Cutting Sequences for Geodesic Flow on the Modular Surface and Continued Fractions

David J. Grabiner[1]

Department of Mathematics and Statistiacs
Bowling Green State University
Bowling Green, OH 43403

Jeffrey C. Lagarias[2]

AT&T Labs
Florham Park, NJ 07932-0971

(July 24, 2000)

Abstract

This paper describes the cutting sequences of geodesic flow on the modular surface $\mathfrak{H}/PSL(2,\mathbb{Z})$ with respect to the standard fundamental domain $\mathcal{F} = \{z = x + iy : -\frac{1}{2} \le x \le \frac{1}{2} \text{ and } |z| \ge 1\}$ of $PSL(2,\mathbb{Z})$. The cutting sequence for a vertical geodesic $\{\theta + it : t > 0\}$ is related to a one-dimensional continued fraction expansion for θ, called the one-dimensional Minkowski geodesic continued fraction expansion, which is associated to a parametrized family of reduced bases of a family of 2-dimensional lattices. We show that the additive ordinary continued fraction expansion of θ can be computed from the cutting sequence for a vertical geodesic by a finite automaton. In the reverse direction, the cutting sequence for a vertical geodesic cannot be computed from the additive continued fraction expansion using any finite automaton, but can be computed by a Turing machine in quadratic time using linear space. The set of cutting sequences for all geodesics forms a two-sided shift in a symbol space $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\}$ which has the same set of forbidden blocks as for vertical geodesics. We show that this shift is not a sofic shift, and that it characterizes the fundamental domain $\mathcal{F}$ up to an isometry of the hyperbolic plane $\mathfrak{H}$.

[1]Supported by an NSF Postdoctoral Fellowship. Some of this work was done while visiting the Mathematical Sciences Research Institute; MSRI is supported by NSF Grant DMS-9022140.

[2]Some of this work was done while visiting the Mathematical Sciences Research Institute.

Cutting Sequences for Geodesic Flow on the Modular Surface and Continued Fractions

David J. Grabiner

Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, OH 43403

Jeffrey C. Lagarias

AT&T Labs
Florham Park, NJ 07932-0971

1. Introduction

This paper describes the symbolic dynamics of cutting sequences of geodesics for the standard fundamental domain $\mathcal{F} = \{z = x + iy : -\frac{1}{2} \leq x < \frac{1}{2}, |z| \geq 1\}$ of the modular group $PSL(2, \mathbb{Z})$ acting on the upper half plane $\mathfrak{H} = \{z \in \mathbb{C} : Im(z) > 0\}$.

We study in particular the cutting sequences of vertical geodesics $\{\theta + it : t > 0\}$ for $\theta \in \mathbb{R}$. These particular geodesics are related to a continued fraction expansion introduced in 1850 by Hermite [19] in terms of quadratic forms, and studied by Humbert [21, 22]. Our motivation for studying them was that they appear in the one-dimensional case of a multidimensional continued fraction introduced in Lagarias [29]. This expansion, called the *Minkowski geodesic continued fraction expansion* (MGCF expansion), is based on following a parametrized family of lattice bases in $GL(d+1, \mathbb{Z}) \backslash GL(d+1, \mathbb{R})$ as the parameter t varies; in the one-dimensional case the family associated to θ is $\begin{bmatrix} 1 & 0 \\ -\theta & t \end{bmatrix}$. The MGCF expansion has the merit that it finds good Diophantine approximations in all dimensions. However, the one-dimensional case of this expansion does not coincide with the ordinary continued fraction expansion. We show in §3.4 that the one-dimensional MGCF expansion of θ is essentially equivalent the cutting sequence expansion of the vertical geodesic $\{\theta + it : 0 < t < \infty\}$. We use this connection to determine the precise relation of the MGCF expansion of θ to the additive continued fraction expansion of θ; the *additive continued fraction* is the variant of the ordinary continued fraction which includes all intermediate convergents.

A second motivation for studying cutting sequences on $\mathfrak{H}/PSL(2, \mathbb{Z})$ arises from their use as symbolic codings of geodesics in the study of geodesic flow on constant negative curvature surfaces of finite volume; we give more background on this at the end of the introduction. Cutting sequence encodings are of special interest because as an ensemble they apparently encode more information about geodesic flow on a Riemann surface than other symbolic encodings of geodesics, which have a simple description (shift of finite type) but only retain topological and not conformal information about the Riemann surface. Adler and Flatto [3, §10] raised the question whether cutting sequence encodings preserve conformal information about the

Riemann surface and also retain information about the lines on the Riemann surface used to define the cuts (i.e. the shape of the fundamental domain in the universal cover.) Consider any hyperbolic polygon $\mathcal{P}$ in $\mathfrak{H}$ which is a fundamental domain for a properly discontinuous group Γ acting on $\mathfrak{H}$, such that $\mathfrak{H}/\Gamma$ has finite volume, and suppose that $\mathcal{P}$ has $|\mathcal{P}|$ sides. Let $\Sigma_{\mathcal{P}}$ denote the closed subshift of the shift on $|\mathcal{P}|$ letters generated by the cutting sequences for $\mathcal{P}$ for geodesic flow on general $\mathfrak{H}/\Gamma$, as defined in §2.2. Adler and Flatto [3, p. 300][1] ask: does $\Sigma_{\mathcal{P}}$ determine $\mathcal{P}$ up to conformal isometry and Γ up to hyperbolic conjugacy? We show here that this is the case for the standard fundamental domain $\mathcal{F}$ of $PSL(2,\mathbb{Z})$.

Our main results are as follows.

(1) The additive continued fraction expansion of θ can be computed from either the cutting sequence expansion or the Minkowski geodesic continued fraction expansion of the vertical geodesic $\theta + it$ by a finite automaton, but *not* vice-versa. (Theorems 4.1, 4.2 and 6.4.) The Minkowski geodesic continued fraction expansion can be computed from the additive continued fraction expansion by a Turing machine using quadratic time and linear space. (Theorem 5.2.)

(2) The set of cutting sequence expansions for (irrational) vertical geodesics $\Pi^0_{\mathcal{F}}$ is characterized in terms of forbidden blocks (Theorem 6.1). A generating set of forbidden blocks is enumerated. (Theorems 5.1 and 6.2.). The number of minimal forbidden blocks of length at most k grows exponentially in k. (Theorem 6.3).

(3) The cutting sequence expansions for all geodesics comprise (essentially) a closed two-sided subshift $\Sigma_{\mathcal{F}}$ of the full shift on three symbols $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\}$. The set of forbidden blocks of $\Sigma_{\mathcal{F}}$ is the same as that for $\Pi^0_{\mathcal{F}}$. (Theorem 7.1). Each symbol sequence in $\sum_{\mathcal{F}}$ corresponds to a unique oriented geodesic on $\mathfrak{H}/PSL(2,\mathbb{Z})$, with the exception of the two sequences $\bar{\mathsf{L}}^{\infty}$ and $\bar{\mathsf{R}}^{\infty}$. In the converse direction, each oriented geodesic gives rise to a finite number of shift-equivalence classes of symbol sequences in $\Sigma_{\mathcal{F}}$. This number is at most eight if it is not a periodic geodesic. (Theorem 7.2). The shift $\Sigma_{\mathcal{F}}$ is not a sofic system. (Theorem 7.3)

(4) The shift $\Sigma_{\mathcal{F}}$ characterizes $\mathcal{F}$ up to an isometry of $\mathfrak{H}$. If $\mathcal{P}$ is a hyperbolic polygon of finite area (possibly with some ideal vertices) which is a fundamental domain of a discrete subgroup Γ of $PSL(2,\mathbb{R})$ and if the subshift $\Sigma_{\mathcal{P}}$ is isomorphic to $\Sigma_{\mathcal{F}}$ by a permutation of symbols, then there is a hyperbolic isometry $g \in PSL(2,\mathbb{R})$ such that $\mathcal{F} = g\mathcal{P}$ and $PSL(2,\mathbb{Z}) = g\Gamma g^{-1}$. (Theorem 8.1).

Result (1) gives a specific sense in which cutting sequences contain information coded in a more concise form than the ordinary continued fraction expansion. To obtain result (2) we show in §5 that vertical geodesics have special features (not shared by general geodesics) that facilitate characterizing forbidden blocks: any vertical geodesic that hits a $PSL(2,\mathbb{Z})$-translate of a corner of the fundamental domain must have rational θ, and each such geodesic hits at most one such corner, with the exception of those $\theta \equiv \frac{1}{2} (\mathrm{mod}\ 1)$, which hits two corners. Furthermore the continued fraction expansion of any rational θ such that $\{\theta + it : t > 0\}$ hits a corner has a symmetry which can be described in terms of the linear fractional transformation $\mathbf{z} \to \mathsf{N}\mathbf{z}$

[1]They raise the question for the $(8g-4)$-sided fundamental polygons discussed in [3], for genus $g \geq 2$.

with $\mathsf{N} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, see Theorem 5.1. In formulating result (3) we *define* cutting sequence expansions for geodesics that hit a corner of the fundamental domain $\mathcal{F}$ of $\mathfrak{H}/PSL(2,\mathbb{Z})$ to be limits of general-position geodesics that hit no corner. This procedure assigns infinitely many different cutting sequences to certain periodic geodesics. Finally, result (4) shows that the shift $\Sigma_{\mathcal{F}}$ encodes conformal information about $\mathcal{F}$.

To conclude this introduction, and to put the results above in a more general context, we recall background on symbolic codings of geodesic flow on a compact Riemann surface with a finite number of punctures. If Γ is a finitely generated discrete group of conformal isometries of the hyperbolic plane $\mathfrak{H}$, such that $\mathfrak{H}/\Gamma$ has finite volume, then $\mathfrak{H}/\Gamma$ is a compact Riemann surface[2] minus a finite number of punctures, and geodesic flow on $\mathfrak{H}/\Gamma$ contains both topological and conformal information about this Riemann surface. Symbolic codings of geodesics were introduced by Hadamard [15] in 1898 as a way of understanding the complicated motions on geodesics on such surfaces. E. Artin [5] showed in 1924 that there exist dense geodesics on the modular surface $\mathfrak{H}/PSL(2,\mathbb{Z})$, using symbolic encodings with continued fractions, and in 1935 G. Hedlund [17] used Artin's coding to show that geodesic flow on this surface was ergodic. Some other symbolic encodings were "boundary expansions" by Nielsen [39] and cutting sequences with respect to a fundamental domain of Γ by Koebe [26] and Morse [37]. Cutting sequence expansions can be viewed as a generalization of the reduction theory of indefinite binary quadratic forms to arbitrary finitely-generated Fuchsian groups, see Katok [23]. More recently Bowen and Series [8] and Series [44], [45] used "modified boundary expansions" to obtain a particularly simple symbolic expansion from which strong forms of ergodicity could be deduced. Adler and Flatto [2], [3] used "rectilinear map" codings of geodesic flow on $\mathfrak{H}/\Gamma$ to explain why certain specific maps of the interval, such as the continued fraction map and backwards continued fraction map, have invariant measures of a simple form. They showed that these invariant measures are inherited from the invariant measure on geodesic flow (Liouville measure) using a cross-section map followed by a factor map. The codings of Bowen and Series [8], [45] and Adler and Flatto [3] have a particularly simple structure – they are shifts of finite type – and they preserve topological information about the Riemann surface, but they lose conformal information, for they are identical for all Riemann surfaces of the same genus. (See Theorem 8.4 of [3].) In this respect it is of interest whether cutting sequence encodings preserve conformal information. Besides all these encodings, two other encodings of geodesic flow on the modular surface $\mathfrak{H}/PSL(2,\mathbb{Z})$ with interesting Diophantine approximation properties appear in Arnoux [4] and Lagarias and Pollington [30].

With regard to this general framework, we note that the proofs in this paper heavily rely on specific facts about $PSL(2,\mathbb{Z})$. The methods used may extend in some form to congruence subgroups of $PSL(2,\mathbb{Z})$, but do not appear likely to apply to general finitely generated Fuchsian groups Γ.

For general terminology in symbolic dynamics we follow Adler and Flatto [3] and Lind and Marcus [32]. From the viewpoint of symbolic dynamics, the associated shift $\Sigma_{\mathcal{F}}$ studied in this paper is an example of a naturally occurring shift more complicated than the ones currently having a well-developed theory, cf.[32].

[2]There are some mild extra conditions needed on Γ for compactness, see Lehner [31, p. 203–205] for sufficient conditions. It suffices for Γ to have a fundamental region $\mathcal{F}$ that is a hyperbolic polygon with a finite number of sides, with only parabolic vertices at infinity.

From the viewpoint of symbolic dynamics, the associated shift $\Sigma_{\mathcal{F}}$ provides a simple example of a naturally occurring shift more complicated than the ones with currently having a well-developed theory, cf.[32].

2. Cutting Sequences

We describe cutting sequences for finite-sided hyperbolic polygons that are the fundamental domain of a given discrete subgroup Γ of $PSL(2,\mathbb{R})$ that acts discontinuously on $\mathfrak{H}$, and then specialize to standard fundamental domain $\mathcal{F}$ of the modular group $PSL(2,\mathbb{Z})$. General references for geodesic flow and for cutting sequences are Adler and Flatto [3] and Series [47].

2.1. Cutting Sequence Shifts

Let $\mathfrak{H}$ denote the hyperbolic plane, represented as the upper half-plane $\mathfrak{H} = \{z = x + iy : y > 0\}$ with hyperbolic line element $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$ and volume $\frac{dxdy}{y^2}$. An oriented geodesic $\gamma = \langle \theta_1, \theta_2 \rangle$ in $\mathfrak{H}$ is uniquely specified by its two ideal endpoints $\theta_1, \theta_2 \in \mathbb{R} \cup \{\infty\}$, with $\theta_1 \neq \theta_2$. It is oriented from its *head* θ_1 to its *foot* θ_2. A *vertical geodesic* $\langle \infty, \theta \rangle$ is the vertical line $\theta + it$, with t decreasing from ∞ to 0. Other geodesics are semicircles perpendicular to the real axis with endpoints θ_1 and θ_2, centered at $\frac{\theta_1+\theta_2}{2}$, and oriented to start at θ_1 and travel to θ_2.

Let Γ be a finitely generated discrete subgroup of $PSL(2,\mathbb{R})$ which acts properly discontinuously on the upper half-plane $\mathfrak{H} = \{x+iy : y > 0\}$ as a group of linear fractional transformations. We suppose that $\mathfrak{H}/\Gamma$ has finite volume with respect to the hyperbolic volume on $\mathfrak{H}$. Let $\mathcal{P}$ be any finite-sided convex hyperbolic polygon which is a *fundamental domain* for Γ, i.e. $\mathcal{P}$ is such that the set of Γ-translates $\{g\mathcal{P} : g \in \Gamma\}$ tessellates $\mathfrak{H}$, and no element $g \in \Gamma$ sends $\mathcal{P}$ to itself except the identity. We allow polygons $\mathcal{P}$ to have one or more ideal vertices ("cusps"), and they may also have extra vertices, called *elliptic vertices*, which have an angle of π and which fall in the middle of a side of $\mathcal{P}$ regarded as a geometric object, as discussed below. Fundamental domains exist for each such finitely generated Γ, cf. [4, Theorem 10.1.4]. Such polygons $\mathcal{P}$ necessarily have an even number of "sides," where by convention each elliptic vertex divides a side of the geometric object $\mathcal{P}$ into two "sides." If $\mathcal{P}$ is to be a fundamental domain for Γ, the "sides" of $\mathcal{P}$ must be identified under the action of Γ. Specifically, there is a pairing of the "sides" of $\mathcal{P}$ which assigns to each "side" s of $\mathcal{P}$ a "side" s' and an isometry $g(s,s') \in \Gamma$ such that $g(s,s')$ maps s to s'. Furthermore $(s')' = s$ and

$$g(s',s) = (g(s,s'))^{-1},$$

while if $s = s'$ then $g(s,s)$ is the identity on s. The pairing indicates how $\mathcal{P}$ tessellates $\mathfrak{H}$. For $g = g(s,s')$, the domain of $\mathcal{P}$ is adjacent to $\mathcal{P}$ in such a way that the "side" labeled s of $g\mathcal{P}$ coincides with the "side" labeled s' of $\mathcal{P}$, while $g(Int(\mathcal{P})) \cap Int(\mathcal{P}) = \emptyset$. If $s = s'$ then necessarily g^2 is the identity on $\mathfrak{H}$, and g is a hyperbolic reflection. The *neighborhood set* of $\mathcal{P}$

$$N_\Gamma(\mathcal{P}) := \{g \in \Gamma : g = g(s,s') \text{ for some side } s \text{ of } \mathcal{P}\} . \tag{2.1}$$

is a set of generators of the group Γ. It will comprise the set of symbols used in the cutting sequence encodings of geodesics described below.

The number of elements of $N_\Gamma(\mathcal{P})$ is generally equal to the number of "sides" of $\mathcal{P}$, but can be less if there are two or more elements $g(s,s')$ that are equal. For convex polygons $\mathcal{P}$ this situation can only arise when there are hyperbolic reflections $g = g(s, s')$ with $s \neq s'$, in which case $g(s, s') = g(s', s)$. In such a case the "sides" s and s' must lie on a common geodesic of $\mathfrak{H}$ and share a common endpoint,[3] which is then a vertex of $\mathcal{P}$ having an angle of π. We call any such vertex an *elliptic vertex.* These two "sides" s and s' are then assigned the same generator, so if we erase the elliptic vertex and glue them together, we get a single geometric side of the polygon $\mathcal{P}$ regarded as a geometric object. We therefore have: For a convex fundamental domain $\mathcal{P}$ of a finitely generated group Γ, the number of elements of $N_\Gamma(P)$ is equal to the number of geometric sides of $\mathcal{P}$. In particular, *the number of elements of $N_\Gamma(P)$ may be odd.*

As an example, the standard fundamental domain $\mathcal{F}$ of $\Gamma = PSL(2, \mathbb{Z})$, is geometrically a triangle with one ideal vertex, but as a fundamental domain for Γ is a quadrilateral having an elliptic vertex at $z = i$. In this case $N_\Gamma(\mathcal{F})$ has three elements.

We associate to geodesic flow on $\mathfrak{H}/\Gamma$ the (two-sided) *cutting sequence subshift* $\Sigma_{\mathcal{P},\Gamma}$ which is a closed subshift of the full shift on $|N_\Gamma(\mathcal{P})|$ letters.

Definition 2.1. The set $G^0_{\mathcal{P},\Gamma}$ of *general position geodesics* for $\mathcal{P}$ consists of those geodesics γ on $\mathfrak{H}$ which intersect the interior of $\mathcal{P}$, and which are transverse to all $\{g(\mathcal{P}) : g \in \Gamma\}$ in the sense that γ contain no vertex nor part of any side of positive length of any translated polygon $\{g\mathcal{P} : g \in \Gamma\}$. If $\mathcal{P}$ contains ideal vertices ("cusps"), then the endpoints of γ must not coincide with any Γ-translate of any such vertex.

To each $\gamma \in G^0_{\mathcal{P},\Gamma}$, we assign the doubly-infinite *cutting sequence*

$$C(\gamma) := (\dots, g_{-2}, g_{-1}, g_0, g_1, g_2, \dots) \tag{2.2}$$

in which all $g_i \in N_\Gamma(\mathcal{P})$, as follows. We first label each "side" s of the fundamental domain $\mathcal{P}$ with the label $g(s, s')$ on its "inside" edge in $\mathcal{P}$. We similarly label all the sides of the translated domain $\{h\mathcal{P} : h \in \Gamma\}$, so that the edge of $h\mathcal{P}$ which is hs has the label $g(s, s')$. Every "side" is now assigned two labels, on its "inside" edge and "outside" edge, because it abuts two copies of a fundamental domain. (Figure 2.1 below indicates such labels.) The geodesic $\gamma = \langle\theta_1, \theta_2\rangle$ visits in order a sequence of translates $\{h_i\mathcal{P}; i \in \mathbb{Z}\}$ in which $h_i \in \Gamma$ and h_0 is the identity. At a crossing of the edge from $h_{i-1}\mathcal{P}$ to $h_i\mathcal{P}$ we assign the symbol g_i on the side of the domain that the geodesic *enters.* (This is the convention of Katok [25]; and opposite to that of Adler and Flatto [3, p. 243], who use the symbol g_i on the exit edge.) This convention yields:

Lemma 2.1. *The cutting sequence $(g_1, \dots, g_j)$ follows a geodesic from $\mathcal{P}$ to the translated domain $\mathcal{P}_j = h_j\mathcal{P}$ with*

$$h_j = g_1 g_2 \cdots g_j. \tag{2.3}$$

Proof. The edge $h_i s$ of $h_i\mathcal{P}$ that borders $h_{i-1}\mathcal{P}$ is assigned the edge label inside $h_i\mathcal{P}$ of $g_i = g(s, s')$. Now we have $(h_{i-1} g(s, s') h_{i-1}^{-1}) h_{i-1}\mathcal{P} = h_i\mathcal{P}$, hence

$$h_{i-1} g_i = h_i \ , \tag{2.4}$$

see Figure 2.1. ■

[3]For non-convex $\mathcal{P}$ this may fail to hold.

