bar{A}}}(f_{\bar{A}}) = h_{\nu_{\bar{A}}}(F_{\bar{A}}) = \frac{\pi^2(2g-2)}{K_{\bar{A}}}$.

Proof. Using the well-known fact that the entropy of the geodesic flow with respect to the normalized Liouville measure $d\widetilde{m} = \frac{dm}{m(SM)}$ on SM is equal to 1 and Abramov's formula, we obtain

$$h_{\widetilde{m}}(\{\varphi^t\}) = 1 = \frac{h_{\nu_{\bar{A}}}(F_{\bar{A}})}{\int_{\Omega_{\bar{A}}} g_{\bar{A}} d\nu_{\bar{A}}}. \tag{33}$$

On the other hand, $d\widetilde{m}$ can be represented by the Ambrose-Kakutani Theorem [5] as a smooth probability measure on the space $\Omega_{\bar{A}}^{g_{\bar{A}}}$ under the ceiling function $g_{\bar{A}}$, so we have

$$d\widetilde{m} = \frac{d\nu_{\bar{A}} ds}{\int_{\Omega_{\bar{A}}} g_{\bar{A}} d\nu_{\bar{A}}} = \frac{d\nu ds}{K_{\bar{A}} \int_{\Omega_{\bar{A}}} g_{\bar{A}} d\nu_{\bar{A}}} = \frac{dm}{m(SM)}. \tag{34}$$

Combining (33) and (34) and using the fact that the two measures dm and $dA d\psi$, which are invariant under the geodesic flow on $S\mathbb{D}$, are related by

$$4dm = dA d\psi,$$

we obtain

$$h_{\nu_{\bar{A}}}(F_{\bar{A}}) = \int_{\Omega_{\bar{A}}} g_{\bar{A}} d\nu_{\bar{A}} = \frac{m(SM)}{K_{\bar{A}}}.$$

But since by the Gauss-Bonnet formula $A(M) = \int_M dA = 2\pi(2g-2)$, we have

$$m(SM) = \int_{SM} dm = \frac{1}{4} \int_{SM} dA d\psi = \frac{1}{4}(2\pi) \int_M dA = \frac{1}{4}(2\pi)\big(2\pi(2g-2)\big) = \pi^2(2g-2),$$

and so $h_{\mu_{\bar{A}}}(f_{\bar{A}}) = h_{\nu_{\bar{A}}}(F_{\bar{A}}) = \frac{\pi^2(2g-2)}{K_{\bar{A}}}$. □

Thus, from the exact shape of the set $\Omega_{\bar{A}}$, the invariant measure and the entropy can be calculated precisely; one just needs to calculate the area $K_{\bar{A}}$ of the attractor. In [25, Proposition 7.1] this invariant was calculated for the case when $\bar{A}$ satisfies the short cycle property.

Chapter IV: Kleinian setting

6. Background

6.1. Complex continued fractions

Combining[4] terminology from Dani-Nogueira [12] and notation from Katok-Ugarcovici [23, 24], we say that a function

$$\lfloor \cdot \rceil_c : \mathbb{C} \to \mathbb{Z}[i]$$

such that w and $\lfloor w \rceil_c$ are a distance at most 1 apart is called a *choice function*. The subscript here is a placeholder for any text or symbols that identify a specific algorithm (e.g., we will discuss $\lfloor z \rceil_{\text{Hur}}$ and $\lfloor \cdot \rceil_{\text{NE}}$ shortly). Each choice function has a *fundamental set* Φ_c given by

$$\Phi_c = \overline{\{ z - \lfloor z \rceil_c \, : \, z \in \mathbb{C} \}}.$$

Below are a few examples of choice functions.

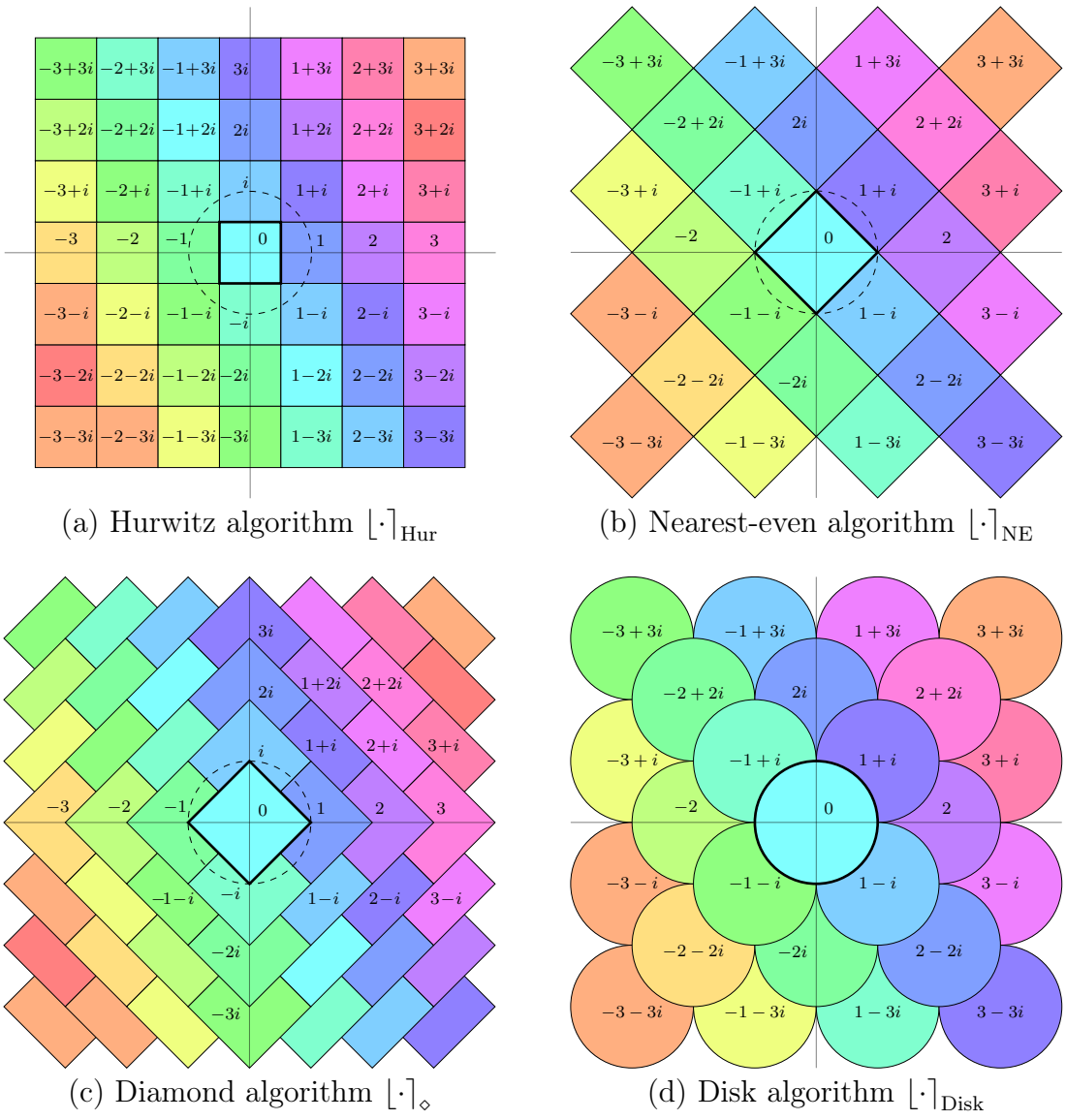
- The *nearest integer algorithm* assigns to $z \in \mathbb{C}$ the Gaussian integer closest to z (the convention to use when z is equidistant to multiple close integers does not have a great effect since all such z's are rational). This algorithm was discussed in detail by Adolf Hurwitz [16] and is also called the *Hurwitz algorithm*. See Figure 16(a).
- The *nearest-even algorithm* chooses the nearest Gaussian integer $x + yi$ for which $x + y$ is even (such a number is called an *even* Gaussian integer in [12]). In this case the fundamental set is a diamond with corners ± 1 and $\pm i$. This algorithm was studied by Julius[5] Hurwitz [17] and by Tanaka [40]. See Figure 16(b).
- In [40], Tanaka discusses a second algorithm which we call here the "disk algorithm." It is described below. See Figure 16(d).
- In Section 7.1, we describe a method to construct a choice function starting from a set Φ, and one example of such a choice function—the "diamond algorithm"—is studied in [46]. See Figure 16(c).

Tanaka [40] describes two complex plus continued fraction algorithms that both use only digits in the ideal (in the ring $\mathbb{Z}[i]$) generated by $\alpha := 1 + i$, that is, the set

$$E = (\alpha) = \{ n\alpha + m\overline{\alpha} \, : \, n, m \in \mathbb{Z} \}.$$

[4] Dani and Nogueira denote choice function by $f(x)$. Katok and Ugarcovici use the notation $\lfloor x \rceil_{a,b}$ for their "generalized integer part" function.

[5] Julius is the older of brother of the more famous Adolf Hurwitz.



(a) Hurwitz algorithm $\lfloor\cdot\rceil_{\text{Hur}}$ (b) Nearest-even algorithm $\lfloor\cdot\rceil_{\text{NE}}$

(c) Diamond algorithm $\lfloor\cdot\rceil_{\diamond}$ (d) Disk algorithm $\lfloor\cdot\rceil_{\text{Disk}}$

FIGURE 16: Regions where choice functions takes different values.

It is worth pointing out that this set can be equivalently defined as

$$E = \{\, x + yi \in \mathbb{Z}[i] \,:\, x + y \text{ is even} \,\}. \tag{35}$$

The first of Tanaka's algorithms is the nearest-even algorithm described above. The second, called here the *disk algorithm*, is described as follows. Define nine (highly overlapping) subsets of the unit disk $\mathbb{D}$:

$$\begin{aligned} &V_0 = \overline{\mathbb{D}} \qquad V_1 = V_2 \cap V_8 \qquad V_2 = \left\{\, w \in \overline{\mathbb{D}} \,:\, |w - \alpha| \geq 1 \,\right\} \\ &V_j = i\, V_{j-2} \text{ for } j = 3, \ldots, 8. \end{aligned} \tag{36}$$

The even integers E are then partitioned into nine regions:

$$E_0 = \{0\} \qquad E_1 = \{\, n\alpha + m\overline{\alpha} : n, m < 0 \,\} \qquad E_2 = \{\, n\alpha : n < 0 \,\}$$
$$E_j = i\, E_{j-2} \text{ for } j = 3, \ldots, 8.$$

Lastly, we denote by $V(a)$ whichever set V_j satisfies $a \in E_j$. The choice function $\lfloor \cdot \rceil_{\text{Disk}}$ is then defined as

$$\lfloor w \rceil_{\text{Disk}} = a \in E \quad \text{if} \quad w \in a + V(a). \tag{37}$$

In Figure 16(d), all the colored regions on the left are translates of V_1, and all the colored regions along $\operatorname{Re} z = \operatorname{Im} z < 0$ are translates of V_2.

6.2. Continued fraction transformations

This section is essentially only definitions of notations.

