SYMBOLIC DYNAMICS FOR THE GEODESIC FLOW ON HECKE SURFACES

DIETER MAYER AND FREDRIK STRÖMBERG

ABSTRACT. In this paper we discuss a coding and the associated symbolic dynamics for the geodesic flow on Hecke triangle surfaces. We construct an explicit cross section for which the first return map factors through a simple (explicit) map given in terms of the generating map of a particular continued fraction expansion closely related to the Hecke triangle groups. We also obtain explicit expressions for the associated first return times.

CONTENTS

1. Introduction 1
2. The geodesic flow on $T^1M$ 3
3. $\lambda$-Continued fraction expansions 4
4. Construction of the cross-section 25
5. Construction of an invariant measure 38
6. Lemmas on continued fraction expansions and reduced geodesics 40

1. INTRODUCTION

Surfaces of negative curvature and their geodesics have been studied since the 1898 work of Hadamard [15] (see in particular the remark at the end of §58). Inspired by the work of Hadamard and Birkhoff [6] Morse [31] introduced a coding of geodesics essentially corresponding to what is now known as „cutting sequences“ and used this coding to show the existence of a certain type of recurrent geodesics [32].

Further ergodic properties of the geodesic flow on surfaces of constant negative curvature given by Fuchsian groups were shown by e.g. Artin [5], Nielsen [36], Koebe [26], Löbell [29], Myrberg [33], Hedlund [16, 17, 18, 19], Morse and Hedlund [30] and Hopf [20, 21]. In this sequence of papers one can see the subject of symbolic dynamics emerging. For a more up-to-date account of the ergodic properties of the geodesic flow on a surface of constant negative curvature formulated in a modern language see e.g. the introduction in Series [47].

2000 Mathematics Subject Classification: Primary: 37D40, Secondary: 37E05, 11A55, 11K50.

This work has been supported by the German Research Foundation (DFG) under contract Ma 633/16-1. The authors appreciate the very helpful remarks of an anonymous referee.
Artin’s [5] approach was novel in that he used continued fractions to code geodesics on the modular surface. After Artin, coding and symbolic dynamics on the modular surface have been studied by e.g. Adler and Flatto [1, 2, 3] and Series [48]. For a recent review of different aspects of coding of geodesics on the modular surface see for example the expository papers by Katok and Ugarcovici [24, 25].

Other important references for the theory of symbolic dynamics and coding of the geodesic flow on hyperbolic surfaces are e.g. Adler-Flatto [4], Bowen and Series [7] and Series [47].

In the present paper we study the geodesic flow on a family of hyperbolic surfaces with one cusp and two marked points, the so-called Hecke triangle surfaces, generalizing the modular surface. Symbolic dynamics for a related billiard has also been studied by Fried [12]. We now give a summary of the paper. Sections 1 and 2 contain preliminary facts about hyperbolic geometry and geodesic flows. In Section 3 we develop the theory of \( \lambda \)-fractions connected to the coding of the geodesic flow on the Hecke triangle surfaces. The explicit discretization of the geodesic flow in terms of a Poincaré section and Poincaré map is developed in Section 4. As an immediate application we derive invariant measures for certain interval maps in Section 5. Some rather technical lemmas are confined to the end in Section 6.

1.1. Hyperbolic geometry and Hecke triangle surfaces. Recall that any hyperbolic surface of constant negative curvature \(-1\) is given as a quotient (orbifold) \( \mathcal{M} = \mathcal{H}/\Gamma \). Here \( \mathcal{H} = \{ z = x + iy \mid y > 0, x \in \mathbb{R} \} \) together with the metric \( ds = \frac{|dz|}{y} \) is the hyperbolic upper half-plane and \( \Gamma \subseteq \text{PSL}_2(\mathbb{R}) \equiv \text{SL}_2(\mathbb{R})/\{ \pm I \} \) is a Fuchsian group. Here \( \text{SL}_2(\mathbb{R}) \) is the group of real two-by-two matrices with determinant 1, \( I = (1,0; 0,1) \) and \( \text{PSL}_2(\mathbb{R}) \) is the group of orientation preserving isometries of \( \mathcal{H} \). The boundary of \( \mathcal{H} \) is \( \partial \mathcal{H} = \mathbb{R}^* = \mathbb{R} \cup \{ \infty \} \), \( \mathcal{H}^* = \mathcal{H} \cup \partial \mathcal{H} \).

If \( g = (a,b; c,d) \in \text{PSL}_2(\mathbb{R}) \) then \( gz = \frac{az+b}{cz+d} \in \mathcal{H} \) for \( z \in \mathcal{H} \), \( gx \in \partial \mathcal{H} \) for \( x \in \partial \mathcal{H} \) and we say that \( g \) is elliptic, hyperbolic or parabolic depending on whether \( |\text{Tr} g| = |a+d| < 2, > 2 \) or \( = 2 \). The same notation applies for fixed points of \( g \).

In the following we identify the elements \( g \in \text{PSL}_2(\mathbb{R}) \) with the map it defines on \( \mathcal{H}^* \). Note that the type of fixed point is preserved under conjugation \( g \mapsto AgA^{-1} \) by \( A \in \text{PSL}_2(\mathbb{R}) \). A parabolic fixed point is a degenerate fixed point, belongs to \( \partial \mathcal{H} \) and is usually called a cusp. Elliptic points appear in pairs, \( z, \bar{z} \) with \( z \in \mathcal{H} \) and \( \bar{z} \) belongs to the lower half-plane \( \mathcal{H}^- \) and \( \Gamma_z \), the stabilizer subgroup of \( z \) in \( \Gamma \), is cyclic of finite order. Hyperbolic fixed points appear also in pairs with \( x, x^* \in \partial \mathcal{H} \), where \( x^* \) is said to be the conjugate point of \( x \). A geodesics \( \gamma \) on \( \mathcal{H} \) is either a half-circle orthogonal to \( \mathbb{R} \) or a line parallel to the imaginary axis and the endpoints of \( \gamma \) are denoted by \( \gamma_+ \in \partial \mathcal{H} \). We identify the set of geodesics on \( \mathcal{H} \) with \( \mathcal{G} = \{ (\xi, \eta) \mid \xi \neq \eta \in \mathbb{R}^* \} \) and use \( \gamma (\xi, \eta) \) to denote the oriented geodesic on \( \mathcal{H} \) with \( \gamma_+ = \xi \) and \( \gamma_- = \eta \). Unless otherwise stated all geodesics are assumed to be parametrized with hyperbolic arc length with \( \gamma (0) \) either at height 1 if \( \gamma \) is vertical or the highest point on the half-circle. The tangent of \( \gamma \) at \( (t) \) is denoted by \( \dot{\gamma} (t) \). It is known that \( z \in \mathcal{H} \) and \( \theta \in [-\pi, \pi) \equiv S^1 \) determine a unique geodesic (cf. Lemma 62) passing through \( z \) whose tangent at \( z \) makes an angle \( \theta \) with the positive real axis. This geodesic is denoted by \( \gamma_{z, \theta} \). It is also well known that a
geodesic $\gamma(\xi, \eta)$ is closed if and only if $\xi$ and $\eta = \xi^*$ are conjugate hyperbolic fixed points.

The unit tangent bundle of $\mathcal{H}$, $T^1\mathcal{H} = \bigcup_{z \in \mathcal{H}} \{ \tilde{v} \in T_z\mathcal{H} \mid |\tilde{v}| = 1 \}$ is the collection of all unit vectors in the tangent planes of $\mathcal{H}$ with base points $z \in \mathcal{H}$ which we denote by $T^1_z\mathcal{H}$. By identifying $\tilde{v}$ with its angle $\theta$ with respect to the positive real axis we can view $T^1\mathcal{H}$ as the collection of all pairs $(z, \theta) \in \mathcal{H} \times S^1$. We may also view this as the set of geodesics $\gamma_{z, \theta}$ on $\mathcal{H}$ or equivalently as $G \subseteq \mathbb{R}^2$.

Let $\pi : \mathcal{H} \to \mathcal{M}$ be the natural projection map, i.e. $\pi(z) = \Gamma z$ and let $\pi^* : T^1\mathcal{H} \to T^1\mathcal{M}$ be the extension of $\pi$ to $T^1\mathcal{H}$. Then $\gamma^* = \pi \gamma$ is a closed geodesic on $\mathcal{M}$ if and only if $\gamma_+$ and $\gamma_-$ are fixed points of the same hyperbolic map $g_\gamma \in \Gamma$. For a more comprehensive introduction to hyperbolic geometry and Fuchsian groups see e.g. [23, 27, 41].

**Definition 1.** For an integer $q \geq 3$ the Hecke triangle group $G_q \subseteq \text{PSL}_2(\mathbb{R})$ is the group generated by the maps $S : z \mapsto -\frac{1}{z}$ and $T : z \mapsto z + \lambda$ where $\lambda = \lambda_q = 2 \cos \left(\frac{\pi}{q}\right) \in [1, 2)$. The corresponding orbifold (Riemann surface) is $\mathcal{M}_q = G_q \backslash \mathcal{H}$, which we sometimes identify with the standard fundamental domain of $G_q$

$$\mathcal{F}_q = \{ z \in \mathcal{H} \mid |\mathbb{R}z| \leq \lambda/2, |z| \geq 1 \}$$

with sides pairwise identified. Let $\rho = \rho_+ = e^{\frac{\pi i}{q}}$ and $\rho_- = -\overline{\rho}$. We define the following oriented boundary components of $\mathcal{F}_q$: $L_0$ is the circular arc from $\rho_-$ to $\rho_+$. $L_1$ is the vertical line from $\rho_+$ to $i\infty$ and $L_{-1}$ is the vertical line from $i\infty$ to $\rho_-$. Thus $\partial \mathcal{F}_q = L_{-1} \cup L_0 \cup L_1$ is the positively oriented boundary of $\mathcal{F}_q$.

**Remark 2.** $G_q$ is a realization of the Schwarz triangle group $\left(\frac{\pi}{q}, \frac{\pi}{q}, \frac{\pi}{2}\right)$ and it is not hard to show (see e.g. [27, VII]) that $G_q$ for $q \geq 3$ is a co-finite Fuchsian group with fundamental domain $\mathcal{F}_q$ and the only relations

$$S^2 = (ST)^3 = Id - \text{the identity in } \text{PSL}_2(\mathbb{R}).$$

Hence $G_q$ has one cusp, that is the equivalence class of parabolic points, and two elliptic equivalence classes of orders 2 and $q$ respectively. Note that $G_3 = \text{PSL}_2(\mathbb{Z})$ --the modular group and $G_4, G_6$ are conjugate to congruence subgroups of the modular group. For $q \neq 3, 4, 6$ the group $G_q$ is non-arithmetic (cf. [23, pp. 151-152]), but in the terminology of [9, 46] it is semi-arithmetic, meaning that it is possible to embed $G_q$ as a subgroup of a Hilbert modular group.

2. The Geodesic Flow on $T^1\mathcal{M}$

We briefly recall the notion of the geodesic flow on a Riemann surface $\mathcal{M} = \Gamma\backslash\mathcal{H}$ with $\Gamma \subseteq \text{PSL}_2(\mathbb{R})$ a Fuchsian group. To any $(z,\theta) \in T^1\mathcal{H} \cong \mathcal{H} \times S^1$ we can associate a unique geodesic $\gamma = \gamma_{z,\theta}$ on $\mathcal{H}$ such that $\gamma(0) = z$ and $\gamma'(0) = e^{i\theta}$.

The geodesic flow on $T^1\mathcal{H}$ can then be viewed as a map $\Phi_t : T^1\mathcal{H} \to T^1\mathcal{H}$ with $\Phi_t(\gamma_{z,\theta}) = \Phi_t(z,\theta) = (\gamma_{\xi(t),\theta}(t)), t \in \mathbb{R}$ satisfying $\Phi_{t+s} = \Phi_t \circ \Phi_s$. The geodesic flow $\Phi^*$ on $T^1\mathcal{M}$ is then given by the projection $\Phi^*_t = \pi^* (\Phi_t)$. 


A more abstract and general description of the geodesic flow, which can be extended to other homogeneous spaces, is obtained by the identification \(T^1\mathcal{H} \cong \text{PSL}_2(\mathbb{R})\). Under this representation the geodesic flow corresponds to right multiplication by the matrix \(a_t^{-1} = \begin{pmatrix} e^t/2 & 0 \\ 0 & e^{-t/2} \end{pmatrix}\) in \(\text{PSL}_2(\mathbb{R})\) (cf. e.g. [11, Ch. 13]).

**Definition 3.** Let \(\Upsilon\) be a set of geodesics on \(\mathcal{H}\). A hypersurface \(\Sigma \subseteq T^1\mathcal{H}\) is said to be a Poincaré section or cross section for the geodesic flow on \(T^1\mathcal{H}\) for \(\Upsilon\) if any \(\gamma \in \Upsilon\) intersects \(\Sigma\)

(P1): transversally i.e. non-tangentially, and

(P2): infinitely often, i.e. \(\Phi_{t_j}(\gamma) \in \Sigma\) for an infinite sequence of \(t_j \to \pm \infty\).

The corresponding first return map is the map \(T : \Sigma \to \Sigma\) such that \(T(z, \theta) = \Phi_{t_0}(z, \theta) \in \Sigma\) and \(\Phi_t(z, \theta) \notin \Sigma\) for \(0 < t < t_0\). Here \(t_0 = t_0(z, \theta) > 0\) is called the first return time.

Poincaré sections were first introduced by Poincaré [40] to show the stability of periodic orbits. For examples of cross section maps in connection with the geodesic flow on hyperbolic surfaces see e.g. [4, 1].

The previous definition extend naturally to \(T^1\mathcal{M}\) with \(\Upsilon\) and \(\Sigma\) replaced by \(\Upsilon^* = \pi(\Upsilon)\) and \(\Sigma^* = \pi^*(\Sigma)\). The first return map \(T\) is used to obtain a discretization of the geodesic flow, e.g. we replace \(\Phi_t(z, \theta)\) by \(\{\Phi_{t_k}(z, \theta)\}\) where \(t_k(z, \theta)\) is a sequence of consecutive first returns. Incidentally this provides a reduction of the dynamics from three to two dimensions and it turns out that in our example the first return map also has a factor map, which allows us to study the three dimensional geodesic flow with the help of an interval map (see Sections 4.3 and 5).

### 3. \(\lambda\)-continued fraction expansions

#### 3.1. Basic concepts

Continued fraction expansions connected to the groups \(G_q\), the so-called \(\lambda\)-fractions, were first introduced by Rosen [42] and subsequently studied by Rosen and others, cf. e.g. [43, 44, 45]. For the purposes of natural extensions (cf. Section 3.4) the results of Burton, Kraaikamp and Schmidt [8] are analogous to ours and we occasionally refer to their results. Our definition of \(\lambda\)-fractions is equivalent to Rosen’s definition (cf. e.g. [42, §2]).

To a sequence of integers, \(a_0 \in \mathbb{Z}\) and \(a_j \in \mathbb{Z}^* = \mathbb{Z}\setminus \{0\}, j \geq 1\) (finite or infinite) we associate a \(\lambda\)-fraction \(\underline{x} = [a_0; a_1, a_2, \ldots]\). This \(\lambda\)-fraction is identified with the point

\[x = a_0 \lambda - \frac{1}{a_1 \lambda - \frac{1}{a_2 \lambda - \cdots}} = \lim_{n \to \infty} T^{a_0} ST^{a_1} \cdots ST^{a_n} 0\]

if the right hand side is convergent. When there is no risk of confusion, we sometimes write \(\underline{x} = x\). For any \(m \geq 1\) we define the head \(\underline{x}(m)\) and the tail \(\underline{x}^{(m)}\) of \(\underline{x}\) by \(\underline{x}(m) = [a_0; a_1, \ldots, a_m]\) and \(\underline{x}^{(m)} = [a_{m+1}, a_{m+2}, \ldots]\). Note that \(-\underline{x} = [-a_0; -a_1, -a_2, \ldots]\). If \(a_0 = 0\), we usually omit the leading \([0];\). Repetitions in a sequence is denoted by a power, e.g. \([a, a, a] = [a^3]\) and an infinite repetition
is denoted by an overline, e.g. \( \overline{[a_1, \ldots, a_k, a_1, \ldots, a_k, \ldots]} = \overline{[a_1, \ldots, a_k]} \). Such a \( \lambda \)-fraction is said to be \emph{periodic} with period \( k \), or an \emph{eventually periodic} \( \lambda \)-fraction has a periodic tail. Two \( \lambda \)-fractions \( x \) and \( y \) are said to be \emph{equivalent} if they have the same tail. In this case it is easy to see that, if the fractions are convergent, then \( x = Ay \) for some \( A \in G_q \).

The sole purpose for introducing \( \lambda \)-fractions is to code geodesics by identifying the \( \lambda \)-fractions of their endpoints with elements of \( \mathbb{Z}^N \). For reasons that will be clear later (Section 3.4), we have to consider also bi-infinite sequences \( \mathbb{Z}^\infty \) and \( \mathbb{Z}^N \) as embedded in \( \mathbb{Z}^\infty \) with a zero-sequence to the left. On \( \mathbb{Z}^\infty \) and \( \mathbb{Z}^N \) we always use the metric \( \hat{h} \) defined by
\[
\hat{h} \left( \{a_i\}_{i=-\infty}^{\infty}, \{b_i\}_{i=-\infty}^{\infty} \right) = \frac{1}{1+n} \quad \text{where } a_i = b_i \text{ for } |i| < n \text{ and } a_n \neq b_n \text{ or } a_{-n} \neq b_{-n}.
\]
In this metric \( \mathbb{Z}^\infty \) and \( \mathbb{Z}^N \) have the topological structure of a Cantor set and the left- and right shift maps \( \sigma^\pm : \mathbb{Z}^\infty \to \mathbb{Z}^\infty, \sigma^\pm \{a_j\} = \{a_{j+1}\} \) are continuous. We also set \( \sigma^+ [a_1, a_2, \ldots] = [a_2, a_3, \ldots] \).

### 3.2. Regular \( \lambda \)-fractions

In the set of all \( \lambda \)-fractions we choose a „good“ subset, in which almost all \( x \in \mathbb{R} \) have unique \( \lambda \)-fractions and in which infinite \( \lambda \)-fractions are convergent. The first step is to choose a „fundamental region“ \( I_q \) for the action of \( T : \mathbb{R} \to \mathbb{R} \), namely \( I_q = \left[ -\frac{\lambda}{2}, \frac{\lambda}{2} \right] \). Then it is possible to express one property of our „good“ subset as follows: If in the fraction \( x = [a_0; a_1, a_2, \ldots] \) the first entry \( a_0 = 0 \), then \( x \in I_q \). That means, we do not allow sequences with \( a_0 = 0 \) to correspond to points outside \( I_q \).

A shift-invariant extension of this property leads to the following definition of regular \( \lambda \)-fractions:

**Definition 4.** Let \( x = [a_0; a_1, a_2, \ldots] \) be a finite or infinite convergent \( \lambda \)-fraction and let \( x_j = \sigma^j x = [0; a_j, a_{j+1}, \ldots] \), \( j \geq 1 \), be the \( j \)-th shift of \( x \). Let \( x_j \) be the corresponding point. Then \( x \) is said to be a \emph{regular \( \lambda \)-fraction} if and only if
\[
(*) \quad x_j \in I_q, \text{ for all } j \geq 1.
\]

A regular \( \lambda \)-fraction is denoted by \( [a_0; a_1, \ldots] \), the space of all regular \( \lambda \)-fractions is denoted by \( \mathcal{A}_q \) and the subspace of \emph{infinite} regular \( \lambda \)-fractions with \( a_0 = 0 \) is denoted by \( \mathcal{A}_{0,q} \).

For a finite fraction \( x = [a_0; a_1, \ldots, a_n] \) we get \( x_j = [0; \ldots] \) and \( x_j = 0 \in I_q \) for \( j > n \).

We will see later that regular \( \lambda \)-fractions can be regarded as \emph{nearest \( \lambda \)-multiple continued fractions}. In the case \( q = 3 \) or \( \lambda = 1 \), nearest integer continued fractions were studied already by Hurwitz [22] in 1889. An account of Hurwitz reduction theory can be found in Fried [13] (cf. also the H-expansions in [24, 25]). For general \( q \) this particular formulation of Rosen’s fractions was studied by Nakada [35].

For the remainder of the paper we let \( h = \frac{q-3}{2} \) if \( q \) is odd and \( h = \frac{q-2}{2} \) if \( q \) is even. The following Lemma is an immediate consequence of [8, (4)].
**Lemma 5.** The points $\mp \frac{\lambda}{2}$ have finite regular $\lambda$-fractions given by

$$\mp \frac{\lambda}{2} = \begin{cases} \left(\frac{\pm 1}{2}\right)^h, & \text{for } q \text{ even}, \\ \left(\frac{\pm 1}{2}, \pm 2, \left(\frac{\pm 1}{2}\right)^h\right), & \text{for } q \text{ odd}. \end{cases}$$

**Lemma 6.** If $q$ is odd, the point $x = 1$ has the finite regular $\lambda$-fraction

$$1 = \left[1; \frac{1}{h}\right].$$

**Proof.** Since $a = \left[1; \frac{1}{h}\right] = T(ST)^{h} 0$, one has also $-a = T^{-1} (ST^{-1})^{h} 0$. From identity (1) we get

$$Sa = (ST)^{h+1} 0 = (T^{-1} S)^{h+2} 0 = T^{-1} (ST^{-1})^{h} ST^{-1} S 0 = T^{-1} (ST^{-1})^{h} 0,$$

and hence $-1/a = -a$. Since $a > 0$, this implies that $a = 1$. 

**Definition 7.** Let $\lfloor x \rfloor$ be the floor function defined by $\lfloor x \rfloor = n \iff n \leq x < n + 1$ for $x > 0$, respectively $n \leq x < n + 1$ for $x \leq 0$, and let $(x)_\lambda = \left[\frac{a}{h} + \frac{1}{2}\right]$ be the corresponding nearest $\lambda$-multiple function. Then define $F_q : I_q \to I_q$ by

$$F_q x = \begin{cases} -\frac{1}{x} - \lfloor -\frac{1}{x} \rfloor \lambda, & x \in I_q \setminus \{0\}, \\ 0, & x = 0. \end{cases}$$

**Lemma 8.** For $x \in \mathbb{R}$ the following algorithm gives a finite or infinite regular $\lambda$-fraction $c_q(x) = [a_0; a_1, \ldots]$ corresponding to $x$:

(i) Set $a_0 := (x)_\lambda$ and $x_1 = x - a_0 \lambda$.

(ii) Set $x_{j+1} := F_q x_j = -\frac{1}{x_j} - a_j \lambda$, $j \geq 1$, with $a_j = \lfloor -\frac{1}{x_j} \rfloor \lambda$, $j \geq 1$.

If $x_j = 0$ for some $j$, the algorithm stops and gives a finite regular $\lambda$-fraction.

**Proof.** By definition we see that $x_{j+1} = T^{-a_j} S x_j, j \geq 1$, and it follows that $x = T^{a_0} S x_1 \ldots S x_n$ for any $n \geq 1$. If $x = [a_0; a_1, \ldots]$ , then for $j \geq 1$ $x_j = \sigma^j x = [0; a_j, \ldots]$ corresponds to the point $x_j$ and condition (*) of Definition 4 is fulfilled, since $F_q$ maps $I_q$ to itself and $x_1 \in I_q$.

**Remark 9.** We say that $F_q$ is a generating map for the regular $\lambda$-fractions. It is also clear from Lemma 8 that $F_q$ acts as a shift map on the space $A_{0,q}$, i.e. $c_q(F_q x) = \sigma c_q(x)$.

An immediate consequence of Lemma 8 is the following corollary:

**Corollary 10.** If $x$ has an infinite regular $\lambda$-fraction, then it is unique and equal to $c_q(x)$ as given by Lemma 8.