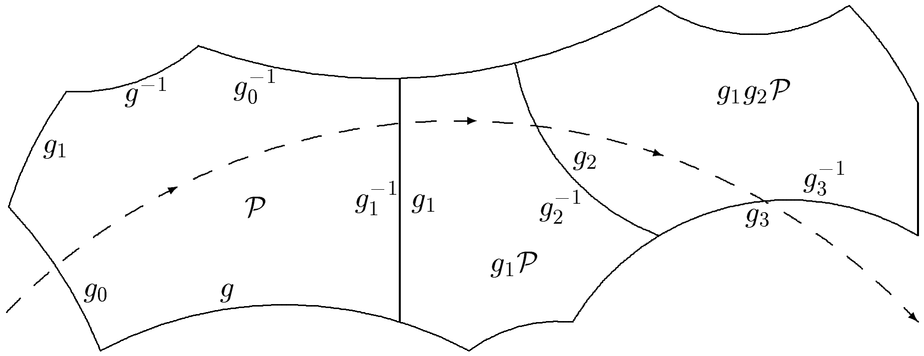


Figure 2.1: The cutting-sequence encoding of a geodesic. The pictured geodesic has cutting sequence $\ldots, g_0, g_1, g_2, \ldots$, starting with g_1 from $\mathcal{P}$.

The transversality hypothesis says that if γ intersects some translate $g(\mathcal{P})$ for $g \in \Gamma$, then γ intersects the interior of $g(\mathcal{P})$, which implies that the expansion (2.2) is well-defined and unique. If $\gamma^{-1} = \langle \theta_2, \theta_1 \rangle$ denotes the reversed geodesic, then $\boldsymbol{\gamma}^{-1} \in G^0_{\mathcal{P},\Gamma}$ has the cutting sequence

$$C(\boldsymbol{\gamma}^{-1}) = (\ldots, g'_{-1}, g'_0, g'_1, \ldots) \tag{2.5}$$

in which $g'_i := (g_{-i})^{-1}$.

Let $\Sigma^0_{\mathcal{P},\Gamma}$ denote the set of all cutting sequences $C(\gamma)$ for $\boldsymbol{\gamma} \in G^0_{\mathcal{P},\Gamma}$. This set is invariant under the (forward) shift operator $\sigma(\{g_i\}) = \{g_{i+1}\}$ since

$$\sigma(C(\gamma)) = C(g_0\gamma) \ ,$$

and the geodesic $g_0\gamma$ is in $G^0_{\mathcal{P},\Gamma}$.

Definition 2.2. The *orbit* or *shift-equivalence class* $[C]$ of a cutting sequence C is the union of all its forward and backward shifts

$$[C] := \bigcup_{k \in \mathbb{Z}} \sigma^k(C); \tag{2.6}$$

i.e., it is the smallest shift-invariant set containing C.

Thus $[C(\gamma)]$ is contained in $\Sigma^0_{\mathcal{P}}$, but is generally not a closed set. The geodesics on the surface $\mathfrak{H}/\Gamma$ are projections of geodesics on $\mathfrak{H}$. Since the shift operator on cutting sequences corresponds to a motion of the geodesic by an element of Γ, the orbit $[C]$ is an invariant of the projected geodesic.

Definition 2.3. The *cutting sequence shift* $\Sigma_{\mathcal{P},\Gamma}$ is the closure of $\Sigma^0_{\mathcal{P},\Gamma}$ in the symbol topology. That is, $\Sigma_{\mathcal{P},\Gamma}$ consists of all symbol sequences

$$(\ldots, g_{-1}, g_0, g_1, \ldots)$$

such that every finite block $(g_i, g_{i+1}, \ldots, g_{i+k})$ occurs in some $C(\gamma)$ for a general position geodesic $\gamma \in G^0_{\mathcal{P},\Gamma}$.

We now extend the definition of cutting sequences to apply to *all* geodesics γ that hit the interior of $\mathcal{P}$. The set of cutting sequences $C(\gamma)$ for a general geodesic $\gamma = \langle\theta', \theta\rangle$ consists of all symbol sequences that can be obtained as a limit point (in the sequence topology) of a sequence of $C(\gamma_j)$ having $\gamma_j \in G^0_{\mathcal{P},\Gamma}$ such that $\gamma_j = \langle\theta'_j, \theta_j\rangle$ with $\theta'_j \to \theta'$ and $\theta_j \to \theta$ as $j \to \infty$. There always exist such convergent sequences $\gamma_j : j \geq 1\}$, because $\Sigma_{\mathcal{P},\Gamma}$ is compact in the symbol topology; hence $C(\gamma) \neq \emptyset$. The set $C(\gamma)$ may be infinite for some geodesics γ. All cutting sequences in $C(\gamma)$ are contained in $\Sigma_{\mathcal{P},\Gamma}$; however, the closure operation used in defining $\Sigma_{\mathcal{P},\Gamma}$ allows the possibility that $\Sigma_{\mathcal{P},\Gamma}$ contains some symbol sequences not coming from any geodesic.

Definition 2.4. The *shift-equivalence class* $[C(\gamma)]$ is the union of all orbits of cutting sequences in $C(\gamma)$.

We show in the specific case $\Gamma = PSL(2,\mathbb{Z})$ and fundamental domain $\mathcal{F}$ that each $[C(\gamma)]$ is a union of a finite number of orbits of individual cutting sequences (Theorem 7.2).

The totality of all possible convex polygons $\mathcal{P}$ that are fundamental domains of some group Γ has an explicit characterization. A general treatment appears in Maskit [33], which covers non-convex fundamental domains, and also covers groups of isometries of $\mathfrak{H}$, which may include orientation-reversing isometries. Maskit [33, section 2] gives a sufficient condition for $\mathcal{P}$ to be a fundamental domain, but for finitely-generated groups where $\mathfrak{H}/\Gamma$ has finite volume and $\mathcal{P}$ is hyperbolically convex it is a necessary condition as well.

2.2. Cutting Sequences for the Modular Surface

We now specialize to cutting sequence expansions for geodesics on the modular surface $\mathfrak{H}/PSL(2,\mathbb{Z})$. The modular group $PSL(2,\mathbb{Z}) = SL(2,\mathbb{Z})/\pm\mathsf{I}$ has many different convex fundamental domains $\mathcal{P}$ that are hyperbolic quadrilaterals; see Beardon [6, Example 9.4.4]. The *standard fundamental domain* for the modular group acting on $\mathfrak{H}$ is the hyperbolic triangle

$$\mathcal{F} = \{z : |z| \geq 1 \quad \text{and} \quad -\frac{1}{2} \leq Re(z) \leq \frac{1}{2}\}\ , \tag{2.7}$$

which has one ideal vertex ("cusp") at $i\infty$. When regarded as a fundamental domain, it is a quadrilateral having an elliptic vertex added at $z = i$, see Figure 2.2.

The four "sides" of $\mathcal{F}$ are indicated in Figure 2.2 along with the two side pairings of s_1 with s_2 and s_3 with s_4, and with elements $g(s, s') \in PSL(2,\mathbb{Z})$, which comprise

$$N(\mathcal{F}) := N_{PSL(2,\mathbb{Z})}(\mathcal{F}) = \left\{ \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}\ . \tag{2.8}$$

The "sides" s_3 and s_4 containing the elliptic vertex together make up one side of $\mathcal{F}$, and the three sides are labeled inside the fundamental domain as follows:

$$\bar{\mathsf{R}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad \Leftrightarrow \qquad s_1 := \left\{ -\frac{1}{2} + it : t > \frac{\sqrt{3}}{2} \right\}$$

$$\bar{\mathsf{L}} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \qquad \Leftrightarrow \qquad s_2 := \left\{ \frac{1}{2} + it : t > \frac{\sqrt{3}}{2} \right\}$$

$$\bar{\mathsf{J}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad \Leftrightarrow \qquad s_3 \cup s_4 := \left\{ |z| = 1 : -\frac{1}{2} < \mathbb{R}e(z) < \frac{1}{2} \right\}\ .$$

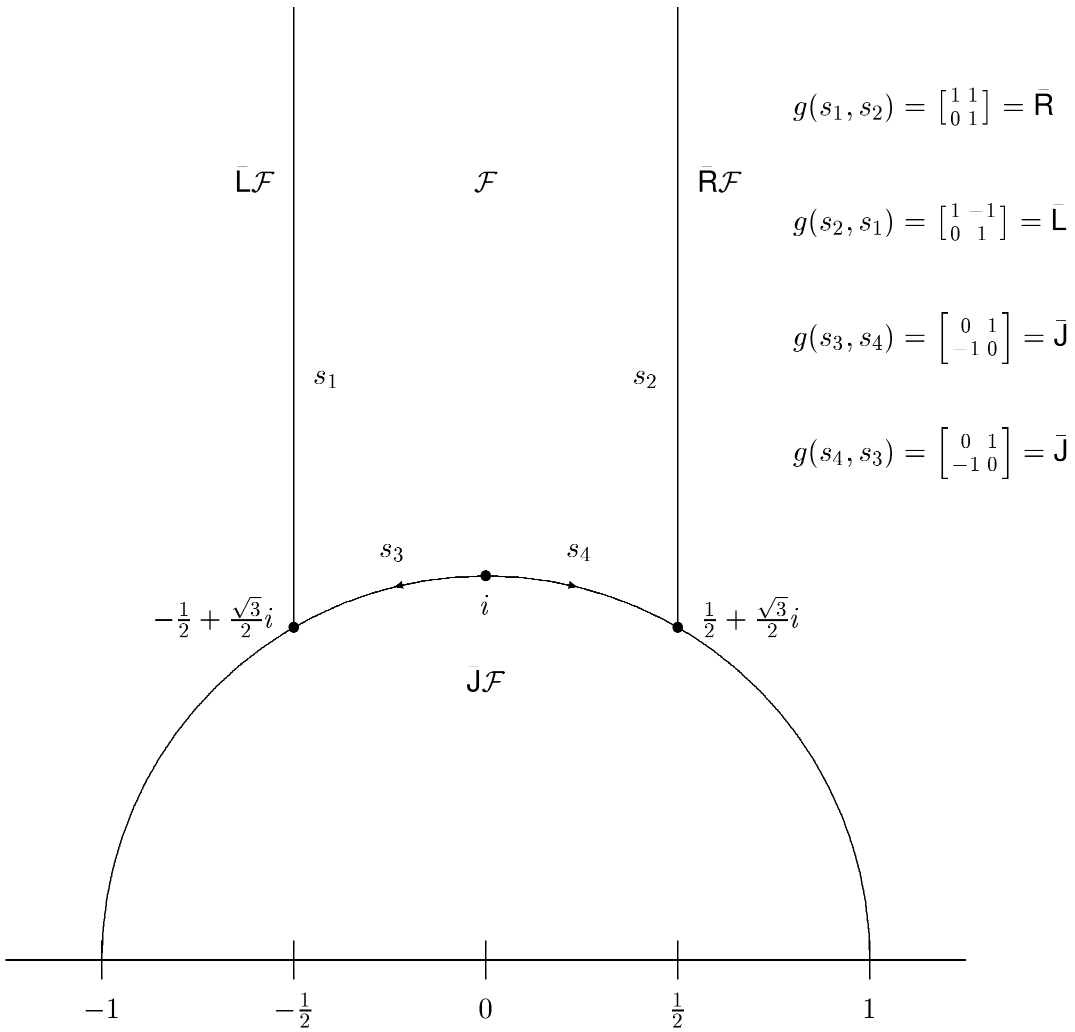


Figure 2.2: The fundamental domain $\mathcal{F}$ with side-pairings.

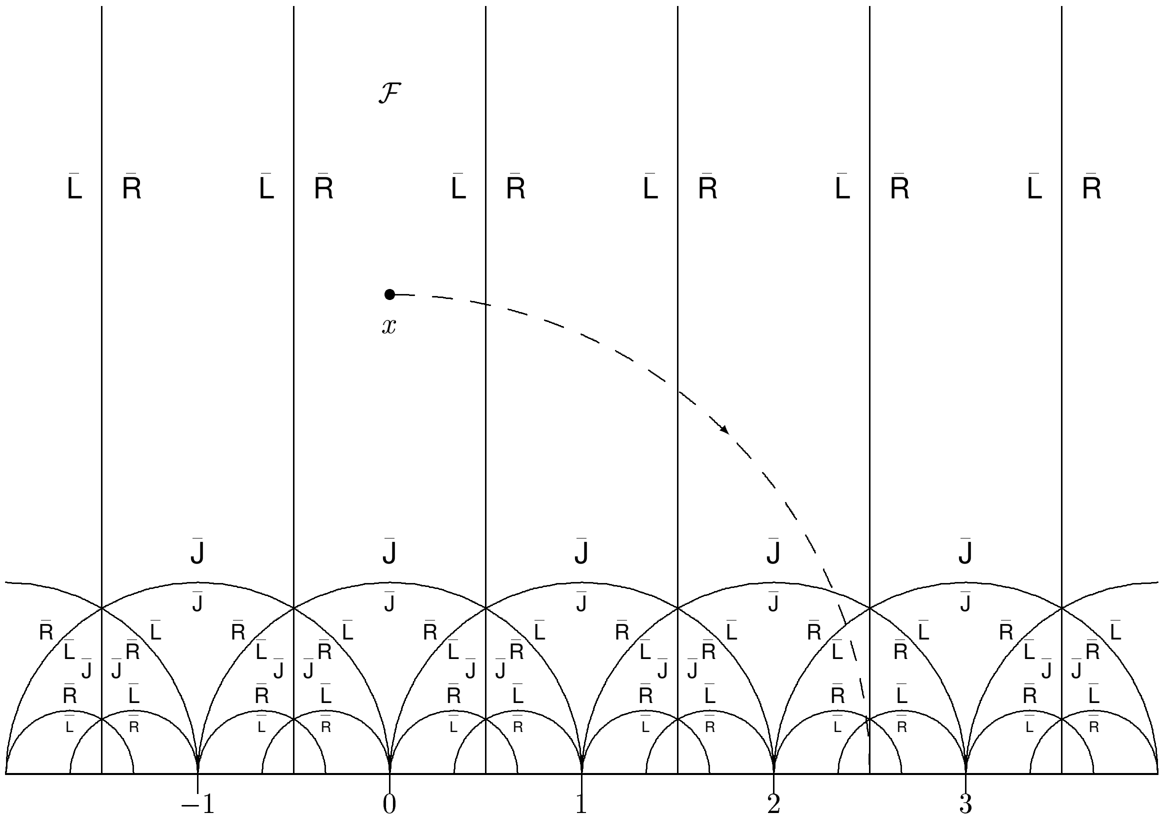


Figure 2.3: Labeled tessellations of $PSL(2,\mathbb{Z})$. The pictured geodesic, starting from x, has cutting sequence beginning $\bar{\mathsf{R}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}, \bar{\mathsf{L}}, \bar{\mathsf{L}}$.

These matrices represent the $PSL(2,\mathbb{Z})$-motions needed to move a neighboring fundamental domain $\mathcal{F}'$ in $\mathfrak{H}$ to $\mathcal{F}$. These neighboring fundamental domains are labeled with the appropriate symbol in Figure 2.2.

The action of $PSL(2,\mathbb{Z})$ tiles $\mathfrak{H}$ with copies of $\mathcal{F}$, and the boundaries of all tiles fit together to cover a countable collection of geodesics of $\mathfrak{H}$, which are exactly those geodesics having two rational endpoints $\langle \frac{a}{b}, \frac{c}{d} \rangle$ such that

$$\det \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \pm 2 \ , \tag{2.9}$$

where we adopt the convention that the cusp $i\infty$ is the rational $\frac{1}{0}$. The resulting labeling of the tiles with the labels $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\}$ is indicated in Figure 2.3.

The *cutting sequence shift* $\Sigma_{\mathcal{F}} := \Sigma_{\mathcal{F},PSL(2,\mathbb{Z})}$ is defined by the method of section 2.1, and is a closed subshift of the shift on three letters $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\}$. For this special case, we present some additional cutting sequence labelings that keep track of corners, which we call *generalized cutting sequences.* We label the two finite corners of $\mathcal{F}$ symbolically by

$$\bar{\mathsf{C}}_1 = \bar{\mathsf{J}}\bar{\mathsf{R}}\bar{\mathsf{J}} = \bar{\mathsf{L}}\bar{\mathsf{J}}\bar{\mathsf{L}} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix} \qquad \Leftrightarrow \qquad \mathcal{S}_{\mathsf{C}}^{-} := z = \frac{1}{2} + i\frac{\sqrt{3}}{2} \tag{2.10}$$

$$\bar{\mathsf{C}}_2 = \bar{\mathsf{J}}\bar{\mathsf{L}}\bar{\mathsf{J}} = \bar{\mathsf{R}}\bar{\mathsf{J}}\bar{\mathsf{R}} = \begin{bmatrix} -1 & 0 \\ -1 & -1 \end{bmatrix} \qquad \Leftrightarrow \qquad \mathcal{S}_{\mathsf{C}}^{+} := z = -\frac{1}{2} + i\frac{\sqrt{3}}{2} \ . \tag{2.11}$$

General geodesics on the modular surface $\mathfrak{H}/PSL(2,\mathbb{Z})$ are given by geodesics on $\mathfrak{H}$ that pass through the fundamental domain. A representative set of lifts covering every geodesic in the modular surface consists of all vertical geodesics with $-\frac{1}{2} < \theta < \frac{1}{2}$, plus all semicircular geodesics in $\mathfrak{H}$ whose maximum point lies in the interior of the fundamental domain. The latter must necessarily hit the vertical line $\{it : t > 0\}$ and have endpoints (θ_1, θ_2) satisfying

$$|\theta_2 - \theta_1| > \sqrt{3}, \text{ and } -1 \leq \theta_1 + \theta_2 < 1.$$

These conditions imply that $\theta_1\theta_2 < 0$.

A *generalized cutting sequence* for a general geodesic γ is the sequence of symbols from $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}, \bar{\mathsf{C}}_1, \bar{\mathsf{C}}_2\}$ that describe the successive sides of the images of the fundamental domains that it hits.

Finally, for the domain $\mathcal{F}$ we define a set of *vertical cutting sequences* which are one-sided cutting sequences associated to the set of oriented geodesics of $\mathfrak{H}$ that emanate from the cusp $\{i\infty\}$ of $\mathcal{F}$ and pass through $\mathcal{F}$, which are exactly the set of vertical geodesics $\{(\infty, \theta) : -\frac{1}{2} < \theta < \frac{1}{2}\}$.

Definition 2.5. The *irrational vertical cutting sequence set* $\Pi_{\mathcal{F}}^0$ consists of cutting sequences of those vertical geodesics $\langle\infty, \theta\rangle$ for irrational θ with $-\frac{1}{2} < \theta < \frac{1}{2}$.

An important result for our analysis is the following simple fact.

Lemma 2.2. *An irrational vertical geodesic cannot hit any finite corner of a $PSL(2,\mathbb{Z})$-translate of $\mathcal{F}$.*

Proof. The generators of $PSL(2,\mathbb{Z})$ acting on the upper half-plane are $z \to z + 1$ and $z \to -1/z$. Both generators preserve the field $\mathbb{Q}(\sqrt{-3})$, in which every element has rational real part. The finite corners $\pm 1/2 + \sqrt{-3}/2$, and thus all of their images, are in this field. ■

Thus the cutting sequences in $\Pi_{\mathcal{F}}^0$ are one-sided infinite sequences which contain no symbol $\bar{\mathsf{C}}_1$ or $\bar{\mathsf{C}}_2$, hence they are all of the form:

$$C^+(\theta) := (g_0, g_1, g_2, \cdots) \quad \text{all} \quad g_i \in N(\mathcal{F}) = \{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\} \ . \tag{2.12}$$

The expansion $C^+(\theta)$ is a kind of continued fraction of θ, and we call it the *cutting sequence expansion* of θ. Note that, for a given point $\theta + it$ corresponding to a fundamental domain $h_i\mathcal{F}$, that there is some point $z \in \mathcal{F}$ such that

$$g_0 g_1 \cdots g_{i-2} g_{i-1}(z) = \theta + it, \tag{2.13}$$

according to Lemma 2.1.

Definition 2.6. The *vertical cutting sequence space* $\Pi_{\mathcal{F}}$ is the closure of $\Pi_{\mathcal{F}}^0$ in the one-sided sequence topology on the one-sided shift space on three symbols $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\}$.

Note that the set $\Pi_{\mathcal{F}}^0$ is not closed under the one-sided shift operator σ^+, because the first symbol of every element of $\Pi_{\mathcal{F}}^0$ is J.

We can now define cutting sequences $C(\theta)$ for rational θ with $-\frac{1}{2} < \theta < \frac{1}{2}$ in $\Pi_{\mathcal{F}}$ as limits of cutting sequences $C(\theta_i)$ of irrational $\theta_i \to \theta$, as described earlier; in this case $C(\theta)$ is a set of one-sided infinite cutting sequences.

We define *generalized cutting sequence expansions* for vertical geodesics, drawn from the alphabet $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}, \bar{\mathsf{C}}_1, \bar{\mathsf{C}}_2\}$, similarly to the case of general geodesics.

3. Continued Fraction Expansions and Cutting Sequence Expansions

This section describes symbolic dynamics for vertical cutting sequences associated to additive continued fraction (ACF) expansions, and for Minkowski geodesic continued fraction expansions (MGCF) for vertical geodesics. It shows that the cutting sequence expansion and MGCF expansions for a geodesic can each be computed from the other using a finite automaton.