Given a choice function $\lfloor \cdot \rceil_c : \mathbb{C} \to \mathbb{Z}[i]$, the "Gauss-like" function $\hat{f}_c : \Phi_c \to \Phi_c$ is given by

$$\hat{f}_c(z) = \frac{-1}{z} - \left\lfloor \frac{-1}{z} \right\rceil_c . \tag{38}$$

It is conjugate to the map

$$r_c(w) = \frac{-1}{w - \lfloor w \rceil_c}$$

via $r_c = S \circ \hat{f}_c \circ S$, where $S(z) = -1/z$. The natural extension of r_c is the map $R_c : \mathbb{C} \times S\Phi_c \circlearrowleft$ given by

$$R_c(u, w) = \left(\frac{-1}{u - a}, \frac{-1}{w - a} \right), \qquad a = \lfloor w \rceil_c . \tag{39}$$

Later we will show that u will often be in $\mathbb{D} \subset \mathbb{C}$, and we will often consider R_c to be a map on $\mathbb{D} \times S\Phi_c$ rather than $\mathbb{C} \times S\Phi_c$.

We define the notation

$$T^a(z) = z + a \qquad \text{for any } a \in \mathbb{Z}[i].$$

With this notation we can express $R_c : \mathbb{D} \times S\Phi_c \circlearrowleft$ as

$$R_c(u, w) = (ST^{-a}u, ST^{-a}w), \qquad a = \lfloor w \rceil_c .$$

With the change of coordinates $\psi(u, w) = (Sw, u)$ or $\psi^{-1}(x, y) = (y, Sx)$, we get a new map $\hat{F}_c : \Phi_c \times \mathbb{D}$ given by $\hat{F}_c = \psi \circ R_c \circ \psi^{-1}$ or by the explicit formula

$$\hat{F}_c(x, y) = (T^{-a}Sx, ST^{-a}y), \qquad a = \lfloor Sx \rceil_c , \tag{40}$$

which is the natural extension of $\hat{f}_c$. Since they differ only by a coordinate change, both $\hat{F}_c$ and R_c are called natural extensions of both $\hat{f}_c$ and r_c. Also note that $R_c(\cdot, w) = (\cdot, r_c(w))$ and that $\hat{F}_c(x, \cdot) = (\hat{f}_c(x), \cdot)$.

Remark 6.1. Generally, in this thesis, the notation $(u, w) \in \mathbb{C}^2$ is used when the geodesic from u to w in $\mathcal{H}^3$ is relevant, and often (u, w) will actually be in $\mathbb{C} \times S\Phi_c$. The notation $(x, y) \in \mathbb{C}^2$ is used for points in $\Phi_c \times \mathbb{D}$ or $\Phi_c \times \mathbb{C}$, with $(x, y) = (-1/w, u)$.

7. New results

7.1. Continued fractions

Recall that given any sequence $(a_0, a_1, a_2, \ldots)$, we can define sequences $\{p_k\}$ and $\{q_k\}$ by

$$\begin{aligned} p_{-1} &= 1, & p_0 &= a_0, & p_n &= a_n p_{n-1} - p_{n-2} \quad \forall\ n \geq 1, \\ q_{-1} &= 0, & q_0 &= 1, & q_n &= a_n q_{n-1} - q_{n-2} \quad \forall\ n \geq 1, \end{aligned} \tag{41}$$

and then for any $0 \leq n < \infty$ we have

$$\frac{p_n}{q_n} = a_0 - \cfrac{1}{a_1 - \cfrac{1}{a_2 - \cfrac{1}{\ddots - \frac{1}{a_n}}}},$$

which we denote by $[a_0; a_1, a_2, \ldots, a_n]$. If the sequence $\{\frac{p_n}{q_n}\}$ converges, we write $[a_0; a_1, a_2, \ldots]$ for the limit. Given $z \in \mathbb{C}$ and a choice function $\lfloor \cdot \rceil_c$, we can construct the digit sequence $\{a_n\}$ by

$$a_0, = \lfloor z \rceil_c \qquad a_n = \left\lfloor -1/\hat{f}_c^{n-1}(z) \right\rceil_c \quad \forall\ n \geq 1, \tag{42}$$

where $\hat{f}_c(z) = \frac{-1}{z} - \left\lfloor \frac{-1}{z} \right\rceil_c$. After then defining $\{p_n\}$ and $\{q_n\}$ by (41), the sequence $\{\frac{p_n}{q_n}\}$ will either terminate with $\frac{p_n}{q_n} = z$ for some n or will converge to z. Terminating sequences only occur for $z \in \mathbb{Q}[i]$, so we will focus only on *irrational* complex numbers, meaning those in $\mathbb{C} \setminus \mathbb{Q}[i]$.

Remark 7.1. All results in this thesis focus on "minus continued fractions" (see Section 1.2). Relating the digits and convergents by

$$\begin{aligned} \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} \qquad \text{for } + \text{ c.f. if } a_0 = 0 \\ \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} &= \begin{pmatrix} 0 & -1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & -1 \\ 1 & a_n \end{pmatrix} \qquad \text{for } - \text{ c.f. if } a_0 = 0, \end{aligned}$$

one sees that for minus c.f. the matrices all have determinant $+1$, while for plus c.f. there are matrices with determinant -1. This is one argument in favor of minus continued fractions over plus.

Lemma 7.2. *For $n \geq 1$, $\dfrac{q_{n-1}}{q_n} = -[0; a_n, a_{n-1}, \ldots]$.*

Proof. First, note that for finite or infinite formal continued fractions,

$$-[0; a_n, a_{n-1}, \ldots] = -\left(0 - \cfrac{1}{a_n - \cfrac{1}{a_{n-1} - \cfrac{1}{\ddots}}}\right) = \cfrac{1}{a_n - \cfrac{1}{a_{n-1} - \cfrac{1}{\ddots}}} = \frac{1}{a_n - [0; a_{n-1}, \ldots]}.$$

From (41), we have $q_0 = 1$ and $q_1 = a_1$. Thus for $n = 1$, $\frac{q_0}{q_1} = -(0 - \frac{1}{a_1})$. Now assume $q_{k-1}/q_k = -[0; a_k, a_{k-1}, \ldots, a_1]$ for some $k \geq 1$. Then

$$\begin{aligned}\frac{q_k}{q_{k+1}} &= \frac{q_k}{a_{k+1}q_k - q_{k-1}} = \frac{1}{a_{k+1} - \frac{q_{k-1}}{q_k}} \\ &= \frac{1}{a_{k+1} - [0; a_k, a_{k-1}, \ldots, a_1]} = -[0; a_{k+1}, a_k, \ldots, a_1].\end{aligned}$$

□

The plus continued fraction version of Lemma 7.2 is mentioned at least as early as 1887 [16, Eqns. 18 and 19].

Lemma 7.3. *Let $u, v \in \mathbb{Z}^d$. For any $C > 0$, there exists $R > 0$ such that $|u| > |v| > R$ implies $|u| - |v| > \frac{C}{|u|}$.*

Proof. For $d = 1$ this is trivial because $|u| \neq |v|$ implies $|u| - |v| \geq 1$ for $u, v \in \mathbb{Z}$.

For $d \geq 2$, however, $|u| - |v| > 0$ can be arbitrarily small. Fortunately, small differences in norms require the norms themselves to be large.

Let $u = (u_1, \ldots, u_d)$ and assume by symmetry that $u_1 \geq u_2 \geq \cdots \geq u_d \geq 0$. Then the largest possible norm of $v \in \mathbb{Z}^d$ with $|v| < |u|$ is

$$\begin{aligned}\sqrt{(u_1 - 1)^2 + u_2^2 + \cdots u_d^2} &= \sqrt{(u_1 - 1)^2 - u_1^2 + u_1^2 + \cdots u_d^2} \\ &= \sqrt{u_1^2 - 2u_1 + 1 - u_1^2 + |u|^2} \\ &= \sqrt{|u|^2 - 2u_1 + 1}\end{aligned}$$

Let $r = |u|$, so then

$$|u| - |v| \geq r - \sqrt{r^2 - 2u_1 + 1}.$$

Now suppose that $u_1 > C + 1$; then

$$\begin{aligned}2(u_1 - C) + \tfrac{C^2}{r^2} &> 1 \\ 2u_1 - 2C + \tfrac{C^2}{r^2} &> 1 \\ 1 + 2C - 2u_1 - \tfrac{C^2}{r^2} &< 0\end{aligned}$$

$$
\begin{aligned}
1 - 2u_1 &< \tfrac{C^2}{r^2} - 2C \\
r^2 + 1 - 2u_1 &< r^2 - 2C + \tfrac{C^2}{r^2} \\
r^2 + 1 - 2u_1 &< \left(r - \tfrac{C}{r}\right)^2 \\
\sqrt{r^2 + 1 - 2u_1} &< r - \tfrac{C}{r} \\
r - \sqrt{r^2 + 1 - 2u_1} &> \tfrac{C}{r}
\end{aligned}
$$

This shows that $u_1 > C + 1$ implies $|u| - |v| > \frac{C}{|u|}$. Since $r \leq u_1\sqrt{d}$, it is sufficient to have $|u| > (C+1)\sqrt{d}$, so the proof is completed with $R = (C+1)\sqrt{d}$. □

Proposition 7.4. *Let $u, w \in \mathbb{C}$ with $u \neq w$ and $w \notin \mathbb{Q}[i]$, and let a_n, p_n, q_n be as in (41) and (42). Then there exists $N < \infty$ such that $R_c^N(u, w) \in \mathbb{D} \times S(\Phi_c)$.*

Proof. Let $(u_n, w_n) = R_c^n(u, w)$, that is,

$$
\begin{aligned}
u_{n+1} &= ST^{-a_n} \cdots ST^{-a_1} ST^{-a_0} u, \\
w_{n+1} &= ST^{-a_n} \cdots ST^{-a_1} ST^{-a_0} w,
\end{aligned}
$$

where $a_n = \lfloor w_n \rceil_c$. Because $w \notin \mathbb{Q}[i]$, these sequences do not terminate [12]. Then

$$
u = T^{a_0} ST^{a_1} S \cdots T^{a_k} S(u_{k+1}) = \frac{p_k u_{k+1} - p_{k-1}}{q_k u_{k+1} - q_{k-1}}.
$$

and thus

$$
u_{k+1} = \frac{q_{k-1}u - p_{k-1}}{q_k u - p_k} = \frac{q_{k-1}}{q_k} + \frac{1}{q_k^2(\frac{p_k}{q_k} - u)}.
$$

Let $C = \frac{1}{|w-u|}$ and note that $\frac{1}{|p_k/q_k - u|} \to C > 0$ since $w \neq u$. Then

$$
\begin{aligned}
u_{k+1} &= \frac{q_{k-1}}{q_k} + \frac{\frac{1}{p_k/q_k - u}}{q_k^2} \\
|u_{k+1}| &\leq \frac{|q_{k-1}|}{|q_k|} + \frac{C}{|q_k|^2} = 1 - \frac{|q_k| - |q_{k-1}| - C/|q_k|}{|q_k|}
\end{aligned}
$$

and so $|u_{k+1}| < 1$ if $|q_k| - |q_{k-1}| - C/|q_k| > 0$ or equivalently

$$
|q_k| - |q_{k-1}| > \frac{C}{|q_k|}.
$$

By Lemma 7.3 there exists $R > 0$ such that this is satisfied for $|q_k| > |q_{k-1}| > R$. The sequence $|q_n|$ must be unbounded since there are infinitely many distinct convergents, so there will be k for which $|q_k| > |q_{k-1}| > R$. □

Remark 7.5. Under the change of coordinates $(x, y) = (Sw, u)$, Proposition 7.4 asserts that $\hat{F}_c^N(x, y)$ is in $\Phi_c \times \mathbb{D}$. The orbit under $\hat{F}_c$ of almost every point will enter $\Phi_c \times \mathbb{D}$, but the orbit may then leave this product. Thus $\Phi_c \times \mathbb{D}$ is not a trapping region, but the set

$$\bigcup_{n=0}^{\infty} \hat{F}_c^n(\Phi_c \times \mathbb{D})$$

will be a trapping region.