The above choice of floor function implies that $F_q$ is an odd function and that $(\pm \frac{1}{2})_\lambda = 0$ in agreement with Lemma 5. The ambiguity connected to the choice of floor function at integers affects only the points $x = \frac{1}{\lambda(1-2k)}$ where $\frac{1}{\lambda} + \frac{1}{2} = k \in \mathbb{Z}$ and $F_q x = (k - \lfloor k \rfloor)_\lambda - \frac{1}{2} = \pm \frac{1}{2}$. By Lemma 5 we conclude, that any point, which has more than one regular $\lambda$-fraction, is $F_q$-equivalent to $\pm \frac{1}{2}$ and hence has a finite $\lambda$-fraction.
Let $q = 2h + 2$. Then the blocks $[(a, (±1)^h, ±1)]$ for $m \geq 1$ are forbidden. The block $[(a, (±1)^h, ±, b)]$ with $a \neq ±1$ and $b \neq ±1$ if $m = 1$, can be rewritten as

$$
[(a, (±1)^h, ±, m, b)] \rightarrow \begin{cases}
[a \neq 1, (±1)^h, ±, m \neq 1, b], & \text{if } m \geq 2, \\
[a \neq 1, (±1)^{h-1}, b \neq 1], & \text{if } m = 1 \text{ and } h \geq 2, \\
[a \neq 1, b \neq 1], & \text{if } m = 1 \text{ and } h = 1.
\end{cases}
$$

Proof. Since by Lemma 5 $[(±1)^h]$, the blocks $[(±1)^h, ±m]$ for $m \geq 1$ are forbidden. Using then the relation $(ST)^{2h+2} = I$ we can restrict ourselves to the case where the number of consecutive 1’s is smaller than or equal to $h + 1$. Then the rewriting rules follow immediately from the relation $(ST)^{2h+2} = I$.

Remark 12. Obviously the rewritten block $[(a, b \neq 1)]$ is forbidden itself if $h = 1$, i.e. $q = 4$ and $a = 2$ and $b \geq 2$ or $a = -2$ and $b \leq -2$ in the case of minus and plus sign respectively. How to get allowed blocks in this case after repeated rewriting will be discussed in Lemma 19.

Lemma 13. Let $q \geq 5$ be odd with $q = 2h + 3$. Then the blocks $[(±1)^{h+1}]$ and $[(±1)^h, ±2, (±1)^h, ±m]$ with $m \geq 1$ are forbidden. The block $[(a, (±1)^{h+1}, b)]$ for $a \neq ±1, b \neq ±1$ can be rewritten as

$$
[(a, (±1)^{h+1}, b)] \rightarrow [a \neq 1, (±1)^h, b \neq 1],
$$

the blocks $[(a, (±1)^h, ±2, (±1)^h, ±m, b)]$ with $a \neq ±1$ can be rewritten as

$$
[(a, (±1)^h, ±2, (±1)^h, ±m, b)] \rightarrow \begin{cases}
[a \neq 1, (±1)^h, ±2, (±1)^h, ±m \neq 1, b], & m \geq 2, \\
[a \neq 1, (±1)^{h-1}, ±2, (±1)^{h-1}, b \neq 1], & m = 1.
\end{cases}
$$

For $q = 3$ the blocks $[(±1)]$ and $[(±2, ±m)]$ with $m \geq 1$ are forbidden. The block $[(a, ±1, b)]$ with $a \neq ±1, b \neq ±1$ and $b \neq ±1$ can be rewritten as

$$
[(a, ±1, b)] \rightarrow [a \neq 1, b \neq 1].
$$

The block $[(a, ±2, ±m, b)]$ with $a \neq ±1, ±2$ can be rewritten as

$$
[(a, ±2, ±m, b)] \rightarrow \begin{cases}
[a \neq 1, ±2, ±m \neq 1, b], & m \geq 2, \\
[a \neq 1, b = 2], & m = 1.
\end{cases}
$$

Proof. Since $(ST)^{2h+3} = I$ we can restrict ourselves again to blocks with no more than $h+1$ consecutive 1’s. From Lemma 5 we know that $[(±1)^h, ±2, (±1)^h]$. 


and hence the forbidden blocks follow immediately. Using then the relation \((ST)^{2h+3}\) = I gives the rewriting rules.

**Remark 14.** In the first rewriting rule for \(q = 3\) the rewritten block \([a \mp 1, b \mp 1]\) is forbidden if \(a = 3\) and \(b \geq 2\) or \(a = -3\) and \(b \leq -2\) (in the case of minus and plus signs respectively) or if \(b = \pm 2\). The rewritten block \([a \mp 1, \mp 2, \pm m \mp 1, b]\) is itself forbidden if \(m = 2\) or \(m = 3\) and \(b \geq 1\) respectively \(b \leq -1\) (in the case of minus and plus signs respectively) or \(b = \pm 1\). The rewritten block \([a \mp 1, b \mp 1]\) is forbidden in this case if \(a = 3\) and \(b \geq 3\) or if \(a = -3\) and \(b \leq -3\) (in the case of minus and plus signs respectively). How to get allowed blocks in these cases after repeated rewriting will be discussed in the proofs of Lemmas 18 and 20.

**Remark 15.** It is easy to see that Rosen’s \(\lambda\)-fractions [42] can be expressed as words in the generators \(T, S\) and \(JS\) of the group \(G_q^* = \langle G_q, J \rangle \subseteq \text{PGL}_2(\mathbb{R})\), where \(J : z \mapsto -\bar{z}\) is the reflection in the imaginary axis, i.e. \(JSx = \frac{1}{2}x\) for \(x \in \mathbb{R}^*\). Since \(J\) is an involution of \(G_q\), e.g. \(JTJ = T^{-1}\) and \(JSJ = S\), it is easy to see that Rosen’s and our notions of \(\lambda\)-fractions are equivalent: e.g. in the \(G_q\)-word identified with our \(\lambda\)-fraction we replace any \(T^{-a}\) by \(JT^aJ\), \(a \geq 1\). Algorithmically this means for a \(\lambda\)-fraction with entries \(a_j\) that the corresponding Rosen fraction has entries \((\epsilon_j, |a_j|)\) where \(\epsilon_1 = -\text{sign}(a_1)\) and \(\epsilon_j = -\text{sign}(a_{j-1}a_j)\) for \(j \geq 2\).

From the definition of regular \(\lambda\)-fractions it is clear that Rosen’s reduced \(\lambda\)-fractions [42, Def. 1] correspond to a fundamental interval \([0, \frac{\lambda}{2}\]) for the action of the group \(\langle T, J \rangle\) together with the choices made for finite fractions in [42, Def. 1 (4)-(5)]. It is easy to verify, for example using the forbidden blocks, that a finite fraction not equivalent to \(\pm \frac{\lambda}{2}\) or an infinite regular \(\lambda\)-fraction correspond to a reduced \(\lambda\)-fraction of Rosen. The main difference between our regular and Rosen’s reduced \(\lambda\)-fractions is that any \(\lambda\)-fraction equivalent to \(\pm \frac{\lambda}{2}\) has two valid regular \(\lambda\)-fractions. The root of this non-uniqueness is our choice of a closed interval \(I_q\) which is in turn motivated by our Markov partitions in Section 3.5.

It is then clear, that those results of [42] and [8] pertaining to infinite reduced \(\lambda\)-fractions can be applied directly to our regular \(\lambda\)-fractions.

**Lemma 16.** An infinite \(\lambda\)-fraction without forbidden blocks is convergent.

*Proof.* This follows from [42, Thm. 5] and Remark 15.

An immediate consequence of Definition 4 and Lemmas 11, 13 and 16 is the following

**Corollary 17.** A \(\lambda\)-fraction is regular if and only if it does not contain any forbidden block.

Rewriting a forbidden block may produce new f.b.’s and by rewriting a f.b. completely we mean that we rewrite also all new f.b.’s that arises. For the reduction procedure cf. Section 3.6 it is important that rewriting a f.b. completely can be done in most cases without affecting the head of the \(\lambda\)-fraction (up to some point).

**Lemma 18.** Suppose that the \(\lambda\)-fraction \(x = [a_0; a_1, \ldots]\) has the first forbidden block beginning at \(a_n\), \(n \geq 2\). If \(q \geq 5\) or \(q = 4\) and the forbidden block is not of
the type $[1^2]$, then the head of $x$ up to $n - 2$, i.e. $x_{(n-2)} = [a_0; a_1, \ldots, a_{n-2}]$ is not affected by rewriting this forbidden block completely.

**Proof.** For simplicity consider an initial f.b. containing $+1$’s, the blocks with $-1$’s are treated analogously. By using the relation $(ST)^q = 1$ we may assume that there are no blocks of consecutive $\pm 1$’s of length greater than $\frac{q}{2}$. The analogue of Lemmas 11 and 13 in this case is very simple: $[a, 1^j, b] \rightarrow [a - 1, (-1)^{q-j-2}, b - 1]$ for any $j > \frac{q}{2}$.

Suppose first that $q$ is even and that the first f.b. begins with $a_n$, i.e.

$$x = [a_0; a_1, \ldots, a_{n-2}, a_{n-1}, 1^h, a_{n+h}, a_{n+h+1}, \ldots], a_{n-1} \neq 1, a_{n+h} \geq 1.$$

By applying Lemma 11 we rewrite $x$ into either $[a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, a_{n+h} - 1, a_{n+h+1}, \ldots]$ if $a_{n+h} \geq 2$ or $[a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, a_{n+h+1} - 1, a_{n+h+2}, \ldots]$ if $a_{n+h} = 1$ (here $a_{n+h+1} \neq 1$). If this rewriting did not produce a new f.b. we are done so suppose that a new f.b. was created. Note that unless $q = 4$ and $a_n = a_{n+1} = 1$ (this case will be treated in Lemma 19) we have $a_{n+h} \geq 2$ and $h \geq 1$ or $a_{n+h} = 1$ and $h \geq 2$. In this case there is a non-empty block of $-1$’s starting at position $n$ and any new f.b. has to either end before or begin after the block of $-1$’s.

If the new f.b. appears to the left of position $n$ it clearly has to end with $a_{n-1} - 1$ but sign $(a_{n-1}) = \text{sign} (a_{n-1} - 1)$ so any such f.b. had to be forbidden also before the rewriting, contradicting the assumption about the position of the first f.b. in $x$. Any new f.b. beginning directly after the $-1$’s has to begin with $+1$ so rewriting it will only change the last $-1$ to $-2$ and there are at least two digits between $a_{n-2}$ and any new f.b.’s.

Now suppose that $q \geq 5$ is odd. There are two different types of f.b.’s but their treatments are very similar. Assume first that we have

$$x = [a_0; a_1, \ldots, a_{n-2}, a_{n-1}, 1^h, 1, a_{n+h}, a_{n+h+1}, \ldots], a_{n-1}, a_{n+h+1} \neq 1,$$

then by Lemma 13 we rewrite $x$ into $x_1 = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, a_{n+h+1} - 1, \ldots]$. There are four possibilities to create a new f.b. in $x_1$:

1. If the f.b. ends at $a_{n-1} - 1$ then it was also forbidden in $x$, since sign $(a_{n-1}) = \text{sign} (a_{n-1} - 1)$, contradicting the assumption that the first f.b. began with $a_n$.
2. If the f.b. ends with the $(-1)^h$ then $x_1 = [a_0; a_1, \ldots, (-1)^h, -2, (-1)^h, a_{n+h+1} - 1, \ldots]$ with $a_{n+h+1} \leq -2$ and $a_{n-1} = \cdots = a_{n-2} = -1$ so the f.b. $\ll (-1)^h \rr$ beginning at $a_{n-2} = -1$ would have been present in $x$, also contradicting the assumption on the first f.b.
3. If the new f.b. begins with the $(-1)^h$ then $a_{n+h+2} = -2$ and $x_2 = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, -2, (-1)^h, a_{n+h+2}, \ldots]$. Rewriting gives $x_2 = [a_0; a_1, \ldots, a_{n-2}, a_{n-1}, 1^h, 2, 1^h, a_{n+h+2} + 1, \ldots]$. 


Since \( a_{n-1} \neq 1 \) any new f.b. in \( x_2 \) must begin with \( a_{n+2h+2} = 1 = -1 \) in which case rewriting it only changes \([\ldots, a_{n-1}, 1^h, 2, 1^h, \ldots]\) to \([\ldots, a_{n-1}, 1^h, 2, 1^h-1, 2, \ldots]\) and there are at least four digits between \( a_{n-2} \) and any new f.b.

4. If the f.b. begins after the \((-1)^h\) then \( x_1 = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, 1^{k+1}, \ldots] \) or \( x_1 = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, 1^h, 2, 1^h, a_{n+3h+2} \ldots] \)

with \( a_{n+3h+2} \geq 1 \). Rewriting this f.b. thus changes the last \(-1\) to \(-2\) and there are at least two digits between \( a_{n-2} \) and any new f.b.

In the second case, assume that we have

\[ x = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, 2, (-1)^h, a_{n+2h+1} - 1, \ldots], \]

then by Lemma 13 we can rewrite \( x \) into either

\[ x_3 = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, -2, (-1)^h, a_{n+2h+1} - 1, \ldots], \]

if \( a_{n+2h+1} \geq 2 \), or

\[ x_4 = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, (-1)^h, -2, (-1)^h-1, a_{n+2h+2} - 1, \ldots], \]

if \( a_{n+2h+1} = 1 \). If \( x_3 \) or \( x_4 \) contains a new forbidden block, a similar argument as above tells us that it must begin with either \( a_{n+2h+1} - 1 = 1 = a_{n+2h+2} - 1 = 1 \) in which case rewriting it will only change the last \(-1\) to \(-2\) if \( h \geq 2 \) or the \(-2\) to \(-3\) if \( h = 1 \). In all cases there are at least 3 digits between \( a_{n-2} \) and any new f.b.

We have shown that for any f.b. we can rewrite it without changing \( a_{n-2} \) or any digit to the left of it and after this rewriting any new f.b.’s are separated from \( a_{n-2} \) by at least 2 digits. A recursive application of the above argument thus shows that in any case any further rewriting will not change the head \( x_{(m-2)} \).

**Lemma 19.** Let \( q = 4 \) and suppose that

\[ x = [a_0; a_1, a_2, \ldots, a_{n-l-1}, (\pm 2)^l, (\pm 1)^k, (\pm 2)^k, a_{n+k+2}, \ldots] \]

has precisely one forbidden block, which begins with \( a_n = a_{n+1} = \pm 1 \) and \( a_{n-l-1}, a_{n+k+2} \neq \pm 2 \). Then a complete rewriting of this forbidden block does not affect the head \( x_{(n-m)} = [a_0; a_1, \ldots, a_{n-m}] \) where \( m = 0 \) if \( l = 0 \), \( m = 3 \) if \( l \geq 1 \) and \( k = 0 \) and \( m = \min (k+3, l+3) \) otherwise. In all cases we assume that \( n \geq m \).

**Proof.** Suppose without loss of generality that \( x = [a_0; a_1, \ldots, a_{n-l-1}, 2^l, 1^2, a_{n+2}, \ldots] \) with \( n \geq l + 1 \geq 1 \), \( a_n = a_{n+1} = 1 \), \( a_{n-l-1} \neq 2 \) and \( a_{n-l} \neq 1 \). If \( l = 0 \) then \( a_{n-1} \neq 1 \). The case of \((-1)^2\) is analogous.

If \( l = 0 \) then one rewriting of \( x \) produces \( x_0 = [a_0; a_1, \ldots, a_{n-1} - 1, a_{n+2} - 1, \ldots] \) which does not contain a new f.b. unless \( a_{n+2} = 2 \). If \( a_{n+2} = 2 \) and \( a_{n+3} \geq 2 \) the new f.b. is of the type covered in Lemma 18 and rewriting it completely does not change \( x_{(n-2)} \). If \( a_{n+2} = 2 \) and \( a_{n+3} = 1 \) then \( a_{n+4} \leq -1 \) and \( x_0 = [a_0; a_1, \ldots, a_{n-1} - 1, 1^2, a_{n+4} \ldots] \) which is rewritten into \( x_1 = [a_0; a_1, \ldots, a_{n-1} - 2, a_{n+4} - 1, \ldots] \) which contains no f.b., and hence \( x_{(n-2)} \) is not changed.
Suppose that \( l \geq 1 \). If \( a_{n+2} \leq -1 \) one rewriting produces \([a_0; a_1, \ldots, a_{n-l-1}, 2^{l-1}, 1, a_{n+2} - 1, \ldots]\) which contains no f.b. and does not change \( \mathcal{E}(n-2) \).

If \( a_{n+2} \geq 3 \) then two rewritings produce \([a_0; a_1, \ldots, a_{n-l-1} - 1, -1, a_{n+2} - 2, \ldots]\) if \( l = 1 \) and \( a_{n+2} = [a_0; a_1, \ldots, a_{n-l-1}, 2^{l-2}, 1, -1, a_{n+2} - 2, \ldots]\) if \( l \geq 2 \). The first rewritten \( \lambda \)-fraction does not contain any f.b. but the second may contain a new f.b. if \( a_{n+2} = 3 \) and \( a_{n+3} \geq 1 \). If \( a_{n+2} = 3 \) and \( a_{n+3} \geq 2 \) this is a f.b. of the type \([1, m]\) with \( m \geq 2 \) and by Lemma 18 we can rewrite it completely without changing \( \mathcal{E}(n-3) \). If \( a_{n+2} = 3 \) and \( a_{n+3} = 1 \) then \( a_{n+4} \leq 1 \) and we rewrite \( a_2 \) into \([a_0; a_1, \ldots, a_{n-l-1}, 2^{l-2}, 1, -2, a_{n+4} - 1, \ldots]\) which contains no f.b., without changing \( \mathcal{E}(n-3) \).

The remaining case is when \( a_{n+2} = 2 \). Assume that \( \mathcal{E} = \{a_0; a_1, \ldots, a_{n-l-1}, 2^l, 1^2, 2^k, a_{n+k+2}, \ldots\}\) with \( a_{n-l-1}, a_{n+k+2} \neq 2 \) and \( k \geq 1 \). It is easy to see that \([2, 1^2, 2]\) is rewritten into \([1^2]\) so if \( l \geq k \) we rewrite \( \mathcal{E} \) in \( k \) steps into \( \mathcal{E}_1 = \{a_0; a_1, \ldots, a_{n-l-1}, 2^{l-k}, 1^2, a_{n+k+2}, \ldots\}\) which, as has been shown above, can be rewritten completely without changing \( \mathcal{E}(n-k-3) \) (the last 2 before the \( 1^2 \) is \( a_{n-k+1} \)). If \( l < k \) we rewrite \( \mathcal{E} \) in \( l \) steps into \([a_0; a_1, \ldots, a_{n-l-1}, 2^l, 2^k, a_{n+k+2}, \ldots\]\) which is then rewritten into \( \mathcal{E}_2 = \{a_0; a_1, \ldots, a_{n-l-1} - 1, 1, 2^{k-l-1}, a_{n+k+2}, \ldots\} \). If either \( k \geq l + 2 \) or \( k = l + 1 \) and \( a_{n+k+2} \geq 3 \) then \( \mathcal{E}_4 \) contains a new f.b. of the type \([1, m]\) with \( m \geq 2 \) which, by Lemma 18, can be rewritten completely without changing \( \mathcal{E}(n-l-2) \). If \( k = l + 1 \) and \( a_{n+k+2} = 1 \) then \( a_{n+k+3} \leq -1 \) and \( \mathcal{E}_4 = \{a_0; a_1, \ldots, a_{n-l-1} - 1, 1^2, a_{n+k+3}, \ldots\}\) which we have shown above can be rewritten completely without changing \( \mathcal{E}(n-l-2) \).

We have shown that the f.b. in \( \mathcal{E} = \{a_0; a_1, \ldots, a_{n-l-1}, 2^l, 1^2, 2^k, a_{n+k+2}, \ldots\}\) can be rewritten completely without changing \( \mathcal{E}(n-m) \) where \( m = 2 \) if \( l = 0 \) and \( m = 3 \) if \( l \geq 1 \) and \( k = 0 \) respectively \( m = \min(l + 3, k + 3) \) otherwise. (Of course we can do better in certain cases but we are mainly interested in whether \( m \) is finite or not.)

**Lemma 20.** Let \( q = 3 \) and suppose that \( \mathcal{E} = \{a_0; a_1, \ldots\} \) has only one forbidden block, which begins at \( a_n \). Assume further that \( \mathcal{E} \) contains at most one digit \( \pm 1 \). Then a complete rewriting of \( \mathcal{E} \) does not change the head \( \mathcal{E}(n-m) \) where \( m = \min(l + 3, k + 4) \) if \( \mathcal{E} = \{a_0; a_1, \ldots, a_{n-l-1}, (\pm 3)^l, \pm 2, \pm 1, (\pm 3)^k, a_{n+k+2}, \ldots\}\), \( l, k \geq 0 \), with \( a_{n-l-1} \neq \pm 2, \pm 3 \) and \( a_{n+k+2} \neq \pm 3 \) and \( m = 3 \) otherwise. We assume in all cases that \( n \geq m \).

**Proof.** Assume without loss of generality that \( a_n \geq 1 \). The case of \( a_n \leq -1 \) is analogous. There are three different f.b.’s to consider. Suppose first that

\[ \mathcal{E} = \{a_0; a_1, \ldots, a_{n-2}, a_{n-1}, 1, a_{n+1}, \ldots\}, \]

where \( a_{n-1} \neq 1, 2, a_{n+1} \neq \pm 1 \) and if \( a_{n+1} = \pm 2 \) then \( (a_{n+2}) = \pm (a_{n+1}) \).

By Lemma 13 we rewrite \( \mathcal{E} \) into \( \mathcal{E}_0 = \{a_0; a_1, \ldots, a_{n-2}, a_{n-1} - 1, a_{n+1} - 1, a_{n+2}, \ldots\} \).

Since \( (a_{n+1}) = \pm (a_{n+1} - 1) \) there can be no new f.b. ending at \( a_{n-1} - 1 \) so there are only three possibilities to create a new f.b., either \( a_{n-1} = 3 \) and \( a_{n+1} \geq 2 \) or \( a_{n+1} = 3 \) and \( a_{n+2} \geq 2 \) or \( a_{n+1} = 2 \):
1. If \( a_{n-1} = 3 \) then \( x_0 = \left[ a_0; a_1, \ldots, a_{n-2}, 2, a_{n+1} - 1, a_{n+2} \ldots \right] \) and if \( a_{n+1} \geq 3 \) we rewrite \( x_0 \) into \( x_1 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, -2, a_{n+1} - 2, a_{n+2} \ldots \right] \) and if \( a_{n+1} = 2 \) into \( x_2 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, a_{n+2} - 2, a_{n+3} \ldots \right] \). Since \( a_{n-2} \neq -1 \) and \( a_{n+2} \leq -2 \) in \( x_2 \) there can not be any new f.b.'s in \( x_2 \). Since the tail \( \left[ a_{n+1}, a_{n+2} \ldots \right] \) does not contain any more forbidden blocks we can only obtain a new f.b. in \( x_1 \) if \( a_{n+1} = 4 \) and \( a_{n+2} \geq 2 \) or \( a_{n+1} = 3 \).

(a) If \( a_{n+1} = 4 \) and \( a_{n+2} \geq 2 \) then \( x_1 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, -2, 2, a_{n+2} \ldots \right] \) which is rewritten into \( \left[ a_0; a_1, \ldots, a_{n-2} - 1, -3, -2, a_{n+2} - 1 \ldots \right] \) and there is at least 3 digits between \( a_{n-3} \) and any new f.b.

(b) If \( a_{n+1} = 3 \) then \( x_1 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, -2, 1, a_{n+2} \ldots \right] \) which is rewritten into \( \left[ a_0; a_1, \ldots, a_{n-2} - 1, -3, a_{n+2} - 1, a_{n+3} \ldots \right] \) and there is at least 2 digits between \( a_{n-3} \) and any new f.b.