3.1. Additive Continued Fraction Expansions

The ordinary continued fraction (OCF) expansion for a real θ, written $\theta = [a_0, a_1, a_2, \cdots]$, can be represented using 2×2 matrices. The use of such matrices to represent continued fractions appears in Frame [12] and Kolden [27], and is described in Stark [49]. Namely, one has

$$\begin{bmatrix} a_0 & 1 \\ 1 & 0 \end{bmatrix} \cdots \begin{bmatrix} a_n & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{bmatrix}, \tag{3.1}$$

where $\frac{p_n}{q_n} = [a_0, a_1, \ldots, a_n]$ is the n-th convergent of θ. The ordinary continued fraction expansion for $\theta > 0$ has a symbolic dynamics which is a one-sided shift on infinitely many symbols, using the alphabet

$$\mathsf{L}_k = \begin{bmatrix} k & 1 \\ 1 & 0 \end{bmatrix}, \; k \geq 0. \tag{3.2}$$

The *additive continued fraction* (ACF) expansion of $\theta > 0$ is a symbolic expansion using the two symbols

$$\mathsf{R} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathsf{F} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{3.3}$$

which is obtained from (3.2) by expanding the symbol L_k as

$$\begin{bmatrix} k & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^k \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{3.4}$$

The additive continued fraction expansion of $\theta > 0$ is thus

$$\mathsf{F}\mathsf{R}^{a_1}\mathsf{F}\mathsf{R}^{a_2}\mathsf{F}\mathsf{R}^{a_3}\cdots, \tag{3.5}$$

regarded as a string of letters in the alphabet $\{\mathsf{R}, \mathsf{F}\}$. The finite truncations of this expression are 2×2 matrices that encode the convergents and intermediate convergents of θ.

We next construct a related continued fraction for real $\theta > 0$ which uses the symbols

$$\mathsf{R} := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathsf{D} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \tag{3.6}$$

The *Farey tree expansion* of a real number $\theta > 0$ that has OCF expansion $\theta = [a_0, a_1, a_2, \ldots]$ is:

$$\mathsf{R}^{a_0}\mathsf{D}^{a_1}\mathsf{R}^{a_2}\mathsf{D}^{a_3}\cdots \tag{3.7}$$

regarded as a sequence of letters in the alphabet $\{\mathsf{R}, \mathsf{D}\}$. This sequence encodes the steps of subtraction needed in the division process to encode the ordinary continued fraction of θ; it also essentially gives all the intermediate convergents to θ, cf. Richards [42], Theorem 2.1. For

$\theta > 0$ the set of allowable symbol sequences for the additive expansion is the full one-sided shift on two letters $\{\mathsf{R}, \mathsf{D}\}$. The association to the Farey tree is described in Lagarias [29].

The symbolic expansions (3.5) and (3.7) are equivalent in the sense that each is convertible to the other using a finite automaton; see section 3.5.

3.2. Cutting Sequences and Minkowski lattice basis reduction

We now develop a correspondence between the cutting sequence expansion of the geodesic $\theta + it$ for $-\frac{1}{2} < \theta < \frac{1}{2}$ and the construction of a Minkowski-reduced lattice basis for the parametrized series of lattice bases

$$\mathsf{B}_t(\theta) = \begin{bmatrix} 1 & 0 \\ -\theta & t \end{bmatrix} \tag{3.8}$$

for the parametrized family of lattices $\Lambda_t = \mathbb{Z}[(1,0), (-\theta, t)]$.

To demonstrate the equivalence, we first transform lattice bases in $GL(2,\mathbb{R})$ to the cone of positive definite symmetric matrices $\mathcal{P}_2$ (equivalently, to positive definite quadratic forms), and then we relate these to elements of $\mathfrak{H}$.

Definition 3.1. A lattice basis $\mathbf{v}_1, \mathbf{v}_2$ is *Minkowski-reduced* if $\mathbf{v}_1$ is a shortest vector in the lattice, and $\mathbf{v}_2$ is a shortest vector not a multiple of $\mathbf{v}_2$.

For each t, there is a matrix $\mathsf{P} = \mathsf{P}_t$ (possibly several) for which $\mathsf{PB}_t(\theta)$ is Minkowski-reduced. General properties of Minkowski-reduction can be found in Cassels [9] and Gruber and Lekkerkerker [14, p. 149ff].

A basis $\mathsf{B} \in GL(2,\mathbb{R})$ corresponds to the positive definite symmetric matrix $\mathsf{M} = \mathsf{BB}^T \in \mathcal{P}_2$ with associated quadratic form

$$Q(x_1, x_2) = [x_1 x_2]\mathsf{M} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = m_{11}x_1^2 + (m_{12} + m_{21})x_1x_2 + m_{22}x_2^2 \ . \tag{3.9}$$

For the basis matrix $\mathsf{B}_t(\boldsymbol{\theta})$ in (3.8) this gives

$$\mathsf{M} := \mathsf{M}_t(\theta) = \begin{bmatrix} 1 & -\theta \\ -\theta & \theta^2 + t^2 \end{bmatrix} \ . \tag{3.10}$$

with associated quadratic form

$$Q_t(x_1, x_2) = (x_1 - \theta x_2)^2 + t^2 x_2^2 = (x_1 - (\theta + it)y_1)(x_1 - (\theta - it)x_2) \ . \tag{3.11}$$

The action of $\mathsf{P} \in GL(2,\mathbb{Z})$ which sends the matrix $\mathsf{B}_t(\theta)$ to $\mathsf{PB}_t(\boldsymbol{\theta})$ is transformed to the $GL(2,\mathbb{Z})$-action on $\mathcal{P}_2$ that takes M to PMP^T, and this takes the associated quadratic form $Q(x_1, x_2)$ to

$$Q'(x_1, x_2) = Q(p_{11}x_1 + p_{21}x_2,\ p_{21}x_1 + p_{22}x_2) \ . \tag{3.12}$$

The Minkowski reduction domain $\bar{\mathcal{M}}$ for $GL(2,\mathbb{Z})\backslash\mathcal{P}_2$ (see Cassels [9]) is the set of quadratic forms (3.9) such that

$$\begin{aligned} Q(1,0) &\le Q(0,1) \ , \\ Q(0,1) &\le Q(1,1) \ , \\ Q(0,1) &\le Q(-1,1) \ . \end{aligned} \tag{3.13}$$

This is equivalent to the conditions

$$|m_{12} + m_{21}| \leq m_{11} \leq m_{22} \ . \tag{3.14}$$

on the form Q in (3.9). The subgroup

$$H_2 := \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

of $GL(2,\mathbb{Z})$ maps the domain $\bar{\mathcal{M}}$ into itself, hence we actually let $H_2\backslash GL(2,\mathbb{Z})$ act on $\mathcal{P}_2$. By definition the Minkowski reduction domain $\mathcal{M}$ of $GL(2,\mathbb{Z})\backslash GL(2,\mathbb{R})$, is the preimage of $\bar{\mathcal{M}}$ under the mapping $\bar{\mathsf{B}} \mapsto \mathsf{B}\mathsf{B}^T$, and the group actions by $GL(2,\mathbb{Z})$ coincide. An algorithm for choosing Minkowski-reduced lattice bases will choose one element of $GL(2,\mathbb{Z})$ out of the four-element coset of $H_2\backslash GL(2,\mathbb{Z})$. A natural canonical choice is to require $\det \mathsf{P} = 1$; this still leaves two choices for P, but multiplying P by -1 will not change the quadratic form and can be done arbitrarily. A different choice will be made by the Minkowski geodesic continued fraction of the next section.

For the second transformation from $\mathcal{P}$ to $\mathfrak{H}$, note that a positive definite quadratic form can be uniquely written

$$Q(x,y) = a(x_1 - \theta x_2)(x_1 - \bar{\theta} x_2) \ , \tag{3.15}$$

with θ and $\bar{\theta}$ complex conjugates and $Im(\theta) > 0$. We map $\mathcal{P}$ to $\mathfrak{H}$ by sending the quadratic form Q to the unordered pair of roots $\{\theta, \bar{\theta}\}$ and then identify this with $\theta \in \mathfrak{H} = \{z : Im(z) > 0\}$. If $\mathsf{P} \in GL(2,\mathbb{Z})$ sends Q to Q' given by (3.14), then a calculation yields

$$Q' = a'(x_1 - \phi x_2)(x_1 - \bar{\phi} x_2) \ ,$$

with $a' = Q(p_{11}, p_{12})$ and with a root ϕ given by

$$\phi = (\mathsf{P}^T)^{-1}\theta = \begin{bmatrix} p_{22} & -p_{21} \\ -p_{12} & p_{11} \end{bmatrix} (\theta) \ , \tag{3.16}$$

where the action of the matrix is a linear fractional transformation on $\mathbb{C}$. If $\det(\mathsf{P}) = 1$, then $\phi \in \mathfrak{H}$, while if $\det(\mathsf{P}) = -1$ then $Im(\phi) < 0$ is the complex conjugate of the root we want, i.e. it reverses the ordering of the roots. Now each four element coset of $H_2\backslash GL(2,\mathbb{Z})$ contains two elements in $SL(2,\mathbb{Z})$ which form a coset of $PSL(2,\mathbb{Z}) = \{\pm \mathsf{I}\}\backslash SL(2,\mathbb{Z})$, so the action (3.16) of $H_2\backslash GL(2,\mathbb{Z})$ on unordered pairs $\{\theta, \bar{\theta}\}$ is equivalent to the standard $PSL(2,\mathbb{Z})$-action on $\mathfrak{H}$. Finally, the conditions (3.13) for a form Q to be Minkowski reduced translate via (3.15) to

$$\theta\bar{\theta} \geq 1, \quad -1 \geq \theta + \bar{\theta} \geq 1 \ ,$$

which is exactly the fundamental domain $\mathcal{F}$ of $PSL(2,\mathbb{Z})$.

The mapping from $GL(2,\mathbb{R})$ to $\mathfrak{H}$ obtained by composing these two transformations sends the set of matrices $\{\mathsf{B}_t(\theta) : t > 0\}$ in $GL(2,\mathbb{R})$ to the geodesic $\{\theta + it : t > 0\}$ in $\mathfrak{H}$. Since the $H_2\backslash GL(2,\mathbb{Z})$ action on $GL(2,\mathbb{R})$ is equivalent to the $PSL(2,\mathbb{Z})$ action on $\mathfrak{H}$ and the reduction domains correspond, the sequence of Minkowski-reduced lattice bases for θ is essentially the same as the cutting sequence expansion for θ. In the reverse direction, $\theta \in \mathfrak{H}$ determines a positive ray $\{aQ : a > 0\}$ in the cone $\mathcal{P}_2$, i.e. an element of $\mathcal{P}_2/\mathbb{R}^+$, and this in term determines a family $\{aBQ : a \in \mathbb{R}^+, Q \in O(n,\mathbb{R})\}$ in $GL(2,\mathbb{R})$, i.e. an element of $GL(2,\mathbb{R})/\mathbb{R}^+O(2,\mathbb{R})$. This does not affect the symbolic dynamics because the Minkowski domain $\bar{\mathcal{M}}$ is invariant under the $\mathbb{R}^+$-action, and because the Minkowski reduction domain $\mathcal{M}$ is invariant under the $R^*O(2,\mathbb{R})$-action.

3.3. Minkowski Geodesic Continued Fraction Expansions

The Minkowski geodesic multidimensional continued fraction (MGCF) expansion is introduced in Lagarias [29]. We consider here the one-dimensional case.

This is a specific algorithm for Minkowski-reduction of the lattice bases

$$\mathsf{B}_t(\theta) = \begin{bmatrix} 1 & 0 \\ -\theta & t \end{bmatrix} \tag{3.17}$$

There exists a sequence of *critical values*

$$\infty = t_0 > t_1 > t_2 > \cdots$$

with $\lim_{n\to\infty} t_n = 0$, and an associated sequence of *convergent matrices* $\{\mathsf{P}^{(n)} : n = 0, 1, 2, \cdots\}$ in $GL(2, \mathbb{Z})$ such that $\mathsf{P}^{(n)}\mathsf{B}_t(\theta)$ is a Minkowski-reduced lattice basis of the lattice when $t_n > t > t_{n+1}$. We write

$$\mathsf{P}^{(n)} := \begin{bmatrix} p_1^{(n)} & q_1^{(n)} \\ p_2^{(n)} & q_2^{(n)} \end{bmatrix} , \tag{3.18}$$

so that

$$\mathsf{P}^{(n)}\mathsf{B}_t(\theta) = \begin{bmatrix} p_1^{(n)} - q_1^{(n)}\theta & q_1^{(n)}t \\ p_2^{(n)} - q_2^{(n)}\theta & q_2^{(n)}t \end{bmatrix} . \tag{3.19}$$

This motivates the name "continued fraction"; as t goes to zero, the approximations $p_i^{(n)} - q_i^{(n)}\theta$ must also go to zero, and thus the p_i/q_i play the role of convergents.

Changing the sign of one or both basis vectors does not affect Minkowski-reduction, so we have four choices of $\mathsf{P}^{(n)}$. For defining the MGCF as a continued fraction, it is natural to require positive denominators; that is, we require that $q_i^{(n)} \geq 0$ for all $n \geq 1$, and if $q_i^{(n)} = 0$ then $p_i^{(n)} > 0$. For lattice basis reduction, it is natural to require $\det \mathsf{P}^{(n)} = 1$; multiplying by $\pm\mathsf{I}$ is still possible but will not affect the process.

In either case, rhe associated *partial quotient matrices* $\mathsf{A}^{(n)}$ are defined by

$$\mathsf{P}^{(n)} = \mathsf{A}^{(n)}\mathsf{P}^{(n-1)} ,$$

so that

$$\mathsf{P}^{(n)} = \mathsf{A}^{(n)}\mathsf{A}^{(n-1)} \cdots \mathsf{A}^{(1)}\mathsf{P}^{(0)} . \tag{3.20}$$

Here $\mathsf{P}^{(0)} = \mathsf{I}$ is the identity matrix if $-\frac{1}{2} \leq \theta < \frac{1}{2}$, since the lattice basis $\{(1,0), (-\theta, t)\}$ is Minkowski-reduced for large t.

We now show that the partial quotient matrices $\mathsf{A}^{(n)}$ are drawn from a finite set.

Lemma 3.1 (Minkowski partial quotient set) . *The allowed partial quotients for the one-dimensional Minkowski geodesic continued fraction algorithm are*

$$\left\{ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\} . \tag{3.21}$$

Proof. Let $\mathsf{M} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}$ be a 2×2 matrix with $\det(\mathsf{M}) > 0$, and associate to it the positive definite quadratic form with coefficient matrix

$$\mathsf{MM}^T = \begin{bmatrix} \|\mathbf{v}_1\|^2 & \langle \mathbf{v}_1, \mathbf{v}_2 \rangle \\ \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \|\mathbf{v}_2\|^2 \end{bmatrix}. \tag{3.22}$$

The Minkowski-reduction conditions for the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T(\mathsf{MM}^T)\mathbf{x}$ are

$$\begin{aligned} \|\mathbf{v}_1\| &\leq \|\mathbf{v}_2\| & (3.23)\\ \|\mathbf{v}_2\| &\leq \|\mathbf{v}_1 + \mathbf{v}_2\| & (3.24)\\ \|\mathbf{v}_2\| &\leq \|\mathbf{v}_1 - \mathbf{v}_2\|, & (3.25) \end{aligned}$$

see Cassels [9, p. 257].

We must now choose our algorithm; we will describe both the natural algorithm for Minkowski lattice basis reduction and the Minkowski geodesic continued fraction algorithm,

For lattice basis reduction, it is natural to apply a new matrix $\mathsf{A}^{(n)}$ of determinant 1. At any critical time t_n, all of the above inequalities are satisfied with the current matrix $\mathsf{P}^{(n-1)}$ for $t_{n-1} > t > t_n$, and at least one holds with equality at $t = t_n$,

We first consider "generic" convergents which occur when exactly two of $\|\mathbf{v}_1(t)\|$, $\|\mathbf{v}_2(t)\|$, $\|\mathbf{v}_1(t) + \mathbf{v}_2(t)\|$ and $\|\mathbf{v}_1(t) - \mathbf{v}_2(t)\|$ become equal at $t = t_n$.

If at $t = t_n$ we have $\|\mathbf{v}_1(t)\| = \|\mathbf{v}_2(t)\|$ then the partial quotient matrix is $\tilde{\mathsf{J}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$; it exchanges $\mathbf{v}_1$ and $\mathbf{v}_2$ and changes the sign of the new $\mathbf{v}_2$ to keep determinant 1. If $\|\mathbf{v}_2(t)\| = \|\mathbf{v}_1(t) + \mathbf{v}_2(t)\|$ then the partial quotient matrix is $\tilde{\mathsf{R}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$; it replaces $\mathbf{v}_2$ with $\mathbf{v}_1 + \mathbf{v}_2$. If $\|\mathbf{v}_2(t)\| = \|\mathbf{v}_1(t) - \mathbf{v}_2(t)\|$ then the partial quotient matrix is $\tilde{\mathsf{L}} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$; it replaces $\mathbf{v}_2$ with $\mathbf{v}_2 - \mathbf{v}_1$.

We might also have more than one of the inequalities (3.23)–(3.25) holding with equality at the same t. If $\|\mathbf{v}_1 + \mathbf{v}_2\| = \|\mathbf{v}_1 - \mathbf{v}_2\|$, then neither one can be equal to $\|\mathbf{v}_1\|$ unless $\mathbf{v}_2 = 0$, nor to $\|\mathbf{v}_2\|$ unless $\mathbf{v}_1 = 0$. We could have $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \|\mathbf{v}_1 + \mathbf{v}_2\|$. In that case, for $t < t_n$, we would have $\|\mathbf{v}_1(t)\| > \|\mathbf{v}_2(t)\| > \|\mathbf{v}_1(t) + \mathbf{v}_2(t)\|$, and to correct these inequalities, we must replace $\mathbf{v}_1$ by $\mathbf{v}_1 + \mathbf{v}_2$; we can take partial quotient matrix $\tilde{\mathsf{C}}_1 = \tilde{\mathsf{J}}\tilde{\mathsf{L}}\tilde{\mathsf{J}} = \begin{bmatrix} -1 & -1 \\ 0 & -1 \end{bmatrix}$ by also changing signs of both vectors Similarly, if $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \|\mathbf{v}_1 - \mathbf{v}_2\|$, we get partial quotient matrix $\tilde{\mathsf{C}}_2 = \tilde{\mathsf{J}}\tilde{\mathsf{R}}\tilde{\mathsf{J}} = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$.

These five matrices (generated from just three) give the natural algorithm for Minkowski lattice basis reduction.

For the Minkowski geodesic continued fraction, we follow a similar process except that we choose signs to keep the q_i positive. We use the fact that all vectors in the lattice $\Lambda_t(\theta)$ have the special form

$$\mathbf{v}_i = \mathbf{v}_i(t) := (p_i - q_i\theta, q_i t) \text{ with } q_i \geq 0. \tag{3.26}$$

Whenever a critical value t_n occurs with $\|(p - q\theta, qt_n)\| = \|(p' - q'\theta, q't_c)\|$, then the inequality

$$\|(p' - q'\theta, q't)\| > \|(p - q\theta, qt)\| \quad \text{for} \quad 0 < t < t_c$$

holds exactly when $|q'| > |q|$. Thus when the shortest vector in the basis $\mathsf{P}^{(n-1)}$ is replaced by a vector in $\mathsf{P}^{(n)}$ the associated denominator must increase.

Again, we first consider "generic" convergents.

Suppose first that $q_1 < q_2$. If at $t = t_n$ we have $\|\mathbf{v}_1(t)\| = \|\mathbf{v}_2(t)\|$ then the partial quotient is $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$; it exchanges $\mathbf{v}_1$ and $\mathbf{v}_2$. If $\|\mathbf{v}_2(t)\| = \|\mathbf{v}_1(t) + \mathbf{v}_2(t)\|$ then the partial quotient is $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$; it replaces $\mathbf{v}_2$ with $\mathbf{v}_1 + \mathbf{v}_2$. The case $\|\mathbf{v}_2(t)\| = \|\mathbf{v}_1(t) - \mathbf{v}_2(t)\|$ cannot occur.

Suppose next that $q_1 \geq q_2$. At $t = t_n$ we cannot have $\|\mathbf{v}_1(t)\| = \|\mathbf{v}_2(t)\|$ since an exchange of $\mathbf{v}_1$ and $\mathbf{v}_2$ would not increase the denominator. We may have $\|\mathbf{v}_1(t) + \mathbf{v}_2(t)\| = \|\mathbf{v}_2(t)\|$, giving the partial quotient $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. If $q_1 > 2q_2$ then $q' = q_1 - q_2 > q_2$ hence the partial quotient $\begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$ is possible; it replaces $\mathbf{v}_2$ with $\mathbf{v}_1 - \mathbf{v}_2$.