In [12], Dani and Nogueira define the fundamental set for a choice function. Here, we show that certain sets in $\mathbb{C}$ can be used to construct choice functions.

Proposition 7.6. *Let $\Phi \subset \overline{\mathbb{D}}$. Let $s : \mathbb{C} \to \mathbb{C}$ be*

$$s(z) = \begin{cases} 0 & \text{if } z = 0 \\ z - 1 & \text{if } -\pi/4 \leq \arg z < \pi/4 \\ z - i & \text{if } \pi/4 \leq \arg z < 3\pi/4 \\ z + 1 & \text{if } 3\pi/4 \leq \arg z \text{ or } \arg z < \pi/4 \\ z + i & \text{if } -3\pi/4 \leq \arg z < -\pi/4. \end{cases}$$

If for any $z \in \mathbb{C} \setminus \Phi$ there exists $n_z \in \mathbb{N}$ such that $s^{n_z}(z) \in \Phi$, then the function

$$f(z) = \begin{cases} 0 & \text{if } z \in \Phi \\ z - s^{n_z}(z) & \text{if } z \notin \Phi \end{cases}$$

will be a valid choice function.

Proof. We need to show that $f(z) \in \mathbb{Z}[i]$ and $|z - f(z)| \leq 1$ for all $z \in \mathbb{C}$.

Since $\Phi \subset \overline{\mathbb{D}}$, using $f(z) = 0 \in \mathbb{Z}[i]$ for $z \in \Phi$ will satisfy $|z - f(z)| \leq 1$.

For $z \notin \Phi$, we prove by induction. If $n_z = 1$, then $z = s(z) + i^k$ for some $k \in \{0, 1, 2, 3\}$, and so $f(z) = z - s(z) = i^k$ is in $\mathbb{Z}[i]$ and $|z - f(z)| = |s(z)| \leq 1$ because $s(z) \in \Phi \subset \overline{\mathbb{D}}$. For $n_z > 1$, we use that $n_{s(z)} = n_z - 1$ repeatedly until we reach $n_{s^{n_z-1}(z)} = 1$ and by induction recover that $f(z) \in \mathbb{Z}[i]$ and $|z - f(z)| \leq 1$ for any n_z. □

Note that if (Gaussian) integer translates of Φ tile the (complex) plane without overlap, then Φ satisfies the condition of Proposition 7.6, and the choice function obtained by this construction will be equivalent to

$$\lfloor z \rceil_\Phi = a \quad \Longleftrightarrow \quad z \in a + \Phi, \quad a \in \mathbb{Z}[i].$$

This is the case for the Hurwitz algorithm, with $\Phi = [-\frac{1}{2}, \frac{1}{2}] + [-\frac{1}{2}, \frac{1}{2}]i$. If integer translates of Φ tile the plane with overlap, then Φ will still satisfy the condition of Proposition 7.6, but it may not have a form like this.

Definition. Let $\lfloor\cdot\rceil_\diamond$ be the choice function constructed as in Proposition 7.6 starting with the diamond-shaped region

$$\{\, z \,:\, |\mathrm{Re}\, z| + |\mathrm{Im}\, z| \le 1 \,\}$$

as Φ. This is referred to here (and in [46]) as the *diamond algorithm*.

Remark 7.7. The sets $\Phi_\diamond := \{\, z \,:\, \lfloor z\rceil_\diamond = 0 \,\}$ and $\Phi_{\mathrm{NE}} := \{\, z \,:\, \lfloor z\rceil_{\mathrm{NE}} = 0 \,\}$ are both equal to the diamond with vertices ± 1 and $\pm i$, but the two choice functions are not the same. See Figures 16(b) and 16(c) for a clear comparison.

7.2. The partition property

Before discusing the dynamics of $\hat{F}_\diamond$, $\hat{F}_{\mathrm{NE}}$, etc., we first introduce a useful set of properties satisfied by these algorithms.

Let $\mathbb{A}$ be the set of possible digits for a particular algorithm. For example, $\mathbb{A} = \mathbb{Z}[i]$ for the Hurwitz algorithm, and $\mathbb{A} = E$ for the nearest-even and disk algorithms (in fact, this construction can be carried out in the $\mathrm{PSL}(2,\mathbb{Z})$ setting with $\mathbb{A} = \mathbb{Z}$; see Chapter II).

For any $a \in \mathbb{A}$, denote

$$X_c(a) = \{\, x \in \Phi_c \,:\, \lfloor -1/x\rceil_c = a \,\},$$

and more generally for any sequence of digits let

$$X_c(a_1, a_2, \ldots, a_n) = \left\{ x \in \Phi_c \,:\, \left\lfloor \frac{-1}{\hat{f}^{k-1}(x)} \right\rceil_c = a_k \text{ for } 1 \le k \le n \right\}. \tag{43}$$

Definition. A choice function $\lfloor\cdot\rceil_c$ is said to have the *partition property* if there exists a finite partition $\mathcal{P}_c$ of Φ_c such for every $a \in \mathbb{A}$ and every $\mathscr{X} \in \mathcal{P}_c$, the set

$$\hat{f}_c\big(\mathscr{X} \cap X_c(a)\big) = T^{-a} S\big(\mathscr{X} \cap X_c(a)\big)$$

is a (possibly empty) union of elements from $\mathcal{P}_c$.

When such a finite partition $\mathcal{P}_c$ exists, we denote its cardinality by N and its elements by

$$\mathcal{P}_c = \{\mathscr{X}_c^{(1)}, \ldots, \mathscr{X}_c^{(N)}\} \tag{44}$$

and define the related countable partition

$$\mathcal{C}_c = \left\{\, \mathscr{X}_c^{(j)} \cap X_c(a) \,:\, 1 \le j \le N, a \in \mathbb{A} \,\right\} \tag{45}$$

(that is, $\mathcal{C}_c$ is the join[6] $\mathcal{P}_c \vee \{ X_c(a) : a \in \mathbb{A} \}$). The partition property then states that for any $A \in \mathcal{C}_c$, $\hat{f}_c(A)$ is a union of elements from $\mathcal{P}_c$.

Note that for any algorithm the partition $\mathcal{P}_c$ satisfying the required property is not unique (indeed, any finite refinement of a valid partition will also work). The specific examples $\mathcal{P}_{\mathrm{NE}}$, $\mathcal{P}_\diamond$, etc., used here will also be maximal in the sense that no other valid partition has them as a refinement, but this is not strictly required.

In practice, it is often easier to work with $S(\Phi_c)$ that with Φ_c itself. Denoting

$$W_c(a) = S\big(X_c(a)\big) = \{ w \in S(\Phi_c) : \lfloor w \rceil_c = a \} \tag{46}$$

and $\mathscr{W}_c^{(j)} = S\big(\mathscr{X}_c^{(j)}\big)$, we require that for every $a \in \mathbb{A}$ and every $1 \le j \le N$, either $W_c(a) \subset \mathscr{W}_c^{(j)}$, $W_c(a) \cap \mathscr{W}_c^{(j)} = \emptyset$, or $T^{-a}\big(W_c(a) \cap \mathscr{W}_c^{(j)}\big)$ is a union of some elements $\mathscr{X}_c^{(k)}$.

Figures 17 through 20 show the partitions $\mathcal{C}_c$ and $\mathcal{P}_c$ for several algorithms on the left of each figure. The thin outlines show the elements of $\mathcal{C}_c$, and the color of each set correspond to which element of $\mathcal{P}_c$ contains it (thus the elements of $\mathcal{P}_c$ appear as large colored regions). The elements of $\mathcal{P}_c$ are denoted $\mathscr{X}_c^{(j)}$, where the values of j are just indices (their order and numeric value are irrelevant). The right of each figure shows the image of the left under $S(z) = -1/z$.

Proposition 7.8. *The nearest-even algorithm has the partition property.*

It will be helpful to define the following sets:

$$\begin{aligned} U_1 &= \{ z \in \mathbb{C} : |\mathrm{Re}\, z| + |\mathrm{Im}\, z| \le 1 \} \\ U_2 &= \left\{ z \in U_1 : \left|z - \tfrac{1+i}{2}\right| \ge \tfrac{1}{\sqrt{2}} \right\} \\ U_j &= i\, U_{j-1}, \quad j = 3, 4, 5. \end{aligned} \tag{47}$$

Proof of Proposition 7.8. The sets $W_{\mathrm{NE}}(a)$, where $a \in E$ is even, are easily described:

$$W_{\mathrm{NE}}(a) = \begin{cases} \emptyset & \text{if } a = 0 \\ a + U_5 & \text{if } a = -1 + i \\ a + U_4 & \text{if } a = 1 + i \\ a + U_3 & \text{if } a = 1 - i \\ a + U_2 & \text{if } a = -1 - i \\ a + U_1 & \text{otherwise.} \end{cases}$$

[6] All "partitions" in this thesis contain closed sets whose boundaries may intersect. Technically we use $A \vee B = \{ \overline{\mathrm{int}(a) \cap \mathrm{int}(b)} : a \in A, b \in B \}$ to avoid elements that are only pieces of boundaries.

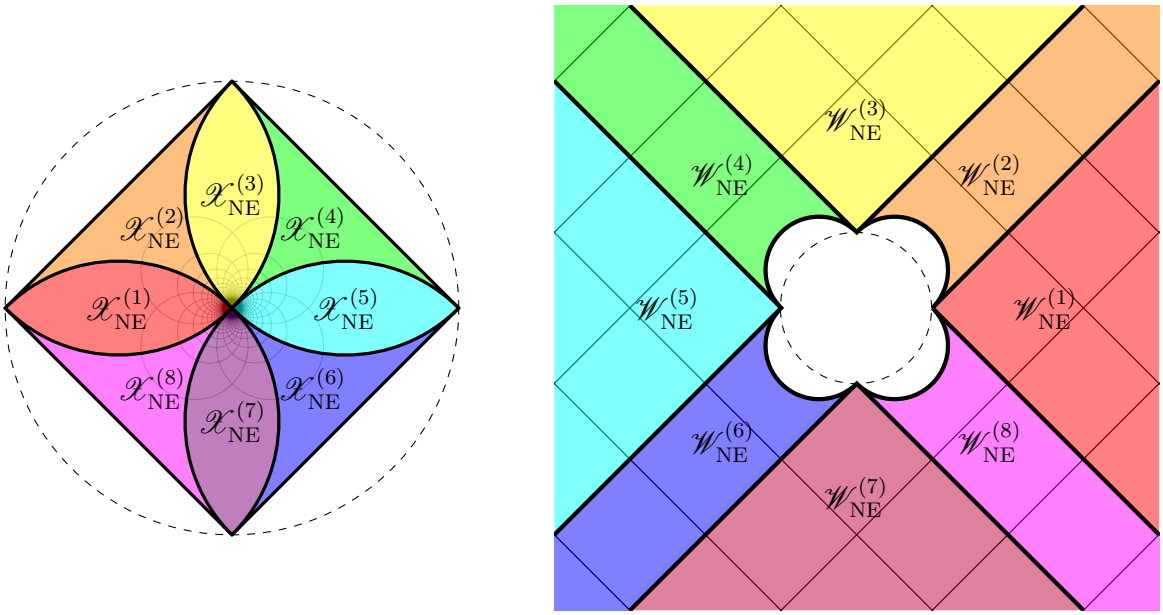


FIGURE 17: Left: $\mathcal{P}_{\mathrm{NE}}$ in colors with thin outlines of $\mathcal{C}_{\mathrm{NE}}$. Right: image under S.