2. If \( a_{n-1} \neq 3 \), \( a_{n+1} = 3 \) and \( a_{n+2} \geq 2 \) then \( x_0 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, 2, a_{n+2} \ldots \right] \) is rewritten into \( x_1 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, a_{n-1} - 2, -2, a_{n+2} - 1, \ldots \right] \) and there is at least 3 digits between \( a_{n-3} \) and any new f.b.

3. If \( a_{n-1} \neq 3 \), \( a_{n+1} = 2 \) then \( x_0 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, 1, a_{n+2} \ldots \right] \) is rewritten into \( x_1 = \left[ a_0; a_1, \ldots, a_{n-2} - 1, -2, a_{n+2} - 1, a_{n+3} \ldots \right] \) and since \( a_{n+2} \leq -2 \) there is at least 3 digits between \( a_{n-3} \) and any new f.b.

We have shown that a single rewriting of a f.b. of the type \( \pm 1 \) does not change the \( \lambda \)-fraction more than two steps to the left of the beginning of the f.b.

The second type of forbidden block is:

\[
\mathcal{P} = \left[ a_0; a_1, \ldots, a_{n-1} - 2, a_{n+1}, a_{n+2} \ldots \right], \quad a_{n-1} \neq \pm 1, 2,
\]

where \( a_{n+1} \geq 2 \) and we rewrite \( \mathcal{P} \) into \( x_5 = \left[ a_0; a_1, \ldots, a_{n-1} - 1, -2, a_{n+1} - 1, a_{n+2} \ldots \right] \). Then any new f.b. in \( x_5 \) must begin with \( a_{n+1} = 2 \) or \( a_{n+1} = 3 \) and \( a_{n+2} \geq 2 \). For \( a_{n+1} = 2 \) rewriting leads to \( x_6 = \left[ a_0; a_1, \ldots, a_{n-1} - 1, -3, a_{n+2} - 1, a_{n+3} \ldots \right] \) which does not contain any new f.b. since \( a_{n+2} \leq -2 \). For \( a_{n+1} = 3 \) rewriting leads to \( x_7 = \left[ a_0; a_1, \ldots, a_{n-1} - 1, -3, -2, a_{n+2} - 1, a_{n+3} \ldots \right] \), hence any more rewriting of newly appearing f.b.'s does not change the head \( \mathcal{P}(n-2) \).

We have shown that rewriting a single f.b. of the type \( [1] \) or \( [2, b] \) with \( b \geq 2 \) does not change the head \( \mathcal{P}(n-3) \).

The third and most complicated type of f.b. is when \( a_n = 2 \) and \( a_{n+1} = 1 \). Suppose that \( \mathcal{P} = \left[ a_0; a_1, \ldots, a_{n-1 - l}, 3^l, 2, 1, a_{n+2+k} \ldots \right], \quad l, k \geq 0, \quad a_{n-1 - l} \neq 2, 3, \quad a_{n+2+k} \neq 3 \). By Lemma 13 it is clear that \( \left[ 3, 2, 1, 3^k \right] \) is rewritten into \( \left[ 2, 1 \right] \) and using this recursively we get different cases depending on if \( l < k \), \( k = l \) or \( k > l \).

1. If \( l > k \) then \( \mathcal{P} \) is rewritten first into \( \left[ a_0; a_1, \ldots, a_{n-1 - l}, 3^{l-k} 2, 1, a_{n+2+k} \ldots \right] \) and then into \( y_0 = \left[ a_0; a_1, \ldots, a_{n-1 - l}, 3^{l-k-1} 2, a_{n+2+k} - 2 \ldots \right] \) which may contain a new f.b. beginning with the 2 but \( a_{n+2+k} - 2 \neq 1 \) and we have shown above that an f.b. of the type \( [2, m] \) with \( m \geq 2 \) can be rewritten without changing any element in the sequence more than two steps to the left of the f.b. Hence we can rewrite any new f.b. without changing the head \( \mathcal{P}(n-k-4) \).
2. If $l < k$ then $\mathbf{x}$ is first rewritten into $[a_0; a_1, \ldots, a_{n-1-l}, 2, 1, 3^{k-l}, a_{n+2+k}, \ldots]$, then into $[a_0; a_1, \ldots, a_{n-1-l} - 1, 1, 3^{k-l-1}, a_{n+2+k}, \ldots]$ and next into $y_1 = [a_0; a_1, \ldots, a_{n-1-l} - 2, 2, 3^{k-l-2}, a_{n+2+k}, \ldots]$ if $k \geq l + 2$ or $y_2 = [a_0; a_1, \ldots, a_{n-1-l} - 1, 1, \ldots]$. If $k > l + 2$ we rewrite $y_2$ into $y_3 = [a_0; a_1, \ldots, a_{n-1-l} - 3, -2, 3^{k-l-3}, a_{n+2+k}, \ldots]$, which can only contain a new f.b. if $k = l + 3$ and $a_{n+2+k} \leq -2$ and we have shown that rewriting such a f.b. does not affect the sequence more than two steps to the left so any further rewriting can not change the head $\mathcal{E}(n-3-l)$.

(b) If $k = l + 2$ and $a_{n+2+k} \geq 2$ we rewrite $y_1 = [a_0; a_1, \ldots, a_{n-1-l} - 2, 2, a_{n+2+k}, \ldots]$ into $y_4 = [a_0; a_1, \ldots, a_{n-1-l} - 3, -2, a_{n+2+k} - 1, \ldots]$ so any new f.b. in $y_4$ is either beginning with $a_{n+2+k} - 1 = 1$ or with $a_{n-1-l} - 3 = 1$. If it is beginning with $a_{n-1-l} - 3 = 1$ then $y_4$ is rewritten into $y_5 = [a_0; a_1, \ldots, a_{n-2-l} - 1, -3, a_{n+2+k} - 1, \ldots]$ which can only have a new f.b. beginning with $a_{n+2+k-1} = 1$. In case $a_{n+2+k-1} - 1 = 1$ we have shown that rewriting such an f.b. does not affect the sequence more than two steps to the left and thus, in both cases, a complete rewriting will not change $\mathcal{E}(n-3-l)$.

3. If $k = l$ then $\mathbf{x}$ is first rewritten into $[a_0; a_1, \ldots, a_{n-1-l}, 2, 1, a_{n+2+k}, \ldots]$, then into $z_0 = [a_0; a_1, \ldots, a_{n-1-l} - 1, a_{n+2+k} - 2, \ldots]$ and since $a_{n-1-l} \neq 2, 3$ and $a_{n+2+k} \neq 3$ any new f.b. must either begin with $a_{n+2+k} - 2 = 0, 2$. In case $a_{n+2+k} = 2$ we get $z_1 = [a_0; a_1, \ldots, a_{n-1-l} + a_{n+3+k} - 1, a_{n+4+k}, \ldots]$ and $a_{n+4+k} \leq -2$ so any new f.b. has to be of the form $a_{n-1-l} + a_{n+3+k} - 1 = \pm 1$ or $a_{n-1-l} + a_{n+3+k} - 1 = -2$. In both cases we have a f.b. which we have shown is possible to rewrite without changing the sequence more than two steps to the left hence rewriting does not change the head $\mathcal{E}(n-4-l)$.

We conclude that for all three types of f.b.'s, a complete rewriting of the initial f.b. leaves the head $\mathcal{E}(n-m)$ unchanged with $m = 3$ unless $\mathbf{x} = [a_0; a_1, \ldots, a_{n-1-l}, 3^l, 2, 1, 3^k, a_{n+2+k}, \ldots]$ in which case $m = \min(l + 3, k + 4)$.

3.3. **Dual regular λ-fractions.** To encode the orbits of the geodesic flow in terms of a discrete invertible dynamical system it turns out that we still need another kind of λ-fraction, the so-called **dual regular λ-fraction**. In the case $q = 3$ this was already introduced by Hurwitz [22], see also [25, p. 102].

Consider the set of λ-fractions $y = [0; b_1, \ldots]$ which do not contain any reversed forbidden block, i.e. a forbidden block given in Definitions 11 or 13 read in reversed order.

Let $R$ be the largest number in this set and define $r = R - \lambda$ and $I_R = [-R, R]$. To give an explicit expression for $R$ we need to investigate the connection between ordering of points and their corresponding λ-fractions.

**Lemma 21.** Let $x, y \in I_q$ with $c_q(x) = [a_1, a_2, \ldots]$ and $c_q(y) = [b_1, b_2, \ldots]$ and suppose that $x$ and $y$ are not $G_q$-equivalent to $\pm \frac{\lambda}{2}$ and that $a_1 \neq b_1$. Then $x < y$
if and only if \( b_1 < 0 < a_1 \) or \( b_1 a_1 > 0 \) and \( a_1 < b_1 \). For one-digit \( \lambda \)-fractions this ordering is simply given by

\[
[1] < [2] < \cdots < 0 < \cdots < [-2] < [-1].
\]

**Proof.** Consider \( \varphi_n(x) = ST^n(x) = \frac{-1}{n \lambda + x} \) then \( \varphi_n(I_q) = \left[ \frac{-2}{\lambda(2n-1)}, \frac{-2}{\lambda(2n+1)} \right] \) and the intervals \([l_n, r_n]\) only overlap at their endpoints which are images of \( \pm \frac{\lambda}{2} \). It is also easy to verify that \( r_n < r_m \) if and only if \( m < 0 \) or \( mn > 0 \) and \( n < m \). Since \( x \in \varphi_{a_1} (I_q) \) and \( y \in \varphi_{b_1} (I_q) \) the lemma follows. \(\square\)

**Lemma 22.** Let \( x = [a_1, \ldots, a_{n-1}, a] \), \( y = [a_1, \ldots, a_{n-1}, b] \) and \( x, y \) the corresponding points. Assume that \([a_2, \ldots, a_{n-1}, a] \) and \([a_2, \ldots, a_{n-1}, b]\) are both regular. Then \( x < y \) if and only if \( b < 0 < a \) or \( ab > 0 \) and \( a < b \).

Furthermore, if \( y = [a_1, \ldots, a_{n-1}] \) then \( x < y \) if and only if \( a > 0 \).

**Proof.** First consider \( \varphi_m(x) = ST^m x = \frac{-1}{m \lambda + x} \) then \( \varphi'_m(x) = (m \lambda + x)^{-2} > 0 \) for all \( x \neq -m \lambda \) so \( \varphi_m \) is increasing and positive in \( (-\infty, -m \lambda) \) and increasing and negative in \( (-m \lambda, \infty) \). Hence \( \varphi_m(x) = ST^m x < \varphi_m(y) = ST^m y \iff \) either \( y < -m \lambda < x < y \).

Define \( A_j = ST^{a_j} ST^{a_{j+1}} \cdots ST^{a_n-1} x_j = A_j ST^{a_0}(0), j = 2, \ldots, n \). Then \( x_1 = x, y_1 = y \) and \( x_j, y_j \in I_q \).

Hence, if \( m \in \mathbb{Z}^+ \) then \( \frac{1}{m} < |m \lambda| \) so Lemma 21 implies that \( ST^m x_j < ST^m y_j \) if and only if \( j \leq 1 \) or \( j \geq n \). Using the fact that \( x_j = ST^{a_j} x_j+1 \) and \( y_j = ST^{a_j} y_j+1 \) we see that \( ST^{a_j} x_j < ST^{a_j} y_j \iff x_j < y_j \iff ST^2 x_j < ST^2 y_j \iff \cdots \iff \varphi^m(0) < \varphi^m(0) \iff b < 0 < a \) or \( ab > 0 \) and \( a < b \).

To prove the last equivalence, define \( z_j = A_j(0) \) and proceed as above: \( A_j(0) \iff A_1) \iff \varphi^m x_j < \varphi^m y_j \iff \cdots \iff \varphi^m(0) \iff b < 0 < a \) or \( ab > 0 \) and \( a < b \).

Using Lemmas 21 and 22 it is easy to prove the following Lemma.

**Lemma 23.** Let \( x, y \in \mathbb{R}^\infty \) with \( c_q(x) = [a_0; a_1, a_2, a_3, \ldots] \), and \( c_q(y) = [b_0; b_1, b_2, b_3, \ldots] \). Then \( x < y \) if and only if either \( a_0 < b_0 \) or \( a_i = b_i \) for \( i = 0, \ldots, n-1 \) and either \( b_n < 0 < a_n \) or \( a_n b_n > 0 \) and \( a_n < b_n \) for some \( n \geq 1 \).

**Proof.** Define \( x_j = \lim_{m \to -\infty} ST^{a_j} ST^{a_{j+1}} \cdots ST^m \) and \( y_j = \lim_{m \to -\infty} ST^{b_j} ST^{b_{j+1}} \cdots ST^m \) for \( j \geq 1 \) and observe that we have \( x_j, y_j \in I_q \) for all \( j \geq 1 \). It is clear that if \( a_0 < b_0 \) then \( x < y \) and if \( a_0 > b_0 \) then \( x > y \). Suppose that \( a_0 = b_0 \). If \( x \neq y \) then there exists a smallest \( n \geq 1 \) such that \( a_n \neq b_n \) and \( a_i = b_i \) for \( 0 \leq i \leq n-1 \). Hence, just as in the previous proof of Lemma 22, \( x < y \) if and only if \( x_1 < y_1 \iff \cdots \iff x_n < y_n \). Since \( x_n, y_n \in I_q \) have infinite \( \lambda \)-fractions we can apply Lemma 21 and see that \( x_n < y_n \iff b_n < 0 < a_n \) or \( a_n b_n > 0 \) and \( a_n < b_n \). \(\square\)
Lemma 24. The number $R$ is given by the following regular $\lambda$-fraction

$$R = \begin{cases} 
1; \frac{1^{h-1}}{2}, & \text{for } q \text{ even,} \\
1; \frac{1^{h}}{2}, \frac{1^{h-1}}{2}, & \text{for } q \geq 5 \text{ odd,} \\
1;3 & \text{for } q = 3.
\end{cases}$$

Proof. To obtain the largest number without reversed forbidden blocks we use Lemma 22 recursively. I.e. the largest one-digit $\lambda$-fraction is $\lfloor -1 \rfloor$ and $\lfloor (1)^{2} \rfloor$ is the largest two-digit $\lambda$-fraction etc. For $q \neq 3 \lfloor (1)^{h} \rfloor$ is the largest $h$-digit $\lambda$-fraction without reversed forbidden blocks but since $\lfloor (1)^{h+1} \rfloor$ is forbidden and reversely forbidden the largest $h + 1$-digit $\lambda$-fraction without reversed f.b.'s is $\lfloor (1)^{h}, -2 \rfloor$. Continuing like this inductively and observing that the $\lambda$-fraction without the first $-1$ is always regular we obtain the following expressions for $R$: $R = \lfloor 0; (1)^{h}, -2, (1)^{h-1} \rfloor$ for even $q$, $R = \lfloor 0; -2, -3 \rfloor$ for $q = 3$ and $R = \lfloor 0; (1)^{h}, -2, (1)^{h}, -2, (1)^{h-1} \rfloor$ for odd $q \geq 5$. The Lemma then follows by rewriting these $\lambda$-fractions recursively into regular $\lambda$-fractions using Lemmas 11 and 13.

Lemma 25. For even $q$ we have the identity $R = 1$.

Proof. Consider the action of $S$ on $R$:

$$SR = \lfloor 0; 1^{h}, \frac{2}{1^{h-1}} \rfloor = \lfloor -1; (1)^{h-1}, -2 \rfloor = -R.$$  

Hence $-1/R = -R$ and since $R > 0$ we must have $R = 1$.

Lemma 26. For odd $q$ we have $\frac{1}{2} < R < 1$ and

a) $-R = (TS)^{h+1} R$

b) $R^{2} + (2 - \lambda) R - 1 = 0$

Proof. From the explicit expansions of $\frac{1}{2}$ and 1 in Lemmas 5 and 6 together with the the expansions of $R$ in Lemma 24 it follows from Lemmas 22 and 23 that $\frac{1}{2} < R < 1$. By rewriting as in Lemma 13 for $q \geq 5$ we get

$$SR = \lfloor 0; 1^{h+1}, \frac{2}{1^{h-1}}, 2, 1^{h} \rfloor = \lfloor -1; (1)^{h-1}, -2, (1)^{h}, -2, (1)^{h-1}, -2 \rfloor$$

and deduce that $R = ST^{-1} (ST)^{h-1} ST^{-2} T (-R) = (ST^{-1})^{h+1} (-R)$ and hence $-R = (TS)^{h+1} R$, which is identity a). A similar rewriting works for $q = 3$. Using the following explicit formula for the matrix $(TS)^{n}$ (cf. e.g. [8, p. 1279])

$$(TS)^{n} = \frac{1}{\sin^{2} \frac{\pi}{q}} \left( B_{n+1} \frac{B_{n}}{B_{n-1}}, -B_{n-1} \right), \text{ where } B_{n} = \sin \frac{n\pi}{q}$$

and some elementary trigonometry gives $(TS)^{h+1} R = \frac{-R^{h+1}}{R+\lambda-1} = -R$ which implies identity b). See also Lemma 3.3 in [8].

Remark 27. Using the representation (2) one can also show that the map $A_{r}$ fixing $r = R - 1$ is given by $A_{r} = (ST)^{h+1} T (ST)^{h} T = \frac{1}{2 - 2\lambda} \left( 2 - 2\lambda \lambda - 2\lambda^{2} \right)$ for odd $q$ and $A_{r} = (ST)^{h-1} ST^{2} = \frac{1}{4 \sin^{3} \frac{\pi}{q}} \left( 2 - \lambda \lambda - 3\lambda^{3} \right)$ for even $q$. 
Let $A$ be a set. It is easy to verify that $A = \sigma^j \cdot A_1$ if $b_0 = 0$ and $y_j \in \sin \left( b_0 \right) \left[ r, R \right]$ for all $j \geq 1$.

Proof. A dual regular $\lambda$-fraction is denoted by $\left[ b_0; b_1, \ldots \right]^*$, the space of all dual regular $\lambda$-fractions by $A_q^*$ and the subspace of all infinite sequences in $A_q^*$ with leading 0 by $A_{0,q}^*$.

Uniqueness of a subset of dual regular $\lambda$-fractions is again asserted using a generating map.

**Definition 28.** Let $y = \left[ b_0; b_1, \ldots \right]$ be a finite or infinite $\lambda$-fraction. Set $y_1 = \left[ 0; b_1, \ldots \right]$, $y_j = \sigma^{j-1} y_1 = \left[ 0; b_1, \ldots \right]$ and $y_j$, $j \geq 1$, the corresponding point in $\mathbb{R}$. Then $y$ is said to be a dual regular $\lambda$-fraction if and only if it has the following properties:

$\left( D_1 \right)$ if $b_0 = 0 \implies y \in I_R$.

$\left( D_2 \right)$ if $b_0 \neq 0 \implies y_1 \in \sin \left( b_0 \right) \left[ r, R \right]$, and

$\left( D_3 \right)$ $y_{j+1} \in \sin \left( -y_j \right) \left[ r, R \right]$ for all $j \geq 1$.

A dual regular $\lambda$-fraction is denoted by $\left[ b_0; b_1, \ldots \right]^*$, the space of all dual regular $\lambda$-fractions by $A_q^*$ and the subspace of all infinite sequences in $A_q^*$ with leading 0 by $A_{0,q}^*$.

**Lemma 30.** For $y \in \mathbb{R}$ the following algorithm produces a finite or infinite dual regular $\lambda$-fraction $c_q^* \left( y \right) = \left[ b_0; b_1, \ldots \right]^*$ corresponding to $y$:

(i) Let $b_0 = \langle y \rangle^*_\lambda$ and $y_1 = y - b_0 \lambda$.

(ii) Set $y_{j+1} = F_q^* y_j = -\frac{1}{y_j} - b_j \lambda$, i.e. $b_j = \langle -\frac{1}{y_j} \rangle^*_\lambda$, $j \geq 1$.

If $y_j = 0$ for some $j$ the algorithm stops and one obtains a finite dual regular $\lambda$-fraction.

**Proof.** It is easy to verify that $\langle y \rangle^*_\lambda = 0 \iff y \in I_R$ and that in general $x - \langle x \rangle^*_\lambda \in \left[ r, R \right]$ for $x \geq R$ and $x - \langle x \rangle^*_\lambda \in \left[ -R, -r \right]$ for $x \leq -R$. It is thus clear that (D1) and (D2) are automatically fulfilled and it follows that $F_q^*$ maps $\left[ -R, 0 \right]$ into $\left[ r, R \right]$ and $\left[ 0, R \right]$ into $\left[ -R, -r \right]$. Hence condition (D3) is also satisfied.

**Remark 31.** We say that $F_q^*$ is a generating map for the dual regular $\lambda$-fractions and it is easily verified that $F_q^*$ acts as a left shift map on $A_{0,q}^*$.

It is easily seen that the points affected by the choice of floor function appearing in $\langle -\frac{1}{y_j} \rangle^*_\lambda$ (cf. $\pm \frac{1}{2}$ in the regular case) are exactly those that are equivalent to $\pm r$.

Thus we obtain the following corollary.

**Corollary 32.** If $y$ has an infinite dual regular $\lambda$-fraction expansions which is not equivalent to the expansion of $\pm r$ then it is unique and is equal to $c_q^*(y)$.

**Lemma 33.** A $\lambda$-fraction $y = \left[ b_0; b_1, \ldots \right]$ is dual regular if and only if the sequence $y_0$ does not contain any reversed forbidden blocks. Thereby $y_0 = y$ if $b_0 = 0$ and $y_0 = S y = \left[ 0; b_0, b_1, \ldots \right]$ if $b_0 \neq 0$. 
Proof. For both even and odd \( q \) we use Lemma 21, 22 and 23 to compare points based on their \( \lambda \)-fractions. We give the details for the case of forbidden blocks containing \(+1\)’s. The case of \(-1\)’s is analogous.

Consider \( q \) even and a reversed forbidden block of the form \([m, 1^h]\) with \( m \geq 1 \). If \( y_0 \) contains such a forbidden block, we have \( y_j = [b_j, 1^h, b_{j+1}, \ldots] < 0 \) for some \( j \geq 0 \) and \( b_j \geq 1 \). Hence \( y_{j+1} = [1^h, b_{j+1}, \ldots] < [1^{h-1}, 2] = r \), i.e. \( y_{j+1} \notin [r, R] \) so by Definition 28 (D2) (in case \( j = 0 \)) or (D3) (in case \( j > 0 \) \( y \) is not dual regular.

Consider \( q \) odd. If \( y_0 \) contains a reversed f.b. of the form \([1^{h+1}]\) then \( y_0 \) is not dual regular since \([1^{h+1}] < [1^h, 2, 1^h, 2, 1^{h-1}] \) = \(-R\). If \( y_0 \) contains a reversed f.b. of the form \([m, 1^h, 2, 1^h]\) for some \( m \geq 1 \) then \( y_j = [b_j, 1^h, 2, 1^h, b_{j+2h+1}, \ldots] < 0 \) for some \( b_j \geq 1 \) and \( j \geq 0 \). Hence \( y_{j+1} = [1^h, 2, 1^h, b_{j+2h+1}, \ldots] < r \), i.e. \( y_{j+1} \notin [r, R] \) so by Definition 28 (D2) (in case \( j = 0 \)) or (D3) (in case \( j > 0 \) \( y \) is not dual regular.