There remain situations in which three or more of $\|\mathbf{v}_1(t)\|$, $\|\mathbf{v}_2(t)\|$, $\|\mathbf{v}_1(t) + \mathbf{v}_2(t)\|$ and $\|\mathbf{v}_1(t) - \mathbf{v}_2(t)\|$ simultaneously becoming equal at $t = t_c$. Only one case is possible; it is

$$\|\mathbf{v}_1(t)\| = \|\mathbf{v}_2(t)\| = \|\mathbf{v}_1(t) + \mathbf{v}_2(t)\| ,$$

which can only occur when $q_1 < q_2$. In this case the partial quotient is $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. The other case consistent with an increasing denominator is $\|\mathbf{v}_2(t)\| = \|\mathbf{v}_1(t) + \mathbf{v}_2(t)\| = \|\mathbf{v}_1(t) - \mathbf{v}_2(t)\|$, but this implies $\|\mathbf{v}_1(t)\| = 0$, which is impossible unless $t = 0$. ∎

We define the *Minkowski geodesic continued fraction expansion* of any real θ satisfying $-\frac{1}{2} < \theta < \frac{1}{2}$ to be the symbol sequence

$$(\mathsf{A}^{(1)}, \mathsf{A}^{(2)}, \mathsf{A}^{(3)}, \dots) . \tag{3.27}$$

We write this sequence left-to-right, although the matrix product in (3.20) runs right-to-left. The associated *Minkowski geodesic symbol set* is

$$\mathsf{L} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}, \mathsf{R} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \mathsf{J} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \mathsf{C} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} . \tag{3.28}$$

The symbol C occurs only for "exceptional" geodesics, and we will usually be concerned with symbolic expansions drawn from the smaller symbol set $\{\mathsf{L}, \mathsf{R}, \mathsf{J}\}$.

The Minkowski geodesic continued fraction expansion corresponding to a general geodesic on $\mathcal{F}$ is the expansion attached to Minkowski reduction of the parametrized lattice bases

$$\mathsf{B}_t(\theta_1, \theta_2) = \begin{bmatrix} 1 & (\theta_1)^{-1} t \\ -\theta_2 & t \end{bmatrix} .$$

The associated quadratic form is

$$Q_t(x, y) = (x - \theta_2 y)^2 + t^2 \left(\frac{1}{\theta_1} x - y \right)^2 , \tag{3.29}$$

and the partial quotients are obtained by the same formula.

3.4. Correspondence between the continued fraction and cutting sequence

We will now develop a correspondence between the three different symbol sequences for θ, obtained by cutting sequences, the natural Minkowski basis reduction algorithm, and the Minkowski geodesic continued fraction algorithm.

The precise correspondence between the symbol sequences involves specifying the relation among the four elements of the H_2-coset in $H_2\backslash GL(2,\mathbb{Z})$ for the Minkowski geodesic continued fraction expansion and the two elements in the $\{\pm\mathsf{I}\}$-coset in $PSL(2,\mathbb{Z})$ for the cutting sequence expansion.

Definition 3.2. The *parity* of a word $\mathsf{W} = \mathsf{S}_1 \dots \mathsf{S}_n$ in the alphabet $\{\mathsf{L},\mathsf{R},\mathsf{J}\}$ is even or odd according to whether there are an even or odd number of L and J. That is,

$$\det(\mathsf{W}) = (-1)^{\text{parity}(\mathsf{W})}. \tag{3.30}$$

We obtain:

Theorem 3.1. *For irrational θ with $-\frac{1}{2} < \theta < \frac{1}{2}$, the one-sided cutting sequence expansion of the vertical geodesic $\langle \infty, \theta\rangle = \{\theta + it : t > 0\}$ in $\Pi^0_{\mathcal{F}}$ is obtained from the Minkowski geodesic continued fraction expansion of θ by the following procedure: If the current initial word of the MGCF expansion has even parity, on the next symbol make the letter replacement $\mathsf{L} \to \bar{\mathsf{L}}, \mathsf{R} \to \bar{\mathsf{R}}$ and $\mathsf{J} \to \bar{\mathsf{J}}$; if it has odd parity, make the letter replacement $\mathsf{L} \to \bar{\mathsf{R}}, \mathsf{R} \to \bar{\mathsf{L}}$ and $\mathsf{J} \to \bar{\mathsf{J}}$. The MGCF can be obtained from the cutting sequence by the reverse process, in which the parity is determined by the current symbols of the MGCF expansion.*

Proof. By Lemma 2.2, the assumption of irrational θ ensures that the MGCF expansions and cutting sequence expansions of θ are both infinite and never use a symbol C. The MGCF expansion $\cdots \mathsf{A}^{(3)}\mathsf{A}^{(2)}\mathsf{A}^{(1)}$ uses the symbols

$$\mathsf{L} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}, \mathsf{R} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \ \mathsf{J} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

drawn from $GL(2,\mathbb{Z})$ and moves right to left, while the cutting sequence expansion uses the symbols

$$\bar{\mathsf{L}} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \ \bar{\mathsf{R}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \bar{\mathsf{J}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

drawn from $SL(2,\mathbb{Z})$ and moves left to right by Lemma 2.1. For notational convenience, we introduce the matrix

$$\mathsf{K} := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

so H_2 is $\{\mathsf{I}, \mathsf{K}, -\mathsf{K}, -\mathsf{I}\}$. We first arrange for matrices to multiply in the same direction. The matrices in the MGCF are chosen so that $\mathsf{P}^{(n)}B_t(\theta) \in \mathcal{F}$ for a given t, while the matrices in the cutting sequence are chosen so that the cutting sequence product $h_n = \mathsf{S}_1 \dots \mathsf{S}_n$ has $\gamma(t) \in h_n\mathcal{F}$. Therefore, we have $h_n = (\mathsf{P}^{(n)})^{-1}$. We expand $(\mathsf{P}^{(n)})^{-1}$ in terms of symbols $\mathsf{L},\mathsf{R},\mathsf{J},\mathsf{K}$ using the relations $\mathsf{L}^{-1} = \mathsf{L}$, $\mathsf{R}^{-1} = \mathsf{KRK}$ and $\mathsf{J}^{-1} = \mathsf{J}$ to obtain an expansion which proceeds left to right. To convert this expansion to the cutting sequence expansion, we must convert to symbols $\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}$ and remove the symbols K, which encode the parity. We first convert to the symbols of the natural Minkowski basis reduction algorithm of Section 3.3.

$$\tilde{\mathsf{R}} := \mathsf{R}, \quad \tilde{\mathsf{L}} := \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad \tilde{\mathsf{J}} := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

in $SL(2, \mathbb{Z})$. To convert a product of matrices from the form $\mathsf{L}, \mathsf{R}, \mathsf{J}$ to $\tilde{\mathsf{L}}, \tilde{\mathsf{R}}, \tilde{\mathsf{J}}$, starting from the right end of the product, we use the relations $\mathsf{LK} = \tilde{\mathsf{R}}$, $\mathsf{RK} = \mathsf{K}\tilde{\mathsf{L}}$, and $\mathsf{JK} = -\tilde{\mathsf{J}}$. In doing this we pick up or lose a factor of K whenever we encounter a matrix L or J and this multiplies the determinant by -1; meanwhile, every R in the MGCF becomes $\mathsf{R}^{-1} = \mathsf{KRK}$ when inverted, and thus is itself encoded with the opposite parity but leaves the parity unchanged for the next matrix. (We sometimes pick up a matrix factor of $-\mathsf{I}$, but this commutes with everything and may be ignored.) Similarly, to reverse this, we use $\tilde{\mathsf{L}} = \mathsf{KL}$ and $\tilde{\mathsf{J}} = \mathsf{KJ}$.

To convert from the natural Minkowski basis reduction algorithm symbols $\tilde{\mathsf{L}}, \tilde{\mathsf{R}}, \tilde{\mathsf{J}}$ to the cutting sequence symbols $\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}$, we use the relations $(\tilde{\mathsf{L}}^T)^{-1} = \bar{\mathsf{R}}$, $(\tilde{\mathsf{R}}^T)^{-1} = \bar{\mathsf{L}}$ and $(\tilde{\mathsf{J}}^T)^{-1} = \bar{\mathsf{J}}$. Thus the conversion from $\tilde{\mathsf{L}}, \tilde{\mathsf{R}}, \tilde{\mathsf{J}}$ to $\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}$ interchanges R with L. ■

3.5. Finite Automata

By a *finite automaton* we mean a deterministic finite-state automaton, as defined in Hopcroft and Ullman [20]. We consider finite automata used as transducers, as in Raney [41], in which each transition edge of the automaton specifies the printing of a specific finite string of output letters, possibly empty. The automaton starts in an initial state, and the input sequence is presented as a sequence of letters, and each letter determines the transition edge to the next state of the automaton.

In this paper we present a number of results asserting the existence of finite automata to convert one-sided infinite symbol sequences of one form to another form. In the proofs we only indicate the "finite-state" character of the conversion process, and generally omit details of the (routine but sometimes involved) construction of the automaton.

As a simple example, the discussion in section 3.1 yields:

Theorem 3.2. *For real $\theta > 1$, the additive continued fraction expansion of θ can be converted to the Farey tree expansion of θ by a finite automaton, and vice versa.*

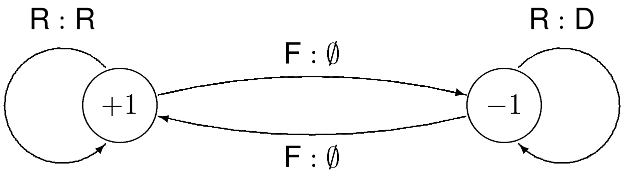
Proof. To convert from the Farey shift expansion (3.7) to the additive continued fraction expansion (3.5), we use $\mathsf{D} = \mathsf{FRF}$ and $\mathsf{F}^2 = \mathsf{I}$.

A finite automaton which converts the additive continued fraction expansion (3.5) to the Farey tree expansion (3.7) must keep two states to keep track of the sign ± 1 of $\det(\mathsf{S}_1 \cdots \mathsf{S}_m)$ of the symbols $\mathsf{S}_i = \mathsf{F}$ or R examined so far. The initial state is $+1$. ■

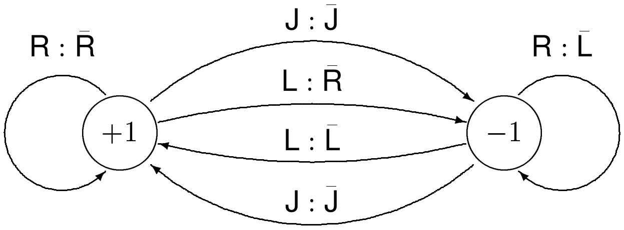
The discussion in section 3.4 yields:

Theorem 3.3. *For irrational θ with $-\frac{1}{2} < \theta < \frac{1}{2}$ there exists a finite automaton to convert the Minkowski geodesic continued fraction expansion of θ to that of the vertical cutting sequence expansion of $\langle \infty, \theta \rangle$ and vice-versa.*

Proof. This follows from Theorem 3.1. For each direction, the finite automaton constructed needs two states, to keep track of whether there a factor of K present; i.e., to keep track of the sign of $\det(\mathsf{S}_1 \cdots \mathsf{S}_n)$. The initial state is $+1$. ■



(a) Additive continued fraction to Farey tree



(b) MGCF to cutting sequence

Figure 3.1: Finite Automata–Transducers (initial state for both automata is $+1$).

We illustrate two of these automata in Figure 3.1. The edge symbol $S : W$ specifies that this edge is taken if the current input symbol is S, and W denotes a symbol sequence to be output. The initial states are the vertices labeled $+1$.

4. Vertical Cutting Sequences to Additive Continued Fractions

The ordinary continued fraction expansion of an irrational real number $-\frac{1}{2} < \theta < \frac{1}{2}$ is easily determined from either its cutting sequence expansion or its Minkowski geodesic continued fraction expansion. We begin with the latter case.

Theorem 4.1. *Given θ with $-\frac{1}{2} < \theta < \frac{1}{2}$, with ordinary continued fraction expansion $\theta = [a_0; a_1, a_2, a_3, \cdots]$, let $\mathsf{S}_0\mathsf{S}_1\mathsf{S}_2\cdots$ be the Minkowski geodesic continued fraction expansion of θ in the alphabet $\{\mathsf{L}, \mathsf{R}, \mathsf{J}, \mathsf{C}\}$. The MGCF expansion can be uniquely parsed into segments $\mathsf{B}_0\mathsf{B}_1\cdots$ where $\mathsf{B}_0 = \mathsf{J}$ or JL and each B_i for $i \geq 1$ is $\mathsf{R}^k\mathsf{J}$ for some $k \geq 1$, $\mathsf{R}^{k+1}\mathsf{JL}$ for some $k \geq 1$, or $\mathsf{R}^k\mathsf{C}$ for some $k \geq 1$. Each segment encodes exactly one or two partial quotients of the OCF expansion. For general segments, $\mathsf{R}^k\mathsf{J}$ encodes a partial quotient $a_n = k$, while $\mathsf{R}^{k+1}\mathsf{JL}$ or $\mathsf{R}^k\mathsf{C}$ encode two partial quotients $a_n = k$, $a_{n+1} = 1$. For the segment B_0, the symbol J encodes $a_0 = 0$, while JL encodes the partial quotients $a_0 = -1$, $a_1 = 1$. The case $\mathsf{R}^k\mathsf{C}$ can only occur when θ is rational.*

Proof. We say that p/q is a *best approximation* to θ if $|q\theta - p| < |q'\theta - p'|$ for $0 < q' < q$. Using this definition, it is a basic result about ordinary continued fractions that the convergents are the complete set of best approximations; see Hardy and Wright [16, Theorem 182]. We say that p/q is a *better approximation* to θ than p'/q' if $|q\theta - p| < |q'\theta - p'|$. The best approximation property thus proves that p_n/q_n is a better approximation than any fraction with denominator less than q_{n+1}.

We start with $\mathsf{P}^{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, the identity matrix. Thus $\mathsf{P}^{(0)}\mathsf{B}_t(\theta)$ is $\begin{bmatrix} 1 & 0 \\ -\theta & t \end{bmatrix}$. Neither of the Minkowski inequalities (3.24) and (3.25) can hold with equality for any t, since $|\theta| < \frac{1}{2}$. We get $t_1^2 = 1 - \left(\lfloor \theta + \frac{1}{2} \rfloor - \theta\right)^2$ as the value at which (3.23) holds with equality. The first partial quotient matrix is thus J. That makes $\mathsf{P}^{(1)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Now $\mathsf{P}^{(1)}\mathsf{B}_t(\theta) = \begin{bmatrix} -\theta & t \\ 1 & 0 \end{bmatrix}$.

We check the Minkowski inequalities. Inequality (3.23) is satisfied, with equality only at $t = t_1$. Inequality (3.24) is an equality if $1 = (1-\theta)^2 + t_2^2$, and (3.25) is an equality if $1 = (-1-\theta)^2 + t_2^2$. If $\theta = 0$, neither one holds with equality for any $t > 0$, and J is thus the whole MGCF, encoding $\theta = [0]$. If $0 < \theta < \frac{1}{2}$, then only (3.24) can hold with equality, so the next partial quotient matrix will be R. In this case, we have

$$\mathsf{P}^{(1)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} p_0 & q_0 \\ p_{-1} & q_{-1} \end{bmatrix},$$

and we let $\mathsf{B}_0 = \mathsf{J}$. If $-\frac{1}{2} < \theta < 0$, then only (3.25) can hold with equality, so the next partial quotient matrix will be L. For $-\frac{1}{2} < \theta < 0$, the continued fraction for θ begins $[-1, 1, \dots]$, and thus $p_1 = 0, q_1 = 1$. This gives

$$\mathsf{P}^{(2)} = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} p_1 & q_1 \\ p_0 & q_0 \end{bmatrix},$$

and we let $\mathsf{B}_0 = \mathsf{J}, \mathsf{L}$.

In either case, after the first step, we have encoded all coefficients through a_{n-1}, and our current matrix is

$$\mathsf{P}^{(j)} = \begin{bmatrix} p_{n-1} & q_{n-1} \\ p_{n-2} & q_{n-2} \end{bmatrix}. \tag{4.1}$$

The rest of the theorem is proved by induction, with the induction hypothesis that after each segment B_k, the matrix $\mathsf{P}^{(j)}$ is in the form (4.1), with the coefficients through a_{n-1} encoded.

First, assume that the current $\mathsf{P}^{(j)}$ is in the form (4.1). Then

$$\mathsf{P}^{(j)}\mathsf{B}_t(\theta) = \begin{bmatrix} p_{n-1} - q_{n-1}\theta & tq_{n-1} \\ p_{n-2} - q_{n-2}\theta & tq_{n-2} \end{bmatrix}.$$

Since $q_{n-1} > q_{n-2}$, the requirement of an increasing denominator makes J impossible for the next partial quotient. The Minkowski inequality for R is

$$\begin{aligned} &\left((p_{n-1} - q_{n-1}\theta) + (p_{n-2} - q_{n-2}\theta)\right)^2 - (p_{n-2} - q_{n-2}\theta)^2 \\ &\qquad + t^2\left((q_{n-1} + q_{n-2})^2 - q_{n-2}^2\right) \ \geq \ 0, \end{aligned} \tag{4.2}$$

and for L, it is

$$\begin{aligned} &\left((p_{n-1} - q_{n-1}\theta) - (p_{n-2} - q_{n-2}\theta)\right)^2 - (p_{n-2} - q_{n-2}\theta)^2 \\ &\qquad + t^2\left((q_{n-1} - q_{n-2})^2 - q_{n-2}^2\right) \ \geq \ 0. \end{aligned} \tag{4.3}$$

By the best approximation condition, p_{n-1}/q_{n-1} is a better approximation to θ than $(p_{n-1} - p_{n-2})/(q_{n-1} - q_{n-2})$. Thus (4.3) is satisfied for small t, and thus for all $t < t_j$, so we cannot have L next. If θ is rational and a_{n-1} is the last term, then (4.2) is satisfied for all t; thus the MGCF terminates. Otherwise, since every even convergent is less than θ, and every odd convergent is greater, $p_{n-1} - q_{n-1}\theta$ and $p_{n-2} - q_{n-2}\theta$ have opposite signs, and the absolute value of their sum is less than the absolute value of $p_{n-2} - q_{n-2}\theta$. Thus (4.2) is not satisfied for sufficiently small t, and t_{j+1} is the value of t at which it holds with equality. Thus the next matrix is R; hence

$$\mathsf{P}^{(j+1)} = \mathsf{R}\mathsf{P}^{(j)} = \begin{bmatrix} p_{n-1} & q_{n-1} \\ p_{n-2} + p_{n-1} & q_{n-2} + q_{n-1} \end{bmatrix}.$$

The remainder of the proof considers the case where $\mathsf{P}^{(j)}$ is of the form

$$\mathsf{P}^{(j)} = \begin{bmatrix} p_{n-1} & q_{n-1} \\ p_{n-2} + mp_{n-1} & q_{n-2} + mq_{n-1} \end{bmatrix}, \tag{4.4}$$

which can be reached after m applications of R to the form (4.1). Now we have

$$\mathsf{P}^{(j)}\mathsf{B}_t(\theta) = \begin{bmatrix} p_{n-1} - q_{n-1}\theta & tq_{n-1} \\ (p_{n-2} - q_{n-2}\theta) + m\,(p_{n-1} - q_{n-1}\theta) & t(q_{n-2} + mq_{n-1}) \end{bmatrix}.$$

Since $q_{n-2} + mq_{n-1} > q_{n-1}$, the next symbol $A^{(j+1)}$ cannot be L. The other two Minkowski inequalities, for J and R, are

$$\begin{aligned} \left((p_{n-2} - q_{n-2}\theta) + m\,(p_{n-1} - q_{n-1}\theta)\right)^2 - (p_{n-1} - q_{n-1}\theta)^2 & \\ + t^2((q_{n-2} + mq_{n-1})^2 - q_{n-1}^2) &\geq 0, \qquad (4.5) \\ \left((m+1)\,(p_{n-1} - q_{n-1}\theta) + (p_{n-2} - q_{n-2}\theta)\right)^2 & \\ - \left(m\,(p_{n-1} - q_{n-1}\theta) + (p_{n-2} - q_{n-2}\theta)\right)^2 & \\ + t^2(((m+1)q_{n-1} + q_{n-2})^2 - (mq_{n-1} + q_{n-2})^2) &\geq 0. \qquad (4.6) \end{aligned}$$

Thus the next symbol is $\mathsf{A}^{(j+1)} = \mathsf{J}$ if (4.5) holds with equality for a larger $t < t_j$ than (4.6), $\mathsf{A}^{(j+1)} = \mathsf{R}$ if (4.6) holds with equality for a larger $t < t_j$, and $\mathsf{A}^{(j+1)} = \mathsf{C}$ if both hold with equality at the same t. Recall that we have, for $0 \leq m \leq a_n$, and n even,

$$\frac{p_{n-1} + p_n}{q_{n-1} + q_n} = \frac{(a_n + 1)p_{n-1} + p_{n-2}}{(a_n + 1)q_{n-1} + q_{n-2}} > \theta \geq \frac{mp_{n-1} + p_{n-2}}{mq_{n-1} + q_{n-2}}, \tag{4.7}$$

with the reverse inequalities holding for n odd. The last inequality holds with equality only if $\theta = p_n/q_n$ and $m = a_n$. If $m < a_n$, then $mq_{n-1} + q_{n-2} < q_n$, so we must have

$$(p_{n-1} - q_{n-1}\theta)^2 < \left((p_{n-2} - q_{n-2}\theta) + m\,(p_{n-1} - q_{n-1}\theta)\right)^2,$$

and thus the left-hand side of (4.5) is positive for small t, so $\mathsf{A}^{(j+1)}$ cannot be J or C. Thus, if $m < a_n$, the only possible case is R, and the matrix $\mathsf{P}^{(j+1)}$ is still of the form (4.4), with m replaced by $m + 1$.