Define the following 8 regions of the diamond:

$$\begin{aligned}
\mathscr{X}_{\mathrm{NE}}^{(1)} &= \left\{ u \in \Phi_{\mathrm{NE}} \,:\, \left|z - \tfrac{-1+i}{2}\right| \le \tfrac{1}{\sqrt{2}}, \left|z - \tfrac{-1-i}{2}\right| \le \tfrac{1}{\sqrt{2}} \right\} \\
\mathscr{X}_{\mathrm{NE}}^{(2)} &= \left\{ u \in \Phi_{\mathrm{NE}} \,:\, \left|z - \tfrac{-1+i}{2}\right| \le \tfrac{1}{\sqrt{2}}, \left|z - \tfrac{-1-i}{2}\right| \ge \tfrac{1}{\sqrt{2}}, \left|z - \tfrac{1+i}{2}\right| \ge \tfrac{1}{\sqrt{2}} \right\} \\
\mathscr{X}_{\mathrm{NE}}^{(j)} &= -i\, \mathscr{X}_{\mathrm{NE}}^{(j-2)}, \quad j = 3, \ldots, 8.
\end{aligned} \tag{48}$$

The nearest-even algorithm is especially elegant because each $X_{\mathrm{NE}}(a)$, $a \neq 0$ even, is already contained within some $\mathscr{X}_{\mathrm{NE}}^{(j)}$. Thus we have

$$\mathcal{C}_{\mathrm{NE}} = \{\, X_{\mathrm{NE}}(a) \,:\, a \in E \setminus \{0\} \,\},$$

where E is the set of even Gaussian integers.

For $|a| \ge 2$, we have $W_c(a) = a + X_{\mathrm{NE}}$ and therefore $T^{-a}(W_c(a)) = \bigcup_{j=1}^{8} \mathscr{X}_{\mathrm{NE}}^{(j)}$.

For $a = 1+i$, we have $T^{-\alpha} W_c(\alpha) = U_4 = \bigcup_{j=2}^{6} \mathscr{X}_{\mathrm{NE}}^{(j)}$. For $a = 1-i, -1+i, -1-i$, we get a similar result by symmetry. □

Proposition 7.9. *The disk algorithm has the partition property.*

Proof. Let

$$\begin{aligned}
\mathscr{X}_{\mathrm{Disk}}^{(2)} &= \left\{ u \in \overline{\mathbb{D}} \,:\, |u - (-1+i)| \le 1 \right\} \\
\mathscr{X}_{\mathrm{Disk}}^{(j)} &= \left\{ -i\, u \,:\, u \in \mathscr{X}_{\mathrm{Disk}}^{(j-1)} \right\} \quad \text{for } j = 3, 4, 5 \\
\mathscr{X}_{\mathrm{Disk}}^{(1)} &= \overline{\mathbb{D}} \setminus \left(\mathscr{X}_{\mathrm{Disk}}^{(2)} \cup \mathscr{X}_{\mathrm{Disk}}^{(3)} \cup \mathscr{X}_{\mathrm{Disk}}^{(4)} \cup \mathscr{X}_{\mathrm{Disk}}^{(5)} \right).
\end{aligned} \tag{49}$$

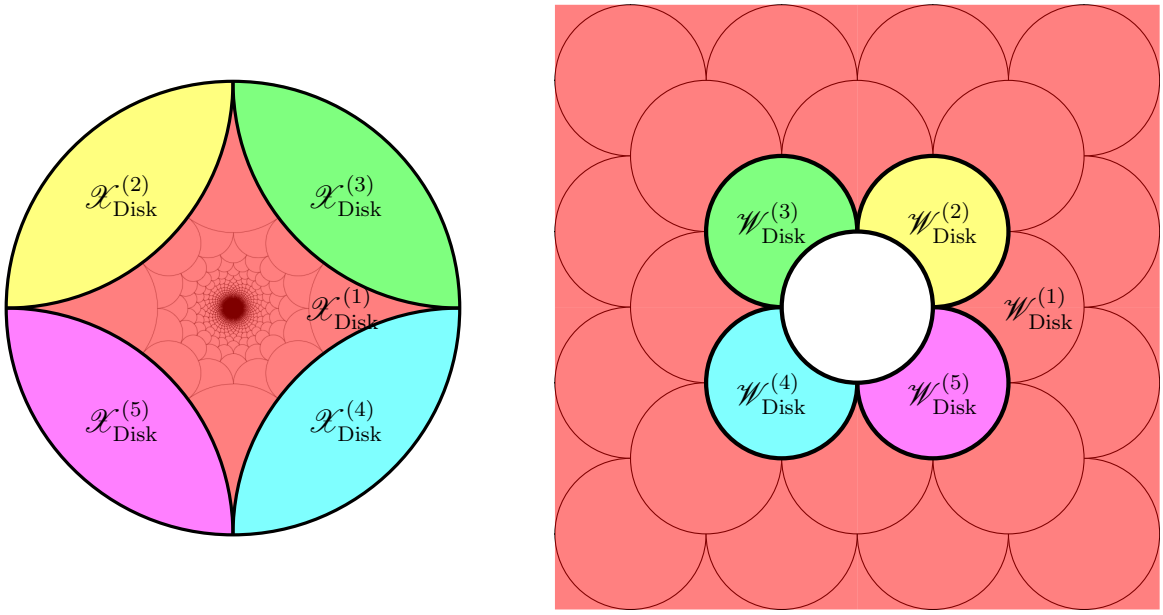


FIGURE 18: Left: $\mathcal{P}_{\mathrm{Disk}}$ in colors with thin outlines of $\mathcal{C}_{\mathrm{Disk}}$. Right: image under S.

Then

$$\begin{aligned}
W_{\mathrm{Disk}}(1+i) &= \mathscr{W}^{(2)}_{\mathrm{Disk}} = (1+i) + \left(\mathscr{X}^{(1)}_{\mathrm{Disk}} \cup \mathscr{X}^{(2)}_{\mathrm{Disk}} \cup \mathscr{X}^{(3)}_{\mathrm{Disk}} \cup \mathscr{X}^{(4)}_{\mathrm{Disk}}\right) \\
W_{\mathrm{Disk}}(-1+i) &= \mathscr{W}^{(3)}_{\mathrm{Disk}} = (-1+i) + \left(\mathscr{X}^{(1)}_{\mathrm{Disk}} \cup \mathscr{X}^{(2)}_{\mathrm{Disk}} \cup \mathscr{X}^{(3)}_{\mathrm{Disk}} \cup \mathscr{X}^{(5)}_{\mathrm{Disk}}\right) \\
W_{\mathrm{Disk}}(-1-i) &= \mathscr{W}^{(4)}_{\mathrm{Disk}} = (-1-i) + \left(\mathscr{X}^{(1)}_{\mathrm{Disk}} \cup \mathscr{X}^{(2)}_{\mathrm{Disk}} \cup \mathscr{X}^{(4)}_{\mathrm{Disk}} \cup \mathscr{X}^{(5)}_{\mathrm{Disk}}\right) \\
W_{\mathrm{Disk}}(1-i) &= \mathscr{W}^{(5)}_{\mathrm{Disk}} = (1-i) + \left(\mathscr{X}^{(1)}_{\mathrm{Disk}} \cup \mathscr{X}^{(3)}_{\mathrm{Disk}} \cup \mathscr{X}^{(4)}_{\mathrm{Disk}} \cup \mathscr{X}^{(5)}_{\mathrm{Disk}}\right).
\end{aligned}$$

For $|a| > 1$, each set $W_{\mathrm{Disk}}(a)$ is a subset of $\mathscr{W}^{(1)}_{\mathrm{Disk}}$ and will be $a + V_k$ for some V_k from (36). Each V_k is a union of some elements $\mathscr{X}^{(j)}_{\mathrm{Disk}}$. □

Proposition 7.10. *The diamond algorithm has the partition property.*

Proof. Define the following 12 regions of the diamond:

$$\begin{aligned}
\mathscr{X}^{(1)}_{\diamond} &= \left\{ u \in \Phi_{\diamond} \,:\, \left|z - \tfrac{-1+i}{2}\right| \le \tfrac{1}{\sqrt{2}}, \left|z - \tfrac{-1-i}{2}\right| \le \tfrac{1}{\sqrt{2}} \right\} \\
\mathscr{X}^{(2)}_{\diamond} &= \left\{ u \in \Phi_{\diamond} \,:\, \left|z - \tfrac{-1+i}{2}\right| \le \tfrac{1}{\sqrt{2}}, \left|z - \tfrac{-1-i}{2}\right| \ge \tfrac{1}{\sqrt{2}}, \operatorname{Im} u \le -\operatorname{Re} u \right\} \\
\mathscr{X}^{(3)}_{\diamond} &= \left\{ u \in \Phi_{\diamond} \,:\, \left|z - \tfrac{-1+i}{2}\right| \le \tfrac{1}{\sqrt{2}}, \left|z - \tfrac{1+i}{2}\right| \ge \tfrac{1}{\sqrt{2}}, \operatorname{Im} u \ge -\operatorname{Re} u \right\} \\
\mathscr{X}^{(j)}_{\diamond} &= \left\{ -i\,u \,:\, u \in \mathscr{X}^{(j-3)}_{\diamond} \right\}, \quad j = 4, \ldots, 12.
\end{aligned} \tag{50}$$

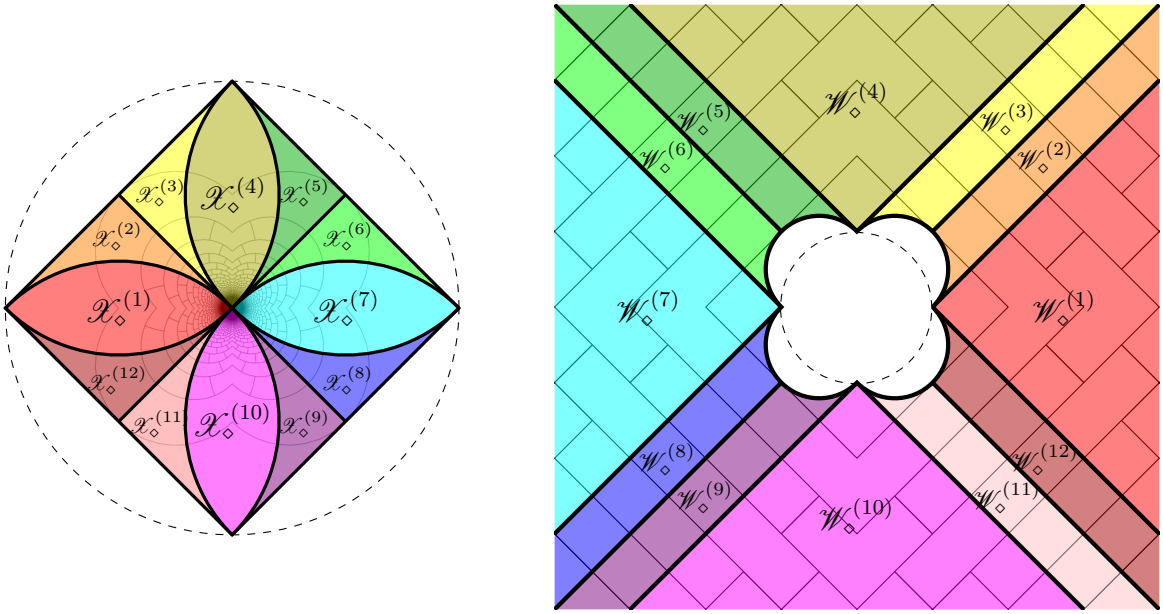
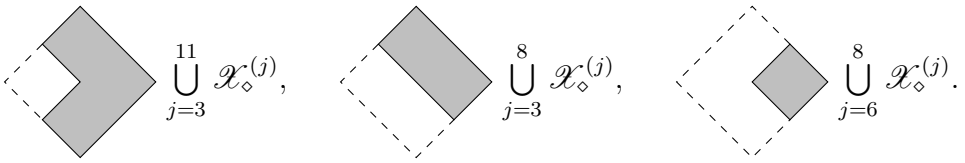


FIGURE 19: Left: $\mathcal{P}_\diamond$ in colors with thin outlines of $\mathcal{C}_\diamond$. Right: image under S.