In the other direction, suppose that \( y = [b_0; b_1, \ldots] \) does not contain any reversed forbidden blocks. It is clear that if \( b_0 = 0 \) then \( y_0 = y \) and \( y = y_0 \in [-R, R] \) so \( y \) is dual regular. Suppose that \( b_0 \neq 0 \). Then \( y_0 = [b_0, b_1, b_2, \ldots] \) and \( y_j = [b_j, b_{j+1}, \ldots] \). Suppose that \( y_0 \) does not contain any reversed f.b. and that \( b_j > 0 \). If \( q \) is odd then \( y_{j+1} > [1^{h-1}, 2, \ldots] \geq r \) and if \( q \) is even then \( y_{j+1} > [1^{h-1}, 2, \ldots] \geq r \). Hence \( y_{j+1} \in [r, R] \). Similarly if \( b_j < 0 \) we see that \( y_{j+1} \in [-R, -r] \). □

**Lemma 34.** An infinite \( \lambda \)-fraction without reversed forbidden blocks converges.

Proof. This follows from the convergence of infinite regular \( \lambda \)-fractions using rewriting. Note that the only case in which rewriting gives a non-convergent \( \lambda \)-fraction is we produce an infinite sequence of new forbidden blocks all beginning at the same position. From the proofs of Lemmas 18 19 and 20 it is easy to see that this only occurs for the \( \lambda \)-fraction \( \underline{x} \) if either \( q = 4 \) and \( \underline{x} \) has the tail \([1^2, 2, 1]\) or \( q = 3 \) and \( \underline{x} \) has the tail \([1, \underline{3}] \). These two tails both contain reversed forbidden blocks and hence can not occur. □

**Remark 35.** Just as the regular \( \lambda \)-fractions are equivalent to the reduced Rosen \( \lambda \)-fractions, one can show that the dual regular \( \lambda \)-fractions are essentially equivalent to a particular instance of so-called \( \alpha \)-Rosen \( \lambda \)-fractions, see [10] and [34] (in the case \( q = 3 \)). Note that \( (y)^* = \lfloor \frac{y}{\alpha} + 1 - \frac{1}{\alpha} \rfloor \) for \( y > 0 \). Hence \( F_q^* x = T_\alpha (x) \) with \( \alpha = \frac{p}{q} \) for \( x < 0 \) where \( T_\alpha \) is the generating map of the \( \alpha \)-Rosen fractions of [10].

### 3.4. Symbolic dynamics and natural extensions.

An introduction to symbolic dynamics and coding can be found in e.g. [28]. See also [4, 47] or [4, Appendix C]. Our underlying alphabet is infinite, \( \mathcal{N} = \mathbb{Z}^+ = \mathbb{Z} \setminus \{0\} \). The dynamical system \( (\mathcal{N}^{\mathbb{Z}^+}, \sigma^+) \) is called the *one-sided full \( \mathcal{N} \)-shift*. Since the forbidden blocks (cf. Definitions 11 and 13) imposing the restrictions on \( \mathcal{A}_{0,q} \) and \( \mathcal{A}_{0,q}^* \) all have finite length it follows that \( (\mathcal{A}_{0,q}, \sigma^+) \) and \( (\mathcal{A}_{0,q}^*, \sigma^+) \) are both *one-sided subshifts of finite type* (cf. [4, Thm. C7]).
One can show that \( \epsilon_q : I_q \to A_q \) and \( \epsilon_q^* : I_R \to A_q^* \) as given by Lemmas 8 and 30 are continuous (with respect to the metric \( h \) defined in Section 3.1) and we call these the regular and dual regular coding map respectively. Let \( \mathbb{R}^\infty = \{ x \in \mathbb{R} \mid \epsilon_q^* (x) \text{ infinite} \} = \mathbb{R} \setminus G_q(\infty) \) be the set of „\( G_q \)’irrational points“ and set \( I_\alpha^\infty = I_\alpha \cap \mathbb{R}^\infty \) for \( \alpha = q, R \). Since the set \( G_q(\infty) \) of cusps of \( G_q \) is countable it is clear, that the Lebesgue measure of \( I_\alpha^\infty \) is equal to that of \( I_\alpha \), \( \alpha = q, R \). By Corollaries 10 and 32 it follows that the restrictions \( \epsilon_q : I_q^\infty \to A_{0,q} \) and \( \epsilon_q^* : I_R^\infty \to A_{0,q}^* \) are homeomorphisms. Since \( \sigma^+ = \epsilon_q \circ F_q \circ \epsilon_q^{-1} \) on \( A_{0,q} \) and \( \sigma^+ = \epsilon_q^* \circ F_q^* \circ \epsilon_q^*^{-1} \) on \( A_{0,q}^* \) it follows that the one-sided subshifts \( (A_{0,q}, \sigma^+) \) and \( (A_{0,q}^*, \sigma^+) \) are topologically conjugate to the abstract dynamical systems \((I_q^\infty, F_q)\) and \((I_R^\infty, F_q^*)\) respectively (see [4, p. 319]).

Consider the set of regular bi-infinite sequences \( \mathcal{B}_q \subset A_{0,q}^* \times A_{0,q} \subset \mathbb{Z}^\infty \) consisting of precisely those \( \ldots, b_2, b_1, a_1, a_2, \ldots \) which do not contain any forbidden block. Then \((\mathcal{B}_q, \sigma)\) is a two-sided subshift of finite type extending the one-sided subshift \((A_{0,q}, \sigma^+)\), where \( \sigma = \sigma^+ + \sigma^- = \sigma^- \). If \( \epsilon_q^* (y) = [b_1, b_2, \ldots] \in A_{0,q}^* \) and \( \epsilon_q (x) = [a_1, a_2, \ldots] \in A_{0,q} \) we define the coding map \( \mathcal{C} : I_q \times I_R \to \mathbb{Z}^\infty \) by \( \mathcal{C} (x, y) = \epsilon_q^* (y) \circ \epsilon_q (x) = \ldots, b_2, b_1, a_1, a_2, \ldots \). For a given bi-infinite sequence \( \zeta = [a_1, a_2, \ldots] \) and \( \zeta = [b_1, b_2, \ldots] \) denote the „future“ and „past“ respectively. In the next section we will see that there exists a domain \( \Omega \subset I_q \times I_R \) such that \( \mathcal{C}_\| \Omega^\infty : \Omega^\infty \to \mathcal{B}_q \) is one-to-one and continuous (here \( \Omega^\infty = \Omega \cap I_q^\infty \times I_R^\infty \), i.e. we neglect points \((x, y)\) where either \( x \) or \( y \) has a finite \( \lambda \)-fraction). The natural extension, \( \tilde{F}_q \), of \( F_q \) to \( \Omega^\infty \) is defined by the condition that \((\mathcal{B}_q, \sigma)\) is topologically conjugate to \( (\Omega^\infty, \tilde{F}_q) \), i.e. by the relations \( \sigma^+ = \mathcal{C} \circ \tilde{F}_q \circ \mathcal{C}^{-1} \) and \( \sigma^- = \mathcal{C} \circ \tilde{F}_q^{-1} \circ \mathcal{C}^{-1} \), meaning that \( \tilde{F}_q (x, y) = (F_q x, \frac{-1}{y + a_1 \lambda}) \) with \( a_1 = \langle \frac{-1}{x} \rangle \lambda \) and \( \tilde{F}_q^{-1} (x, y) = \langle \frac{-1}{x + b_1 \lambda} \rangle \lambda \).
the case \( q = 3 \) respectively for \( n = 1 \) in case \( q > 3 \) not onto the entire interval \( I_q \), that means the Markov partition is not proper.

From the explicit formula of \( F_q^{-1} \) it is clear that we also need to consider the orbits of the endpoints of \( \pm [r, R] \). From Lemmas 25 and 26 we see that \( \# \{O^s(-R)\} = \kappa + 1 \). Set \( r_0 = -R \) and let \( 0 > r_1 > r_2 > \cdots > r_\kappa = r > -R = r_0 \) be an ordering of \( O^s(-R) = \{r_j\} \). One can verify that \( r_{\kappa+1-j} \in I_j \), \( 1 \leq j \leq \kappa \). Define the intervals \( R_j = [r_j, R] = -R_{-j}, 1 \leq j \leq \kappa \), the rectangles \( \Omega_j = I_j \times R_j, 1 \leq |j| \leq \kappa \) and finally the domain \( \Omega = \bigcup_{|j| \leq \kappa} \Omega_j \). We also set \( \Omega^\infty = \Omega \cap I_q^\infty \cap I_R^\infty \).

**Remark 36.** For even \( q \) we have \( \phi_0 = -\frac{1}{2} = [1^h], r = [1^{h-1}, 2] \) and \( \kappa = h \)
where \( h = \frac{q-2}{2} \) (see Lemma 5 and 24). It is then easy to verify that
\[
\phi_j = F_q^j (\phi_0) = -\phi_{-j} = [1^{h-j}], 0 \leq j \leq h,
\]
\[
r_j = F_q^{h-j} (r) = [1^{j-1}, 2, 1^{h-1}], 1 \leq j \leq h,
\]
\[
I_j = \overline{\phi_{j-1}, \phi_j} = \left[ [1^{h+1-j}], [1^{h-j}] \right], -I_{-j}, 1 \leq j \leq h,
\]
\[
R_j = [r_j, R] = -R_{-j}, 1 \leq j \leq h.
\]

**Remark 37.** For odd \( q \geq 5 \) we have \( \phi_0 = -\frac{1}{2} = [1^h, 2, 1^h], r = [1^h, 2, 1^{h-1}, 2] \) and \( \kappa = 2h + 1 \) where \( h = \frac{q-3}{2} \) (see Lemma 5 and 24). It is then easy to verify that
\[
\phi_{2j} = F_q^j (\phi_0) = [1^{h-j}, 2, 1^h], 0 \leq j \leq h,
\]
\[
\phi_{2j-1} = F_q^{h+1-j} (\phi_0) = [1^{h+1-j}], 1 \leq j \leq h + 1,
\]
\[
r_{2j} = [1^j, 2, 1^{h-1}, 2, 1^h, 2], 0 \leq j \leq h
\]
\[
r_{2j-1} = [1^{j-1}, 2, 1^{h-1}, 2, 1^h], 0 \leq j \leq h - 1
\]
\[
r_{2j+1} = [1^j, 2, 1^{h-1}, 2, 1^h], 1 \leq j \leq h.
\]

Hence
\[
I_{2j+1} = [\phi_{2j}, \phi_{2j+1}] = \left[ [1^{h-j}, 2, 1^h], [1^{h-j}] \right], 0 \leq j \leq h,
\]
\[
I_{2j} = [\phi_{2j-1}, \phi_{2j}] = \left[ [1^{h+1-j}], [1^{h-j}, 2, 1^h] \right], 1 \leq j \leq h,
\]
\[
R_k = [r_k, R] = -R_{-k}, 1 \leq k \leq 2h + 1.
\]

For \( q = 3 \) we have \( \kappa = 1, \phi_0 = -\frac{1}{2} = [2], \phi_1 = 0 \) and \( r_1 = r = [3] \). Hence
\[
I_1 = [-\frac{1}{2}, 0] = -I_{-1} \text{ and } R_1 = [r, R] = -R_{-1}.
\]

To establish the sought correspondence between the domain \( \Omega^\infty \) and \( B_q^0 \) we first need a Lemma.

**Lemma 38.** \( r \) is the smallest number \( y \) in \( I_R \) such that \( C(x, y) \in B_q \) for all \( x \in I_\kappa \).

**Proof.** Let \( q \) be even. We know from Lemma 24 and its proof that \( r = [1^{h-1}, 2] \), \( -R = [1, 1^{h-1}, 2] \) and \( \phi_{-1} = \phi_{h-1} = [1] \). Hence \( C(\phi_{h-1}, r) \in B_q^0 \) and for \( -R \leq y < r \) then \( C_q^* (y) = [1^h, b_1, \ldots] \) and \( C(\phi_{h-1}, y) \) contains the forbidden block \( [1^{h+1}] \).
Let $q \geq 5$ be odd. Then $r = \lfloor 1^h, 2, 1^{h-1}, 2 \rfloor^*$, $-R = \lfloor 1, 1^{h-1}, 2, 1^h, 2 \rfloor^*$ and $\phi_{q-1} = \phi_{2h} = \lfloor 2, 1^h \rfloor^*$. Hence $\mathcal{C}(\phi_{2h}, r) \in B_q$ and if $-R \leq y < r$ then $c_q(y) = [1^h, 2, 1^h, b_{2h+2}, \ldots]$ and $\mathcal{C}(\phi_{2h}, y)$ contains the forbidden block $\llbracket 1^h, 2, 1^h, 2 \rrbracket$. We have shown that $r$ is the smallest number such that $\mathcal{C}(\phi_{q-1}, r)$ does not contain a forbidden block and it is easy to show that also $\mathcal{C}(x, r) \in B_q$ for any $x \in I_q = \phi_{q-1}, 0)$. The same argument applies to $q = 3$.

**Lemma 39.** $(x, y) \in \Omega^\infty \iff \mathcal{C}(x, y) \in B_q$.

**Proof.** Just as in the proof of Lemma 38 it is not hard to verify that $r_j$ is the smallest number in $I_R$ with a dual regular expansion which can be prepended to the regular expansion of $\phi_j$ and hence of all $x \in [\phi_j, 0)$.

**Definition 40.** To determine the first return map we introduce a „conjugate“ region $\Omega^* = \tilde{S}(\Omega^\infty)$ where $\tilde{S}(x, y) = (Sx, -y)$, i.e. setting $I_j^* = S(I_j \cap I_j^\infty)$, $\mathcal{R}_j^* = -\mathcal{R}_j \cap I_j^\infty$ and $\Omega_j^* = I_j^* \times \mathcal{R}_j^*$ we get $\Omega^* = \cup \Omega_j^*$.

Thus $\Omega$ with $\Omega^\infty$ as a dense subset is the domain of the natural extension $\tilde{F}_q$ of $F_q$. An example of $\Omega$ and $\Omega^*$ is given in Figure 1. See [35] for another choice of a „conjugate“ $\Omega^*$ of $\Omega$ using the maps $(x, y^{-1})$ and also [8] for the corresponding domain for the reduced Rosen fractions. In the case $q = 3$ the domain $\Omega^*$ was considered also by Hurwitz in his reduction process for pairs of points under his continued fraction expansion.

### 3.6. Reduction of $\lambda$-fractions.

In a first step in our construction of a cross-section for the geodesic flow we select a set of geodesics on $\mathcal{H}$ which contains at least one lift of each geodesic on $\mathcal{M}_q = \mathcal{H}/G_q$, i.e. a set of „representative” or „reduced” geodesics modulo $G_q$. For an overview and a discussion of different reduction procedures in the case of PSL$_2(\mathbb{Z})$ see [25, Sect. 3].

**Lemma 41.** Let $u \neq v \in \mathbb{R}^\infty$ both have infinite $\lambda$-fractions. Then there exists $B \in G_q$ such that $(Bu, Bv) \in \Omega^*$.

**Proof.** Set $u' = Su = \frac{1}{u}$ and $v' = Jv = -v$. Assume without loss of generality that $c_q(u') = [a_0; a_1, a_2, \ldots]$ and $c_q(v') = [b_0; b_1, b_2, \ldots]$. We now extend the domain of definition of $\tilde{F}_q^{-1}$ from $I_q \times I_R$ to $\mathbb{R}^\infty \times I_R$ by setting...
\[
\hat{F}_q^{-1}(x,y) = \left(\frac{-1}{x+bx}, F_q^*(y)\right), \quad b = \left(-\frac{1}{y}\right)^*.
\]

Since \(x \in \mathbb{R}^\infty x + b\lambda \neq 0\) so \(\hat{F}_q^{-1}(x,y)\) is well-defined. Then \(\hat{F}_q^{-1}(u', v') = (ST^{b_1}u', T^{-b_1}Sv')\), \(\hat{F}_q^{-1}^2(u', v') = (ST^{b_2}ST^{b_1}u', T^{-b_2}T^{-b_1}ST^{b_1}Sv')\), etc. In fact, if \(n \geq 1\) then \(\hat{F}_q^{-1}^n(u', v') = (SA_nSu', JA_nJv') = \hat{S} \circ A_n(u, v)\) where \(A_n = T^{b_n}ST^{b_{n-1}}S \cdots T^{b_1}S\). It is easy to verify that \(JT^mJ = T^{-m}\) for any \(m \in \mathbb{Z}\).

If \(a_j = -b_{j+1}, j = 0, \ldots, k - 1\) and \(a_k + b_{k+1} \neq 0\) and we take \(n > k\) it is clear that \(u'_n = SA_nSu' \) has the formal \(\lambda\)-fraction \(u'_n = [0; b_n, \ldots, b_{k+2}, b_{k+1} + a_k, a_{k+1}, \ldots]\) and \(v'_n = JA_nJv'\) has the dual regular \(\lambda\)-fraction \(v'_n = [b_{n+1}, b_{n+2}, \ldots]^*\).

To simplify the notation we assume \(k = 0\), i.e. \(a_0 + b_1 \neq 0\). Since \(c_q(u') \in A_q\) and \(c_q(v') \in A_q^*\) it is clear that any f.b. in \(u'_n\) must include the term \(a_0 + b_1\). Let \(n_0 \geq 1\) be larger than the length of the largest f.b. If \(q \geq 5\) then by Lemma 18 it is clear that we may choose \(n \geq n_0\) large enough so that rewriting \(u'_n\) into a regular \(\lambda\)-fraction does not change the beginning of \(u'_t\), up to \(b_{n-n_0}\).

If \(q = 4\) and it is not the case that \(u' = [\pm 1 - b_1; \pm 2]\) and \(v' = [b_1, \pm 1, (\pm 2)]^*\) for some \(\pm b_1 \leq -1\) (i.e. \(u'\) and \(v'\) have the same tail as \(r\)), then by Lemma 19 we know that we can choose \(n\) large enough so we can rewrite \(u'_n\) completely without changing the head \(u'_{n-n_0}\) for some \(n_0 \geq 0\).

If \(q = 3\) then \(a_j \neq \pm 1\) for \(j \neq 0\) and \(b_j \neq \pm 1\) for all \(j\). Hence there are only three kinds of f.b.'s possible in \(u'_n\) for \(q = 3\). Either the f.b. starts with \(b_2 = \pm 2\) and \(\pm (b_1 + a_0) \geq 1\) or it starts with \(b_1 + a_0 = \pm 1\) or \(b_1 + a_0 = \pm 2\) and \(\pm a_0 \geq 2\). We can thus apply Lemma 20 and in case \(u'\) and \(v'\) are not of the form \([b_1, 2, 3]^*\) and \(u' = [1 - b_1; 3]\) for some \(b_1 \leq -1\) we can choose \(n\) large enough that rewriting of \(u'_n\) does not change the head \(u'_{n-n_0}\) for some \(n_0 \geq 0\).

After a complete rewriting (if necessary) which does not change the head of \(u'_n\) up to \(b_{n-n_0}\) we may assume that \(u'_n = [b_n, b_{n-1}, \ldots, b_{n-n_0}, a_1, a_2, \ldots] \in A_q\) and \(v'_n = [b_{n+1}, b_{n+2}, \ldots]^* \in A_q^*\). Hence \(\mathcal{C}(u'_n, v'_n) = [\ldots, b_{n+2}, b_{n+1}, b_n, \ldots, b_{n-n_0}, a_1, a_2, \ldots]\) and since \(n_0\) is larger than the length of any f.b. it is clear that \(\mathcal{C}(u'_n, v'_n) \in B_q\), i.e. \((u'_n, v'_n) \in \Omega\) or equivalently \(\hat{S} (u'_n, v'_n) \in \Omega^*\) and hence \(\hat{S} (u'_n, v'_n) = \hat{S} (\hat{S} (A_nu, A_nv)) = (A_nu, A_nv) \in \Omega^*\) with \(A_n = T^{b_n}ST^{b_{n-1}}S \cdots T^{b_1}S\).}

We now have to treat the special cases which are left. Without loss of generality we assume that we have f.b. containing a 1, i.e. we assume the plus sign in the statement of the theorem. The case of minus sign is analogous.

Suppose that \(q = 3\), \(y' = [b_1, 2, 3]^*\) and \(u' = [1 - b_1; 3]\) with \(b_1 \leq -2\) (since \(y' \in A_q^*\)). Then we can rewrite \(y'\) into the regular \(\lambda\)-fraction \([b_1 - 1, -3]\). Hence \(u = Su' = -v' = v = ST^{1-b_1}(r)\) but we assumed that \(u \neq v\) so this case can not happen.

Suppose that \(q = 4\), \(y' = [b_1, 1, 2]^*\) and \(u' = [1 - b_1; 2]\) with \(b_1 \leq -1\) (since \(y' \in A_q^*\)). Then we can rewrite \(y'\) into the following regular \(\lambda\)-fraction \([b_1 - 1, -2]\). Hence \(u = Su' = -v' = v = ST^{1-b_1}(r)\) but we assumed that \(u \neq v\) so this case can not happen.

\(\square\)
**Remark 42.** In the previous lemma we actually showed that for any \( u \in \mathbb{R}^\infty \), which in case \( q = 3 \) or \( 4 \) is not \( G_q \)-equivalent to \( \pm r \), there exists a \( B \in G_q \) such that \( (Bu, Bu) \in \Omega^* \).

**Remark 43.** For \( q = 3 \) the previous lemma and remark should be compared with Hurwitz [22, §7].

In fact, one can do slightly better than in the previous Lemma by using the important property of the number \( r \), namely that \( r \) and \( -r \) are \( G_q \)-equivalent but not orbit-equivalent, i.e. \( O^* (r) \neq O^* (-r) \) (cf. Lemmas 25 and 26). Using the explicit map identifying \( r \) and \( -r \) one can show that it is possible to reduce any geodesic to one with endpoints in \( \Omega^* \) without the upper horizontal boundary.

**Lemma 44.** If \( (x, y) \in \Omega^* \) and \( y \) has a dual regular expansion with the same tail as \( -r \) then there exists \( A \in G_q \) such that \( (Ax, Ay) \in \Omega^* \) and \( Ay \) has the same tail as \( r \).

**Proof.** Using \( \tilde{F} \) we may assume that \( y = -r \) and \( Sx \in I_\kappa \). Consider even \( q \). Let \( A := T^{-1}ST^{-1} \) with \( A(-r) = r \) (recall that \( -R = S R \)). Set \( y' = Ay = r \) and \( x' = Ax \). By Remark 36 \( I_\kappa = I_\kappa = [-\frac{1}{\lambda}, 0] \) hence if \( Sx \in I_\kappa \) then \( x \geq \lambda \), \( T^{-1}x \geq 0 \), \( ST^{-1}x < 0 \), \( T^{-1}ST^{-1}x < -\lambda \) and finally \( ST^{-1}ST^{-1}x \in (0, \frac{1}{\lambda}] \subseteq I_{-h} \). Hence \( (Sx', -y') \in \Omega^* \).

Consider odd \( q \geq 5 \). There are three cases to consider if \( y = -r \) and \( Sx \in I_\kappa = I_{2h+1} = (\phi_{2h}, 0) \) (cf. Remark 37): We have \( c_q(Sx) = [a_1, a_2, \ldots] \) with either \( a_1 \geq 3 \) or \( c_q(Sx) = [2, 1^j, a_{j+2}, \ldots] \) for some \( 0 \leq j \leq h - 1 \) for \( a_{j+2} \neq 1 \) or \( c_q(Sx) = [2, 1^h, a_{h+2}, \ldots] \) for \( a_{h+2} \leq -1 \). In the first case, set \( A = T^{-1}, y' = T^{-1}y = -R \) and \( x' = T^{-1}x \). Then \( Sx \in [-\frac{1}{\lambda}, 0], x \in [3\lambda, \infty) \), \( x' \in [2\lambda, \infty) \) and \( Sx' \in [\frac{1}{\lambda}, 0] \subseteq [-\frac{1}{\lambda}, 0] \). Hence \( (Sx', R) \in \Omega^* \).

In the second case, unless \( j = h - 1, a_{h+1} = 2, a_{h+2} = \cdots = a_{2h+1} = 1 \) and \( a_{2h+2} \geq 1 \) we also set \( A = T^{-1}, y' = -R \) and \( x' = Ax \). Thus \( c_q(Sx') = [1^j, a_{j+2}, \ldots] \in A_{0, q} \) and it follows that \( Sx' \in [-\frac{1}{\lambda}, 0] \) and \( (Sx', R) \in \Omega^* \).