We next consider the case when $m = a_n$ in (4.4). In this case, the next symbol $\mathsf{A}^{(j+1)}$ may be any of R, J, and C. First suppose $\mathsf{A}^{(j+1)} = \mathsf{J}$. If so, we have finished our segment $\mathsf{B}_k = \mathsf{R}^{a_n}\mathsf{J}$, and $\mathsf{P}^{(j+1)}$ is now in the correct form (4.1), since it is

$$\begin{bmatrix} p_n & q_n \\ p_{n-1} & q_{n-1} \end{bmatrix}.$$

Next suppose that $m = a_n$ and $\mathsf{A}^{(j+1)} = \mathsf{R}$. For this to happen, we need (4.6) to be true only for $t \geq t_c$. Since $m = a_n$, we have $mq_{n-1} + q_{n-2} = q_n$, and similarly for p_n. For (4.6) to be true only for sufficiently large t, $(p_{n-1}+p_n)/(q_{n-1}+q_n)$ must be a better approximation to θ than p_n/q_n. This cannot happen if $\theta = p_n/q_n$, hence a_{n+1} is defined. By the best approximation property, we must have $q_{n-1}+q_n \geq q_{n+1}$, and it follows that $a_{n+1} = 1$, $p_{n+1} = p_n + p_{n-1}$, and $q_{n+1} = q_n + q_{n-1}$, so that in this case

$$\mathsf{P}^{(j+1)} = \mathsf{R}\mathsf{P}^{(j)} = \begin{bmatrix} p_{n-1} & q_{n-1} \\ p_{n+1} & q_{n+1} \end{bmatrix}.$$

This matrix is in the form (4.4) but with $m = a_n + 1$. Now $\mathsf{A}^{(j+2)}$ cannot be R or C, because the left-hand side of (4.6) is positive for small t, so $\mathsf{A}^{(j+2)}$ must be J, which gives

$$\mathsf{P}^{(j+2)} = \mathsf{J}\mathsf{R}\mathsf{P}^{(j)} = \begin{bmatrix} p_{n+1} & q_{n+1} \\ p_{n-1} & q_{n-1} \end{bmatrix}. \tag{4.8}$$

This gives

$$\mathsf{P}^{(j+2)}\mathsf{B}_t(\theta) = \begin{bmatrix} p_{n+1} - q_{n+1}\theta & tq_{n+1} \\ p_{n-1} - q_{n-1}\theta & tq_{n-1} \end{bmatrix},$$

The following symbol $\mathsf{A}^{(j+3)}$ cannot be J, because $q_{n+1} > q_{n-1}$ and we must have an increasing denominator. The Minkowski inequality for R is

$$\begin{aligned} &((p_{n+1} - q_{n+1}\theta) + (p_{n-1} - q_{n-1}\theta))^2 - (p_{n-1} - q_{n-1}\theta)^2 \\ &\qquad + t^2((q_{n+1} + q_{n-1})^2 - q_{n-1}^2) \;\geq\; 0. \end{aligned} \tag{4.9}$$

The constant term here is positive because $p_{n+1} - q_{n+1}\theta$ and $p_{n-1} - q_{n-1}\theta$ have the same sign, so this inequality holds for all t. The Minkowski inequality for L is

$$\begin{aligned} &((p_{n+1} - q_{n+1}\theta) - (p_{n-1} - q_{n-1}\theta))^2 - (p_{n-1} - q_{n-1}\theta)^2 \\ &\qquad + t^2((q_{n+1} - q_{n-1})^2 - q_{n-1}^2) \;\geq\; 0, \end{aligned} \tag{4.10}$$

which is not satisfied for small enough t because $(p_{n+1} - q_{n+1}\theta) - (p_{n-1} - q_{n-1}\theta) = p_n - q_n\theta$, and thus its constant term is negative. Thus $\mathsf{A}^{(j+3)} = \mathsf{L}$, so that

$$\mathsf{P}^{(j+3)} = \mathsf{L}\mathsf{J}\mathsf{R}\mathsf{P}^{(j)} = \begin{bmatrix} p_{n+1} & q_{n+1} \\ p_n & q_n \end{bmatrix}. \tag{4.11}$$

This matrix has the form (4.1), and we have encoded the two coefficients, a_n and $a_{n+1} = 1$, with a segment $\mathsf{B}_k = \mathsf{R}^{a_n+1}\mathsf{J}\mathsf{L}$.

Finally, suppose that $m = a_n$ and $\mathsf{A}^{(j+1)} = \mathsf{C}$. Again, (4.6) must fail to hold for sufficiently small t, and it follows that $a_{n+1} = 1$. Thus, since $q_{n+1} = q_n + q_{n-1}$ and $p_{n+1} = p_n + p_{n-1}$, we have

$$\mathsf{P}^{(j+1)} = \mathsf{C}\mathsf{P}^{(j)} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_{n-1} & q_{n-1} \\ p_n & q_n \end{bmatrix} = \begin{bmatrix} p_{n+1} & q_{n+1} \\ p_n & q_n \end{bmatrix}. \tag{4.12}$$

Here $\mathsf{P}^{(j+1)}$ is in the form (4.1), and we have again encoded the two coefficients, a_n and $a_{n+1} = 1$ as $\mathsf{R}^{a_n}\mathsf{C}$. This completes the induction step.

Lemma 2.2 shows that the last case $\mathsf{R}^k\mathsf{C}$ can occur only for rational θ. ■

There is an analogous conversion method from the cutting sequence expansion to the additive ordinary continued fraction expansion, as follows.

Theorem 4.2. *Given θ with ordinary continued function expansion $\theta = [a_0; a_1, a_2, a_3, \cdots]$, let $\mathsf{S}_0^*\mathsf{S}_1^*\mathsf{S}_2^* \cdots$ be the cutting sequence expansion for the geodesic $\langle\infty, \theta\rangle = \{\theta + it : t > 0\}$ in the alphabet $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}, \bar{\mathsf{C}}_1, \bar{\mathsf{C}}_2\}$. It can be uniquely parsed into segments $\bar{\mathsf{B}}_0\bar{\mathsf{B}}_1\bar{\mathsf{B}}_2 \cdots$ where $\bar{\mathsf{B}}_0 = \bar{\mathsf{J}}$ or $\bar{\mathsf{J}}\bar{\mathsf{R}}$, encoding $a_0 = 0$ or $a_0 = -1$, $a_1 = 1$, respectively, and each succeeding segment is $\bar{\mathsf{R}}^k\bar{\mathsf{J}}$ or $\bar{\mathsf{L}}^k\bar{\mathsf{J}}$ encoding a_n for n even and odd, respectively, or is $\bar{\mathsf{R}}^{k+1}\bar{\mathsf{J}}\ \bar{\mathsf{R}}$, or $\bar{\mathsf{R}}^k\bar{\mathsf{C}}_1$ encoding $a_n = k, a_n + 1 = 1$ for n even, or is $\bar{\mathsf{L}}^{k+1}\bar{\mathsf{J}}\ \bar{\mathsf{L}}$ or $\bar{\mathsf{L}}^k\bar{\mathsf{C}}_2$ encoding $a_n = k$, $a_{n+1} = 1$ for n odd. The symbols $\bar{\mathsf{C}}_1, \bar{\mathsf{C}}_2$ can only occur in expansions of rational θ.*

Proof. This follows from Theorem 4.1 by noticing that the parity of the initial word $\mathsf{W}_i = \mathsf{B}_0 \ldots \mathsf{B}_i$ changes after each segment B_i encoding one term, and does not change after any segment B_i encoding two terms, while the segment B_0 has odd parity if it encodes no terms, and even parity if it encodes a_1. ■

An immediate consequence of Theorem 4.1 and 4.2 is the following result.

Theorem 4.3. *There exists a finite automaton which converts the Minkowski geodesic continued fraction expansion of each irrational θ with $-\frac{1}{2} < \theta < \frac{1}{2}$ to the additive continued fraction expansion of θ. There exists a finite automaton that converts the vertical cutting sequence expansion of $\langle\infty, \theta\rangle$ to the additive ordinary continued fraction expansion of θ.*

Proof. The segment-partition of the MGCF expansion given in Theorem 4.1 permits the Farey tree expansion (3.7) to be computed by a finite automaton, because the necessary information to decide on the symbol R versus L depends only on the determinant ± 1 of the product of the MGCF matrices scanned plus the values of the last four MGCF symbols in the expansion. Next, Theorem 4.2 guarantees that a finite automaton exists to convert the cutting sequence expansion as well. Finally, Theorem 3.2 applies to give the additive continued fraction from the Farey tree expansion. ■

Theorem 4.1 implies that the Minkowski geodesic continued fraction expansion of θ can represented in an abbreviated form

$$\theta := [\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3, ...]$$

similar to its ordinary continued fraction expansion

$$\theta = [a_0; a_1, a_2, a_3, ...] \ ,$$

with the change that each symbol 1 in the OCF expansion is to be replaced by one of three possible symbols 1_h, 1_m and 1_c. Here 1_h means that the continued fraction partial quotient $a_n = 1$ begins a new segment RJ or RRJL in the MGCF expansion (so that the previous convergent p_{n-1}/q_{n-1} was "hit" at the end of a segment), 1_m means that it combines with the previous partial quotient in a block $\mathsf{R}^{k+1}\mathsf{JL}$ (so that p_{n-1}/q_{n-1} was "missed"), and 1_c means that it combines with the previous partial quotient in a block with a C-symbol (a "corner"). However it does not specify which sequence of symbols actually occur as legal expansions. We study this next.

5. Additive Continued Fractions to Vertical Cutting Sequences

In this section, we show how to construct the Minkowski geodesic continued fraction expansion of θ given the additive continued fraction expansion of θ. Let θ have the ordinary continued fraction expansion

$$\theta = [a_0; a_1, a_2, a_3, \cdots],$$

in which $a_0 = 0$ or -1. In view of the results of section 4, it suffices to determine for each $a_n = 1$ whether or not it has label $1_h, 1_m$ or 1_c in the partition of the MGCF expansion given in Theorem 4.1.

Theorem 5.1. *Given $\theta = [a_0; a_1, a_2, a_3, \cdots]$ with $-\frac{1}{2} < \theta < \frac{1}{2}$, suppose $a_{n+1} = 1$. Set $\alpha_n = [0, a_n, a_{n-1}, \ldots, a_1] = \frac{q_{n-1}}{q_n}$, and $\beta_n = [a_{n+1}, a_{n+2}, \cdots]$, so that $\alpha_n \in [0,1]$ and $\beta_n \in [1,2]$. In terms of the linear fractional transformation*

$$\mathsf{N}(z) := \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} (z) = \frac{z+2}{2z+1} , \tag{5.1}$$

the Minkowski geodesic continued fraction expansion of θ has

$$\tilde{a}_{n+1} = \begin{cases} 1_h & \text{if } \beta_n > \mathsf{N}(\alpha_n) , \\ 1_c & \text{if } \beta_n = \mathsf{N}(\alpha_n) , \\ 1_m & \text{if } \beta_n < \mathsf{N}(\alpha_n) . \end{cases} \tag{5.2}$$

Remark. The matrix N acting as a linear fractional transformation maps $[0,1]$ to $[1,2]$ while reversing orientation. Its inverse $\mathsf{N}^{-1} = \frac{1}{3}\begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}$ is not integral. The theorem could also be formulated in terms of the linear fractional transformation $-\mathsf{N}$ which maps $[0,1]$ to $[-2,-1]$ and is an involution, i.e. $-\mathsf{N}(-\mathsf{N}(z)) \equiv z$. In particular, since N sends $\mathbb{R}^+$ into itself,

$$\beta < \mathsf{N}(\alpha) \iff \alpha > -\mathsf{N}(-\beta) . \tag{5.3}$$

The symbol 1_c can occur only when θ is rational, because $\mathsf{N}(\alpha_n)$ is always rational, while β_n is rational if and only if the expansion of θ terminates.

Proof of Theorem 5.1. Note that if $a_1 = 1$, it always becomes 1_m, and we have $\alpha_n = 0$ and thus $\mathsf{N}(\alpha_n) = 2 > \beta_n$ as required. The following discussion assumes that $a_{n+1} = 1$, with $n \geq 1$, so that we are not in the segment B_0.

By the argument in the proof of Theorem 4.1, whether a_{n+1} becomes 1_h, 1_c, or 1_m depends on which of (4.5) and (4.6) holds with equality for a larger value of t; we have 1_m if it is (4.6), and 1_c if both hold with equality for the same t. Since $a_{n+1} = 1$, we have

$$\begin{aligned} p_{n+1} &= p_n + p_{n-1}. \\ q_{n+1} &= q_n + q_{n-1}. \\ p_{n+1} - q_{n+1}\theta &= (p_n - q_n\theta) + (p_{n-1} - q_{n-1}\theta), \end{aligned}$$

The condition for 1_m is thus

$$\frac{(p_n - q_n\theta)^2 - (p_{n+1} - q_{n+1}\theta)^2}{q_{n+1}^2 - q_n^2} > \frac{(p_{n-1} - q_{n-1}\theta)^2 - (p_n - q_n\theta)^2}{q_n^2 - q_{n-1}^2}, \tag{5.4}$$

or, equivalently,

$$\frac{(p_n - q_n\theta)^2 - (p_{n+1} - q_{n+1}\theta)^2}{(p_{n-1} - q_{n-1}\theta)^2 - (p_n - q_n\theta)^2} > \frac{q_n^2 - q_{n+1}^2}{q_{n-1}^2 - q_n^2}, \tag{5.5}$$

and the condition for 1_c is equality. Putting everything in terms of p_{n-1}, p_n, q_{n-1}, and q_n gives

$$\frac{(p_{n-1} - q_{n-1}\theta)^2 + 2(p_{n-1} - q_{n-1}\theta)(p_n - q_n\theta)}{(p_{n-1} - q_{n-1}\theta)^2 - (p_n - q_n\theta)^2} > \frac{q_{n-1}^2 + 2q_{n-1}q_n}{q_{n-1}^2 - q_n^2}. \tag{5.6}$$

Now let

$$\alpha_n = \frac{q_{n-1}}{q_n}, \tag{5.7}$$

$$\beta_n = -\frac{p_{n-1} - q_{n-1}\theta}{p_n - q_n\theta}. \tag{5.8}$$

These quantities have ordinary continued-fraction expansions $\alpha_n = [0, a_n, a_{n-1}, \ldots, a_1]$, and $\beta_n = [a_{n+1}, a_{n+2}, \cdots]$; see Venkov [50, section 2.4]. Clearly $0 < \alpha_n < 1$ and $\beta_n > 1$, and, since $a_{n+1} = 1$, $\beta_n < 2$. Now (5.6) becomes

$$\frac{\beta_n^2 - 2\beta_n}{\beta_n^2 - 1} > \frac{\alpha_n^2 + 2\alpha_n}{\alpha_n^2 - 1}, \tag{5.9}$$

or, by subtracting 1 from each side,

$$\frac{-2\beta_n + 1}{\beta_n^2 - 1} > \frac{2\alpha_n + 1}{\alpha_n^2 - 1}. \tag{5.10}$$

Let $C = (2\alpha_n + 1)/(\alpha_n^2 - 1)$. Then (5.10) holds if and only if $-\beta_n$ is between the roots of

$$Cx^2 - 2x - 1 - C = 0. \tag{5.11}$$

The sum of the roots of (5.11) is $2/C$, and the larger root is α_n, so the other root is

$$x = \frac{2\alpha_n^2 - 2}{2\alpha_n + 1} - \alpha_n = \frac{-\alpha_n - 2}{2\alpha_n + 1} = -\mathsf{N}(\alpha_n). \tag{5.12}$$

This gives our desired condition; we have $\tilde{a}_{n+1} = 1_m$ if $\beta_n < \mathsf{N}(\alpha_n)$. Since $-\mathsf{N}$ is an involution and is increasing on $[-2, -1]$ and on $[0, 1]$, we can also write this condition as $\alpha_n > -\mathsf{N}(-\beta_n)$. ■

Theorem 5.1 suffices to classify all θ containing the symbol 1_c, i.e. all vertical geodesics that hit a corner of a translate of a fundamental domain.

Corollary 5.1. *Let* $-\frac{1}{2} < \theta < \frac{1}{2}$ *and suppose that the MGCF expansion of* θ *contains a symbol* 1_c. *Then* θ *is rational and has ordinary continued fraction expansion of the form*

$$\theta = \begin{cases} [0, a_1, a_2, , \ldots ,a_n, 1_c, b_1, \ldots ,b_m], & \text{if } 0 \le \theta < \frac{1}{2}. \\ [-1, 1, a_1 - 1, a_2, , \ldots ,a_n, 1_c, b_1, \ldots ,b_m], & \text{if } -\frac{1}{2} < \theta < 0. \end{cases} \tag{5.13}$$

in which the additive continued fraction expansion of $\alpha_n^* := [1, b_1, \ldots ,b_m]$ *is computable from the additive continued fraction expansion of* $\alpha_n = [0, a_n, \ldots ,a_1]$ *by a finite automaton, and vice-versa. Furthermore there is an absolute constant* c_0 *such that*

$$\textit{ACF-length } (\alpha_n^*) \le c_0 \quad (\textit{ ACF-length } (\alpha_n)) \ . \tag{5.14}$$

Proof. Raney [41] proves that given any fixed linear fractional transformation $\tilde{\mathsf{M}}(z) := \frac{az+b}{cz+d}$ with integer coefficients a, b, c, d and $ad - bc \neq 0$, the ACF expansion of $\tilde{\mathsf{M}}(\theta)$ can be computed from the ACF expansion of θ using a finite automaton. Furthermore the conversion process inflates the ACF-length by at most a multiplicative constant (depending on $\tilde{\mathsf{M}}$). Apply this to $\beta_n = \mathsf{N}(\alpha_n)$. The specific automaton for $\mathsf{N}(z) = \frac{z+2}{2z+1}$ is given in Raney [41, pp. 274–275], and the constant in this case is 3. ■

These results give a bound on the computational complexity of computing the MGCF expansion from the additive continued fraction expansion.

Theorem 5.2. *The Minkowski geodesic continued fraction expansion of* θ *can be computed from the additive continued fraction expansion by a Turing machine in quadratic time using linear space. That is, there are absolute constants* c_1, c_2, c_3 *such that for any* θ *with* $-\frac{1}{2} < \theta < \frac{1}{2}$, *the first* ℓ *symbols of the MGCF expansion of* θ *can be computed using the first* $c_1\ell$ *symbols of the additive continued fraction expansion of* θ *using at most* $c_2\ell^2$ *time steps and* $c_3\ell$ *space locations.*

Proof. Theorem 4.1 shows that the main problem is to resolve whether a given symbol $a_{n+1} = 1$ which appears as the ℓth symbol in the ordinary continued fraction expansion of θ is to be 1_m, 1_h, or 1_c in the MGCF expansion. For this, we use Theorem 5.1. At worst, we must look all the way back to the beginning of the MGCF expansion. Corollary 5.1 shows that the ACF expansion for $\mathsf{N}(\alpha_n)$ is of length at most $c_0\ell$, and thus comparing it to β_n requires looking at no more than $c_0\ell + 1$ symbols, and thus $O(\ell)$ time and $O(\ell)$ space. Since there may be $O(\ell)$ different 1's to be resolved, the total time required is $O(\ell^2)$. ■

There are examples which do require $\Omega(\ell^2)$ time steps; for example,

$$\theta = (\sqrt{3} - 1)/2 = [0, \overline{2, 1_h}],$$

has this property. For this θ, it is necessary to backtrack all the way to the first symbol to determine that each 1 is 1_h, because the sequence $[0, (2, 1_h)^j, 2, 1_c, (2, 1_m)^j, 4]$ is a 1_c-sequence for each $j \ge 1$.

In the next section, we will show that the MGCF expansion cannot be computed from the additive continued fraction expansion using a finite automaton, or even using a pushdown automaton with one stack, as defined in Hopcroft and Ullman [20].

6. Vertical Geodesics: Forbidden Blocks

In this section, we characterize the allowable cutting sequences of $\Pi^0_{\mathcal{F}}$.

Definition 6.1. A finite word W in the symbol set $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\}$ is a *forbidden block* of $\Pi^0_{\mathcal{F}}$ if it occurs in no cutting sequence of $\Pi^0_{\mathcal{F}}$; otherwise, it is an *admissible block.* It is an *excluded initial block* of $\Pi^0_{\mathcal{F}}$ if does not occur as an initial segment of any cutting sequence of $\Pi^0_{\mathcal{F}}$; otherwise, it is an *included initial block.*

It is easy to see that excluded initial blocks alone can be used to characterize *any* subset of the one-sided shift on $\{\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}\}$. We show that $\Pi^0_{\mathcal{F}}$ has the stronger property that it is determined by its set of forbidden blocks, as follows.