For $a > 1$, each set $T^{-a}W_\diamond(a)$ is one of the following shapes or a rotated version:



When a block $W_\diamond(a)$ that is rectangular (the middle shape above) intersects multiple $\mathscr{W}_\diamond^{(j)}$, the intersections are quarter-diamonds (like the third shape above).

For $a = 1$, there are three non-empty intersections:

$$W_\diamond(1) \cap \mathscr{W}_\diamond^{(1)} = 1 + \left(\mathscr{X}_\diamond^{(6)} \cup \mathscr{X}_\diamond^{(7)} \cup \mathscr{X}_\diamond^{(8)}\right)$$
$$W_\diamond(1) \cap \mathscr{W}_\diamond^{(2)} = 1 + \mathscr{X}_\diamond^{(5)}$$
$$W_\diamond(1) \cap \mathscr{W}_\diamond^{(12)} = 1 + \mathscr{X}_\diamond^{(9)}$$

The cases $a = i, -1, -i$ are similar. □

Proposition 7.11. *The Hurwitz algorithm has the partition property.*

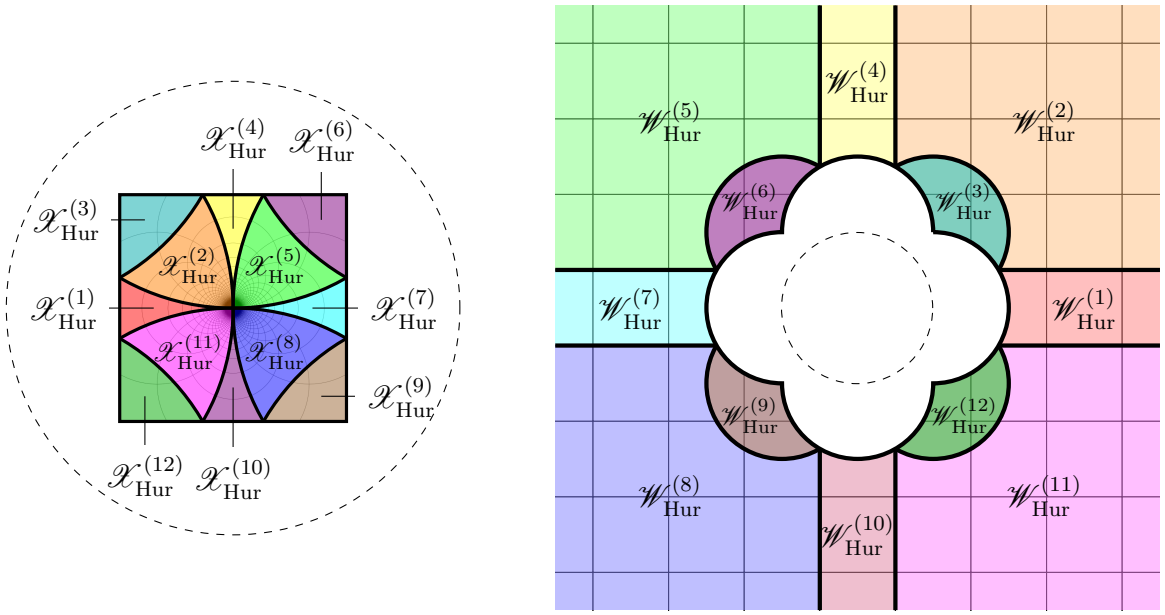


FIGURE 20: Left: $\mathcal{P}_{\mathrm{Hur}}$ in colors with thin outlines of $\mathcal{C}_{\mathrm{Hur}}$. Right: image under S.

Proof. Define the following 12 regions of the unit square centered at the origin:

$$\begin{aligned}
\mathscr{X}_{\mathrm{Hur}}^{(1)} &= \{\, u \in \Phi_{\mathrm{Hur}} \,:\, |z-i| \geq 1, |z+i| \geq 1 \,\} \\
\mathscr{X}_{\mathrm{Hur}}^{(2)} &= \{\, u \in \Phi_{\mathrm{Hur}} \,:\, |z-i| \leq 1, |z+1| \leq 1, |z-(-1+i)| \geq 1 \,\} \\
\mathscr{X}_{\mathrm{Hur}}^{(3)} &= \{\, u \in \Phi_{\mathrm{Hur}} \,:\, |z-(-1+i)| \leq 1 \,\} \\
\mathscr{X}_{\mathrm{Hur}}^{(j)} &= \left\{\, -i\,u \,:\, u \in \mathscr{X}_{\mathrm{Hur}}^{(j-3)} \,\right\}, \quad j = 4, \ldots, 12.
\end{aligned} \tag{51}$$

If $|\mathrm{Re}\, a| > 2$ or $|\mathrm{Im}\, a| > 2$, then we have $W_{\mathrm{Hur}}(a) = a + X_{\mathrm{Hur}}$, and this square will be entirely contained in exactly one set $\mathscr{W}_{\mathrm{Hur}}^{(j)}$.

We discuss $a = 2$, $2{+}i$, $2{+}2i$, $1{+}i$ here, and then the remaining cases are handled by symmetry. First, note that $W_{\mathrm{Hur}}(2) \subset \mathscr{W}_{\mathrm{Hur}}^{(1)}$ and $W_{\mathrm{Hur}}(1+i) \subset \mathscr{W}_{\mathrm{Hur}}^{(3)}$ but that $W_{\mathrm{Hur}}(2+i)$ and $W_{\mathrm{Hur}}(2+2i)$ are contained in any single $\mathscr{W}_{\mathrm{Hur}}^{(j)}$. For those we must look at intersections with $\mathscr{W}_{\mathrm{Hur}}^{(2)}$ and $\mathscr{W}_{\mathrm{Hur}}^{(3)}$ separately.

$$W_\diamond(2) = 2 + \bigcup_{j=4}^{10} \mathscr{X}_{\mathrm{Hur}}^{(j)}$$

$$W_\diamond(2+i) \cap \mathscr{W}_{\mathrm{Hur}}^{(2)} = (2+i) + \bigcup_{j=4}^{10} \mathscr{X}_{\mathrm{Hur}}^{(j)}$$

$$W_\diamond(2+i) \cap \mathscr{W}_{\text{Hur}}^{(3)} = (2+i) + \bigcup_{j=1}^{3} \mathscr{X}_{\text{Hur}}^{(j)} \cup \mathscr{X}_{\text{Hur}}^{(11)}$$

$$W_\diamond(2+2i) \cap \mathscr{W}_{\text{Hur}}^{(2)} = (2+2i) + \bigcup_{j=1}^{11} \mathscr{X}_{\text{Hur}}^{(j)}$$

$$W_\diamond(2+2i) \cap \mathscr{W}_{\text{Hur}}^{(3)} = (2+2i) + \mathscr{X}_{\text{Hur}}^{(12)}$$

$$W_\diamond(1+i) = (1+i) + \bigcup_{j=4}^{7} \mathscr{X}_{\text{Hur}}^{(j)}. \qquad \square$$

7.3. Finite product structure

Definition. A set $\Omega \in \mathbb{C} \times \mathbb{C}$ is said to have *finite product structure* if there exist $N \in \mathbb{N}$ and sets $X_1, \ldots, X_N, Y_1, \ldots, Y_N \subset \mathbb{C}$ such that $\Omega = \bigcup_{k=1}^{N} X_k \times Y_k$.

In the following discussion, we drop all subscript "c" for readability. The results will apply to any algorithm satisfying the partition property.

Let $\mathcal{P}$ be the finite partition from (44) and $\mathcal{C}$ be the countable partition from (45). Without subscripts, we write

$$\mathcal{P} = \{\mathscr{X}^{(1)}, \mathscr{X}^{(2)}, \ldots, \mathscr{X}^{(N)}\}.$$

Each element of the refinement $\mathcal{C}$ has the form $X(a) \cap \mathscr{X}^{(j)}$ for some $a \in \mathbb{Z}[i]$ and some $1 \le j \le N$. For any $A \in \mathcal{C}$, denote by $j(A)$ the unique index $k \in \{1, \ldots, N\}$ for which $A \subset \mathscr{X}^{(k)}$, and define

$$J(A) = \left\{ 1 \le j \le N \,:\, \mathscr{X}^{(j)} \subset \hat{f}(A) \right\} \tag{52}$$

$$\mathcal{C}(j) = \left\{ A \in \mathcal{C} \,:\, \mathscr{X}^{(j)} \subset \hat{f}(A) \right\} = \{\, A \in \mathcal{C} \,:\, j \in J(A) \,\}. \tag{53}$$

Note that since (52) and (53) both use the condition $\mathscr{X}^{(j)} \subset \hat{f}(A)$, we have that

$$\{\,(A,j) \,:\, A \in \mathcal{C}, j \in J(A)\,\} \qquad \text{and} \qquad \{\,(A,j) \,:\, 1 \le j \le N, A \in \mathcal{C}(j)\,\}$$

are exactly equal. This allows us to equate unions of the following two forms:

$$\bigcup_{A \in \mathcal{C}} \bigcup_{k \in J(A)} \qquad \text{and} \qquad \bigcup_{k=1}^{N} \bigcup_{A \in \mathcal{C}(k)}.$$

This is used, for example, in the proof of Theorem 7.12 below.

Theorem 7.12. *Let $\mathscr{Y}^{(1)}, \ldots, \mathscr{Y}^{(N)}$ be arbitrary complex sets. The map*

$$\hat{F}(x,y) = (T^{-a}Sx, ST^{-a}y), \quad a = \lfloor Sx \rceil,$$

is bijective a.e. on the set $\bigcup_{k=1}^{N} \mathscr{X}^{(k)} \times \mathscr{Y}^{(k)}$ if and only if the following system of equations holds:

$$\mathscr{Y}^{(k)} = \bigcup_{A \in \mathcal{C}(k)} ST^{-a_1(A)} \mathscr{Y}^{(j(A))}, \qquad 1 \leq k \leq N, \tag{54}$$

where $a_1(A)$ is the constant value of $\lfloor -1/x \rceil$ for all $x \in A$.

We will prove Theorem 7.12 by first defining the set

$$\Omega := \bigcup_{A \in \mathcal{C}} (A \times \mathscr{Y}^{(j(A))}) \tag{55}$$

(this depends on the $\mathscr{Y}^{(k)}$, which are still arbitrary) and then showing its relation to the conditions of the theorem. At first, it may not be clear that the set Ω in (55) has finite product structure, but in fact it does:

Lemma 7.13. $\Omega = \bigcup_{k=1}^{N} (\mathscr{X}^{(k)} \times \mathscr{Y}^{(k)})$.