In the remaining two cases, set \( A = T^{-1} \left( ST^{-1} \right)^h ST^{-2} \), \( y' = Ay = r \) and \( x' = Ax \). Then \( Sx' = SAx = \left( ST^{-1} \right)^h ST^{-2} T^2 \left( ST^{-1} \right)^h ST^{-2} \cdots \) \( = \left( ST^{-1} \right)^{h+1} - \cdots \). If \( j = h \) then \( c_q(Sx') = [a_{j+2} - 1, \ldots] \in A_{0, q} \) and if \( j = h - 1, a_{h+1} = 2, a_{h+2} = \cdots = a_{2h+1} = 1 \) and \( a_{2h+2} \geq 1 \) then \( Sx' = [-1, 1^{h+1}, a_{2h+1}, \ldots] \) which is rewritten into \( [-2, (-1)^h, a_{2h+1} - 1, \ldots] \) which is either regular or contains a new f.b. beginning at \( a_{2h+1} - 1 = 1 \). In any case, by Lemma 18 it is clear that after rewriting any f.b. in \( Sx' \) completely we either get \( c_q(Sx') = [-2, (-1)^h, a_{2h+1} - 1, \ldots] \) or \( c_q(Sx') = [-2, (-1)^{h-1}, -2, \ldots] \) and in both cases \( C(Sx', -y') \in \mathcal{E}_q^0 \).

If \( q = 3 \) and \( y = -r \) then there are two possibilities, either \( Sx \in \left( -\frac{1}{\lambda}, -\frac{1}{3} \right) \) or \( Sx \in \left( -\frac{1}{2}, 0 \right) \). In the first case we let \( A = T^{-1}ST^{-2} \) so that \( y' = Ay = r \) and \( Sx' = ST^{-1}ST^{-2}x \). Since \( x \in (2, 3) \) it is clear that \( ST^{-1}ST^{-2}x \in S(0, 1) = ST^{-1}(-\infty, -1) = S(-\infty, -2) = (0, \frac{1}{2}) \) hence \( (Sx', -y') \in \Omega^* \).
For an oriented geodesic arc 

$\Omega^\infty$. In the second case, let $A = T^{-1}$. Then $y' = -r - 1 = -R$ and $Sx' = ST^{-1}x \in ST^{-1}(3, \infty) = S(2, \infty) = \left(-\frac{1}{2}, 0\right)$ so $(Sx', -y') \in \Omega^\infty$.

In all cases we have shown that $(Sx', -y') \in \Omega^\infty$, i.e. $(x', y') \in \Omega^*$ and $y' = -R$ or $r$. □

3.7. Geodesics and geodesic arcs. If $\gamma (\xi, \eta)$ is a geodesic in $\mathcal{H}$ oriented from $\eta$ to $\xi$ (cf. Section 1.1) and $A \in \text{PSL}_2(\mathbb{R})$ we define the geodesic $A\gamma$ as $A\gamma = \gamma (A\xi, A\eta)$. If $(\xi, \eta) \in \Omega^*$ we associate to $(\xi, \eta)$ a bi-infinite sequence (code) to $\gamma$, $\mathcal{C} (\gamma) = \mathcal{C} \circ \mathcal{S} (\xi, \eta) = \epsilon_q (-\eta) . c_{q}^\gamma (S\xi) \in \mathcal{B}_q$.

**Definition 45.** For an oriented geodesic arc $e$ on $\mathcal{H}$ we let $\tau$ denote the unique geodesic containing $e$ and preserving the orientation, e.g. $L_1 = \{\frac{\pi}{2} + iy | y > 0\}$ oriented upwards. Let $e^\pm$ denote the forward and backward end points of $e$ and let $-e$ denote the geodesic arc with endpoints $-e^\pm$. Here $-e$ should not be confused with the geodesic $e$ with reversed orientation, denoted by $e^{-1}$.

For $z, w \in \mathcal{H} \cup \partial \mathcal{H}$ denote by $[z, w]$ the geodesic arc oriented from $z$ to $w$ including the endpoints in $\mathcal{H}$.

**Definition 46.** Let $\mathcal{B}_q = \{\xi \in \mathcal{B}_q | (\sigma^n \xi)(-) \neq \epsilon^* (-r), \forall n \geq 0\}$ be the set of bi-infinite sequences which does not have the same past as $-r$ and let

$\Upsilon = \{\gamma | \mathcal{C} (\gamma) \in \mathcal{B}_q^0\}$.

**Lemma 47.** The coding map $\tilde{\mathcal{C}} : \Upsilon \to \mathcal{B}_q^0$ is a homeomorphism.

**Proof.** $\Upsilon$ is identified with the set of endpoints in $\Omega^*$, the map $\tilde{\mathcal{S}} : \Omega^* \to \Omega^\infty$ in Definition 40 is a continuous bijection and by Corollary 10 and 32 respectively Lemmas 39 and 44 it is clear that $\mathcal{C} : \Omega^\infty \to \mathcal{B}_q$ is well-defined, continuous and that each point $(\xi, \eta) \in \Omega^\infty$ has a unique code unless $\eta$ is equivalent to $\pm r$ in which case it might have two codes. Since one of these is disregarded in the definition of $\mathcal{B}_q^0$ it is clear that $\mathcal{C} = \mathcal{C} \circ \tilde{\mathcal{S}} : \Upsilon \to \mathcal{B}_q^0$ is a homeomorphism. For the case of $\mathcal{G}_q = \text{PSL}_2(\mathbb{Z})$ see also [25, p. 105]. □

The set $\Upsilon$ contains representatives of all geodesics on $\mathcal{H}/\mathcal{G}_q$. This property is an immediate corollary to the following Lemma which additionally also provides a reduction algorithm.

**Lemma 48.** Let $\gamma$ be a geodesic on the hyperbolic upper half-plane with endpoints $\gamma_-, \gamma_+ \in \mathbb{R}^\infty$, $\gamma_- \neq \gamma_+$ and $\gamma_- = [b_0, b_1, \ldots]^*$. Then there exists an integer $n \geq 0$ and $A \in \mathcal{G}_q$ such that

$\gamma' = AT^{-bn} S \ldots T^{-b_1} ST^{-b_0} \gamma$

and $\tilde{\mathcal{C}} (\gamma') \in \mathcal{B}_q^0$. Here $A$ is one of the maps $Id$, $T^{-1}$ and $T^{-1}ST^{-1}$ for even $q$, respectively $T^{-1} (ST^{-1})^h ST^{-2}$ for odd $q$.

**Proof.** This is an immediate consequence of Lemmas 41 and 44. Note that $T^{-b_0} \gamma_- = [b_1, b_2, \ldots]^* \mathcal{F}_n^* T^{-b_0} \gamma_- = T^{-b_0} \gamma_- \ldots T^{-b_1} ST^{-b_0} \gamma_-$. □
COROLLARY 49. If $\gamma^*$ is a geodesic on $\mathcal{H}/G_q$ with all lifts having endpoints in $\mathbb{R}^\infty$ then $\gamma$ contains an element of $\pi^{-1}(\gamma^*)$.

LEMMA 50. If for $q \in I_q c_q(\xi) = [a_1, \ldots, a_n] \in A_{0,q}$, then $\xi$ is the attractive fixed point of the hyperbolic map $A = ST^{a_1}ST^{a_2} \cdots ST^{a_n}$ with conjugate fixed point $\xi^* = \eta^{-1}$, where $c_q^*(\eta) = [a_n, a_{n-1}, \ldots, a_1]^* \in A_{0,q}^*$. Conversely, if $\xi$ is an hyperbolic fixed point of $B \in G_q$, then $c_q(\xi)$ is eventually periodic.

Proof. It is not hard to show, that $A \xi = q \xi$ and $A \eta^{-1} = \eta^{-1}$. Since $q \in I_q$, with $R \leq 1 < \frac{2}{3}$ it is clear that $\xi^* \neq \xi$ and hence $A$ is hyperbolic. The other statement in the Lemma is easy to verify by writing $B$ in terms of generators, rewriting any forbidden blocks and going through all cases of non-allowed sequences, e.g. if $B$ ends with $S$.

Since the geodesic $\gamma(\xi, \eta)$ is closed if and only if $\xi$ and $\eta$ are conjugate hyperbolic fixed points and since $r$ and $-r$ are $G_q$-equivalent we conclude from Lemma 50 that there is a one-to-one correspondence between closed geodesics on $\mathcal{M}_q = \mathcal{H}/G_q$ and the set of equivalence classes, under the shift map, of purely periodic regular $\lambda$-fractions except for the one containing $-r$.

REMARK 51. Because $\Omega^\infty$ only contains points with infinite $\lambda$-fractions, the set $\mathcal{T}$ does not contain lifts of geodesics which disappear out to infinity, i.e. with one or both endpoints equivalent to $\infty$. The neglected set however corresponds to a set in $T^1 M$ of measure zero with respect to any probability measure invariant under the geodesic flow. See e.g. the introduction of [24].

The subshift of finite type $(B^0_2, \sigma)$ is conjugate to the invertible dynamical system $(\mathcal{T}, \tilde{F}_q)$. Here $\tilde{F}_q : \mathcal{T} \rightarrow \mathcal{T}$ is the map naturally induced by $\tilde{F}_q$ acting on $\Omega^\infty$: i.e. if $\gamma = \gamma(\xi, \eta)$ then $\tilde{F}_q(\gamma) = \tilde{F}_q(\gamma')$ where $(\xi', \eta') = \tilde{S} \circ \tilde{F}_q \circ \tilde{S}^{-1}(\xi, \eta)$. Using the same notation for both maps should not lead to any confusion.

3.8. Reduced geodesics.

DEFINITION 52. A geodesic $\gamma(\xi, \eta)$ with $\xi, \eta \in \mathbb{R}^\infty$ is said to be reduced if $\xi(\gamma) \in B^0_2$ and $|\xi| > \frac{3}{2}$ or $\eta \xi < 0$. Denote by $\mathcal{T}_r$ the set of reduced geodesics and by $\Omega_r^*$ the corresponding set of $(\xi, \eta) \in \Omega^\infty$. Then $\Omega_r^\infty = \tilde{S}(\Omega_r^*) \subseteq \Omega^\infty$ and $\Omega_r = \Omega_r^\infty \cap \Omega^\infty$, the closure of $\Omega_r^\infty$ in $\mathbb{R}^2$, i.e. $\Omega_r = \mathcal{F}(\mathcal{T}, \lambda)$, and $B_q^0 = \mathcal{C}(\mathcal{T}) \subseteq B^0_2$.

REMARK 53. We observe that for odd $q$, by Lemma 6, $\phi_{n-1} = [2, 1^h] = \frac{-1}{\lambda+1}$ and thus $-\frac{2}{\lambda+1} < \phi_{n-1}$. For even $q$ on the other hand $\phi_{n-1} = \frac{1}{\lambda-1}$ and since $-\frac{1}{\lambda-1} < -\frac{2}{\lambda+1}$ we have $\phi_{n-1} < -\frac{2}{\lambda+1}$. Hence the shape of $\Omega_n^\infty = \{ (u, v) \in \Omega^\infty \mid u \leq \frac{2}{\lambda+1} \text{ or } uv < 0 \}$ differs slightly between even and odd $q$. Set $\Lambda_1 := (-\frac{1}{\lambda+1}, 0) \times [0, R]$ and $\Lambda_2 := (0, \frac{2}{\lambda+1}) \times [0, -r]$ for even $q$ respectively $\Lambda_2 := (0, \phi_{n-1}) \times [0, -r]$ and $\Lambda_3 := (\phi_{n-1}, \frac{2}{\lambda+1}) \times [0, -r_{n-1}]$ for odd $q$. Then we have:

$$\Omega_r = \bigcup_{j=1}^k \Lambda_j \cup -\Lambda_j,$$
with $k = 2$ for even and $3$ for odd $q$. See also Figure 2 where $\Omega$ is displayed for $q = 5$ and $q = 6$ as a subset of $\Omega$. An even more convenient description of the set of reduced geodesics is in terms of the bi-infinite codes of their base points $(\xi, \eta)$

$$B_{q,r}^0 = \{ \ldots b_2, b_1, a_0, a_1, \ldots \} \in B_q^0 \mid |a_0| \geq 2, \text{ or } a_0b_1 > 0 \}.$$

**Lemma 54.** If $\gamma$ is a geodesic on $\mathcal{H}$ with $\mathcal{E}(\gamma) \in B_q^0$ then there exists an integer $k \geq 0$ such that $F_q^k \gamma$ is reduced.

**Proof.** For $k \geq 0$ let $\gamma^k := F_q^k \gamma$, $c_q(\gamma^k) = [a_k; a_{k+1}, \ldots]$ and $c_q^*(\gamma^k) = [a_{k-1}, \ldots, a_0, b_1, \ldots]^*$. If $|a_k| \geq 2$ then $|\gamma^k| > \frac{3\lambda}{2}$ and if $a_{k-1}a_k < 0$ then $\gamma^k \gamma^k < 0$. In both cases by definition $F_q^k \gamma \in \Upsilon_r$. Since an infinite sequence of $1$’s or $-1$’s is forbidden, it is clear that there exists a $k \geq 0$ such that one of these conditions apply. 

Combining the above lemma with Lemma 41 we have shown the following lemma.

**Lemma 55.** Let $\gamma (\xi, \eta)$ be a geodesic on the hyperbolic upper half-plane with $\xi \neq \eta \in \mathbb{R}^\infty$. Then $\gamma$ is $G_q$-equivalent to a reduced geodesic.

4. **Construction of the cross-section**

As a cross section for the geodesic flow on the unit tangent bundle $T^1M$ of $M$ which can be identified with $\mathcal{F}_q \times S^1$ modulo the obvious identification of points on $\partial \mathcal{F}_q \times S^1$, we will take a set of vectors with base points on the boundary $\partial \mathcal{F}_q$ directed inwards with respect to $\mathcal{F}_q$. The precise definition will be given below. For a different approach to a cross section related to a subgroup of $G_q$ see [14]. For the sake of completeness we include the case $q = 3$ in our exposition but it is easy to verify that our results in terms of the cross-section, first return map and return time agree with the statements in [24, 25].
4.1. The cross section.

**Definition 56.** Consider the fundamental domain $\mathcal{F}_q$ and its boundary arcs $L_0$ and $L_{\pm 1}$ as in Definition 1. For the construction of the cross section we use the additional arcs $L_{\pm 2} = \pm [\rho, \lambda] = T^{\pm 1}L_{\pm 1}$ and $L_{\pm 3} = \pm [\rho, \rho + \lambda] = T^{\pm 1}L_0$.

**Definition 57.** We define the following subsets of $T^1\mathcal{M}$:

\[
\Gamma_r = \{(z, \theta) \in L_r \times S^1 | \dot{\gamma}_{z, \theta}(s) \text{ directed inwards at } z \}, \quad r = 0, \pm 1,
\]

\[
\Sigma^j = \{(z, \theta) \in \Gamma_j | \gamma_{z, \theta} = \gamma(\xi, \eta) \in \Upsilon, |\xi| > \frac{3\lambda}{2} \text{ or } \xi \eta < 0 \}, \quad -1 \leq j \leq 1,
\]

\[
\Sigma^{\pm 2} = \{(z, \theta) \in \Gamma_{\pm 1} | \gamma_{z, \theta} \notin \Upsilon, \gamma = T^{\pm 1}S\gamma_{z, \theta} \in \Upsilon, \frac{3\lambda}{2} < \gamma_+ \leq \lambda + 1 \},
\]

\[
\Sigma^{\pm 3} = \{(z, \theta) \in \Gamma_0 | \gamma_{z, \theta} \notin \Upsilon, \gamma = T^{\pm 1}\gamma_{z, \theta} \in \Upsilon \}.
\]

If $q$ is even let $\Sigma := \bigcup_{j=-2}^{2} \Sigma^j$ and if $q \geq 5$ is odd let $\Sigma := \bigcup_{j=-3}^{2} \Sigma^j$. If $q = 3$ we drop the restriction on $\gamma_+$ in the definition of $\Sigma^{\pm 2}$ and set $\Sigma = \bigcup_{j=-2}^{2} \Sigma^j$.

We will show that there is a one-to-one correspondence between $\Sigma$ and the set of reduced geodesics.

**Lemma 58.** There exists a bijection $\mathcal{P} : \Upsilon_r \to \Sigma$ defined through $\mathcal{P}(\gamma) := (z, \theta) \in \Sigma$ with $z = \gamma(s) \in \partial \mathcal{F}_q$ for some $s \in \mathbb{R}$ and $\theta$ given by $\text{Arg} \dot{\gamma}(s) = \theta$.

**Proof.** Consider Figure 3 and a geodesic $\gamma_1(\xi, \eta)$ from $\eta \in [-R, -r]$ to $\xi > \frac{3\lambda}{2}$. It is clear that either $\gamma_1$ intersects $L_{-1} \cup L_0$ inwards or $L_2$ from the left to the right.

Let $z = \gamma(s)$ be this intersection and set $\theta = \text{Arg} \dot{\gamma}(s)$.

In the first case we get $\mathcal{P}\gamma = (z, \theta) \in \Sigma^{-1} \cup \Sigma^0$. In the second case we either get $\mathcal{P}\gamma = (z, \theta) \in \Sigma^2$ if $\xi \in \left(\frac{3\lambda}{2}, \lambda + 1\right)$ or, if $q$ is odd and $\xi > \lambda + 1$, we get $\mathcal{P}\gamma = (z', \theta') \in \Sigma^3$, where $z' = \gamma(s') \in L_3$ and $\theta' = \text{Arg} \dot{\gamma}(s')$, since by Lemma 93 $\gamma$ must intersect $L_3$.

Remember, that for $q$ even, $r = 1 - \lambda$, so $\gamma$ can not intersect $L_3$ to the right of $\frac{3\lambda}{2}$.
SYMBOLIC DYNAMICS FOR HECKE SURFACES

Figure 4. Illustration of the first return map

Consider next a geodesic \( \gamma(\xi, \eta) \) with \( \eta \in [-R, 0) \) and \( \xi \in \left( \frac{2}{3}, \frac{33}{32} \right) \). Since the geodesic \( SL_{-1} \) from 0 to \( \frac{3}{2} \) intersects \( \rho \), the intersection point of \( \gamma \) and \( T_1 \) must lie above \( \rho \). Hence \( \gamma \) intersects either \( L_{-1} \) or \( L_0 \) inwards, i.e. \( P\gamma \in \Sigma^{-1} \cup \Sigma^0 \).

The case \( \xi < 0 \) is analogous and the inverse map \( P^{-1} : \Sigma \to \Upsilon \) is clearly given by \( P^{-1}(z, \theta) = \gamma_{z, \theta} \) if \( (z, \theta) \in \Sigma_j \), \( |j| \leq 1 \), respectively \( P^{-1}(z, \theta) = T^{\pm 1} \gamma_{z, \theta} \) if \( (z, \theta) \in \Sigma^{\pm 2} \) and \( P^{-1}(z, \theta) = T^{\pm 1} \gamma_{z, \theta} \) if \( (z, \theta) \in \Sigma^{\pm 3} \).

Definition 59. For \( (\xi, \eta) \in \Omega^*_r \) we define \( \tilde{P} : \Omega^*_r \to \Sigma \) by \( \tilde{P}(\xi, \eta) := P\gamma(\xi, \eta) \).

A consequence of Lemma 54 is that for any reduced geodesic we can find an infinite number of reduced geodesics in its forward and backward \( F_q \)-orbit (with infinite repetitions if the geodesic is closed). Furthermore, since the base-arcs \( L_{\pm 1} \) and \( L_0 \) of \( \Sigma \) consist of geodesics, none of whose extensions are in \( \Upsilon_r \), it is clear that any reduced geodesic intersects \( \Sigma \) transversally. The set \( \Sigma \) thus fulfills the requirements (P1) and (P2) of Definition 3 and is a Poincaré (or cross-) section with respect to \( \Upsilon_r \). Since any geodesic \( \gamma^* \) on \( M_q \) which does not go into infinity has a reduced lift we also have the following lemma.

Lemma 60. \( \pi^*(\Sigma) \) is a Poincaré section for the part of the geodesic flow on \( T^1M \) which does not disappear into infinity.

From the identification of \( \Sigma \) and \( \Omega^*_r \) via the map \( \tilde{P} \) we see that the natural extension \( \tilde{F}_q \) of the continued fraction map \( F_q \) induces a return map for \( \Sigma \), i.e. if \( z = (z, \theta) \in \Sigma \) then \( \tilde{P} \circ \tilde{F}_q^n \circ \tilde{P}^{-1}z \in \Sigma \) for an infinite number of \( n \neq 0 \). We give a geometric description of the first return map for \( \Sigma \) and we will later see that this map is in fact also induced by \( \tilde{F}_q \).

Definition 61. The first return map \( T : \Sigma \to \Sigma \) is defined as follows (cf. Figure 4): If \( z_0 \in \Sigma \) and \( \gamma = P^{-1}z_0 \in \Upsilon_r \) let \( \{w_n\}_{n \in \mathbb{Z}} \) be the ordered sequence of intersections in the direction from \( \gamma_- \) to \( \gamma_+ \) between \( \gamma \) and the \( G_q \)-translates of \( \partial F \) with \( w_0 \) given by \( z_0 \). Since \( \gamma_+ \) and \( \gamma_- \) have infinite \( \lambda \)-fractions they are not cusps
of \( G_n \) and the sequence \( w_n \) is bi-infinite. For each \( w_n \) let \( A_n \in G_n \) be the unique map such that \( w_n' = A_n w_n \in \partial F \) and \( \gamma' = A_n \gamma \) intersects \( \partial F \) at \( w_n' \) in the inwards direction.

If \( \gamma' \in \Upsilon_r \) and \( \mathcal{P}\gamma' = z' \) we say that \( z' \in \Sigma \) is a return of \( \gamma \) to \( \Sigma \). If \( n_0 > 0 \) is the smallest integer such that \( w_{n_0} \) gives a return to \( \Sigma \) we say that the corresponding point \( \mathcal{P}A_{n_0} \gamma = z_1 \in \Sigma \) is the first return and the first return map \( T : \Sigma \to \Sigma \) is defined by \( Tz_0 = z_1 \) where \( z_1 \) is the first return after \( z_0 \). Sometimes \( \tilde{T} : \Omega^*_r \to \Omega^*_r \) given by \( \tilde{T} = \mathcal{P}^{-1} \circ T \circ \mathcal{P} \) is also called the first return map.

After proving some useful geometric lemmas in the next section we will show in Section 4.3 that the first return map \( T \) is given explicitly by powers of \( F_q \).

4.2. Geometric lemmas.

**Lemma 62.** The map \( z = (x, \theta, \gamma) \mapsto (\gamma x, \theta, s) \cong (\xi, \eta, s) \) where \( \gamma x, \theta = \gamma (\xi, \eta) \), \( \gamma x, \theta (s) = z \) and \( \text{Arg} \gamma x, \theta (s) = \theta \) is a diffeomorphism for \( \theta \neq \pm \pi/2 \).