Theorem 6.1. *A finite word* W *in the symbols* $\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}$ *is an excluded initial block in* $\Pi^0_{\mathcal{F}}$ *if and only if at least one of the blocks* $\bar{\mathsf{L}}\mathsf{W}$ *and* $\bar{\mathsf{R}}\mathsf{W}$ *is a forbidden block of* $\Pi^0_{\mathcal{F}}$.

Proof. We prove the contrapositive. First, we show that for every included initial block W, both $\bar{\mathsf{R}}^n\mathsf{W}$ and $\bar{\mathsf{L}}^n\mathsf{W}$ are admissible blocks of $\Pi^0_{\mathcal{F}}$ for all $n \geq 1$. Geometrically, this encodes the fact that the geodesic $\langle \infty, \theta \rangle$ is a limit of geodesics $\langle \theta', \theta \rangle$ as $|\theta'| \to \infty$, where $\theta' \to -\infty$ is associated to $\bar{\mathsf{R}}^n\mathsf{W}$ and $\theta \to +\infty$ is associated to $\bar{\mathsf{L}}^n\mathsf{W}$. The block W corresponds to a finite initial segment of the cutting sequence of $\theta + it$ for some irrational θ, say for $t_0 \leq t \leq \infty$ and it has first symbol $\bar{\mathsf{J}}$. By Lemma 2.2, this geodesic cannot hit a corner of any translate of $\mathcal{F}$, hence there is some positive ϵ such that it is at distance at least ϵ from any corner for $t_0 \leq t \leq \theta$. Pick p/q a large negative rational number, and observe that the geodesic $\langle \theta', \theta \rangle$ contains the word $\bar{\mathsf{R}}^n\mathsf{W}$ in its one-sided infinite cutting sequence, in which it enters the domain $\mathcal{F}$ just after the $\bar{\mathsf{R}}^n$. Here we require that the radius $r = \theta - p/q$ is at least $n + 2$ so that it produces the sequence $\bar{\mathsf{R}}^n$, and satisfies $r - \sqrt{r^2 - 1} < \epsilon$ so that the geodesic is within ϵ of $\langle \infty, \theta \rangle$ over the range $t_0 \leq t < 1$ and thus produces the sequence W. Now take a matrix $\mathsf{M} = \begin{bmatrix} q' & p' \\ q & -p \end{bmatrix} \in SL(2, \mathbb{Z})$, so that $\mathsf{M}(\gamma) = \langle \infty, \theta' \rangle$ with $\theta' = \mathsf{M}(\theta)$, and note that θ' is necessarily irrational. The $PSL(2, \mathbb{Z})$-action doesn't affect cutting sequences, so the cutting sequence of $\mathsf{M}(\gamma)$ still contains the word $\bar{\mathsf{R}}^n\mathsf{W}$. There is a unique choice of q' and p' such that $-\frac{1}{2} < \theta' < \frac{1}{2}$. This vertical geodesic has the same cutting sequence as γ. Thus $\bar{\mathsf{R}}^n\mathsf{W}$ and likewise $\bar{\mathsf{L}}^n\mathsf{W}$ are admissible blocks of $\Pi^0_{\mathcal{F}}$ for all $n \geq 1$.

Second, we prove that if $\mathsf{W} = \mathsf{S}_1 \cdots \mathsf{S}_n$ is a word in $\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}$ for which $\bar{\mathsf{R}}\mathsf{W}$ and $\bar{\mathsf{L}}\mathsf{W}$ are both admissible blocks of $\Pi^0_{\mathcal{F}}$, there is a vertical geodesic whose cutting sequence has W as an initial segment. To show this, note that the first symbol in W is necessarily $\bar{\mathsf{J}}$, since $\bar{\mathsf{L}}\bar{\mathsf{R}}$ and $\bar{\mathsf{R}}\bar{\mathsf{L}}$ are forbidden blocks. Let γ_1 and γ_2 be irrational vertical geodesics whose cutting sequences contain the words $\bar{\mathsf{R}}\mathsf{W}$ and $\bar{\mathsf{L}}\mathsf{W}$, respectively. Translate each of them by the appropriate elements of $PSL(2, \mathbb{Z})$ to geodesics $\gamma_1' = \langle p_1/q_1, \theta_1' \rangle$ and $\gamma_2' = \langle p_2/q_2, \theta_2' \rangle$ in such a way that the translated geodesics enter the fundamental domain $\mathcal{F}$ at the symbol immediately preceding W in their cutting sequences. Since the cutting sequence of γ_1' has the letter $\bar{\mathsf{R}}$ as it enters $\mathcal{F}$, it is oriented in the direction of increasing real part and has $p_1/q_1 < -1/2$, while γ_2' is oriented in the direction of decreasing real part and has $p_2/q_2 > 1/2$. The first symbol in W is $\bar{\mathsf{J}}$, hence each geodesic exits $\mathcal{F}$ through its bottom edge at a point with real part between $-1/2$ and $1/2$. If γ_1' and γ_2' do not intersect, that is if $\theta_1' < \theta_2'$, then pick any irrational θ between θ_1' and θ_2', say

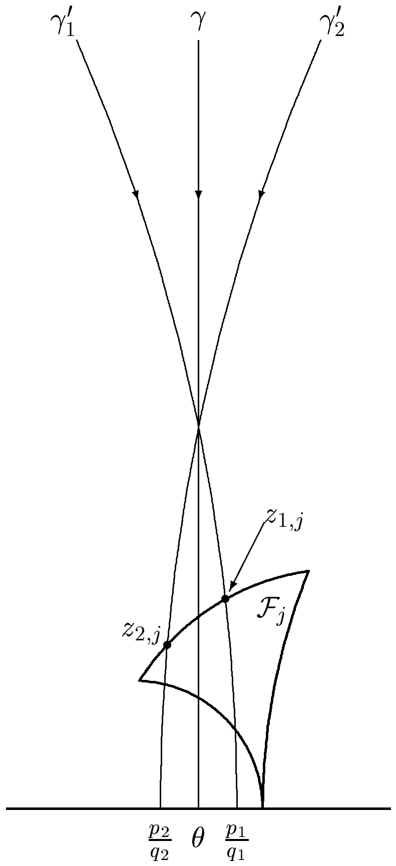


Figure 6.1: Intersecting geodesics: γ is trapped between γ_1' and γ_2' within $\mathcal{F}_j$.

$\theta = \theta_1'$, and let $\gamma = \langle\infty, \theta\rangle$. If γ_1' and γ_2' do intersect, let θ be the real part of their intersection and let $\gamma = \langle\infty, \theta\rangle$. In either case, γ passes through $\mathcal{F}$ since $-\frac{1}{2} < \theta_1' \leq \theta \leq \theta_2' < \frac{1}{2}$. Thus the intersection must have real part θ with $-\frac{1}{2} < \theta < \frac{1}{2}$. In either case, we show that γ has initial word W in its cutting sequence. After leaving $\mathcal{F}$, both γ_1' and γ_2' pass through the same sequence of translated fundamental domains $\mathcal{F}_1, \ldots, \mathcal{F}_n$, entering along the same edge of each, corresponding to the symbols in W. The vertical geodesic γ must then enter $\mathcal{F}_j$ on the same edge at a point z_j between the points $z_{1,j}$ and $z_{2,j}$ where γ_1' and γ_2' hit it, because the domains $\mathcal{F}_j$ are hyperbolically convex; see Figure 6.1. Also, it cannot hit any other fundamental domain between z_{j-1} and z_j, because every point in that interval on γ is on a geodesic between points of γ_1' and γ_2' which are in $\mathcal{F}$. Thus the initial cutting sequence of γ is W. Finally, if the finite endpoint θ of γ is rational, then since γ hits no corners up to hitting the edge W_n, it is a Euclidean distance at least ϵ away from every corner on the W-edges. For every irrational θ' with $|\theta' - \theta| < \epsilon$, the irrational vertical geodesic $\gamma' = \{\theta' + it : t > 0\}$ has the same initial segment W. ■

There are two main types of forbidden blocks (including initial blocks). Some blocks are *edge-forbidden*; they cannot occur in a cutting sequence because they correspond to a geodesic hitting two edges which are part of the same hyperbolic line. Others are *whole-forbidden* because a geodesic can hit any pair of edges in the sequence, but no geodesic can hit the whole sequence.

Series [46, Theorem 3.1] shows that any minimal edge-forbidden block for any geodesic flow must correspond to geodesic hitting a sequence of domains which are all adjacent to the same boundary line, crossing that line as the first and last step. This allows us to find the edge-forbidden blocks from Figure 2.3; without loss of generality, we can let the boundary line be the unit circle, with the forbidden block starting inside the unit circle, going outside, and coming back inside. Since the unit circle has three segments, there are nine such blocks: $\bar{\mathsf{J}}\bar{\mathsf{J}}$, $\bar{\mathsf{L}}\bar{\mathsf{R}}$, $\bar{\mathsf{R}}\bar{\mathsf{L}}$, $\bar{\mathsf{L}}\bar{\mathsf{J}}\bar{\mathsf{L}}\bar{\mathsf{J}}$, $\bar{\mathsf{R}}\bar{\mathsf{J}}\bar{\mathsf{R}}\bar{\mathsf{J}}$, $\bar{\mathsf{J}}\bar{\mathsf{L}}\bar{\mathsf{J}}\bar{\mathsf{L}}$, $\bar{\mathsf{J}}\bar{\mathsf{R}}\bar{\mathsf{J}}\bar{\mathsf{R}}$, $\bar{\mathsf{L}}\bar{\mathsf{J}}\bar{\mathsf{L}}\bar{\mathsf{L}}\bar{\mathsf{J}}\bar{\mathsf{L}}$, $\bar{\mathsf{R}}\bar{\mathsf{J}}\bar{\mathsf{R}}\bar{\mathsf{R}}\bar{\mathsf{J}}\bar{\mathsf{R}}$. These also lead to excluded initial blocks. Since the edges $\bar{\mathsf{R}}$ and $\bar{\mathsf{L}}$ meet at infinity, any forbidden block beginning with either one corresponds to an excluded initial block. These blocks are exactly those which are forbidden by Theorem 4.2.

We next obtain from Theorem 5.1 a characterization of the the whole-forbidden blocks and whole-excluded initial blocks in $\Pi^0_{\mathcal{F}}$. These conditions involve a *critical symbol* 1_h or 1_m which is mislabeled in a cutting sequence expansion.

Definition 6.2. A finite sequence $[d_1, \ldots, d_n, 1, b_1, \ldots, b_m]$ of positive integers is *ambiguous* if it does not suffice to determine the MGCF-label on the critical symbol 1. More precisely, if

$$\begin{aligned} \delta_0 &:= [0, d_n, \ldots, d_1], \text{ and } \delta_1 := [0, d_n, \ldots, d_1 + 1], \\ \beta_0 &:= [1, b_1, \ldots, b_m], \text{ and } \beta_1 := [1, b_1, \ldots, b_m + 1] \end{aligned} ,$$

then $[d_1, \ldots, d_n, 1, b_1, \ldots, b_n]$ is *ambiguous* if and only if

$$\mathrm{int}[\beta_0, \beta_1] \cap \mathrm{int}[\mathsf{N}(\delta_0), \mathsf{N}(\delta_1)] \neq \emptyset, \tag{6.1}$$

Here $\mathrm{int}[\delta_0, \delta_1]$ denotes the closed interval determined by δ_0 and δ_1 with any ordering of the endpoints, i.e. $\delta_1 < \delta_0$ may occur.

Note that the interval $\mathrm{int}[\mathsf{N}(\delta_0), \mathsf{N}(\delta_1)]$ contains all real numbers whose continued-fraction expansions begin with the same partial quotients as δ_0.

Definition 6.3. A *central sequence* $[d_1, \ldots, d_n, 1, b_1, \ldots, b_n]$ is an ambiguous sequence such that $p/q = [0 \text{ or } -1, d_1, \ldots, d_n, 1, b_1, \ldots, b_m]$ is a 1_c sequence. We say that p/q is the rational number *associated to* this central sequence.

We designate initial words of $\Pi^0_{\mathcal{F}}$ using the initial symbol $d_1 = \infty$.

Definition 6.4. An *initial sequence* $[\infty, d_2, \ldots, d_n, 1, b_1, \ldots, b_m]$ denotes an initial word, which if $-\frac{1}{2} < \theta < 0$ has $d_2 = 1$, $d_3 = a_1 - 1$, and $d_i = a_{i-2}$ for $i \geq 4$, while if $0 < \theta < \frac{1}{2}$, $d_i = a_{i-1}$ for all $i \geq 2$.

We define ambiguous initial sequences as in Definition 6.2. If we set

$$\delta_0 = \delta_1 = [0, d_n, \ldots, d_2] = [0, d_n, \ldots, d_2, \infty], \tag{6.2}$$

then the initial sequence is *ambiguous* if

$$\mathrm{int}[\beta_0, \beta_1] \cap \mathrm{int}[\mathsf{N}(\delta_0), \mathsf{N}(\delta_1)] \neq \emptyset, \tag{6.3}$$

where $\mathrm{int}[\mathsf{N}(\delta_0), \mathsf{N}(\delta_1)]$ is the single point $\mathsf{N}(\delta_0)$.

Theorem 6.2 (Characterization of $\Pi^0_{\mathcal{F}}$) *The set $\Pi^0_{\mathcal{F}}$ of cutting sequences of irrational θ in $-\frac{1}{2} < \theta < \frac{1}{2}$ uses the alphabet $\{\bar{\mathsf{R}}, \bar{\mathsf{L}}, \bar{\mathsf{J}}\}$. It consists of all sequences that factorize in segments $\mathsf{B}_0\mathsf{B}_1\mathsf{B}_2\ldots$ as in Theorem 4.2, with the additional property that no sequence of consecutive segments corresponds to a Minkowski geodesic continued fraction expansion*

$$[d_1, \ldots, d_n, 1_*, b_1, \ldots, b_m]$$

with $d_1 = \infty$ allowed, with associated $\delta_0, \delta_1, \beta_0, \beta_1$, such that:

(i) Both of $[d_2, \ldots, d_n, 1, b_1, \ldots, b_m]$ and $[d_1, d_2, \ldots, d_n, 1, b_1, \ldots, b_{m-1}]$ are ambiguous.

(ii) Either

$$1_* = 1_h \text{ and } \operatorname{int}[\beta_0, \beta_1] < \operatorname{int}[\mathsf{N}(\delta_0), \mathsf{N}(\delta_1)], \tag{6.4}$$

or

$$1_* = 1_m \text{ and } \operatorname{int}[\beta_0, \beta_1] > \operatorname{int}[\mathsf{N}(\delta_0), \mathsf{N}(\delta_1)]\ . \tag{6.5}$$

Proof. Immediate from Theorem 5.1 using Theorems 4.1 and 4.2. ■

Theorem 6.2 characterizes a large set of forbidden blocks which are given by conditions (i) and (ii) when d_1 is finite. It also gives an additional set of excluded initial blocks when $d_1 = \infty$. These are sufficient to determine $\Pi^0_{\mathcal{F}}$ because they identify each 1 as 1_h or 1_m as appropriate. However, Theorem 6.2 does not give a complete set of minimal forbidden blocks, because there are extra forbidden blocks in which several symbols $d_i = 1$ and $b_i = 1$ are replaced by 1_h and 1_m, which if left as indeterminate 1's would not be forbidden.

The complete set of minimal forbidden blocks of $\Pi^0_{\mathcal{F}}$ seems harder to characterize. Consider a given block of symbols $\bar{\mathsf{L}}, \bar{\mathsf{R}}, \bar{\mathsf{J}}$; this can be parsed by Theorem 4.2 into a block of complete segments $[a_1, \ldots, a_m]$ in which each symbol a_i is either 1_h, 1_m, or an integer at least 2, together with some conditions on the adjacent incomplete segments; for example, an incomplete segment $\bar{\mathsf{L}}^4$ can be part of a complete segment encoding either $a_{m+1} \geq 4$, or $a_{m+1} = 3$, $a_{m+2} = 1_m$. Then $[a_1, \ldots, a_m]$ is a forbidden block if and only if a certain finite set of linear fractional conditions on two real numbers α, β of the form

$$\alpha < \frac{a\beta + b}{c\beta + d} \tag{6.6}$$

with $ad - bc = \pm 1$ are inconsistent. There is one such inequality for each symbol 1_h or 1_m in the block, which encodes the condition that $[\alpha, a_1, \ldots, a_m, \beta]$ produces the correct symbol 1_h or 1_m. There may also be one or two inequalities on α alone (or on β alone) if there is an incomplete segment at that end, and one more condition (6.6) if the possible encoding of that segment includes another symbol 1_h or 1_m. The total number of inequalities is linear in m.

Tables 6.1, 6.2, and 6.3 below list some central sequences, ambiguous sequences and forbidden blocks, respectively; these were obtained by applying the transformation N to simple δ_i. Table 6.3 illustrates the computation of forbidden sequences from a central sequence; any change to the terms in a central sequence which increases either α_n (and thus decreases $\mathsf{N}(\alpha_n)$) or β_n forces the 1 to be 1_h, and conversely for 1_m.

Theorem 6.2 implies that $\Pi^0_{\mathcal{F}}$ is complicated in the sense that its set of minimal forbidden blocks is very large.

$[2, 1_c, 4]$	$[2, 1_c, 3, 1]$	$[1, 1, 1_c, 4]$	$[1, 1, 1_c, 3, 1]$
$[3, 1_c, 2, 2]$	$[3, 1_c, 2, 1, 1]$	$[1, 2, 1_c, 2, 2]$	$[1, 2, 1_c, 2, 1, 1]$
$[4, 1_c, 2]$	$[4, 1_c, 1, 1]$	$[1, 3, 1_c, 2]$	$[1, 3, 1_c, 1, 1]$
$[3j+1, 1_c, 1, j]$	$[3j+1, 1_c, 1, j-1, 1]$	$[1, 3j, 1_c, 1, j]$	$[1, 3j, 1_c, 1, j-1, 1]$
$[3j+2, 1_c, 1, j, 3]$	$[3j+2, 1_c, 1, j, 2, 1]$	$[1, 3j+1, 1_c, 1, j, 3]$	$[1, 3j+1, 1_c, 1, j, 2, 1]$
$[2, 2, 1_c, 3]$	$[2, 2, 1_c, 2, 1]$	$[1, 1, 2, 1_c, 3]$	$[1, 1, 2, 1_c, 2, 1]$
$[2, 3, 1_c, 2, 5]$	$[2, 3, 1_c, 2, 4, 1]$	$[1, 1, 3, 1_c, 2, 5]$	$[1, 1, 3, 1_c, 2, 4, 1]$
$[3, 2, 1_c, 3, 4]$	$[3, 2, 1_c, 3, 3, 1]$	$[1, 2, 2, 1_c, 3, 4]$	$[1, 2, 2, 1_c, 3, 3, 1]$
$[4, 3, 1_c, 2, 3]$	$[4, 3, 1_c, 2, 2, 1]$	$[1, 3, 3, 1_c, 2, 3]$	$[1, 3, 3, 1_c, 2, 2, 1]$
$[5, 2, 1_c, 3, 2]$	$[5, 2, 1_c, 3, 1, 1]$	$[1, 4, 2, 1_c, 3, 2]$	$[1, 4, 2, 1_c, 3, 1, 1]$
$[j, 1, 1_c, 3j+1]$	$[j, 1, 1_c, 3j, 1]$	$[1, j-1, 1, 1_c, 3j+1]$	$[1, j-1, 1, 1_c, 3j, 1]$
$[3, j, 1, 1_c, 3j+2]$	$[3, j, 1, 1_c, 3j+1, 1]$	$[1, 2, j, 1, 1_c, 3j+2]$	$[1, 2, j, 1, 1_c, 3j+1, 1]$

Table 6.1: Some central sequences, including all with at most five terms.

$[\ldots, j, 1, \underline{1}, 3j+1, \text{any}, \ldots]$	$[\ldots, \text{any}, 3j+1, \underline{1}, 1, j, \ldots]$
$[\ldots, j, 1, \underline{1}, 3j+2, \text{any}, \ldots]$	$[\ldots, \text{any}, 3j+2, \underline{1}, 1, j, \ldots]$
$[\ldots, j, 1, \underline{1}, 3j+3, \text{any}, \ldots]$	$[\ldots, \text{any}, 3j+3, \underline{1}, 1, j, \ldots]$
$[\ldots, 1, 2, \underline{1}, 2, 1, \ldots]$	
$[\ldots, 2, 2, \underline{1}, 3, \geq 4, \ldots]$	$[\ldots, \geq 4, 3, \underline{1}, 2, 2, \ldots]$
$[\ldots, 3, 2, \underline{1}, 3, 2, \ldots]$	$[\ldots, 2, 3, \underline{1}, 2, 3, \ldots]$
$[\ldots, 3, 2, \underline{1}, 3, 3, \ldots]$	$[\ldots, 3, 3, \underline{1}, 2, 3, \ldots]$
$[\ldots, 4, 2, \underline{1}, 3, 2, \ldots]$	$[\ldots, 2, 3, \underline{1}, 2, 4, \ldots]$
$[\ldots, \geq 5, 2, \underline{1}, 3, 1, \ldots]$	$[\ldots, 1, 3, \underline{1}, 2, \geq 5, \ldots]$

Table 6.2: Non-central ambiguous sequences which go two terms forward and two terms back from the underlined 1.