Proof. Instead of just a union over all $A \in \mathcal{C}$, first group the products $A \times \mathscr{Y}^{(j(A))}$ by the value of $j(A)$.

$$\Omega = \bigcup_{k=1}^{N} \bigcup_{\substack{A \in \mathcal{C} \\ j(A)=k}} \left(A \times \mathscr{Y}^{(k)}\right) = \bigcup_{k=1}^{N} \left(\left(\bigcup_{\substack{A \in \mathcal{C} \\ j(A)=k}} A \right) \times \mathscr{Y}^{(k)} \right).$$

Since $j(A) = k$ is equivalent to $A \subset \mathscr{X}^{(k)}$, the union $\bigcup A$ above is exactly $\mathscr{X}^{(k)}$. $\square$

Proof of Theorem 7.12. First, assume (54) holds, and use the original form of Ω from (55). The image of Ω under $\hat{F}$ is

$$\begin{aligned}
\hat{F}(\Omega) &= \bigcup_{A \in \mathcal{C}} \left(T^{-a_1(A)} SA \times ST^{-a_1(A)} \mathscr{Y}^{(j(A))}\right) \\
&= \bigcup_{A \in \mathcal{C}} \left(\left(\bigcup_{j \in J(A)} \mathscr{X}^{(j)} \right) \times ST^{-a_1(A)} \mathscr{Y}^{(j(A))} \right) \\
&= \bigcup_{A \in \mathcal{C}} \bigcup_{k \in J(A)} \left(\mathscr{X}^{(k)} \times ST^{-a_1(A)} \mathscr{Y}^{(j(A))}\right) \\
&= \bigcup_{k=1}^{N} \bigcup_{A \in \mathcal{C}(k)} \left(\mathscr{X}^{(k)} \times ST^{-a_1(A)} \mathscr{Y}^{(j(A))}\right)
\end{aligned}$$

$$= \bigcup_{k=1}^{N} \left(\mathscr{X}^{(k)} \times \left(\bigcup_{A \in \mathcal{C}(k)} ST^{-a_1(A)} \mathscr{Y}^{(j(A))} \right) \right)$$

The union in the second coordinate is exactly the union in (54), so we have $\hat{F}(\Omega) = \bigcup_{k=1}^{N} (\mathscr{X}^{(k)} \times \mathscr{Y}^{(k)})$, and by Lemma 7.13 this is exactly Ω. A function is always surjective onto its image. Because all of the unions here are disjoint except on boundaries, and because each transformation ST^{-a} is bijective, $\hat{F}$ is injective except on a set of zero Lebesgue measure. Thus $\hat{F}$ is bijective almost everywhere on Ω.

Now assume that $\hat{f}$ is bijective a.e. on Ω, so $\hat{f}(\Omega)$ must equal Ω (both are closed). By the same steps as above, we have

$$\hat{F}(\Omega) = \bigcup_{k=1}^{N} \left(\mathscr{X}^{(k)} \times \left(\bigcup_{A \in \mathcal{C}(k)} ST^{-a_1(A)} \mathscr{Y}^{(j(A))} \right) \right)$$

and in order for this to equal $\bigcup_{k=1}^{N} \mathscr{X}^{(k)} \times \mathscr{Y}^{(k)}$ it must be that

$$\bigcup_{A \in \mathcal{C}(k)} ST^{-a_1(A)} \mathscr{Y}^{(j(A))} = \mathscr{Y}^{(k)}$$

for each $1 \leq k \leq N$. This is exactly the system (54). □

The question remains what kind of sets $\mathscr{Y}^{(k)}$ could satisfy (54). We will describe $\mathscr{Y}^{(k)}$ for some specific algorithms in Section 7.4.

Theorem 7.14. *Let $\Omega = \bigcup_{k=1}^{N} \mathscr{X}^{(k)} \times \mathscr{Y}^{(k)}$ be a bijectivity domain for $\hat{F}$. If each $\mathscr{Y}^{(k)}$ contains an open neighborhood of the origin and $|a_n|$ is unbounded for all irrational $x \in \Phi$, then for every $(x, y) \in \mathbb{C} \times \mathbb{C}$ with $x \notin \mathbb{Q}[i]$ there exists an $N < \infty$ such that $\hat{F}^N(x, y) \in \Omega$.*

Proof. Proposition 7.4 shows that for every $(x, y) \in \mathbb{C} \times \mathbb{C}$ with $x \notin \mathbb{Q}[i]$ there exists $m < \infty$ such that $\hat{F}^m(x, y) \in \Phi \times \mathbb{D}$. Thus we can assume $(x, y) \in \Phi \times \mathbb{D}$.

Suppose $(x, y) \notin \Omega$. Let $M \in \mathbb{N}$ be such that each $\mathscr{Y}^{(k)}$ contains the ball $B(0, \frac{1}{M})$ of radius $\frac{1}{M}$ centered at the origin in the complex plane. Because $|a_k|$ is unbounded, we can assume $a_1 = \lfloor -1/x \rceil$ has absolute value at least $M+2$ (replacing if necessary x by some iterate $\hat{f}^k x$). Because $y \in \mathbb{D}$, we have

$$\left| T^{-a_1} y \right| > \left| T^{-(M+1)} y \right| > M,$$

and so $|ST^{-a_1} y| < \frac{1}{M}$. This means that

$$\hat{F}(x, y) = (T^{-a_1} Sx, ST^{-a_1} y)$$

will be inside $\mathscr{X}^{(j)} \times B(0, \frac{1}{M})$ for some $1 \leq j \leq N$, and this product is contained in $\mathscr{X}^{(j)} \times \mathscr{Y}^{(j)} \subset \Omega$. □

7.4. Domains for specific algorithms

We now return to the specific algorithms that were shown to have the partition property in Section 7.2. Recall that

- $\mathscr{X}_{\mathrm{NE}}^{(1)}, \ldots, \mathscr{X}_{\mathrm{NE}}^{(8)}$ are given in (48) and shown on the left of Figure 17 (page 54).
- $\mathscr{X}_{\mathrm{Disk}}^{(1)}, \ldots, \mathscr{X}_{\mathrm{Disk}}^{(5)}$ are given in (49) and shown on the left of Figure 18 (page 55).
- $\mathscr{X}_{\diamond}^{(1)}, \ldots, \mathscr{X}_{\diamond}^{(12)}$ are given in (50) and shown on the left of Figure 19 (page 56).
- $\mathscr{X}_{\mathrm{Hur}}^{(1)}, \ldots, \mathscr{X}_{\mathrm{Hur}}^{(12)}$ are given in (51) and shown on the left of Figure 20 (page 57).

We now describe the associated sets $\mathscr{Y}_c^{(k)}$ for three of these algorithms and show that (54) holds for these sets, that is,

$$\mathscr{Y}_c^{(k)} = \bigcup_{A \in \mathcal{C}_c(k)} ST^{-a_1(A)} \mathscr{Y}_c^{(j(A))},$$

where $a_1(z) = \lfloor -1/z \rceil_c$, $j(A)$ is the unique index j for which $A \subset \mathscr{X}_c^{(j)}$, and

$$\mathcal{C}_c(k) = \left\{ A \in \mathcal{C}_c \,:\, \mathscr{X}_c^{(k)} \subset \hat{f}_c(A) \right\}.$$

For the nearest-even and disk algorithms, we will make use of the sets V_k and U_k defined in (36) and (47), respectively. We also denote by $E' = E \setminus \{0\}$ the set of non-zero even Gaussian integers.

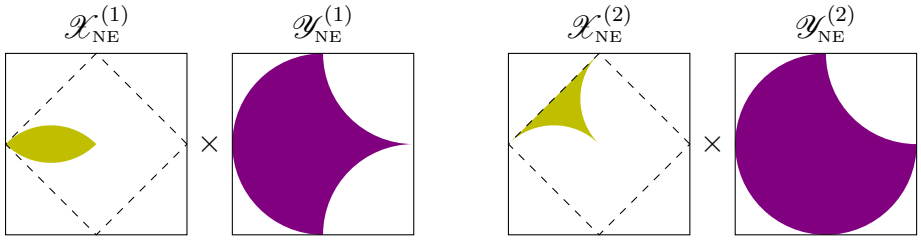


FIGURE 21: Two of the products in Ω_{NE}. The other products are rotations of these.

Proposition 7.15. *$\mathscr{Y}_{\mathrm{NE}}^{(j)} = V_j$ for $j = 1, \ldots, 8$, satisfy (54).*

Proof. Recall that $\mathcal{C}_{\mathrm{NE}} = \{ X_{\mathrm{NE}}(a) \,:\, a \in E' \}$. This means that

$$\mathcal{C}_{\mathrm{NE}}(k) = \left\{ A \in \mathcal{C}_{\mathrm{NE}} \,:\, \mathscr{X}_{\mathrm{NE}}^{(k)} \subset \hat{f}_{\mathrm{NE}}(A) \right\} = \left\{ X_{\mathrm{NE}}(a) \,:\, \mathscr{X}_{\mathrm{NE}}^{(k)} \subset T^{-a} W_{\mathrm{NE}}(a) \right\}.$$

As given in the proof of Proposition 7.8, the sets $W_{\mathrm{NE}}(a)$ for $a \in E'$ are of the form $a + U_j$. For $|a| \geq 2$, the U_j in question is $U_1 = \Phi_{\mathrm{NE}}$. For $a = \pm 1 \pm i$, the set U_j will be a subset of the diamond.

Consider $k = 1$. The set $\mathscr{X}_{\mathrm{NE}}^{(1)}$ is a subset of U_1, U_2, and U_5; its interior is disjoint with U_3 and U_4. Using that $W_{\mathrm{NE}}(-1-i) = (-1-i)+U_2$, $W_{\mathrm{NE}}(-1+i) = (-1+i)+U_5$, and that $W_{\mathrm{NE}}(a) = a + U_1$ for $|a| \geq 2$, we get that $\mathcal{C}_{\mathrm{NE}}(1)$ contains exactly those $X_{\mathrm{NE}}(a)$ for which $a = -1 - i$, $a = -1 + i$, or $|a| \geq 2$. Equivalently,

$$\mathcal{C}_{\mathrm{NE}}(1) = \{\, X_{\mathrm{NE}}(a) \,:\, a \in E',\ a \neq 1 + i,\ a \neq 1 - i \,\}\,.$$

Therefore (54) for $k = 1$ for the nearest-even algorithm becomes

$$\mathscr{Y}_{\mathrm{NE}}^{(1)} = \bigcup_{\substack{a \in E' \\ a \neq 1+i, 1-i}} ST^{-a} \mathscr{Y}_{\mathrm{NE}}^{(j)}, \qquad \text{where } a \in \mathscr{W}_{\mathrm{NE}}^{(j)}.$$

Note that the condition $a \in \mathscr{W}_{\mathrm{NE}}^{(j)}$ is equivalent to $a \in E_j$ or to $\mathscr{Y}_{\mathrm{NE}}^{(j)} = V(a)$ in the language used to define the disk algorithm at the end of Section 6.1. Since

$$S(V_1) = \{\, z \in \mathbb{C} \,:\, |z| \geq 1, |z - (-1 + i)| \geq 1, |z - (-1 - i)| \geq 1 \,\}\,,$$

(see Figure 16(d)) we indeed have that

$$S(V_1) = \bigcup_{\substack{a \in E' \\ a \neq 1+i, 1-i}} T^{-a} V(a),$$

which proves (54) for $k = 1$.