**Proof.** Let \( z = x + iy, y > 0 \), and \( \theta \in (-\pi, \pi) \) be given. First we want to show that there exist \( \xi, \eta \in \mathbb{R}^* \) and \( s \in \mathbb{R} \) such that for the geodesic \( \gamma = \gamma (\xi, \eta) \) one finds \( \gamma (s) = z \) and \( \text{Arg} \gamma (s) = \theta \). Without loss of generality we may assume that \( \theta \in (-\pi/2, \pi/2) \) so that \( \eta < \xi \). Set \( c = \frac{1}{2} (\eta + \xi) \) and \( r = \frac{1}{2} (\xi - \eta) \) and parametrize \( \gamma \) as \( \gamma (t) = c + re^{it}, 0 < t < \pi \). It is easy to verify that if \( c = x + y \tan \theta, r = \frac{y}{\cos \theta} \) and \( t_0 = \theta + \frac{\pi}{2} \) then \( \gamma (t_0) = z \) and \( \text{Arg} \gamma (t_0) = \theta \). See Figure 5. To find the arc length parameter \( s \) we use the isometry \( A : z \mapsto \frac{z + (c - r)}{x + (c + r)} \) mapping \( \gamma \) to \( i\mathbb{R}^* \), \( A \gamma (t) = i (\tan \frac{t}{2})^{-1} \) and \( A (c + ir) = i \). It is then an easy computation to see that \( s (\theta) = d (z, c + ir) = d (i/\tan \frac{t}{2}, i) = \ln \tan \frac{t}{2} = \ln \tan \left( \frac{\theta}{2} + \frac{\pi}{4} \right) \). From the above formulas one can easily deduce differentiability of the map \( (x, \theta) \mapsto (\xi, \eta, s) \) as well as its inverse away from \( \theta = \pm \pi/2 \). \( \square \)

**Corollary 63.** The map \( (x, y, \theta) \mapsto (\xi, \eta, s) \) of Lemma 62 gives a change of variables on \( T^1 \mathcal{H} \) which is diffeomorphic away from \( \theta = \pm \pi/2 \). Explicitly, \( \xi = x + y \tan \theta \) and \( \eta = x + y \tan \theta - \frac{y}{\cos \theta}, s = \ln \tan \left( \frac{\theta}{2} + \frac{\pi}{4} \right) \) and the corresponding Jacobian is \( \frac{\partial (x, y, \theta)}{\partial (\xi, \eta, s)} = \frac{1}{2} \cos^2 \theta \).

**Proof.** This follows from the proof of Lemma 62 and a trivial computation. \( \square \)
Definition 64. For \( z \in \mathcal{H} \) and \( \xi \in \mathbb{R} \) define
\[
g(z, \xi) = \frac{|z - \xi|^2}{\Im z}
\]
and for \( \gamma \) a geodesic with endpoints \( \gamma_+ \) and \( \gamma_- \) set \( g(z, \gamma) = g(z, \gamma_+) \).

Lemma 65. If \( A \in PSL_2(\mathbb{R}) \) then \( g(Az, A\xi) = g(z, \xi) A' (\xi) \).

Proof. Let \( A = \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \), then \( \Im Az = \frac{3z}{|cz+d|^2} \), \( |Az - A\xi|^2 = \frac{|z-\xi|^2}{|cz+d|^2} \) and since \( \xi \in \mathbb{R} \)
\[
g(Az, Ax) = g(z, \xi) (c\xi + d)^{-2} = g(z, \xi) A' (\xi).
\]

Lemma 66. Let \( \gamma = \gamma(\xi, \eta) \) be a geodesic with \( \xi, \eta \in \mathbb{R} \), \( \eta < \xi \) and suppose that \( z_j, j = 1, 2 \), with \( \eta \leq Rz_1 < Rz_2 \leq \xi \) are two points on \( \gamma \). Then
\[
d(z_1, z_2) = \ln g(z_1, \xi) - \ln g(z_2, \xi) = \ln \left( \frac{\Im z_2}{\Im z_1} \right) \left( \frac{z_1 - \xi}{z_2 - \xi} \right)^2.
\]

Proof. It is easy to verify, that the hyperbolic isometry \( B : w \mapsto \frac{w - b}{\xi - w} \) maps \( i\mathbb{R}^+ \) and if \( a < b \) then \( d(ia, ib) = \int_a^b \frac{du}{y} = \ln \left( \frac{b}{a} \right) \). Thus if \( \Im z_j = y_j \) we get
\[
d(z_1, z_2) = d(Bz_1, Bz_2) = d(i\Im Bz_1, i\Im Bz_2) = \ln \left( \frac{y_2}{y_1} \right) \left( \frac{z_1 - \xi}{z_2 - \xi} \right)^2.
\]

Lemma 67. Let \( \gamma = \gamma(\xi, \eta) \) be a geodesic with \( \xi, \eta \in \mathbb{R} \), \( \eta < 0 < \xi \) and let \( z = \gamma \cap i\mathbb{R} \) be the intersection of \( \gamma \) with the imaginary axis. Then
\[
z = i\sqrt{-\xi}\eta.
\]

Proof. With \( r = \frac{1}{2} (\xi - \eta) \) and \( c = \frac{1}{2} (\xi + \eta) \) any point on \( \gamma \) is given by \( \gamma(t) = c + re^{it} \) for some \( 0 \leq t \leq \pi \). Suppose \( z = \gamma(t_0) \), then \( \Re \gamma(t_0) = c + r \cos t_0 = 0 \) and hence \( \cos t_0 = -\frac{\xi}{r} \). But then \( \sin^2 t_0 = 1 - \frac{\xi^2}{r^2} \) and therefore \( z = ir \sin t_0 = i\sqrt{r^2 - \xi^2} = i\sqrt{-\xi}\eta \).

Lemma 68. Let \( \gamma = \gamma(\xi, \eta) \) be a geodesic with \( \xi, \eta \in \mathbb{R} \), \( \eta < \xi \). For \( \omega \) a geodesic intersecting \( \gamma \) at \( w \in \mathcal{H} \) let \( A = \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \in PSL_2(\mathbb{R}) \) be such that \( A\omega = i\mathbb{R}^+ \). Then
\[
w = w(\xi, \eta) = \frac{1}{ad + bc + ac(\xi + \eta)} \left[ ac\xi\eta - bd + ci\sqrt{-l_A(\xi)} l_A(\eta) \right], \text{ and}
\]
\[
g(w, \gamma) = \frac{|w - \xi|^2}{\Im w} = (\xi - \eta) \sqrt{-l_A(\xi)} l_A(\eta)
\]
where \( l_A(\xi) = (a\xi + b)(c\xi + d) \) and \( \epsilon = \text{sign} (ad + bc + ac(\xi + \eta)) \).
Proof. According to Lemma 65 \( g (w, \gamma) = g (A^{-1} z, A^{-1} \gamma') = g (z, \gamma') A^{-1} (A \xi) \) where \( \gamma' = A \gamma \) and \( z = Aw \in i \mathbb{R} \). By Lemma 67 \( z = i \sqrt{\xi' \eta'} \) with \( \xi' = A \xi \) and \( \eta' = A \eta \). We choose \( A \), i.e. the orientation of \( \gamma' \) such that \( \eta' < 0 < \xi' \) and hence 
\[
\text{sign} (a \xi + b) = \text{sign} (c \xi + d) \quad \text{and} \quad \text{sign} (a \eta + b) = - \text{sign} (c \eta + d). \]

Then 
\[
w = A^{-1} \left( i \sqrt{-A \xi A \eta} \right) = \left( di \sqrt{\frac{a \xi + b \eta + b}{c \xi + d c \eta + d}} - b \right) \left( a - ci \sqrt{\frac{a \xi + b \eta + b}{c \xi + d c \eta + d}} \right)^{-1}
\]
\[
= \left[ i \sqrt{\frac{a \xi + b \eta + b}{c \xi + d c \eta + d}} - \left( ab + dc \left( \frac{a \xi + b \eta + b}{c \xi + d c \eta + d} \right) \right) \right] \left( c^2 \left| a \xi + b \eta + b \right| + a^2 \right)^{-1}
\]
\[
= - \frac{(ab)(c \xi + d)(c \eta + d) + dc |a \xi + b||a \eta + b| + i \sqrt{-l_A (\xi \eta)} l_A (\eta) }{c^2 (a \xi + b)(a \eta + b) + a^2 c \xi + d c \eta + d}
\]
\[
= \frac{\epsilon (ab)(c \xi + d)(c \eta + d) - dc (a \xi + b)(a \eta + b) + i \sqrt{-l_A (\xi \eta)} l_A (\eta) }{ac (\xi + \eta) + (ad + bc)}
\]
\[
= \frac{\epsilon [(\xi + \eta)ca + (cb + ad)]}{ac (\xi + \eta) + (ad + bc)}
\]

where \( \epsilon = \text{sign} ((a \xi + b)(a \eta + b)) = - \text{sign} (ac (\xi + \eta) + ad + bc) \) since \( w \in H \).

For the function \( g \) we now have
\[
g (z, \gamma') = \frac{|z - \xi'|^2}{3z} = \frac{\xi'^2 - \xi' \eta'}{\sqrt{-\xi' \eta'}} = (\xi' - \eta') \sqrt{\frac{\xi'}{-\eta'}}.
\]

Since \( A^{-1} (A \xi) = (c \xi + d)^2 \), Lemma 65 implies that 
\[
g (w, \gamma) = (\xi' - \eta') \sqrt{\frac{\xi'}{-\eta'}} (c \xi + d)^2 = \left( \frac{a \xi + b}{c \xi + d} - \frac{a \eta + b}{c \eta + d} \right) \sqrt{\frac{a \xi + b}{c \xi + d} + \frac{a \eta + b}{c \eta + d}} (c \xi + d)^2
\]
\[
= \left( (a \xi + b)(c \eta + d) - (a \eta + b)(c \xi + d) \right) \sqrt{\frac{(a \xi + b)(c \xi + d)}{(a \eta + b)(c \eta + d)}}
\]
\[
= (\xi - \eta) \sqrt{\frac{(a \xi + b)(c \xi + d)}{(a \eta + b)(c \eta + d)}} = (\xi - \eta) \sqrt{- \frac{l_A (\xi)}{l_A (\eta)}}.
\]

Application of the previous Lemma to vertical or circular geodesics yields the following corollaries:

**Corollary 69.** Let \( \gamma (\xi, \eta) \) be a geodesic with \( \eta < \xi \). Then

\( \text{sign} (a \xi + b) = \text{sign} (c \xi + d) \) and \( \text{sign} (a \eta + b) = - \text{sign} (c \eta + d) \). Then
a) if \( \eta < a < \xi \) for some \( a \in \mathbb{R} \), then \( \gamma \) and the vertical geodesic \( \omega_v = a + i\mathbb{R}^+ \) intersect at

\[
Z_v(\xi, \eta) = a + i\sqrt{(\xi - a)(-\eta + a)} \in \mathcal{H}, \text{ and}
\]

\[
g_v(\xi, \eta) = g(Z_v, \gamma) = (\xi - \eta) \sqrt{\frac{-\xi - a}{\eta - a}};
\]

b) if \( \eta < c - \rho < \xi < c + \rho \) for some \( c \in \mathbb{R} \) and \( \rho \in \mathbb{R}^+ \), then \( \gamma \) and the circular geodesic \( \omega_c \) with center \( c \) and radius \( \rho \) intersect at

\[
Z_c(\xi, \eta) = \xi \eta + \rho^2 - c^2 \left( \frac{\xi + \eta - 2c}{\xi + \eta - 2c} \right) + i \sqrt{\left( (\xi - c)^2 - \rho^2 \right) \left( (\rho^2 - (\eta - c)^2 \right)} \in \mathcal{H}, \text{ and}
\]

\[
g_c(\xi, \eta) = g(Z_c, \gamma) = (\xi - \eta) \sqrt{\frac{-(\xi - c)^2 - \rho^2}{(\rho^2 - (\eta - c)^2)}};
\]

The subscripts „v“ and „c“ above refer to intersections with vertical and circular geodesics respectively.

**Corollary 70.** Let \( \gamma = \gamma(\xi, \eta) \) be an arbitrary geodesic on \( \mathcal{H} \) with \( \xi, \eta \in \mathbb{R} \).

For \( Z_j(\xi, \eta) = \gamma \cap \mathcal{L}_j \) set \( g_j(\xi, \eta) = g(Z_j, \gamma) \), \(-2 \leq j \leq 2\). If \( Z_j(\xi, \eta) \) exists, the following formulas hold:

\[
Z_0(\xi, \eta) = \frac{1}{\xi + \eta} \left( 1 + \xi \eta + \epsilon i \sqrt{\left( \frac{\xi^2}{2} - 1 \right) \left( 1 - \eta^2 \right)} \right), \epsilon = \text{sign} (\xi + \eta),
\]

\[
Z_{\pm 1}(\xi, \eta) = \pm \frac{\lambda}{2} + i \sqrt{\left( \frac{\xi \mp \lambda}{2} \right) \left( -\eta \pm \frac{\lambda}{2} \right)},
\]

\[
Z_{\pm 2, \pm 3}(\xi, \eta) = \frac{1}{\xi + \eta \mp 2c} \left( \xi \eta + \rho^2 - c^2 \mp \epsilon i \sqrt{\left( \frac{\xi + c}{2} - \rho^2 \right) \left( \rho^2 - (\eta \mp c)^2 \right)} \right)
\]

where \( \epsilon = \text{sign} (x + y \mp 2c) \).

Furthermore

\[
g_0(\xi, \eta) = (\xi - \eta) \sqrt{\frac{\xi^2 - 1}{1 - \eta^2}},
\]

\[
g_{\pm 1}(\xi, \eta) = (\xi - \eta) \sqrt{-\frac{\xi \mp \frac{\lambda}{2}}{\eta \mp \frac{\lambda}{2}}},
\]

\[
g_{\pm j}(\xi, \eta) = (\xi - \eta) \sqrt{-\frac{(\xi \mp c)^2 - \rho^2}{(\eta \mp c)^2 - \rho^2}}, j = 2, 3.
\]

Here \( (\rho, c) = \left( \lambda - \frac{1}{\lambda}, \frac{1}{\lambda} \right) \) for \( j = 2 \) and \( (1, \lambda) \) for \( j = 3 \).
Proof. Taking $A_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, $A_{\pm 1} = \left( \begin{pmatrix} 1 \pm \frac{1}{\sqrt{2}} \\ 0 & 1 \end{pmatrix} \right)$ and $A_{\pm 2} = \frac{1}{\sqrt{2}} \left( \begin{pmatrix} 1 \pm e^{-\rho} \\ 1 \end{pmatrix} \right)$ it is easy to verify that $A_j T_j = i \mathbb{R}^+$ preserving the orientation.

**Lemma 71.** Let $\gamma = \gamma(\xi, \eta)$ be a geodesic with $\xi, \eta \in \mathbb{R}$. Then $\gamma$ intersects the vertical arc $L_1$ if and only if $\eta < \frac{1}{2} \xi < \xi$ and $\delta(\xi, \eta) < 0$ where

$$\delta(\xi, \eta) = n\lambda - 2\xi - \lambda.$$ 

For even $q$ we have in particular $\delta(\xi, \eta) = \eta - (TS)^{\lambda} \xi$.

Proof. It is clear, that $\gamma$ intersects $L_1$ if and only if the intersection with $T_1 = \frac{1}{2} + i \mathbb{R}^+$ is at a height above $\sin \frac{x}{q} = \frac{y}{q}$. By Lemma 69 the point of intersection is given by $w = \theta \pm i \sqrt{\left( \frac{1}{2} - \frac{1}{2}\right) \left( -\frac{1}{2} + \frac{1}{2}\right)}$. We thus need to check the inequality $(\xi - \frac{1}{2}) \left( -\frac{1}{2} + \frac{1}{2}\right) > \sin \frac{\pi}{q}$. With $\eta < \frac{1}{2} \xi < \xi$ it is clear that $\exists w$ decreases as $\eta$ increases for $\xi$ fixed. Using $\frac{1}{2} = \cos \frac{\pi}{q}$ we see, that

$$\left( \frac{\xi - \lambda}{2} \right) \left( \frac{\xi}{2} - \eta \right) = \cos^2 \frac{\pi}{q} = 1 - \frac{\lambda^2}{4} \iff \lambda - 2\eta = \frac{4 - \lambda^2}{2\xi - \lambda} \iff \eta = \frac{\lambda \xi - 2}{2\xi - \lambda} = A\xi$$ 

where $A = \frac{1}{4\pi} \left( \frac{\lambda - 3}{2} \right) \in SL(\mathbb{R})$. Hence $\exists w > \sin \frac{\pi}{q} \iff \eta < A\xi \iff \delta(\xi, \eta) < 0$.

Observe, that $A\rho = \rho$ and $A^2 = Id$, i.e. $A$ is elliptic of order 2. The stabilizer $G_{q,\rho}$ of $\rho$ in $G_q$ is a cyclic group with $q$ elements generated by $TS$. For even $q = 2h + 2$ one can use the explicit formula (2) to verify that $A = (TS)^{\lambda} \in G_q$. For odd $q$ on the other hand there is no element of order 2 in $G_{q,\rho}$, so $A \notin G_q$. 

**Corollary 72.** Let $\gamma = \gamma(\xi, \eta)$ be a geodesic with $\xi, \eta \in \mathbb{R}$. Set $\delta_n(\xi, \eta) := \delta(\xi - n\lambda, \eta - n\lambda)$. Then $\gamma$ intersects the line $T^n L_1$ if and only if $\eta < (n + \frac{1}{2}) \lambda < \xi$ and $\delta_n(\xi, \eta) < 0$.

4.3. **The first return map.** Our aim in this section is to obtain an explicit expression for the first return map $T : \Sigma \rightarrow \Sigma$. The notation is as in Definition 61, see also Figure 4. The main idea is to use geometric arguments to identify possible sequences of intersections $\{w_n\}$ and then use arguments involving regular and dual regular $\lambda$-fractions to determine whether a particular $w_n$ corresponds to a return to $\Sigma$ or not.

**Lemma 73.** If $\xi = \left[ 1; 1^{j_1}, a_{j_1+1}, a_{j_2+2}, \ldots \right]$ with $c_q(S\xi) = \left[ 1^{j+1}, a_{j+1}, \ldots \right] \in A_{0,q}$ then $\xi \in \left( (TS)^j \lambda, (TS)^j \frac{3\lambda}{2} \right)$ if $a_{j+1} \leq -1$ respectively $\xi \in \left( (TS)^{j+1} \frac{3\lambda}{2}, (TS)^{j} \lambda \right)$ if $a_{j+1} \geq 2$.

Proof. Note that $\xi = (TS)^{j} \xi'$ where $c_q(\xi') = \left[ 1; a_{j+1}, a_{j+1}, \ldots \right]$. If $a_{j+1} \leq -1$ then $\xi' \in (\lambda, \frac{3\lambda}{2})$ and since $TSx = \lambda - \frac{1}{2}$ is strictly increasing there $\xi = \ldots$
Define the geodesic arcs $\xi' \in (TS)^j (\lambda, \frac{3\lambda}{2}) = ([TS]^j \lambda, [TS]^j \frac{3\lambda}{2})$. If on the other hand $a_{j+1} \geq 2$ then $\xi'' = ST^{-1} \xi' \in \left(\frac{3\lambda}{2}, \infty\right)$ and $\xi = (TS)^j \xi' = (TS)^{j+1} \xi''$ and therefore $\xi \in (TS)^{j+1} \left(\frac{3\lambda}{2}, \infty\right) = ([TS]^{j+1} \frac{3\lambda}{2}, [TS]^{j} \lambda)$.

**Definition 74.** Define the geodesic arcs

$$\chi_j := (TS)^j L_2 = (TS)^{j+1} TL_{-1}, \quad 0 \leq j \leq h,$$

$$\omega_j := (TS)^j L_3 = (TS)^{j+1} L_0, \quad 0 \leq j \leq h + 1$$

and set $\alpha_j := (TS)^j \lambda$, $\beta_j := (TS)^j \frac{3\lambda}{2}$ and $\delta_j := (TS)^j (\lambda + 1)$. Then $\chi_j = [\rho, \alpha_j]$, $\omega_j = \left[\rho, (TS)^j (\rho + \lambda)\right] \in [\rho, \delta_j]$ and $\alpha_j < \beta_j < \delta_j, \quad 0 \leq j \leq h + 1$. Note that $\alpha_h = \frac{1}{2}$ and $\chi_h = L'_1$ for even $q$ while $\delta_{h+1} = \frac{1}{2}$ and $\omega_{h+1} \subseteq L'_1$ for odd $q$ (see Figure 6).

**Lemma 75.** If $\gamma = \gamma (\xi, \eta) \in \Upsilon$ with $c_q (\xi) = [1; 1^j, a_{j+1}, \ldots]$ then $\gamma$ has the following sequence of intersections with $G_q \partial \mathcal{F}$ after passing $L_1$: $\omega_0, \chi_0, \ldots, \omega_{j-1}, \chi_{j-1}, \omega_j$ if $\xi \in (\alpha_j, \beta_j)$ ($a_{j+1} \leq -1$) respectively $\omega_0, \chi_0, \ldots, \omega_j, \chi_j$ if $\xi \in (\beta_{j+1}, \alpha_j)$ ($a_{j+1} \geq 2$).

**Proof.** See Figure 6. Since all arcs involved are hyperbolic geodesics it is clear that $\gamma$ does not intersect any other $\chi_i$’s or $\omega_i$’s than those mentioned. Suppose that $\xi \in (\alpha_j, \beta_j)$, then after $\chi_{j-1}$ the geodesic $\gamma$ may intersect either $\omega_j$ or its extension, i.e. $[(TS)^j (\rho + \lambda), \delta_j]$. If it intersects this extension it has to pass first through the arc $[(TS)^j T \rho, \alpha_{j-1}]$. But the completion of this arc is clearly $(TS)^j TL_1 = [\beta_j, \alpha_{j-1}]$ and hence $\gamma$ can not intersect this arc and must pass through $\omega_j$. The second case is analogous, except that we do not care about whether the next intersection is at $\omega_{j+1}$ or $(TS)^{j+1} TL_1$.

**Lemma 76.** If $\gamma (\xi, \eta) \in \Upsilon_r$ and $c_q (\xi) = [1; 1^j, a_{j+1}, \ldots]$ then $\tilde{\gamma} (\xi, \eta) = \tilde{F}_q^{j+1} \gamma (\xi, \eta)$.

**Proof.** Let $(z, \theta) = \tilde{\mathcal{P}} (\xi, \eta)$, then $z \in L_{-1} \cup L_0$ and by Lemma 75 the subsequent intersections are $w_0, w_1, w_2, \ldots, w_{2j+1}, w_{2j+2}$ if $a_{j+1} \leq -1$ and $w_0, w_1, w_2, \ldots$,
Define

Consider possibilities:

\[ n(\xi) := \begin{cases} 
\varepsilon \cdot 3, & k = h + 1, \quad q \text{ odd}, \\
\varepsilon \cdot 2, & k = h, \quad a_h \geq 2, \quad q \text{ even}, \\
\varepsilon \cdot 1, & k = h + 1, \quad q \text{ even}, \\
0, & \text{else}. 
\end{cases} \]

We also have to consider the return map for the second type of reduced geodesics.

**Lemma 78.** For \( z \in \Sigma \) with \( P^{-1}z = (\xi, \eta) \in \Omega_q \) and \( |\xi| > 3\lambda \) one has \( Tz = P \circ \hat{F}_q \circ P^{-1}z \in \Sigma^h \) where \( k = K(\xi) \) and \( n = n(\xi) \).

**Proof.** Consider \( z_0 = P\gamma \in \Sigma \) with \( \gamma = (\xi, \eta) \in \Upsilon \) and assume without loss of generality that \( \xi > 0 \) with \( c_q(\xi) = [a_0; 1^j, a_{j+1}, \ldots] \) for some \( j \geq 0 \), \( a_{j+1} \neq \pm 1 \), and \( a_{j+1} \neq \pm 1 \) if \( j = 0 \) (the case of \(-1\)'s is analogous). Recall the notation in Definition 61, in particular the sequence \( \{w_n\}_{n \in \mathbb{Z}} \) and the corresponding maps \( A_n \in G_q \). It is clear, that \( w_n \) gives a return if and only if \( A_n \gamma \in \Upsilon \).

There are two cases to consider: Either \( z_0 \in \Sigma^{-1} \cup \Sigma_0 \) respectively \( z_0 \in \Sigma^{-1} \cup \Sigma^0 \) in the case of odd \( q \) or \( z_0 \in \Sigma^2 \). In Figure 7 these different possibilities are displayed, \( P\gamma_A \in \Sigma^{-1}, P\gamma_B \in \Sigma^0, P\gamma_C \in \Sigma^2 \) and \( P\gamma_D \in \Sigma^3 \). It is clear, that if \( z_0 \in \Sigma^2 \) then \( w_0 = TSz_0 \in L_2 \) and the sequence of \( \{w_n\} \) is essentially different from the case \( z_0 \in \Sigma^2 \) when \( w_0 = z_0 \).