$[\geq 3, 1_h, \geq 3]$
$[\geq 3, 1_h, 2, 1]$
$[1, 1_m, 1]$
$[\geq 1, 2, 1_m, 1]$
$[1, 1_m, 2, \geq 2]$
$[\geq 1, 2, 1_m, 2, \geq 2]$

Table 6.3: Forbidden sequences obtained from the central sequence $[1, 2, 1_c, 2, 2]$; the reverses of these sequences are also forbidden.

Theorem 6.3. *The number $n(k)$ of minimal forbidden blocks of $\Pi^0_{\mathcal{F}}$ of length at most k grows exponentially in k; in fact*

$$\liminf_{k\to\infty} n(k)^{1/k} \geq 2^{1/12}. \tag{6.7}$$

Proof. We will show that each central sequence associated to a rational $p/q \neq 1/2$ yields two minimal forbidden blocks. The forbidden blocks are produced by adding one symbol to each end of the central sequence, and by replacing the central 1_c with a three-symbol word. All these forbidden blocks are distinct. Assuming these facts are proved, consider the central sequences

$$[d_1, d_2, \ldots, d_n, 1_c, b_1, \ldots, b_m]. \tag{6.8}$$

in which each $d_i = 1$ or 2, and in which $[b_1, \ldots, b_m]$ is determined from $[d_1, \ldots, d_n]$. The number of symbols in the cutting sequence encoding of $[d_1, \ldots, d_n]$ is at most $3n$, and by Corollary 5.1, the number of symbols in the cutting sequence encoding of $[b_1, \ldots, b_m]$ is at most $9n$. In obtaining forbidden blocks, the 1_c term is encoded by three symbols, and one symbol is added at each end, hence all resulting forbidden blocks contain at most $12n + 5$ symbols. We conclude that there are at least 2^{n+1} minimal forbidden blocks of length at most $12n+5$, which proves (6.7).

We now construct the minimal forbidden blocks. We use the result of Appendix A, which shows that any vertical geodesic for $\theta = p/q$ with $-\frac{1}{2} < \theta < \frac{1}{2}$ hits at most one corner of a fundamental domain. This fact implies that each of the central sequences (6.8) contains only one symbol 1_c, and thus each p/q comes from at most one central sequence. Associated to the central sequence is a word $\mathsf{W}_1\mathsf{S}\mathsf{W}_2$ in which $\mathsf{S} = \bar{\mathsf{C}}_1$ or $\bar{\mathsf{C}}_2$ is a corner symbol. For the corner symbol, there are two choices of three symbols in $\{\bar{\mathsf{R}}, \bar{\mathsf{L}}, \bar{\mathsf{J}}\}$, which replace the 1_c by 1_h or 1_m; for $\bar{\mathsf{C}}_1$, the choices are $\bar{\mathsf{J}}\bar{\mathsf{R}}\bar{\mathsf{J}}$ or $\bar{\mathsf{L}}\bar{\mathsf{J}}\bar{\mathsf{L}}$. We consider the eight words obtained from $\mathsf{W}_1\mathsf{S}\mathsf{W}_2$ by replacing S by either choice of a three-symbol block, and by adding a prefix symbol and a suffix symbol, each of which may be either $\bar{\mathsf{L}}$ or $\bar{\mathsf{R}}$. The continued fraction for p/q can be recovered from any one of these eight blocks, and since it has only a single 1_c, the critical 1_c in the central sequence is also uniquely determined.

We claim that six of these eight blocks are admissible blocks for $\Pi^0_{\mathcal{F}}$ and the other two are forbidden blocks. The vertical geodesic $\gamma = \langle \infty, p/q \rangle$ corresponding to

$$\frac{p}{q} = [0 \text{ or } -1, d_1, \ldots, d_m, 1_c, b_1, \ldots, b_m]$$

can be approximated by geodesics in the symbol topology in six different ways. Two of these consist of approximating geodesics which do not cross γ at all but approach it from the left and right, respectively. The other four consist of approximating geodesics which cross γ, either crossing above or below the corner that γ hits at the 1_c, and initially approaching γ either from the left or from the right. For sufficiently good approximations, these produce the admissible blocks. (Here we again use the fact that the geodesic γ hits exactly one corner, which implies that all geodesics sufficiently close to γ have the same sequences W_1 and W_2.) The other two blocks are forbidden by condition (ii) of Theorem 6.2. They encode cutting sequences for a geodesic that would have to approach γ from the left, pass to the right of the corner, and then return to the left of γ, or vice versa; such a geodesic would have to cross γ twice, which is impossible.

Let $\bar{\mathsf{C}}_a$ and $\bar{\mathsf{C}}_b$ denote the two possible three-symbol encodings of S. Of the four blocks

$$\bar{\mathsf{R}}\mathsf{W}_1\bar{\mathsf{C}}_a\mathsf{W}_2\bar{\mathsf{R}}\ ,\quad \bar{\mathsf{R}}\mathsf{W}_1\bar{\mathsf{C}}_a\mathsf{W}_2\bar{\mathsf{L}}\ ,\quad \bar{\mathsf{L}}\mathsf{W}_1\bar{\mathsf{C}}_a\mathsf{W}_2\bar{\mathsf{R}} \text{ and } \bar{\mathsf{L}}\mathsf{W}_1\bar{\mathsf{C}}_a\mathsf{W}_2\bar{\mathsf{L}}\ ,$$

exactly three are admissible and one is forbidden. Every sub-block of the forbidden block appears in one of the three admissible blocks, hence the forbidden block is minimal. The same argument applies to the other four blocks containing $\bar{\mathsf{C}}_b$, and produces a second minimal forbidden block. ■

Theorem 6.4. *There does not exist a finite automaton which, when given the additive continued fraction expansion of an irrational number θ with $0 < \theta < \frac{1}{2}$ as its input sequence, computes the Minkowski geodesic continued fraction expansion of θ, or, equivalently, the cutting sequence expansion of the geodesic $\langle\infty, \theta\rangle$.*

Proof. Any finite automaton that would compute the Minkowski geodesic expansion of θ must output the n-th term of this expansion after seeing at most a bounded number of symbols following the n-th symbol of the ACF expansion of θ. However the forbidden block criteria of Theorem 6.2 show that it is sometimes necessary to see an arbitrarily large string of symbols after the n-th symbol, to decide if 1_h or 1_m should be used. For example, for each $j \geq 1$, the sequence

$$[2^{4j+2}, 1_c, 3, (8,4)^j] \tag{6.9}$$

is a central sequence. Adding some later terms decreases β_{4j+2}, and thus changes the 1_c to 1_m, while decreasing the final 4 in $(8,4)^j$ to a 3 and then adding some further terms increases β_{4j+2} and thus changes the 1_c to 1_h. A finite automaton cannot look ahead through the $14j+6$ terms which are necessary, since j can be any integer. In fact, even a pushdown automaton (with one stack) cannot correctly compute all such cutting sequences, because it must look $14j+6$ steps ahead, but also $12j+6$ steps back to distinguish $[a_0; 2^{4j}, 1_m, 3, (8,4)^j, 8, 3, \ldots]$ from $[a_0; 2^{4j+2}, 1_h, 3, (8,4)^j, 8, 3, \ldots]$. ■

7. Two-sided Cutting Sequences: Structure of $\Sigma_{\mathcal{F}}$

We now use the information on vertical cutting sequences $\Pi^0_{\mathcal{F}}$ to characterize the two-sided cutting sequences $\Sigma_{\mathcal{F}}$.

Theorem 7.1. *The cutting sequence shift $\Sigma_{\mathcal{F}}$ for the fundamental domain $\mathcal{F}$ of $PSL(2,\mathbb{Z})$ is the closed subshift whose forbidden blocks coincide with the set of forbidden blocks of $\Pi^0_{\mathcal{F}}$.*

Proof. This result follows from the fact that the set of images of the set of irrational vertical geodesics under $PSL(2,\mathbb{Z})$ is dense in the space of all geodesics of $\mathfrak{H}/PSL(2,\mathbb{Z})$.

Let $\mathcal{S}^0_{\Pi}$ and $\mathcal{S}^0_{\Sigma}$ and $\mathcal{S}_{\Sigma}$ denote the complete sets of forbidden blocks of $\Pi^0_{\mathcal{F}}$, $\Sigma^0_{\mathcal{F}}$, and $\Sigma_{\mathcal{F}}$, respectively. Since $\Sigma_{\mathcal{F}}$ is the closure of $\Sigma^0_{\mathcal{F}}$, we have $\mathcal{S}^0_{\Sigma} = \mathcal{S}_{\Sigma}$.

We first show that

$$\mathcal{S}^0_{\Sigma} \subseteq \mathcal{S}^0_{\Pi}. \tag{7.1}$$

For this it suffices to show that every admissible word W in $\Pi^0_{\mathcal{F}}$ is an admissible word in $\Sigma^0_{\mathcal{F}}$. Suppose that the word W appears in the cutting sequence of the irrational vertical geodesic

$\langle\infty,\theta\rangle = \{\theta + it : t > 0\}$. This geodesic is a limit of geodesics $\langle\phi_i,\theta\rangle$ where $\phi_i \to \infty$ through a sequence of values such that $\langle\phi_i,\theta\rangle$ hits no corner of an image of $\mathcal{F}$. The word $\mathsf{W} = \mathsf{S}_1 \dots \mathsf{S}_r$ corresponds to a specific set of edges of translated fundamental domains $\{g_j\mathcal{F} : 0 \le j \le r\}$ with $g_j \in SL(2,\mathbb{Z})$. For all sufficiently large ϕ_i, the geodesic $\langle\phi_i,\theta\rangle$ passes through the same sequence of fundamental domains $\{g_j\mathcal{F} : 0 \le j \le r\}$, hitting the same sequence of edges in the same order. Thus W occurs in the two-sided cutting sequence of $\langle\phi_i,\theta\rangle$, so it is an admissible word of $\Sigma^0_{\mathcal{F}}$.

The reverse inclusion

$$\mathcal{S}^0_{\Pi} \subseteq \mathcal{S}^0_{\Sigma}. \tag{7.2}$$

is proved similarly. Let W be an admissible word in some general position geodesic $\gamma = \langle\phi,\theta\rangle$. There is a sequence of translated fundamental domains $\{\mathcal{F}_j : 0 \le j \le r\}$ which ϕ passes through that corresponds to W. Since γ hits no corners, we can choose a rational number p/q sufficiently close to ϕ such that the geodesic $\gamma' = \langle p/q,\theta\rangle$ passes through the same set of fundamental domains $\{\mathcal{F}_j : 0 \le j \le r\}$, hitting the same sequence of edges in the same order, hence its cutting sequence (which is only one-sided infinite) contains the word W. Now apply to γ' the transformation M in $PSL(2,\mathbb{Z})$ which takes p/q to ∞, and θ to some θ' in $(-\frac{1}{2},\frac{1}{2})$. For this choice, the cutting sequence of $\mathsf{M}(\gamma')$ is in $\Pi^0_{\mathcal{F}}$, and (7.2) follows. ■

Theorem 7.2. *Every element in $\Sigma_{\mathcal{F}}$ is a cutting sequence for a unique oriented geodesic on $\mathfrak{H}/PSL(2,\mathbb{Z})$ which hits $\mathcal{F}$, except for the two sequences $\bar{\mathsf{R}}^\infty$ or $\bar{\mathsf{L}}^\infty$. Every oriented geodesic γ on $\mathfrak{H}/PSL(2,\mathbb{Z})$ has at least one and at most finitely many shift-equivalence classes of cutting sequences in $\Sigma_{\mathcal{F}}$. If γ is not periodic then it has at most eight shift-equivalence classes of cutting sequences in $\Sigma_{\mathcal{F}}$.*

To establish this result, we first prove a preliminary lemma.

Lemma 7.1. *Let $\gamma_j = \langle\theta'_j,\theta_j\rangle$ for $j = 1,2,\dots$ be a sequence of general position geodesics that intersect $\mathcal{F}$ which have cutting sequences $\{\mathsf{S}^{(j)}_i : i \in \mathbb{Z}\}$ such that the symbol $\mathsf{S}^{(j)}_0$ corresponds to the geodesic γ_j entering the fundamental domain $\mathcal{F}$. If the cutting sequences $\{\mathsf{S}^{(j)}_i\}$ converge in the symbol topology as $j \to \infty$ to a limit sequence $\{\mathsf{S}_i\}$ then the endpoints θ'_j and θ_j converge to unequal limiting values*

$$\theta' = \lim_{j\to\infty} \theta'_j \text{ and } \theta = \lim_{j\to\infty} \theta_j. \tag{7.3}$$

in $\mathbb{R} \cup \{-\infty,\infty\}$. The geodesics γ_j converge to a limiting geodesic $\gamma = \langle\theta',\theta\rangle$ if at least one of θ,θ' is finite. The exceptional cases $\langle-\infty,\infty\rangle$ and $\langle\infty,-\infty\rangle$ correspond to the limiting symbol sequences $\bar{\mathsf{R}}^\infty$ and $\bar{\mathsf{L}}^\infty$, respectively.

Proof. Let $\mathsf{S}_+ = \{\mathsf{S}_i : i > 0\}$ and $\mathsf{S}_- = \{\mathsf{S}_i : i \le 0\}$ denote the positive and negative part of the limit sequence, respectively. If S_+ is $\bar{\mathsf{R}}^\infty$, then since any geodesic γ_j which enters $\mathcal{F}$ after S_0 and has $\mathsf{S}^{(j)}_i = \bar{\mathsf{R}}$ for $1 \le i \le n$ must have $\theta_j \ge n + 1/2$, we conclude that $\lim_{j\to\infty}\theta_j = \infty$ in this case. Similarly, if S_+ is $\bar{\mathsf{L}}^\infty$, then $\lim_{j\to\infty}\theta_j = -\infty$; if S_- is $\bar{\mathsf{R}}^\infty$, then $\lim_{j\to\infty}\theta'_j = -\infty$ if S_- is $\bar{\mathsf{L}}^\infty$, then $\lim_{j\to\infty}\theta'_j = -\infty$. In particular, this shows that the two-sided limit sequence $\bar{\mathsf{R}}^\infty$ corresponds to $\langle-\infty,\infty\rangle$, and $\bar{\mathsf{L}}^\infty$ corresponds to $\langle\infty,-\infty\rangle$.

We next observe that since $\bar{\mathsf{R}}\bar{\mathsf{L}}$ and $\bar{\mathsf{L}}\bar{\mathsf{R}}$ are forbidden blocks, if S_+ is not $\bar{\mathsf{R}}^\infty$ or $\bar{\mathsf{L}}^\infty$ it must contain a symbol $\bar{\mathsf{J}}$. Let the initial segment of S_+ up to the first such symbol be $\bar{\mathsf{R}}^n\bar{\mathsf{J}}$

(resp. $\bar{\mathsf{L}}^n\bar{\mathsf{J}}$) Any geodesic γ_j which matches these symbols necessarily has $\theta_j \leq n + 3/2$ (resp. $\theta_j \geq -n - 3/2$) and is positively oriented (resp. negatively oriented), hence $\theta_j \geq -1/2$ (resp. $\theta_j \leq 1/2$). In either case, if a limit (7.3) exists it must be finite. Similar reasoning applies to S_i to show that $\lim_{j\to\infty} \theta'_j$ is infinite if and only if S_- does not contain a symbol $\bar{\mathsf{J}}$.

We now suppose that S_+ contains a symbol $\bar{\mathsf{J}}$, and claim that $\lim_{j\to\infty} \theta_j$ exists and is finite. The argument above shows that $\{\theta_j\}$ is bounded, so to prove this claim it suffices to show that if $\{\theta_j\}$ is not a Cauchy sequence then the one-sided sequences $\{\mathsf{S}_i^{(j)} : i > 0\}$ cannot converge in the symbol topology. If it is not a Cauchy sequence, there is some $\epsilon > 0$ such that for any N, we have $j, k \geq N$ with $|\theta_j - \theta_k| > \epsilon$. The general-position geodesics γ_j and γ_k enter $\mathcal{F}$ and thus have radius at least $\sqrt{3}/2$. Consequently, any point $z = x + iy$ on the geodesic γ_j with $|x - \theta_j| < 1/4$ and $0 < y < \epsilon'$ for $\epsilon' < \epsilon^2/36$ will actually have $|x - \theta_j| < \epsilon/3$, and thus $|x - \theta_k| > 2\epsilon/3$; similarly, points on γ_k with $0 < y < \epsilon'$ and $|x - \theta_k| < 1/4$ must have $|x - \theta_k| > 2\epsilon/3$. This implies an upper bound N_ϵ on the number of fundamental domains which γ_j and γ_k both hit somewhere inside the box $|Re(z) - (\theta_j + \theta_k)/2| \leq 2\epsilon$, $Im(z) \leq \epsilon$. We may assume $\epsilon < 1/2$, hence any such fundamental domain has a cusp at a finite rational point p/q. The rosette $R(p/q)$ of all translated fundamental domains that touch p/q has Euclidean diameter less than $4/q^2$. Thus any translated fundamental domain $\mathcal{F}'$ which is hit by both γ_j and γ_k inside the box must have two points whose real parts differ by at least $\epsilon/3$, hence $q < \sqrt{12/\epsilon}$. In addition, p/q must lie within Euclidean distance $4/q^2$ of both θ_j and θ_k, hence there are only finitely many such rational p/q. Finally, there is a finite bound depending only on q and ϵ on the number of fundamental domains in each rosette $R(p/q)$ which have Euclidean diameter exceeding $\epsilon/3$. Together this yields a finite upper bound N_ϵ on the number of common fundamental domains $\mathcal{F}'$ between γ_j and γ_k in the box. Outside the box, after exiting $\mathcal{F}$ the geodesics γ_j and γ_k traverse the strip $\epsilon' \leq Im(z) \leq 1$, and thus can hit at most N'_ϵ translated fundamental domains $\mathcal{F}'$ in that strip, because there are only finitely many geodesics with Euclidean radius at least ϵ' which intersect the region $-\frac{1}{2} < Re(z) < \frac{1}{2}$. We conclude that the positive cutting sequences γ_j and γ_k cannot agree on all of their first $N_\epsilon + N'_\epsilon + n + 1$ symbols. Thus the sequences $\mathsf{S}_+^{(j)} = \{\mathsf{S}_i^{(j)} : j > 0\}$ cannot converge in the symbol topology, which proves the claim.

A similar argument applies to the limit sequence S_- if it contains the symbol $\bar{\mathsf{J}}$, to prove that $\lim_{j\to\infty} \theta'_j = \theta'$ exists and is finite.

The geodesic $\gamma = \langle \theta', \theta \rangle$ is encoded as a limit of the sequence of geodesics γ_i. If all of the γ_i for $i \geq N$ intersect a particular domain $\mathcal{F}'$, then since the endpoints θ'_i and θ_i converge to θ' and θ, and $\mathcal{F}$ is closed, γ also intersects $\mathcal{F}$, at least on the boundary. If no γ_i for $i \geq N$ intersect a domain $\mathcal{F}'$, then γ does not intersect the interior of $\mathcal{F}'$. Thus, if we consider γ to hit a domain if either it passes through the interior, or it intersects the boundary and all γ_i for sufficiently large i intersect the interior, we can define a cutting sequence as the set of edges which separate these domains, and this sequence is the limit in the symbol topology of the cutting sequences of γ_i. ■

Proof of Theorem 7.2. Since $\Sigma_{\mathcal{F}}$ is a compact set, every oriented geodesic $\gamma = \langle \theta', \theta \rangle$ which hits the interior of $\mathcal{F}$ has at least one cutting sequence in $C(\gamma)$ by taking the limit point of cutting sequences from a family $\gamma_j = \langle \theta'_j, \theta_j \rangle$ with $\theta'_j \to \theta'$ and $\theta_j \to \theta$ as $j \to \infty$.

It remains to show that each oriented geodesic in $\mathfrak{H}$ which hits $\mathcal{F}$ has only a finite number of shift-equivalence classes of cutting sequences in $[C(\gamma)]$. The simplest cases are general position geodesics, for which $C(\gamma)$ is a single cutting sequence, and $[C(\gamma)]$ consists of a single shift-equivalence class.

We determine how many shift-equivalence classes of cutting sequences are possible for a limiting geodesic. First, consider the limiting geodesics with exactly one rational endpoint (considering ∞ as rational). These correspond to vertical geodesics under the $PSL(2,\mathbb{Z})$ action, so by Lemma 2.2, such geodesics cannot hit any internal corners, and thus have two possible encodings; either can be obtained by using approximating geodesics which approach the rational endpoint from either side. Thus $[C(\gamma)]$ contains two shift-equivalence classes.

Next, consider those geodesics which have two rational endpoints. Without loss of generality, we can move one endpoint to ∞ and so have a vertical geodesic $\gamma = \langle\infty, p/q\rangle$. Such a geodesic hits only finitely many corners of fundamental domains. If it hits n corners at finite values of t, then $C(\gamma)$ contains exactly $2n+4$ cutting sequences. Two of them come from geodesics approximating γ from either side without crossing it, and the other $2n+2$ result from approximating geodesics which cross γ between the kth and $(k+1)$st corners with $0 \le k \le n$, either from left to right or from right to left. (Here the 0th corner is ∞, and the $(n+1)$st corner is p/q.) In Appendix A, we prove that $n \le 1$ for all rational p/q, except those $p/q \equiv \frac{1}{2} (\operatorname{mod} 1)$, which have $n = 2$. Thus $[C(\gamma)]$ contains at most eight shift-equivalence classes.