For $k = 2, \ldots, 8$, the argument is very similar, with $\mathcal{C}_{\mathrm{NE}}(k)$ containing all but 2 or 3 elements of $\mathcal{C}_{\mathrm{NE}}$ depending on whether V_j has 1 or 2 parts removed from the disk. □

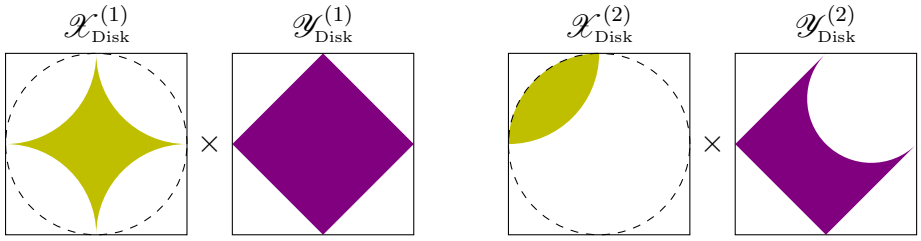


FIGURE 22: Two of the products in Ω_{Disk}. The other products are rotations of $\mathscr{X}_{\mathrm{Disk}}^{(2)} \times \mathscr{Y}_{\mathrm{Disk}}^{(2)}$.

Proposition 7.16. *$\mathscr{Y}_{\mathrm{Disk}}^{(j)} = U_j$ for $j = 1, \ldots, 5$, satisfy (54).*

Proof. Similar to the nearest-even case, $\mathcal{C}_{\text{Disk}} = \{ X_{\text{Disk}}(a) : a \in E' \}$ because each $X_{\text{Disk}}(a)$ is fully contained in some $\mathscr{W}^{(j)}_{\text{Disk}}$, and each $W_{\text{Disk}}(a) = SX_{\text{Disk}}(a)$ is of the form $a + V_j$. Unlike the nearest-even case, we do not break down E' by $|a| = \sqrt{2}$ and $|a| \geq 2$. Instead we have the decomposition $E = \bigcup_{j=1}^{8} E_j$ given in the definition of the disk algorithm at the end of Section 6.1.

Consider $k = 1$. The set $\mathscr{X}^{(1)}_{\text{Disk}}$ is a subset of all V_j, and so $\mathcal{C}_{\text{Disk}}(1) = \mathcal{C}_{\text{Disk}}$ and (54) for $k = 1$ for the disk algorithm is

$$\mathscr{Y}^{(1)}_{\text{Disk}} = \bigcup_{a \in E'} ST^{-a} \mathscr{Y}^{(j)}_{\text{Disk}}$$

where $j = 2, 3, 4, 5$ for $a = 1+i, -1+i, -1-i, 1-i$, respectively, and $j = 1$ for $|a| \geq 2$. The union is exactly the left side of Figure 17.

Now consider $k = 2$. The set $\mathscr{X}^{(2)}_{\text{Disk}}$ is a subset of all V_j except V_3, V_4, and V_5. Thus (54) being true for $k = 2$ for the disk algorithm is equivalent to

$$\mathscr{Y}^{(2)}_{\text{Disk}} = \bigcup_{\substack{a \in E' \\ a \notin E_3 \cup E_4 \cup E_5}} ST^{-a} \mathscr{Y}^{(j)}_{\text{Disk}},$$

where again $j = 2, 3, 4$ for $a = 1+i, -1+i, -1-i$, respectively ($a = 1-i$ is not part of this union), and $j = 1$ for $|a| \geq 2$. Because

$$a \in \mathscr{W}^{(j)}_{\text{NE}} \quad \Longleftrightarrow \quad -a \in E_j,$$

we want that

$$S(U_2) = \bigcup_{\substack{a \in E' \\ a \notin E_3 \cup E_4 \cup E_5}} T^a U(a),$$

where $U(1 + i) = U_4$, $U(1 - i) = U_3$, $U(-1 - i) = U_2$, and $U(a) = U_1$ for $|a| \geq 2$. This is indeed true (see Figure 17).

For $k = 3, 4, 5$, the argument is the same as for $k = 2$, but with a rotation of the entire plane. □

Proposition 7.17. *For the diamond algorithm, the sets*

$$\begin{aligned}
\mathscr{Y}^{(1)}_{\diamond} &= \{ y \in \overline{\mathbb{D}} : |y - 1| \geq 1, -\tfrac{1}{2} \leq \operatorname{Im} y \leq \tfrac{1}{2} \} \\
\mathscr{Y}^{(2)}_{\diamond} &= \{ y \in \overline{\mathbb{D}} : |y - 1| \geq 1, \operatorname{Im} y \leq \tfrac{1}{2} \} \\
\mathscr{Y}^{(3)}_{\diamond} &= \{ y \in \overline{\mathbb{D}} : |y - i| \geq 1, \operatorname{Re} y \leq \tfrac{1}{2} \} \\
\mathscr{Y}^{(j)}_{\diamond} &= -i \, \mathscr{Y}^{(j-3)}_{\diamond}, \quad j = 4, \ldots, 12,
\end{aligned}$$

make $\bigcup_{j=1}^{12} \mathscr{X}^{(j)}_{\diamond} \times \mathscr{Y}^{(j)}_{\diamond}$ *a bijectivity domain for* $\hat{F}_{\diamond}$.

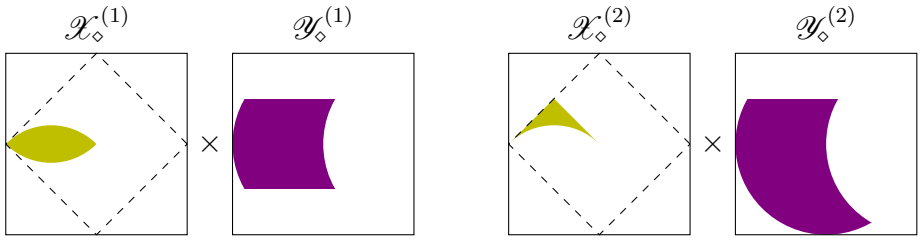


FIGURE 23: Two of the products in $\Omega_\diamond$. The other products are rotations or reflections of these.

Proof. In [46], there are sets denoted $Z_k, W_k \subset \mathbb{C}$ such that

$$D_\diamond = \bigcup_{k=1}^{40} Z_k \times W_k$$

is a bijectivity domain for

$$F_\diamond(u,w) = \begin{cases} (-1/u, -1/w) & \text{if } |\operatorname{Re} w| + |\operatorname{Im} w| \le 1 \\ (u-1, w-1) & \text{if } -\operatorname{Re} w \le \operatorname{Im} w \le \operatorname{Re} w \\ (u-i, w-i) & \text{if } -\operatorname{Im} w \le \operatorname{Re} w \le \operatorname{Im} w \\ (u+i, w+i) & \text{if } \operatorname{Re} w \le \operatorname{Im} w \le -\operatorname{Re} w \\ (u+1, w+1) & \text{if } \operatorname{Im} w \le \operatorname{Re} w \le -\operatorname{Im} w \end{cases}$$

as proven in [46, Theorem 7]. This map $F_\diamond$ is related to $\hat{F}_\diamond$ as follows:

$$R_\diamond(u,w) = F_\diamond^n(u,w), \qquad n = \left|\operatorname{Re}\lfloor w \rceil_\diamond\right| + \left|\operatorname{Im}\lfloor w \rceil_\diamond\right| + 1,$$

which ensures that S is the last map applied by $F_\diamond$. Then, as discussed in Section 6.2, $\hat{F}_\diamond = \psi \circ R_\diamond \circ \psi^{-1}$ for $\psi(u,w) = (Sw, u)$.

The set $\mathscr{X}_\diamond^{(6)}$ here is exactly W_4 in [46], and $\mathscr{X}_\diamond^{(7)}$ here is the union of W_5 and its complex conjugate. Comparing each $\mathscr{X}_\diamond^{(j)}$ here to the various definitions in [46], we find that

$$\bigcup_{j=1}^{12} S(\mathscr{Y}_\diamond^{(j)}) \times S(\mathscr{X}_\diamond^{(j)}) = \{\, (z,w) \in D_\diamond \,:\, w \in \Phi_\diamond \,\}$$

is a bijectivity domain for $R_\diamond$, which means that $\bigcup_{j=1}^{12} \mathscr{X}_\diamond^{(j)} \times \mathscr{Y}_\diamond^{(j)}$ must be a bijectivity domain for $\hat{F}_\diamond$. □

Remark 7.18. With Propositions 7.15 through 7.17, we now have explicit descriptions of sets Ω_{NE}, Ω_{Disk}, and $\Omega_\diamond$ which are bijectivity domains for their respective algorithms. For the nearest-even algorithm, Theorem 7.14 shows that Ω_{NE} is in fact the attractor of $\hat{F}_{\mathrm{NE}}$. Unfortunately, Theorem 7.14 does not apply to the diamond algorithm because each set $\mathscr{Y}_\diamond^{(k)}$ has the origin on its boundary rather than its interior.

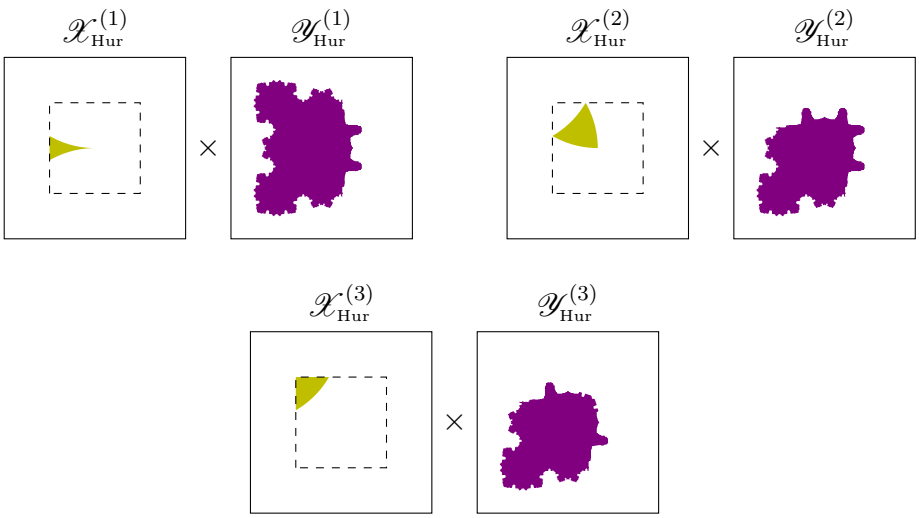


FIGURE 24: Numerical approximation of three of the products in Ω_{Hur}. The other products are rotations of these.

For the Hurwitz algorithm, the sets $\mathscr{Y}_{\mathrm{Hur}}^{(k)}$ appear to be fractal. Figure 24 shows a computer approximation of these sets formed by iterating random points under $\hat{F}_{\mathrm{Hur}}$ and then restricting to points (x, y) with $x \in \mathscr{X}_{\mathrm{Hur}}^{(k)}$ for a given k.

7.5. Sofic shifts

Definition. A finite sequence $(a_1, a_2, \ldots, a_n)$ is called *admissible* for a given continued fraction algorithm $\lfloor \cdot \rceil_c$ if the set $X_c(a_1, a_2, \ldots, a_n)$ from (43) has nonempty interior. This means that there exist complex numbers in Φ_c whose c-continued fraction representations begin with the given sequence.