**Case 1:** If \( z_0 \in \Sigma^2 \) (see geodesics \( \gamma_A, \gamma_B \) and \( \gamma_D \) in Figure 7), then \( w_n \in T^nL_{-1} \) for \( 1 \leq n \leq k - 1 \) and \( k = a_0 - 1, a_0 + 1 \) depending on whether \( z_0 \in \Sigma^0 \cup \Sigma^1 \) or \( \Sigma^2 \) and whether the next intersection is on \( T^aL_0 \) or \( T^{a+1}L_0 \). Then either \( w_k \in T^aL_0 \) or \( w_k \in T^{a+1}L_0 \) (see geodesics \( \gamma_E \) and \( \gamma_F \) in Figure 8). Since \( A_n = T^{-n} \) for \( w_n \in T^nL_{-1} \) and, as we will show in Lemma 88 \( T^{-n} \gamma \notin \Upsilon \) none of the \( w_n \in T^nL_{-1} \) for \( 1 \leq n \leq k - 1 \) gives a return to \( \Sigma \). There are now two possibilities:

(i) If \( w_k \in T^aL_0 \), then \( A_k = ST^{-a} \) and \( \gamma_k = A_k \gamma = \hat{F}_q \gamma \). If \( j = 0 \) it is clear that \( \gamma' \in \Upsilon \) and \( Tz_0 = P \circ \hat{F}_q \gamma \in \Sigma^0 \). If \( j \geq 1 \), by Lemma 76 applied to \( \gamma' \) we get \( Tz_0 = P \circ \hat{F}_q \gamma' \in P \circ \hat{F}_q \gamma' \in \Sigma^0(\xi) \).

(ii) If \( w_k \in T^{a+1}L_0 \), then we will show in Lemma 90 and 94 that none of the arcs emanating from \( T^{a+1} \) gives a return (cf. Figures 8 and 6) except for the next return at \( T^{a+1} \left( ST^{-1} \right)^{h+1}L_1 \). Furthermore it follows, that \( Tz_0 = P \circ \hat{F}_q \circ P^{-1}z_0 \in \Sigma^h \) with \( k = K(\xi) \) (here \( h \) or \( h + 1 \)) and \( n = n(\xi) \).
Case 2: If $z_0 \in \Sigma^2$ then $\frac{3\lambda}{2^2} < \xi < \lambda + 1$ and $\gamma$ must intersect $T\mathcal{L}_1$ below $T\rho$.

By the same arguments as in Case 1 we conclude that the first return is given by $w_{q-1} \in T(S\mathcal{T}^{-1})^{h+1}L_1$ and $Tz_0 = \mathcal{P} \circ \hat{F}_q^k \circ \mathcal{P}^{-1}z_0 \in \Sigma^n$ where $k = K(\xi)$ and $n = n(\xi)$ as in Case 1 (ii). In all cases we see, that the first return map $T : \Sigma \to \Sigma$ is given by $T = \mathcal{P} \circ \hat{F}_q^k \circ \mathcal{P}^{-1}$ or alternatively by $\hat{T} = \hat{F}_q^k$ where $k = 1, h$ or $h + 1$ depending on $\xi$. □
By combining Lemma 76 and 78 it is easy to see, that the first return map \( \tilde{\mathcal{P}} \) is determined completely in terms of the coordinate \( \xi \):

**Proposition 79.** If \( z \in \Sigma \) with \( \tilde{\mathcal{P}}^{-1}z = (\xi, \eta) \in \Omega^k \) then \( Tz = \mathcal{P} \circ \tilde{\mathcal{P}}^k \circ \tilde{\mathcal{P}}^{-1}z \in \Sigma^n \) where \( k = K(\xi) \) and \( n = n(\xi) \).

Having derived explicit expressions for the first return map, in a next step we want to get explicit formulas for the first return time, i.e. the hyperbolic length between the successive returns to \( \Sigma \).

### 4.4. The first return time.

**Lemma 80.** Let \( \gamma = (\xi, \eta) \in \Upsilon_r \) with \( \xi = [a_0; (e)^{k-1}, a_k, \ldots] \) (\( e = \text{sign}(a_0) \)) and \( \mathcal{P}^\gamma = z_0 = (z_0, \theta_0) \) and let \( z_1 = Tz_0 \). For \( w_0 \in \gamma \) the point corresponding to \( z_1 \), i.e. \( w_0 \in G_q z_1 \), one has

\[
d(z_0, w_0) = \ln g(z_0, \gamma) - \ln g(z_1, \tilde{T} \gamma) + 2 \ln F(\gamma)
\]

where \( F(\gamma) = \prod_{j=1}^k |\xi_j| \) with \( k = K(\xi) \) as in Definition 77 and \( \xi_j = SF_q^i S \xi \).

**Proof.** Set \( \gamma_j := \tilde{\mathcal{P}}^j \gamma = B_j \gamma \) and \( w_j := B_j w_0 \). By Lemma 65 \( g(w_0, \gamma) = g(B_0^{-1} w_1, B_0^{-1} \gamma_1) = g(w_1, \gamma_1) \xi_1^{-2} \). Applying the same formula to \( g(w_j, \gamma_j) \) for \( j = 1, \ldots, k \) we get \( \ln g(z_0, \gamma) = \ln g(z_1, \tilde{T} \gamma) - 2 \ln \prod_{j=1}^k |\xi_j| \). The statement then follows by Proposition 79 and Lemma 66.

**Lemma 81.** If \( \gamma = (\xi, \eta) \) is a reduced closed geodesic with \( c_q (S \xi) = [a_1, \ldots, a_n] \) of minimal period \( n \) and \( \xi_j = SF_q^i S \xi \), then the hyperbolic length of \( \gamma \) is given by

\[
l(\gamma) = 2 \sum_{j=1}^n \ln |\xi_j| = -\ln \prod_{j=1}^n |[a_{j+1}, \ldots, a_n, a_1, \ldots, a_j]|^2.
\]

**Proof.** Denote by \( (z_j, \theta_j) \in \Sigma \) the successive returns of \( \gamma \) to \( \Sigma \) and let \( w_{j-1} \in \gamma \) be the point on \( \gamma \) corresponding to \( z_j \). If \( \gamma \) is closed, the set \( \{z_j\}_{j \geq 0} \) is finite with \( N + 1 \) elements for some \( N + 1 \leq n \), i.e. \( z_{N+1} = z_0 \). It is clear that the length of \( \gamma \) is given by adding up the lengths of all pieces between the successive returns to \( \Sigma \) and a repeated application of Lemma 80 gives us

\[
l(\gamma) = \sum_{j=0}^N d(z_j, w_j) = \sum_{j=0}^N \left( \ln g(z_j, \tilde{T}^j \gamma) - \ln g(w_j, \tilde{T}^j \gamma) \right)
\]

\[
= \sum_{j=0}^N \left( \ln g(z_j, \tilde{T}^j \gamma) - \ln g(z_{j+1}, \tilde{T}^{j+1} \gamma) + 2 \ln F(\tilde{T}^j \gamma) \right)
\]

\[
= 2 \sum_{j=0}^N \ln F(\tilde{T}^j \gamma) = 2 \ln \prod_{i=1}^n |SF_q^i S \xi|.
\]

\( \square \)
**Remark 82.** Formula (3) can also be obtained by relating the length of \( \gamma(\xi, \eta) \) to the axis of the hyperbolic matrix fixing \( \xi \) and observing that this matrix must be given by the map \( F^n_\eta \) acting on \( \xi \).

In the case of \( \text{PSL}_2(\mathbb{Z}) \) and the Gauss (regular) continued fractions formula (3) is well-known.

We are now in a position to discuss the first return time. By Lemma 66 it is clear that we need to calculate the function \( g_j(\xi, \eta) = \frac{|w_j - z_j|}{z_j} \), where \( w_j = u_j + iv_j = \mathcal{Z}_j(\xi, \eta) \) for all the intersection points in Corollary 70.

**Definition 83.** Let \( B \in \text{PSL}_2(\mathbb{R}) \) be given by \( Bz = \frac{2 - \lambda z}{\lambda - 2z} \). Set \( \delta_n(\xi, \eta) := \eta - T^nBT^{-n}\xi, \)

\[
\begin{align*}
\Xi_+ &:= \left\{ (\xi, \eta) \in \Omega_r^+ \mid \delta_{(\xi),\lambda-1}(\xi, \eta) \geq 0 \right\}, \\
\Xi_- &:= \left\{ (\xi, \eta) \in \Omega_r^+ \mid \delta_{(\xi),\lambda-1}(\xi, \eta) < 0 \right\}.
\end{align*}
\]

**Proposition 84.** The first return time \( r \) for the geodesic \( \gamma(\xi, \eta) \) is given by the function

\[
r(\xi, \eta) = \ln g_A(\xi, \eta) - \ln g_B(\xi, \eta) + 2 \ln F_{\gamma}(\xi, \eta) \quad \text{for} \quad \xi > 0,
\]

respectively

\[
r(\xi, \eta) = r(-\xi, -\eta), \quad \text{for} \quad \xi < 0.
\]

Thereby \( A = A(\xi, \eta) \) is given by

\[
A = \begin{cases} 
-1, & -R \leq \eta < -\frac{1}{2}, \quad \xi > -B(-\eta), \\
0, & -R \leq \eta < -\frac{1}{2}, \quad \xi < -B(-\eta), \text{ or } -\frac{1}{2} \leq \eta < -r, \quad \xi \geq B(\eta), \\
2, & \frac{3}{2} \lambda - \frac{1}{2} \leq \eta < -r, \quad \frac{3}{2} < \xi < B(\eta) < \lambda + 1, \\
3, & \lambda - 1 < \eta < -r, \quad \lambda + 1 < \xi < B(\eta),
\end{cases}
\]

\( K(\xi) \) and \( n(\xi) \) are defined as in Definition 77, whereas the functions \( g_j(\xi, \eta) = g(z_j, \gamma) \) for \( z_j \in L_j \) are given as in Corollary 70 and \( F_{\gamma}(\xi, \eta) \) is given as in Lemma 80.

**Proof.** Consider \( \gamma = \gamma(\xi, \eta) \in \mathcal{Y}_\tau \) with \( \xi > 0 \) and suppose that \( \mathcal{P}\gamma = \mathcal{Z}_0 \in \Sigma \) and \( \mathcal{T}\mathcal{Z}_0 = \mathcal{Z}_1 \in \Sigma \) with \( w \in \gamma \) corresponding to \( \mathcal{Z}_1 \). Since geodesics are parametrized by arc length the first return time is simply the hyperbolic length between \( \mathcal{Z}_0 \) and \( w \), i.e.

\[
r(\xi, \eta) = d(\mathcal{Z}_0, w) = \ln g(\mathcal{Z}_0, \gamma) - \ln g(w, \gamma) \\
= \ln g(\mathcal{Z}_0, \gamma) - \ln g(\mathcal{Z}_1, \tilde{F}_{\gamma}\gamma) + 2 \ln F(\gamma)
\]

by Lemma 80. If \( \mathcal{Z}_0 \in L_j \), we set \( g(\mathcal{Z}_0, \gamma) = g_j(\xi, \eta) \) as given in Corollary 70. By Corollary 72 it is easy to verify, that the sets in the definition of \( A(\xi, \eta) \) correspond exactly to the cases \( \mathcal{Z}_0 \in L_{-1}, L_0, L_2 \) and \( L_3 \) respectively, where the last set is
empty for even \( q \). It is also easy to see, that \( B \left( \frac{2}{3} \lambda - \frac{1}{2} \right) = \frac{21}{2} \) and \( B ( \lambda - 1) = \lambda + 1 \). The statement of the Proposition now follows from the explicit formula for \( F(\gamma) \) in Lemma 80 and the domains in Proposition 79 for which \( T = F^k_q \). That \( r (\xi, -\eta) = r (\xi, \eta) \) follows from the invariance of the cross-section with respect to reflection in the imaginary axis.

5. Construction of an Invariant Measure

By Liouville's theorem we know that the geodesic flow on \( T^1 \mathcal{H} \) preserves the measure induced by the hyperbolic metric. This measure, the Liouville measure, is given by \( dm = y^{-2}dx dy d\theta \) in the coordinates \((x + iy, \theta) \in \mathcal{H} \times S^1 \) on \( T^1 \mathcal{H} \). Using the coordinates \((\xi, \eta, s) \in \mathbb{R}^3 \) given by Corollary 63 we obtain the Liouville measure in these coordinates

\[
\frac{dm}{g^2} = \frac{|\partial (x, y, \theta)|}{\partial (\xi, \eta, s)} \frac{2}{r^2 \cos^2 \theta} d\xi d\eta ds = \frac{2d\xi d\eta ds}{(\eta - \xi)^2}.
\]

The time discretization of the geodesic flow on \( T^1 \mathcal{M} \) in terms of the cross-section \( \Sigma \) and the first return map \( \tilde{T} : \Sigma \rightarrow \Sigma \) thus preserves the measure \( dm' = \frac{2d\xi d\eta}{(\eta - \xi)^2} \) on \( \Omega^\infty_r \). We prefer to work with the finite domain \( \Omega^\infty_r \subset \Omega^\infty \). Hence the measure \( d\mu (u, v) = dm' (\xi, \eta) \) given by

\[
d\mu = \frac{2dudv}{(1 - uv)^2}
\]

on \( \Omega^\infty_r \) is invariant under \( \tilde{T} (u, v) = F^q_k (u, v) = (f_1 (u), f_2 (v)) \) where \( f_1 (u) = F^q_k (u) u = Au \) and \( f_2 (v) = A^{-1} v \). Because \( d\mu \) is equivalent to Lebesgue measure, we deduce that \( d\mu \) is in fact an \( F^q_k \) invariant measure on \( \Omega^\infty_r \). If \( \pi_x (x, y) = x \) it is clear that \( \pi_x \circ \tilde{T} (u, v) = f_1 (u) = f_1 \circ \pi_x (u, v) \) so \( f_1 \) is a factor map of \( \tilde{T} \). An invariant measure \( \hat{\mu} \) of \( f_1 : I_q \rightarrow I_q \) can be obtained by integrating \( d\mu \) in the \( v \)-direction. We get different alternatives depending on \( q \) being 3, even or odd greater than 3.

5.1. \( q = 3 \). In this case \( F^q_k = F_q \) and if we set \( U_1 = [-\frac{1}{2}, 0] = -U_{-1} \) and \( V_1 = [r, R] \) then \( d\hat{\mu} = \chi_U d\mu_1 + \chi_{U_{-1}} d\mu_{-1} \) where

\[
d\mu_1 (u) = \int_1^R \frac{2dudv}{(1 - uv)^2} = \left[ \frac{1}{u (1 - uv)} \right]_{u = R}^{v = R} du = \frac{1}{u (1 - ur)} - \frac{1}{u (1 - ur)} du = \frac{1}{u (1 - ur) (1 - ur)} du = \frac{1}{(1 - ur) (1 - ur)} du
\]

and \( d\mu_{-1} (u) = -d\mu_1 (-u) = \frac{1}{(1 + ur) (1 + ur)} du \). Here \( r = \frac{\sqrt{3} - 3}{2} \) and \( R = r + 1 = \frac{\sqrt{3} - 1}{2} \).
5.2. Even \( q \geq 4 \). Here \( U_1 = \left[ -\frac{1}{2}, -\frac{2}{3\lambda} \right] \), \( U_2 = \left[ -\frac{2}{3\lambda}, 0 \right] \), \( V_1 = [0, R] \) and \( V_2 = [r, R] \). Hence
\[
\frac{d\mu_1}{du}(u) = \int_0^R \frac{2dv}{(1-u)^2} = \frac{1}{u(1-u)} - \frac{1}{u} = \frac{R}{1-uR},
\]
\[
\frac{d\mu_2}{du}(u) = \int_0^U \frac{2dv}{(1-u)^2} = \frac{1}{u(1-u)} - \frac{1}{u} = \frac{R}{1-uR},
\]
\[
\frac{d\mu_{-j}}{du}(u) = \frac{d\mu_j}{du}(-u)
\]
and the invariant measure \( \tilde{\mu} \) of \( F_q^K \) for even \( q \) is given by
\[
d\tilde{\mu}(u) = \sum_{j=-3}^3 \chi_{U_j} (u) d\mu_j (u)
\]
where \( \chi_{U_j} \) is the characteristic function for the interval \( U_j \). This measure is piece-wise differentiable and finite. The finiteness is clear since \( uR \) and \( ur \neq 1 \) for \( u \in I_q \). If \( \int_{I_q} d\tilde{\mu}(u) = c \) then \( \frac{1}{c} d\tilde{\mu} \) is a probability measure on \( I_q \).

5.3. Odd \( q \geq 5 \). Let \( U_1 = \left[ -\frac{\lambda}{2}, -\frac{2}{3\lambda} \right] \), \( U_2 = \left[ -\frac{2}{3\lambda}, -\frac{1}{2\lambda} \right] \), \( U_3 = \left[ -\frac{1}{2\lambda}, 0 \right] \), \( V_1 = [0, R] \), \( V_2 = [r_{\lambda-1}, R] \) and \( V_3 = [r, R] \). Then
\[
\frac{d\mu_1}{du}(u) = \int_0^R \frac{2dv}{(1-u)^2} = \frac{1}{u(1-u)} - \frac{1}{u} = \frac{R}{1-uR},
\]
\[
\frac{d\mu_2}{du}(u) = \int_{r_{\lambda-1}}^R \frac{2dv}{(1-u)^2} = \frac{1}{u(1-u)} - \frac{1}{u} = \frac{R}{1-uR},
\]
\[
\frac{d\mu_3}{du}(u) = \int_{r_{\lambda-1}}^R \frac{2dv}{(1-u)^2} = \frac{1}{u(1-u)} - \frac{1}{u} = \frac{R}{1-uR},
\]
\[
\frac{d\mu_{-j}}{du}(u) = \frac{d\mu_j}{du}(-u)
\]
and the invariant measure \( \tilde{\mu} \) of \( F_q^K \) for odd \( q \) is given by
\[
d\tilde{\mu}(u) = \sum_{j=-3}^3 \chi_{U_j} (u) d\mu_j (u)
\]
where \( \chi_{U_j} \) is the characteristic function for the interval \( U_j \). This measure is piece-wise differentiable and finite. The finiteness is clear since \( uR \), \( ur \) and \( ur_{\lambda-1} \neq 1 \) for \( u \in I_q \). If \( \int_{I_q} d\tilde{\mu}(u) = c \) then \( \frac{1}{c} d\tilde{\mu} \) is a probability measure on \( I_q \).

Remark 85. The geodesic flow on finite surfaces of constant negative curvature has been shown by Ornstein and Weiss [39] to be Bernoulli with respect to Liouville.
measure. Since our cross section is smooth also the Poincaré map $T$ is Bernoulli [38] with respect to the induced measure $d\mu$. But the factor map of a Bernoulli system is again Bernoulli [37]. Hence also the map $f_1 : I_q \to I_q$ is Bernoulli with respect to the measure $\tilde{\mu}$.

**Remark 86.** For another approach leading to an infinite invariant measure see e.g. Haas and Gröchenig [14].

5.4. Invariant measure for $F_q$. It is easy to verify that $dm (\xi, \eta) = \frac{2d\xi d\eta}{(\xi - \eta)^2}$ is invariant under Möbius transformations, i.e. if $A \in \text{PSL}_2(\mathbb{R})$ then $dm (A\xi, A\eta) = dm (\xi, \eta)$. By considering the action of $F_q$ on $\Omega^*$, i.e.

$$F_q^* (\xi, \eta) = \tilde{S} \circ F_q \circ \tilde{S} (\xi, \eta) = \left( SF_q S\xi, \frac{1}{n\lambda - \eta} \right) = \left( ST^{-n}\xi, ST^{-n}\eta \right)$$

it is clear that $dm$ is invariant under $F_q^* : \Omega^* \to \Omega^*$ and letting $u = S\xi$ and $v = -\eta$ it is easy to verify that $d\mu (u, v) = \frac{2d\nu d\nu}{(1-wv)^2}$ is invariant under $F_q : \Omega \to \Omega$. We can thus obtain corresponding invariant measure $d\tilde{\mu}$ for $F_q$ by projecting on the first variable. Let $d\tilde{\mu} (u) = d\mu_j (u)$ for $u \in I_j$. Then

$$\frac{d\mu_j}{du} (u) = 2 \int_{I_j} \frac{dv}{(1-wv)^2} \left[ \frac{1}{u} \right]^{R_j} = 2 \left[ \frac{1}{1-Ru} - \frac{1}{1-r_j u} \right] = \frac{2(1-R-u)}{(R-u)(1-r_j u)}$$

and the invariant measure $\tilde{\mu}$ of $F_q$ is given by

$$d\tilde{\mu} (u) = \sum_{j=-\infty}^{\infty} \chi_{I_j} (u) d\mu_j (u)$$

where $\chi_{I_j}$ is the characteristic function for the interval $I_j$. This measure is piecewise differentiable and finite. If $c = \int_{I_q} d\tilde{\mu}$ then $\frac{1}{c}d\tilde{\mu}$ is a probability measure on $I_q$. It turns out that $c = \frac{1}{2}C$ where $C^{-1} = \ln \left( \frac{1}{\sin \frac{\pi q}{q}} \right)$ for even $q$ and $C^{-1} = \ln (1 + R)$ for odd $q$ (see Lemma 3.2 and 3.4 in [8]).

6. Lemmas on continued fraction expansions and reduced geodesics

This section contains a collection of rather technical lemmas necessary to show that the first return map on $\Sigma$ in Lemma 78 is given by powers of $F_q$.

**Lemma 87.** If $\gamma \in \Upsilon$ with $\langle \gamma^+ \rangle_\lambda = a_0$ intersects $T^{\alpha_0 - 1} L_0$ then

$$\gamma_+ = \begin{cases} [a_0; (1^{h-1}, 2)^l, 1^h, a_{(l+1)h+1}, \ldots] & \text{for some } 0 \leq l \leq \infty, \\ [a_0; 1^h, a_{h+1}, \ldots] & \text{for } a_{(l+1)h+1} \leq -1, \text{ if } q \text{ is even and } a_{h+1} \geq 2, \text{ if } q \text{ is odd.} \end{cases}$$
Proof. Let \( x = \gamma_+ - a_0 \lambda \). By convexity \( \gamma \) does not intersect \( \overrightarrow{T^{a_0-1}L_0} \) more than once. Hence \( x \in \left( -\frac{1}{2}, 1 - \lambda \right] \). If \( q \) is odd, then \( c_q(1) = [1^h; 1^{h+1}] \) and \( c_q(-\frac{1}{2}) = [1^{h+1}, 2; 1^{h}] \) according to Lemmas 6 and 5 hence \( [1^{h+1}, 2; 1^{h}] < c_q(x) < [1^h; 1^{h+1}] \) and by the lexicographic ordering (see proof of Lemma 24) it is clear, that \( c_q(x) = [1^{h}, a_{h+1}, \ldots] \) with \( a_{h+1} \geq 2 \). If \( q \) is even then \( c_q(-\frac{1}{2}) = [1^h; 1^{h+1}] \) and \( 1 - \lambda = r \) with \( c_q(r) = \left[ 1^{h-1}, 2 \right] \) so that \( [1^{h}] < c_q(x) \leq [1^{h-1}, 2] \). By the lexicographic ordering it is clear, that \( c_q(x) = \left[ (1^{h-1}, 2), 1^h, a_{(l+1)h+1}, \ldots \right] \) for some \( l \geq 0 \) \((l = \infty \) is allowed) and \( a_{(l+1)h+1} \leq -1 \) if \( l < \infty \).