It remains to bound the number of shift-equivalence classes of cutting sequences for geodesics which have two irrational endpoints. If such a geodesic hits no corners, then it has only one cutting sequence in $C(\gamma)$. If it hits exactly one corner, then it has exactly two cutting sequences, which are obtained using geodesics which approach it while staying on opposite sides of the corner. The difficult case occurs with geodesics that hit at least two corners. We show that all such geodesics hit infinitely many corners and are periodic. In this case, $C(\gamma)$ will be an infinite set. To show periodicity, we observe that a corner in the upper half-plane is the intersection of two circles with centers at rational points on the x-axis and rational radii, so its x-coordinate is rational and its y-coordinate is the square root of a rational number. Thus, if the circle with radius r and center $(x_0, 0)$ passes through two such points (x_1, y_1) and (x_2, y_2), then it satisfies the equations

$$\begin{aligned}(x_1 - x_0)^2 + y_1^2 &= r^2, \\ (x_2 - x_0)^2 + y_2^2 &= r^2.\end{aligned}$$

Since y_1^2 and y_2^2 are rational, so is r^2; also, equating the left sides gives a linear equation for x_0 with rational coefficients. Thus the circle intersects the x-axis in two algebraically conjugate real quadratic surds. Pell's equation allows us to find $M \in SL(2,\mathbb{Z})$ which preserves one endpoint (and thus its conjugate). Applying this transformation to our geodesic re-scales it while preserving the orientation; hence the geodesic necessarily is a periodic geodesic on $\mathfrak{H}/PSL(2,\mathbb{Z})$. The set $C(\gamma)$ is then infinite because we can choose a set of approximating geodesics which cross between any pair of consecutive corners of γ, and the resulting limit cutting sequences are all distinct.

We complete this case by showing that if a periodic geodesic γ on $\mathfrak{H}/PSL(2,\mathbb{Z})$ hits exactly n corners of translates of fundamental domains in its period, then there are exactly $2n+2$ shift-equivalence classes in $[C(\gamma)]$. Label the corners which are hit in one period $c_1, \ldots, c_n$. The approximating geodesics γ_i can only be close in the symbol topology if they all cross γ between the same pair of corners, or if none cross γ at all. If they cross, then we can use the periodicity of γ's cutting sequence to shift the crossing point between c_1 and c_{n+1}. With n possible crossing regions, and crossing possible either from inside to outside or vice versa, there are $2n$ encodings; we get two more from approximating geodesics which are completely inside or outside γ, for a total of exactly $2n+2$ shift-equivalence classes. ■

Theorem 7.2 gives no uniform upper bound on the number of shift-equivalence classes of cutting sequences that correspond to periodic geodesics. We formulate the question whether a uniform upper bound exists as open problem (3) in the concluding section. There are periodic geodesics which correspond to 10 cutting sequence shift-equivalence classes because they hit four corners in one period, such as $\langle -\sqrt{13}, \sqrt{13} \rangle$ and $\langle -\sqrt{133}, \sqrt{133} \rangle$. Could this be the maximal number that occurs?

Theorem 7.1 implies that the shift $\Sigma_{\mathcal{F}}$ is a relatively complicated set. Indeed Theorem 6.3 showed that the set of minimal forbidden blocks is very large. We now show that $\Sigma_{\mathcal{F}}$ is not a sofic shift. A *sofic shift* is any shift that is a factor of a shift of finite type (see Marcus and Lind [32].) Alternatively, it is the set of possible bi-infinite walks on an edge-labeled finite graph.

Theorem 7.3. *The cutting sequence shift $\Sigma_{\mathcal{F}}$ for the fundamental domain $\mathcal{F}$ of $PSL(2,\mathbb{Z})$ is not a sofic shift.*

Proof. If Σ is a shift, the *follower set* $F_{\Sigma}(\mathsf{W})$ of a one-sided word $\mathsf{W} = (\ldots, \mathsf{S}_{-2}, \mathsf{S}_{-1})$ is the set of all one-sided words $\mathsf{W}^{+} = (\mathsf{S}_0, \mathsf{S}_1, \ldots)$ such that $\mathsf{W}\mathsf{W}^{+}$ is an element of Σ. Sofic shifts are characterized by the property that the totality of different follower sets $\{F_{\Sigma}(\mathsf{W}) : \text{all } \mathsf{W}\}$ is finite; see Marcus and Lind [32, Theorem 3.2.10].

We show that $\Sigma_{\mathcal{F}}$ has infinitely many follower sets. The sequences $[3, 2^{4j+2}, 1_c, 3, (8,4)^j, 10]$ and $[4, 2^{4j+2}, 1_c, (8,4)^j, 13]$ are central. Thus, in $[\ldots, 3, 2^{4j+2}, 1_*, 3, (8,4)^k, 13, \ldots]$, the 1_* is 1_h if $j \geq k$ but 1_m if $j < k$, regardless of the symbols on either side. Thus the follower sets of one-sided words which end $[3, 2^{4j+2}]$ are different for all j. ■

8. The Shift $\Sigma_{\mathcal{F}}$ Determines $\mathcal{F}$ up to Isometry

Our object is to prove the following result.

Theorem 8.1. *Let Γ be a finitely generated discrete subgroup of $PSL(2,\mathbb{R})$ that acts properly discontinuously on $\mathfrak{H}$, and which has a polygonal fundamental domain $\mathcal{P}$ which is hyperbolically convex. Suppose that the cutting sequence shift $\Sigma_{\mathcal{P}}$ is isomorphic to $\Sigma_{\mathcal{F},PSL(2,\mathbb{Z})}$ by a permutation of symbol alphabets. Then there is an element $g \in PSL(2,\mathbb{R})$ such that $\mathcal{F} = g\mathcal{P}$ and*

$$PSL(2,\mathbb{R}) = g\Gamma g^{-1} .$$

Proof. Since the cutting sequence contains only three symbols, the polygon must be a triangle.

Since a geodesic can hit the $\bar{\mathsf{L}}$ edge an unlimited number of consecutive times, the $\bar{\mathsf{L}}$ and $\bar{\mathsf{L}}^{-1}$ edges of the triangle must intersect at an angle of zero. The $\bar{\mathsf{L}}^{-1}$ edge cannot be the same as the $\bar{\mathsf{L}}$ edge because a geodesic can hit it twice consecutively, nor can it be the $\bar{\mathsf{J}}$ edge because no geodesic can hit that edge twice consecutively; thus it must be the $\bar{\mathsf{R}}$ edge. Since a geodesic cannot hit the $\bar{\mathsf{J}}$ edge twice consecutively, the $\bar{\mathsf{J}}$ edge must equal the $\bar{\mathsf{J}}^{-1}$ edge. Thus the generator at the $\bar{\mathsf{J}}$ edge is an involution, and to have determinant 1, it must be an inversion.

Let the angle between $\bar{\mathsf{J}}$ and $\bar{\mathsf{L}}$ be α, and the angle between $\bar{\mathsf{J}}$ and $\bar{\mathsf{R}}$ be β. If $2\alpha + \beta < \pi$, then a geodesic approaching the JR-corner could hit edges $\bar{\mathsf{J}}, \bar{\mathsf{L}}, \bar{\mathsf{J}}, \bar{\mathsf{L}}$ in sequence; however, this cutting sequence is not possible for our fundamental domain. Likewise, $2\beta + \alpha \geq \pi$. Also, for any hyperbolic triangle, the sum of the angles is less than π, so $\alpha + \beta < \pi$.

The conditions for a given polygon to be a fundamental domain are given in Maskit [33, section 2]. In our case, it is necessary that either $n(\alpha + \beta) = 2\pi$ for some n, or $n\alpha = \pi$ and $m\beta = \pi$ for some m and n. The only cases consistent with the conditions on α and β are $\alpha = \beta = \pi/3$, the desired fundamental domain; and $\alpha = \pi/3, \beta = \pi/2$, which is the half of our fundamental domain with $x > 0$. But that domain is only a fundamental domain for a group including the reflection in the line $x = 0$, which has determinant -1. (That domain also has different symbolic dynamics; a geodesic can hit $\bar{\mathsf{L}}$ and $\bar{\mathsf{R}}$ consecutively.) ■

9. Open Problems

(1). We have shown in one special case that the polygon $\mathcal{P}$ can be recovered from the data $\Sigma_{\mathcal{P}}$ up to isometry. The example $\mathcal{F}$ comes from a Riemann surface of genus 0. Does the result persist for higher-genus Riemann surfaces? Can any $\Sigma_{\mathcal{P}}$ be explicitly determined for a Riemann surface of genus at least one?

(2). Since $\Sigma_{\mathcal{F}}$ determines $\mathcal{F}$ up to isometry, in principle it determines $vol(\mathfrak{H}/PSL(2,\mathbb{Z}))$. Can $vol(\mathfrak{H}/PSL(2,\mathbb{Z})) = \frac{\pi}{3}$ be easily computed directly from $\Sigma_{\mathcal{F}}$?

(3). Is there a universal upper bound on the number of shift equivalence classes of cutting sequences corresponding to any periodic geodesic on $\mathfrak{H}/PSL(2,\mathbb{Z}$? Equivalently, is there a universal upper bound on the number of times that a periodic geodesic can hit a corner of the fundamental domain $\mathcal{F}$, during a single period?

(4). The zeta function $\zeta_\Sigma(z)$ of a shift Σ is defined by

$$\zeta_\Sigma(z) = \exp(\sum_{k=1}^{\infty} N_k \frac{z^k}{k}),$$

in which N_k counts the number of periodic words in Σ of period k. Is there a simple formula for the (dynamical) zeta function of $\Sigma_{\mathcal{F}}$? What is the topological entropy of $\Sigma_{\mathcal{F}}$?

(5). Cutting sequence shifts $\Sigma_{\mathcal{P}}$ can be constructed in higher-dimensional cases along the lines considered in [29], if one restricts to a suitable subclass of geodesics, called "flat" geodesics in [29]. Can any such $\Sigma_{\mathcal{P}}$ be determined explicitly?

Acknowledgment. We are indebted to L. Flatto and M. Sheingorn for helpful comments and references.

A. A Bound on the Number of Corners on a Vertical Geodesic

Lemma A.1. *For a rational θ, the vertical geodesic $\gamma = \{\theta + it : t > 0\}$ has at most one value of t such that $\theta + it$ is a corner of a $PSL(2,\mathbb{Z})$-translate of $\mathcal{F}$, unless $\theta \equiv \frac{1}{2} \pmod 1$, in which case it has exactly two such values, which are $t = \sqrt{3}/2$ and $\sqrt{3}/6$.*

Proof. Since the corner $-1/2 + \sqrt{-3}/2$ of $\mathcal{F}$ is obtained from the corner $1/2 + \sqrt{-3}/2$ by the transformation $z \to z - 1$, every corner is a $PSL(2,\mathbb{Z})$-translate of $1/2 + \sqrt{-3}/2$. Let the element of $SL(2,\mathbb{Z})$ be $\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$, with $ad - bc = 1$. We then have

$$\begin{aligned}\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]\left(\frac{1}{2} + \frac{\sqrt{-3}}{2}\right) &= \frac{a\left(\frac{1}{2} + \frac{\sqrt{-3}}{2}\right) + b}{c\left(\frac{1}{2} + \frac{\sqrt{-3}}{2}\right) + d} \\ &= \frac{(4ac + 2ad + 2bc + 4bd) + (2ad - 2bc)\sqrt{-3}}{4(c^2 + cd + d^2)} \\ &= \frac{(2ac + ad + bc + 2bd) + \sqrt{-3}}{2(c^2 + cd + d^2)}. \end{aligned} \tag{A.1}$$

Let $N = 2ac + ad + bc + 2bd$ and $D = c^2 + cd + d^2$, so that the real part of (A.1) is $N/2D$. We will show that $N/2D$ is either in lowest terms or can be reduced to lowest terms by dividing both N and D by 3. (The factor 3 can occur; for example, take $a = 2$, $b = 1$, $c = 1$, $d = 1$.) Since $ad - bc = 1$, $ad + bc$ is odd and thus N is not divisible by 2. Now substitute $a = (1 + bc)/d$ in (A.1); this gives

$$\begin{aligned}\frac{N}{2D} &= \frac{2\frac{1+bc}{d}c + 2bd + \frac{1+bc}{d}d + bc}{2D} \\ &= \frac{2c + 2bc^2 + 2bd^2 + d + 2bcd}{2d(c^2 + cd + d^2)} \\ &= \frac{b(c^2 + cd + d^2) + 2c + d}{2d(c^2 + cd + d^2)}. \end{aligned} \tag{A.2}$$

Since $(c, d) = 1$, we have $c^2 + cd + d^2 \equiv 1 \pmod 2$, and also $(d, c^2 + cd + d^2) = (d, c^2) = 1$. It follows that

$$\begin{aligned}(N, D) = (dN, D) = (bD + 2c + d, D) = (2c + D, D) = (2c + d, c^2 + cd + d^2) \\ = (2c + d, c^2 + cd + d^2 - d(2c + d)) = (2c + d, c(c - d)).\end{aligned}$$

Now, $(c, 2c + d) = (c, d) = 1$, and $(2c + d, c - d) = (3c, c - d) \leq (3, c - d)(c, c - d) = 1$ or 3. Therefore $(N, D) = 1$ or 3.

Suppose the geodesic is $\theta + it$ with $\theta = v/w$. There are only two possible values of (A.1) on this geodesic; $N/2D$ can only equal v/w in lowest terms if $w = 2D$ or $w = 2D/3$. The imaginary part of (A.1) is either $\sqrt{-3}/w$ or $\sqrt{-3}/3w$. However, these two points are separated by a hyperbolic distance of $\ln 3$, the length of the finite side of the domain $\mathcal{F}$; therefore, they can both be $PSL(2,\mathbb{Z})$-translates of corners only if they are connected by a $PSL(2,\mathbb{Z})$-translate of an edge. This edge is the geodesic $\theta + it$ itself, and the only vertical edges of translates $g\mathcal{F}$ occur when $\theta = (n + 1/2)$ for $n \in \mathbb{Z}$.

For $\theta = n + 1/2$, there are two $SL(2, \mathbb{Z})$-translates of corners. If we take $a = 1$, $b = n$, $c = 0$, $d = 1$, we get the corner $(n + 1/2) + \sqrt{-3}/2$; if we take $a = n + 1$, $b = n$, $c = 1$, $d = 1$, we get the corner $(n + 1/2) + \sqrt{-3}/6$. ■

References

[1] R. Adler, Geodesic flows, interval maps, and symbolic dynamics, in: *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces,* T. Bedford, M. Keane, and C. Series (eds.), Oxford University Press, Oxford 1991, pp. 93–123.

[2] R. Adler and L. Flatto, Cross Section Maps for Geodesic Flows I. The Modular Surface, in: A. Katok, Ed.,*Ergodic Theory and Dynamical Systems II*, Proceedings, Special Year, Maryland 1979–80, Birkhäuser,Boston 1982, pp. 103–161.

[3] R. Adler and L. Flatto, Geodesic flows, interval maps and symbolic dynamics, Bull. Amer. Math. Soc. **25** (1991), 229–334.

[4] P. Arnoux, Le codage des flot geódeśique sur la surface modulaire, Enseign. Math. **40** (1994), 29–48.

[5] E. Artin, Eine mechanische System mit quasiergodischen bahnen, Hamburger Math. Abhandlungen **3** (1924) 170–175. (*Collected Papers*, pp. 499-505.)

[6] A. F. Beardon, *The Geometry of Discrete Groups*, Springer-Verlag, New York 1983.

[7] G. D. Birkhoff, *Dynamical Systems*, AMS Colloquium Publ. No. 9, American Math. Society, Providence, RI 1927. (Revision: 1966, J. Moser).

[8] R. Bowen and C. Series, Markov maps associated with Fuchsian groups, Publ. Math. I.H.E.S. **50** (1979), 153–170.

[9] J. W. S. Cassels, *Rational Quadratic Forms,* Academic Press, New York 1978.

[10] I. P. Cornfeld, S. V. Fomin and Ya. G. Sinai, *Ergodic Theory*, Springer-Verlag, New York 1982.

[11] T. W. Cusick and M. Flahive, *The Markoff and Lagrange Spectra,* American Math. Society, Providence RI 1990.

[12] J. S. Frame, Continued Fractions and Matrices, Amer. Math. Monthly **56** (1949), 98–103.

[13] J. R. Goldman, Hurwitz sequences, the Farey process, and generalized continued fractions, Adv. in Math. **72** (1988), 239–260.

[14] P. M. Gruber and C. G. Lekkerkerker, *Geometry of Numbers*, North-Holland, Amsterdam 1987.

[15] J. Hadamard, Les surfaces a' corbures oppossees et leurs lignes geódésiques, J. de Math. **5** seris IV (1898), 27–73.

[16] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Fourth Edition), Oxford University Press, Oxford 1960.

[17] G. A. Hedlund, A metrically transitive group defined by the modular group, Amer. J. Math. **52** (1935), 668–678.

[18] G. A. Hedlund, The dynamics of geodesic flows, Bull. Amer. Math. Soc. **45** (1939), 241–260.

[19] C. Hermite, Sur l'introduction des variables continues dans la theorie des nombres, J. reine Angew. Math. **41** (1851), 191–216.

[20] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, Reading Mass. 1979.

[21] G. Humbert, Sur la méthode d'approximation d'Hermite, J. Math Pures Appl. (7th Series) **2** (1916), 70–103.

[22] G. Humbert, Sur les fractiones continues ordinaires et les formes quadatique binaires indéfinies, J. Math Pures Appl. (7th Series) **2** (1916), 104–154.

[23] S. Katok, Reduction theory for Fuchsian groups, Math. Annalen **273** (1986), 461–470.

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

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

[26] P. Köebe, Riemannsche Manningfaltigkeiten und nichteucklidische Raumfornen (Vierte Mitteilungen: Verlauf Geodatischer Linien), Sitzb. Preuss. Akad. Wiss. 1929, 414–457.

[27] K. Kolden, Continued fractions and linear substitutions, Arch. Math og Naturvid. **50** (1949), 141–196.

[28] J. C. Lagarias, Number theory and dynamical systems, in *The Unreasonable Effectiveness of Number Theory* (S. A. Burr, ed.), Proc. Symp. Appl. Math. No. 46, Amer.Math.l Soc., Providence RI, 1992, pp. 35–72.

[29] J. C. Lagarias, Geodesic multidimensional continued fractions, Proc. London Math. Soc., **69** (1994), 464–488.

[30] J. C. Lagarias and A. M. Pollington, The continuous Diophantine approximation mapping of Szekeres, J. Australian Math. Soc., Series A, **59** (1995), 148–172.

[31] J. Lehner, *Discontinuous Groups and Automorphic Functions*, Amer. Math. Soc., Providence, RI, 1964.

[32] J. D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding,* Cambridge University Press, Cambridge 1995.

[33] B. Maskit, On Poincaré's theorem for fundamental polygons, Advances in Math. **7** (1971), 219–230.

[34] R. Moeckel, Geodesics on modular surfaces and continued fractions, Ergodic Theory Dyn. Sys. **2** (1982), 69–83.

[35] M. Morse, A one-to-one representation of geodesics on a surface of negative curvature, Amer. J. Math. **43** (1921), 33–51.

[36] M. Morse, Recurrent geodesics on a surface of negative curvature, Trans. Amer. Math. Soc. **22** (1921), 84–100.

[37] M. Morse, Symbolic Dynamics, Institute for Advanced Study Lecture Notes 1966 (unpublished).

[38] M. Morse and G. A. Hedlund, Symbolic dynamics, Amer. J. Math. **60**, (1938), 815–866.

[39] J. Nielsen, Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächer, Acta. Math. **50** (1927), 189–358.

[40] P. J. Nicholls, Diophantine approximation via the modular group, J. London Math. Soc. **17** (1978), 11–17.

[41] G. N. Raney, On continued fractions and finite automata, Math. Annalen **206** (1973), 265–283.

[42] I. Richards, Continued fractions without tears, Math. Mag. **54** (1981), 163–171.

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

[44] C. Series, The modular surface and continued fractions, J. London Math. Soc. **31** (1985), 69–80. (Correction: see [47, p. 148].)

[45] C. Series, The geometry of Markoff numbers, Math. Intelligencer **7**, No. 3 (1985) 20–29.

[46] C. Series, Geometric Markov coding on surfaces of constant negative curvature, Ergod. Th. Dynam. Sys. **6** (1986), 601–625.

[47] C. Series, Geometrical methods of symbolic coding, in: *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces,* (T. Bedford, M. Keane, C. Series (Eds.), Oxford U. Press, Oxford, 1991, pp. 125–151.

[48] H. J. S. Smith, Memoire sur les equations modulaires, Atti,Accad. dei Lincei, Memorie d. Sci. phys. mat. nat. Ser. III, Vol I. 1877, 136–149. in *Collected Mathematical Papers, Volume II,* Chelsea: New York 1965, pp. 224–241.

[49] H. M. Stark, *An Introduction to Number Theory,* Markham, Chicago 1971 (Reprint: M.I.T. Press).

[50] B. A. Venkov, *Elementary Number Theory*, Walters-Noordhoff Publishing: Groningen, 1970.

email: `jcl@research.att.com`
`grabiner@wcnet.org`