In the $\mathrm{PSL}(2, \mathbb{Z})$ setting, algorithms with "dual codes" lead to sofic shifts [24, Theorem 5.8]. In this complex setting, however, all algorithms with the partition property lead to sofic shifts.

Proposition 7.19. *If a continued fraction algorithm satisfies the partition property, then the shift map on the closure of the set of all admissible sequences is sofic.*

Proof. This shift is precisely $\hat{f}_c$ acting on continued fraction representations.

As described in [4, Appendix C], sofic systems are obtained from Markov ones by amalgamation of the alphabet, and Markov systems may be obtained from sofic ones by refinement of the alphabet. In this case the original alphabet $\mathbb{A}$ is either $\mathbb{Z}[i]$ (all Gaussian integers) or E (even Gaussian integers). The countable partition $\mathcal{C}_c$ is essentially a refinement of $\mathbb{A}$ because each element of $A \in \mathcal{C}_c$ can be associated to a

pair $(a, j) \in \mathbb{A} \times \{1, \ldots, N\}$ for which $X_c(a) \cap \mathscr{X}^{(j)}$ is nonempty. Then the fact that

$$\hat{f}_c\big(X_c(a) \cap \mathscr{X}^{(j)}\big) = \bigcup_{k \in J(X_c(a) \cap \mathscr{X}^{(j)})} \mathscr{X}^{(k)}$$

means that each symbol (a, j) dictates precisely which (a', j') are allowed to follow it in an admissible sequence: those for which $j' \in J\big(X_c(a) \cap \mathscr{X}^{(j)}\big)$ and $X_c(a') \cap \mathscr{X}^{(j')}$ is nonempty. Thus the shift map is Markov on the extended alphabet

$$\big\{\, (a, j) \,:\, a \in \mathbb{A}, 1 \leq j \leq N, X_c(a) \cap \mathscr{X}^{(j)} \neq \emptyset \,\big\}$$

and is sofic with respect to the original alphabet $\mathbb{A}$. □

Remark 7.20. For the nearest-even and disk algorithms, the associated shift systems are actually Markov with respect to $\mathbb{A}$. This is equivalent to each set $X_c(a)$ being entirely contained in some element of $\mathcal{P}_c$, since that means that each $a \in \mathbb{A}$ has only one $j \in \{1, \ldots, N\}$ associated to it, and thus the alphabet is not meaningfully extended (equivalently, the semiconjugacy is actually a conjugacy).

References

[1] R. Adler, *Symbolic dynamics and Markov partitions*, Bull. Amer. Math. Soc., **35** (1998), No. 1, 1–56.

[2] R. Adler and L. Flatto, *Cross section maps for geodesic flows, I (The Modular surface)*, Birkhäuser, Progress in Mathematics (ed. A. Katok), 1982, 103–161.

[3] R. Adler and L. Flatto, *Cross section map for geodesic flow on the modular surface*, Contemp. Math., **26** (1984), 9–23. MR0737384 (85j:58128)

[4] R. Adler and L. Flatto, *Geodesic flows, interval maps, and symbolic dynamics*, Bull. Amer. Math. Soc., **25** (1991), No. 2, 229–334.

[5] W. Ambrose and S. Kakutani, *Structure and continuity of measurable flows*, Duke Math. J., **9** (1942), 25–42.

[6] E. Artin, *Ein Mechanisches System mit quasiergodischen Bahnen*, Addison Wesley, Collected Papers, 1965, 499–501.

[7] A. F. Beardon, *The Geometry of Discrete Groups*, Springer, New York, 1983. MR0698777 (85d:22026)

[8] J. Birman and C. Series, *Dehn's algorithm revisited, with applications to simple curves on surfaces*, Combinatorial Group Theory and Topology (AM-111), Princeton University Press, 1987, 451–478.

[9] R. Bowen and C. Series, *Markov maps associated with Fuchsian groups*, Inst. Hautes Études Sci. Publ. Math. No. 50 (1979), 153–170.

[10] A. Cano et al., *Complex Kleinian Groups*, Progress in Mathematics 303 (2013).

[11] W. Cao and K. Gongopadhyay, *Commuting isometries of the complex hyperbolic space*, Proc. Amer. Math. Soc, **139** (2011), 3317–3326. arXiv:1002.2479.

[12] S. G. Dani and A. Nogueira, *Continued fractions for complex numbers and values of binary quadratic forms*, Trans. American Math. Society. 2014.

[13] W. Goldman, *Complex hyperbolic geometry*, Clarendon Press, 1999.

[14] G. A. Hedlund, *A metrically transitive group defined by the modular group*, Amer. J. Math., **57** (1935), 668–678.

[15] D. Hensley, *Continued fractions*. World Sci. Publishing Co. Pte. Ltd. 2006.

[16] A. Hurwitz, *Über die Entwicklung complexer Grössen in Kettenbruche*, Acta Math. **11**, 1887. 187–200.

[17] J. Hurwitz, *Über die Reduction der Binären Quadratischen Formen mit Complexen Coefficienten und Variabeln*, Acta Math. **25**, 1902. 231–290.

[18] A. Katok and B. Hasselblatt, *Intro. to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.

[19] S. Katok, *Fuchsian Groups*, University of Chicago Press, 1992.

[20] S. Katok, *Coding of closed geodesics after Gauss and Morse*, Geometriae Dedicata, **63** (1996), 123–145.

[21] S. Katok and I. Ugarcovici, *Arithmetic coding of geodesics on the modular surface via continued fractions*. Centrum Wiskunde & Informatica. Amsterdam. **135** (2005), 59–77.

[22] S. Katok and I. Ugarcovici, *Symbolic dynamics for the modular surface and beyond*, Bull. Amer. Math. Soc., **44** (2007), 87–132.

[23] S. Katok and I. Ugarcovici, *Structure of attractors for (a,b)-continued fraction transformations*, Journal of Modern Dynamics, **4** (2010), 637–691.

[24] S. Katok and I. Ugarcovici, *Applications of (a,b)-continued fraction transformations*, Ergod. Th. & Dynam. Sys., **32** (2012), 755–777.

[25] S. Katok and I. Ugarcovici, *Structure of attractors for boundary maps associated to Fuchsian groups*, Geometriae Dedicata, **191** (2017), 171–198.

[26] S. Katok and I. Ugarcovici, *Errata: Structure of attractors for boundary maps associated to Fuchsian groups*, to appear in Geometriae Dedicata.

[27] S. Katok and A. J. Zydney. *Adler and Flatto revisited: cross-sections for geodesic flow on compact surfaces of constant negative curvature*, to appear in Studia Mathematica. arXiv:1710.07618.

[28] P. Koebe, *Riemannsche Mannigfaltigkeiten und nicht euklidische Raumformen*, IV, Sitzungsberichte Deutsche Akademie von Wissenschaften, 1929, 414–557.

[29] A. Lukyanenko, *Geometric mapping theory of the Heisenberg group, sub-Riemannian manifolds, and hyperbolic spaces*, Doctoral Thesis, University of Illinois at Urbana-Champaign, 2014.

[30] A. Lukyanenko and J. Vandehey, *Continued fractions on the Heisenberg group*, Acta Arithmetica, **167** (2015), 19–42.

[31] J. Vandehey, *Lagrange's theorem for continued fractions on the Heisenberg group*, Bulletin of the London Math. Soc., **47** (2015), 866–882.

[32] B. Maskit, *On Poincaré's Theorem for fundamental polygons*, Adv. Math., **7** (1971), 219–230.

[33] M. Morse, *Symbolic dynamics* (unpublished), Institute for Advanced Study Notes, Princeton, 1966.

[34] N. Ocwald, *Hurwitz's complex continued fractions: a historical approach and modern perspectives*, Doctoral Thesis, Universität Würzburg, 2014.

[35] J. Parker, *Hyperbolic spaces: the Jyväskylä notes*, 2007.

[36] C. Series, *Symbolic dynamics for geodesic flows*, Acta Math., **146** (1981), 103–128.

[37] C. Series, *The infinite word problem and limit sets in Fuchsian groups*, Ergod. Th. & Dynam. Sys. 1, 1981, 336–360.

[38] C. Series, *The modular surface and continued fractions*, J. London Math. Soc., **31** (1985), 69-80.

[39] C. Series, *Geometrical Markov coding of geodesics on surfaces of constant negative curvature*, Ergod. Th. & Dynam. Sys. 6, 1986, 601–625.

[40] A. Tanaka, *A complex continued fraction transformation and its ergodic properties*, Tokyo J. Math., **8**, 1985. 191–214.

[41] P. Tukia, *On discrete groups of the unit disk and their isomorphisms*, Ann. Acad. Sci. Fenn., Series A, I. Math. 504 (1972), 5–44.

[42] B. Weiss, *On the work of Roy Adler in Ergodic theory and dynamical systems*, in Symbolic dynamics and its applications (New Haven, CT, 1991), 19–32, Contemp. Math., 135, Amer. Math. Soc., Providence, RI, 1992.

[43] D. Zagier, *Zetafunkionen und quadratische Körper: eine Einfuẗhrung in die höhere Zahlentheorie*, Springer-Verlag, 1982.

[44] D. Zagier, *Possible notions of "good" reduction algorithms*, Personal communication, 2007.

[45] A. J. Zydney, *Structure of attractor region for natural extension map associated with p-adic continued fractions*, Undergraduate Honors Thesis, Pennsylvania State University, 2011.

[46] A. J. Zydney, *Bijectivity and trapping regions for complex continued fraction transformation.* Master's Paper, Pennsylvania State University, 2016. arXiv:1608.06351.

Vita
Adam J. Zydney

Education

The Pennsylvania State University

Ph.D. in Mathematics	May 2018
M.A. in Mathematics	May 2017
B.S. with Honors in Mathematics (Computer Science Concentration)	May 2011

Teaching Awards

Charles H. Hoover Memorial Award	April 2015
Departmental Teaching Award	December 2013

Teaching Experience

Private and Penn State Student Support Services Program

Introductory Calculus Tutor	Fall 2015, Spring 2016
Introductory C++ (CSE 121) Tutor	Fall 2008

The Pennsylvania State University

Vector Calculus (MATH 231) Instructor	Spring 2014, Fall 2014, Spring 2015
Vector Calculus (MATH 230) Instructor	Fall 2012, Fall 2013
Business Calculus (MATH 110) Teaching Assistant	Fall 2012
Honors Real Analysis (MATH 312H) Teaching Assistant	Spring 2012

Johns Hopkins Center for Talented Youth

Paradoxes and Infinities (PDOX) Instructor	Summer 2013, 2014, 2015
Cryptology and Adv. Crypto. (CODE, COD2) TA	Summer 2010, 2011, 2012

Additional Math Experience

The Pennsylvania State University

Co-organizer and founder, Graduate Student Colloquium	Fall 2017 - Spring 2018

American Institute of Mathematical Sciences

Editorial Assistant (Copy-editor)	Spring 2013 - 2018

State College Area School District

Elementary Math Curriculum Selection Committee Member	April 2010 - June 2011

Programming Experience

Mandarin Companion and Allset Learning

Python Developer	August 2016 - 2018

Penn State Dance Marathon Technology Committee

Lead Database Developer	Spring 2008 - 2011

Languages: Python, Java, PHP, MySQL, CSS, JavaScript, LaTeX, Ti*k*Z, Mathematica, Matlab, Octave