**Lemma 88.** Let \( \gamma \in \Upsilon \) with \( \langle \gamma_+ \rangle = a_0 \). Then \( \overrightarrow{T^{-\text{sign}(a_0)}\gamma} \notin \Upsilon \) for \( n \geq 1 \).

**Proof.** Without loss of generality assume \( a_0 \geq 1 \) and let \( \gamma^n = T^{-n}\gamma \) for \( n \geq 1 \). Then \( \gamma^n_+ = \gamma_+ + n\lambda > \gamma_+ \) and \( \gamma^n_- = \gamma_- + n\lambda \). Since \( \gamma \in \Upsilon \) \( \gamma_- \geq -R \Rightarrow \gamma^n_- \geq n\lambda - R = -r + (n - 1)\lambda \geq -r \). Hence \( \gamma^n_- \notin [-R, -r] \) so \( \gamma^n \notin \Upsilon \). The case of \( a_0 \leq -1 \) is analogous.

**Lemma 89.** If \( \gamma \in \Upsilon \) then \( \gamma^n = ST^{-n}\gamma \in \Upsilon \) if and only if \( n = a_0 = \langle \gamma_+ \rangle \).

**Proof.** By definition, if \( \gamma^n \in \Upsilon \) then \( S\gamma^n_+ = \gamma_+ - n\lambda \in I^\infty_+ \) and hence \( \gamma_+ \in (n\lambda - \frac{1}{2}, n\lambda + \frac{1}{2}) \Rightarrow n = a_0 \). It is also clear that \( \gamma^{a_0} = F_q\gamma \in \Upsilon \).

**Lemma 90.** Suppose that \( q \) is even. If \( \gamma = \langle \xi, \eta \rangle \in \Upsilon \) with \( a_0 = \langle \gamma_+ \rangle \geq 2 \) intersects \( \overrightarrow{T^{a_0-1}L_0} \) then the first return map is given as \( \overrightarrow{T}(\xi, \eta) = F_q K(\xi, \eta) \in \overrightarrow{P}^{-1}(\Sigma^m(\xi)) \) where \( K(\xi) \) and \( n(\xi) \) are as in Definition 77.

**Proof.** Consider Figures 6 and 8 showing the arcs around the point \( \rho \). The picture is symmetric with respect to \( \Re z = \frac{1}{2} \) and invariant under translation, so it applies in the present case. After passing through \( \overrightarrow{T^{a_0-1}L_0} \) the geodesic \( \gamma \) will intersect a sequence of translates of the arcs \( \overrightarrow{\alpha_j} \) and \( \overrightarrow{\omega_j} \) which are the reflections of \( \chi_j \) and \( \omega_j \) in \( \Re z = \frac{1}{2} \) exactly as in Lemma 66, except that it now passes through every arc. Note, even the argument why \( \gamma \) intersects \( \overrightarrow{\omega_j} \) and not its extension applies. Let \( w_n \) and \( A_n \) be as in Definition 61 except that now \( w_0 \in \overrightarrow{T^{a_0-1}L_0} \).

Then \( w_{2j} \in \overrightarrow{T^{a_0-1} \overrightarrow{\alpha_j}} \) \( = T^{a_0-1}(ST^{-1})^j SL_0 \) for \( 0 \leq j \leq h \) and also \( w_{2j+1} \in \overrightarrow{T^{a_0-1} \overrightarrow{\omega_j}} = (ST^{-1})^{j+1} L_1 \), for \( 0 \leq j \leq h \) with corresponding maps \( A_{2j+1} = (TS)^{j+1} T^{a_0} \) and \( A_{2j} = (ST)^j ST^{a_0-1} \). Set \( \gamma := A_k \gamma \) and \( \xi_k := A_k \xi \) for \( 0 \leq k \leq 2h + 1 \). There are three cases when the point \( w_k \) can define a return to \( \Sigma \):

- a) if \( \gamma_{2j+1} \in \Upsilon \) and \( z = \overrightarrow{\gamma_{2j+1}} \in \Sigma^1 \),
- b) if \( \gamma_{2j-1} \not\in \Upsilon \) but \( \gamma_{2j-1} = TS \gamma_{2j-1} \in \Upsilon \) with \( \xi_{2j-1} \in \left( \frac{3\lambda}{2}, \lambda + 1 \right) \) and \( z = \overrightarrow{\gamma_{2j-1}} \in \Sigma^2 \), or
- c) if \( \gamma_{2j} \in \Upsilon \) and \( z = \overrightarrow{\gamma_{2j}} \in \Sigma^0 \).

According to Lemma 87 we get \( T^{a_0-1} \xi \in \left( \frac{1}{2}, 1 \right] \) and \( c_q(\xi) = [a_0; 1^{h-1}, a_h, a_{h+1}, \ldots] \) with \( a_h = 2 \) or \( a_{h} = 1 \) and \( a_{h+1} \leq -1 \). Also \( \frac{1}{2} = \left( (-1)^h \right) \) and \( 1 = R = \left( (-1)^h , -2, (-1)^{h-1} \right) * \).
Hence $\xi_{2j+1} \in (-\phi_{j+1} - r_{h-j}) \subseteq I_q \Rightarrow \gamma_{2j+1} \notin \Upsilon$ for $1 \leq j \leq h - 1$ (cf. Remark 36). Also note, that $A_{2h+1} = (TS)^{h+1} T^{1-a_0} = (ST^{-1})^h ST^{-a_0}$, therefore $\gamma_{2h+1} = A_{2h+1} \gamma = ST^{-1} F_q \gamma$ and hence $\gamma_{2h+1} \notin \Upsilon$ unless $a_h = 1$ in which case $\gamma_{2h+1} = \tilde{F}_q \gamma \in \Upsilon$ and $P \gamma_{2h+1} \in \Sigma^1$.

For $\gamma'_{2j+1} = TS \gamma_{2j+1} = \gamma_{2j+3}$ we conclude $\gamma'_{2j+1} \notin \Upsilon$ for $0 \leq j \leq h - 2$. For $a_h = 1$ we find $\gamma'_{2h-1} = \gamma_{2h+1} \in \Upsilon$ but $\xi'_{2h-1} < 0$ and hence we do not get a point of $\Sigma^2$. If $a_h = 2$ on the other hand then $\gamma'_{2h-1} \notin \Upsilon$ but $\gamma_{2h+1} = TS \gamma_{2h+1} = \tilde{F}_q \gamma \in \Upsilon$ and $\xi'_{2h+1} > \frac{3\lambda}{2}$ so $P \gamma_{2h+1} \in \Sigma^2$.

For $\gamma_{2j} = S \gamma_{2j-1}$ we get $\xi_{2j} \in S (-\phi_j - r_{h+1-j}) = (-\lambda - \phi_{j+1} - \lambda - r_{h-j})$ and therefore $-\frac{3\lambda}{2} < \xi_{2j} < -\frac{\lambda}{2}$ for $0 \leq j \leq h - 1$. Since $T^{1-a_0} \eta < -R$ we have $\eta_{2j} \in S (TS)^j (-\infty, -R) = (\phi_{h-j}, r_j)$ that is $\eta_{2j} < 0$. Hence $\gamma_{2j} \notin \Upsilon$ for $0 \leq j \leq h - 1$. Note, that $F_q^h \xi > \frac{2}{h}$ implies that $\xi_{2h} = T^{-1} ST^{-1} F_q^h \xi \in T^{-1} ST^{-1} \left( \frac{\lambda}{2}, \infty \right)$ for $q > 4$ and $\xi_{2h} \in (-\infty, -\lambda)$ for $q = 4$. In any case $\xi_{2h} < 0$ and since $\eta_{2h} < r$ it is clear, that $\gamma_{2h} \notin \Upsilon$.

We conclude that the first return is given by $w_{2h+1}$ and $\tilde{T} (\xi, \eta) = F_q^{h+1} (\xi, \eta) \in \tilde{P}^{-1} (\Sigma^1)$ if $a_h = 1$ and $\tilde{T} (\xi, \eta) = F_q^h (\xi, \eta) \in \tilde{P}^{-1} (\Sigma^2)$ if $a_h = 2$. This can be written in the form $\tilde{T} = F_q^{K(\xi)} \in \tilde{P}^{-1} (\Sigma^2)$ with $K (\xi)$ and $n (\xi)$ as in Definition 77.

**Lemma 91.** For $q$ even consider $\xi \in (-\frac{1}{2}, 1 - \lambda)$ with $c_q (\xi) = [0; 1^{-1}, a_h, a_{h+1}, \ldots]$. Then $a_h = 1$ if and only if $\xi < -\frac{\lambda^3}{4 + \lambda^2}$.

**Proof.** By Lemma 87 we know either $a_h = 1$ and $a_{h+1} \leq -1$ or $a_h = 2$. It is clear that the boundary point between these two cases is given by $x_0 = [0; 1, (-1)^h] = (ST)^h \left( \frac{1}{2} \right)$. Using (2) one can show that $\xi_0 = (ST)^h \left( \frac{1}{2} \right) = \frac{(\lambda^2 - 2)\lambda + \lambda}{-\lambda^2 + 2} = \frac{-\lambda^3}{4 + \lambda^2}$.

The following corollary is easy to verify by estimating the intersection of $\gamma (-r, (a_0 - \frac{1}{2}) \lambda)$ and $T^{a_0 - 1} L_0$. It implies that the case $T = F_q^h$ does not occur for $\{ \xi \} \lambda \geq 3$.

**Corollary 92.** Let $q$ be even and suppose that $\gamma = \gamma (\xi, \eta) \in \Upsilon$ with $a_0 = \{ \xi \} \lambda \geq 3$. If $\gamma$ intersects $T^{a_0 - 1} L_0$ then $c_q (\xi) = [a_0; 1^{-1}, a_{h+1}, \ldots]$ with $a_{h+1} \leq -1$.

**Lemma 93.** Let $q$ be odd and suppose that $\gamma \in \Upsilon$. Let $l$ be the geodesic arc $[\rho + \lambda, 1 + \lambda]$, i.e. the continuation of $L_3$. Then $\gamma$ does not intersect $L^\perp$ outwards (i.e. in the direction from $0$ to $\pm \infty$).

**Proof.** Take $\gamma = \gamma (\xi, \eta) \in \Upsilon$ and assume $\xi > 0$. Suppose that $\gamma$ intersects $l$ in the outwards direction. Since $\gamma$ can not intersect the geodesic $T S \tilde{L}_{-1} = [\lambda, \lambda + \frac{2}{\lambda}]$ more than once we have $\xi \in (\lambda + 1, \lambda + \frac{2}{\lambda})$ and because $\gamma \in \Upsilon$ we have $-R \leq \eta < -r$. If $w (\xi, \eta)$ is the intersection between $\gamma$ and the line $T \tilde{L}_1 = \frac{2}{\lambda} + i \mathbb{R}^+$ then $\exists w (\xi, \eta) > \exists w (\xi, -r) \geq \exists w (\lambda + 1, -r)$. To show that $\exists w > \exists T \rho =$
\( \Im \rho = \sin \frac{\pi}{q} \) it is enough to bound \( \Im w(\lambda + 1, -r) \) from below. By Lemma 69 we have

\[
\Im w(\lambda + 1, -r)^2 = \left( \lambda + 1 - \frac{3\lambda}{2} \right) \left( \frac{5\lambda}{2} - R \right) = \left( 1 - \frac{\lambda}{2} \right) \left( 1 + \frac{\lambda}{2} + (2\lambda - R - 1) \right) = \sin^2 \frac{\pi}{q} + \left( 1 - \frac{\lambda}{2} \right) (2\lambda - R - 1) > \sin^2 \frac{\pi}{q}
\]

since \( 2\lambda > R + 1 \) and \( 1 - \frac{\lambda}{2} > 0 \). Hence \( \Im w(\xi, \eta) > \sin \frac{\pi}{q} \) and \( \gamma \) does not intersect \( l \) in the direction from 0 to \( \infty \). An analogous argument for \( \xi < 0 \) concludes the Lemma.

**Lemma 94.** For \( q \) odd, let \( \gamma = (\xi, \eta) \in \Upsilon_r \) be reduced with \( a_0 = \{ \xi \}_\lambda \geq 2 \). If \( \gamma \) intersects \( T^{a_0-1}L_0 \) then \( \hat{T}(\xi, \eta) = F_{q}^{h+1}(\xi, \eta) \in \hat{\Pi}^{-1}(\Sigma^3) \).

**Proof.** Consider once more Figure 6 showing the arcs around \( \rho \). Analogous to the proof of Lemma 90 we have \( w_{2j} \in T^{a_0-1}\eta_j = T^{a_0-1}(ST^{-1})^jSL_0 \), for \( 0 \leq j \leq h + 1 \) and \( w_{2j+1} \in T^{a_0-1}\eta_j = T^{a_0-1}(ST^{-1})^{j+1}L_1 \), for \( 0 \leq j \leq h \) with the corresponding maps \( A_{2j+1} = (TS)^{j+1}T^{1-a_0} \) and \( A_{2j} = (ST)^jST^{1-a_0} \). Set \( \gamma_j := A_j \gamma \) and \( \xi_j := A_j \xi \). There are now four possibilities to produce a return to \( \Sigma \):

a) if \( \gamma_{2j+1} \in \Upsilon_r \) and \( z = P\gamma_{2j+1} \in \Sigma^1 \),
b) if \( \gamma_{2j-1} \notin \Upsilon \) but \( TS\gamma_{2j+1} \in \Upsilon_r \) and \( z = PTS\gamma_{2j+1} \in \Sigma^2 \),
c) if \( \gamma_{2j} \in \Upsilon_r \) and \( z = P\gamma_{2j} \in \Sigma^0 \),
d) if \( \gamma_{2j} \notin \Upsilon \) but \( T^{\pm 1}\gamma_{2j} \in \Upsilon_r \) and \( z = P\gamma_{2j} \in \Sigma^{\pm 3} \).

We will see that most of these cases do not give a return. Since \( T^{1-a_0}\xi \in \left( \frac{1}{2}, 1 \right) \) Lemma 87 shows that \( c_{\pi}(\xi) = [a_0; 1^h, a_{h+1}, \ldots] \) with \( a_{h+1} \geq 2 \). Suppose also, that \( c_{\pi}(\eta) = [0; b_1, b_2, \ldots] \). For the following arguments it is important to remember that the action of \( TS \) on \( \partial \mathcal{H} \cong \mathbb{R}^* \cong \mathbb{S}^1 \) is monotone as a rotation around \( \rho \).

Since \( \gamma_{2j+1} = (TS)^{j+1}T^{1-a_0,\gamma} \) we have \( \xi_{2j+1} \in (TS)^{j+1}\left( \frac{1}{2}, 1 \right) \). But \( \gamma_{2j+1} \notin \Upsilon \) for \( 0 \leq j \leq h - 1 \). Furthermore \( \xi_{2h+1} \in TS(-\phi_{2h}, -\phi_{2h+1}) = (-1, 0) \) and therefore \( |\xi_{2h+1}| < \frac{2}{\pi} \), so that also \( \gamma_{2h+1} \notin \Upsilon \).

If \( \gamma_{2j+1} = TS\gamma_{2j+1} = \gamma_{2j+3} \) then \( \gamma_{2j+3} \notin \Upsilon \) for \( 1 \leq j \leq h - 1 \) and since \( \gamma_{2j+3} = A_{2j+3} \gamma = A_{2j+3}\gamma = \hat{F}^{h+1}_q \gamma \in \Upsilon_r \) and we have a return at \( w_{2h+1} \) with \( z_1 = 1 \).

Since \( \gamma_{2j} = (ST)^jST^{1-a_0,\gamma} \) obviously \( \xi_{2j} \in (TS)^{j}\left( \frac{1}{2}, 1 \right) \) and \( \xi_{2j} < \frac{\pi}{2} \), hence \( \xi_{2j} < -\frac{\pi}{2} \). But \( T^{1-a_0}\eta < -r - \lambda = -R \) and hence \( \eta_{2j} \in (ST)^jS(-\infty, -R) = (\phi_{2h-1} + 1, r_{2j+1}) \) for \( 0 \leq j \leq h \) (cf. Remark 37) respectively \( \eta_{2j} \in \hat{F}_q \gamma_{2h-1} \). But \( \gamma_{2j} \notin \Upsilon_r \) for \( 0 \leq j \leq h + 1 \).
Finally, since $T^{-1} \eta_{2j} < r_{2j+1} - \lambda < -R$ and $T \xi_{2j} \in I_q$ it is clear that $T^{\pm 1} \gamma_{2j} \notin \Upsilon$ for $0 \leq j \leq h + 1$.

Acknowledgments. The authors appreciate also the many helpful discussions with Tobias Mülhenbruch on different aspects of this paper.

References

[1] R. L. Adler and L. Flatto. Cross section maps for geodesic flows. I. The modular surface. In Ergodic theory and dynamical systems, II (College Park, Md., 1979/1980), volume 21 of Progr. Math., pages 103–161. Birkhäuser Boston, Mass., 1982.
[2] R. L. Adler and L. Flatto. The backward continued fraction map and geodesic flow. Ergodic Theory Dynam. Systems, 4(4):487–492, 1984.
[3] R. L. Adler and L. Flatto. Cross section map for the geodesic flow on the modular surface. In Conference in modern analysis and probability (New Haven, Conn., 1982), volume 26 of Contemp. Math., pages 9–24. Amer. Math. Soc., Providence, RI, 1984.
[4] R. L. Adler and L. Flatto. Geodesic flows, interval maps, and symbolic dynamics. Bull. Amer. Math. Soc. (N.S.), 25(2):229–334, 1991.
[5] E. Artin. Ein mechanisches System mit quasiergodischen Bahnen. Hamb. Math. Abh., 3:170–177, 1924.
[6] D. Birkhoff. Quelques théorèmes sur le mouvement des systèmes dynamiques. Bull. Soc. Math. France, 40:305–323, 1912.
[7] R. Bowen and C. Series. Markov maps associated with Fuchsian groups. Inst. Hautes Études Sci. Publ. Math., (50):153–170, 1979.
[8] R. M. Burton, C. Kraaikamp, and T. A. Schmidt. Natural extensions for the Rosen fractions. Trans. Amer. Math. Soc., 352(3):1277–1298, 2000.
[9] Paula Cohen and J. Wolfart. Modular embeddings for some nonarithmetic Fuchsian groups. Acta Arith., 56(2):93–110, 1990.
[10] K. Dajani, C. Kraaikamp, and W. Steiner. Metrical theory for $\alpha$-Rosen fractions. Arxiv:math.NT/0702516v1, Feb 2007.
[11] M. Einsiedler and T. Ward. Ergodic theory: with a view towards number theory. In preparation.
[12] D. Fried. Symbolic dynamics for triangle groups. Invent. Math., 125(3):487–521, 1996.
[13] D. Fried. Reduction theory over quadratic imaginary fields. J. Number Theory, 110(1):44–74, 2005.
[14] K. Gröchenig and A. Haas. Backward continued fractions, Hecke groups and invariant measures for transformations of the interval. Ergodic Theory Dynam. Systems, 16(6):1241–1274, 1996.
[15] J. Hadamard. Les surfaces à courbures opposées et leurs lignes géodésiques. J. Math. Pures et Appl., 5(4):27–73, 1898.
[16] G. A. Hedlund. On the metric of the non-special geodesics on a surface constant negative curvature. Proc. Natl. Acad. Sci. USA, 19:345–348, 1933.
[17] G. A. Hedlund. On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature. Ann. of Math. (2), 35(4):787–808, 1934.
[18] G. A. Hedlund. Two-dimensional manifolds and transitivity. Ann. of Math. (2), 37(3):534–542, 1936.
[19] G. A. Hedlund. Fuchsian groups and mixtures. Ann. of Math. (2), 40(2):370–383, 1939.
[20] E. Hopf. Fuchsian groups and ergodic theory. Trans. Amer. Math. Soc., 39(2):299–314, 1936.
[21] E. Hopf. Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung. Ber. Verh. Sächs. Akad. Wiss. Leipzig, 91:261–304, 1939.
[22] A. Hurwitz. über eine besondere Art der Kettenbruch-Entwicklung reeller Grössen. Acta Math., 12:367–405, 1889.
[23] S. Katok. Fuchsian Groups. The University of Chicago Press, 1992.
S. Katok and I. Ugarcovici. Arithmetic coding of geodesics on the modular surface via continued fractions. In European women in mathematics—Marseille 2003, volume 135 of CWI Tract, pages 59–77. Centrum Wisk. Inform., Amsterdam, 2005.

S. Katok and I. Ugarcovici. Symbolic dynamics for the modular surface and beyond. Bull. Amer. Math. Soc. (N.S.), 44(1):87–132 (electronic), 2007.

P. Koebe. Riemannsche mannigfaltigkeiten und nichteuklidische raumformen. iv: Verlauf geodätischer linien. Sitzungsberichte Akad. Berlin, 1929:414–457, 1929.

J. Lehner. Discontinuous groups and automorphic functions. Mathematical Surveys, No. VIII. American Mathematical Society, Providence, R.I., 1964.

D. Lind and B. Marcus. An introduction to symbolic dynamics and coding. Cambridge University Press, Cambridge, 1995.

F. Löbell. Über die geodätischen Linien der Clifford-Kleinschen Flächen. Math. Z., 30(1):572–607, 1929.

G. A. Morse, M. Hedlund. Symbolic Dynamics. Amer. J. Math., 60(4):815–866, 1938.

H. M. Morse. A One-to-One Representation of Geodesics on a Surface of Negative Curvature. Amer. J. Math., 43(1):33–51, 1921.

H. M. Morse. Recurrent geodesics on a surface of negative curvature. Trans. Amer. Math. Soc., 22(1):84–100, 1921.

P. J. Myrberg. Ein Approximationssatz für die Fuchsschen Gruppen. Acta Math., 57(1):389–409, 1931.

H. Nakada. Metrical theory for a class of continued fraction transformations and their natural extensions. Tokyo J. Math., 4(2):399–426, 1981.

H. Nakada. Continued fractions, geodesic flows and Ford circles. In Algorithms, fractals, and dynamics (Okayama/Kyoto, 1992), pages 179–191. Plenum, New York, 1995.

J. Nielsen. Om geodætiske linier i lukkede mangfoldigheder med konstant negativ krumning. Mat. Tidsskrift B, 1925:37–44, 1925.

D. Ornstein. Ornstein theory. Scholarpedia, 2008, 3(3):3957.

D. Ornstein and B. Weiss. On the Bernoulli nature of systems with some hyperbolic structure. Ergodic Theory Dynam. Systems, 18(2):441–456, 1998.

D. S. Ornstein and B. Weiss. Geodesic flows are Bernoullian. Israel J. Math., 14:184–198, 1973.

H. Poincaré. Mémoire sur les courbes définies par une équation différentielle. J. Math. Pures Appl., 3(8):251–296, 1882. Ch. V-IX.

J. G. Ratcliffe. Foundations of Hyperbolic Manifolds. Springer-Verlag, 1994.

D. Rosen. A class of continued fractions associated with certain properly discontinuous groups. Duke Math. J., 21:549–563, 1954.

D. Rosen and T. A. Schmidt. Hecke groups and continued fractions. Bull. Austral. Math. Soc., 46(3):459–474, 1992.

D. Rosen and C. Towse. Continued fraction representations of units associated with certain Hecke groups. Arch. Math. (Basel), 77(4):294–302, 2001.

T. A. Schmidt. Remarks on the Rosen $\lambda$-continued fractions. In Number theory with an emphasis on the Markoff spectrum (Provo, UT, 1991), volume 147 of Lecture Notes in Pure and Appl. Math., pages 227–238. Dekker, New York, 1993.

P. Schnutz Schaller and J. Wolfart. Semi-arithmetic Fuchsian groups and modular embeddings. J. London Math. Soc. (2), 61(1):13–24, 2000.

C. Series. Symbolic dynamics for geodesic flows. Acta Math., 146(1-2):103–128, 1981.

C. Series. The modular surface and continued fractions. J. London Math. Soc. (2), 31(1):69–80, 1985.
