CODING OF CLOSED GEODESICS AFTER GAUSS AND MORSE

Svetlana Katok
Department of Mathematics
The Pennsylvania State University
University Park, PA 16802, USA
e-mail: katok_s@math.psu.edu

Abstract. Closed geodesics associated to conjugacy classes of hyperbolic matrices in $SL(2,\mathbb{Z})$ can be coded in two different ways. The geometric code, with respect to a given fundamental region, is obtained by a construction universal for all Fuchsian groups, while the arithmetic code, given by "–" continued fractions, comes from the Gauss reduction theory and is specific for $SL(2,\mathbb{Z})$. In this paper we give a complete description of all closed geodesics for which the two codes coincide.

0. Introduction. The group $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm 1\}$ acts on the upper half-plane $\mathcal{H} = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$ by Möbius transformations:

$$z \to \frac{az+b}{cz+d},$$

which are orientation–preserving isometries of $\mathcal{H}$ endowed with the hyperbolic metric. This action extends to the Euclidean boundary of $\mathcal{H}$, $\mathbb{R} \cup \infty$, by the same formula. The fixed points of $\gamma \in PSL(2,\mathbb{R})$ are found by solving $z = \gamma(z) = \frac{az+b}{cz+d}$. If the corresponding matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is hyperbolic (i.e. $|\operatorname{tr} A| > 2$), γ has two hyperbolic fixed points in $\mathbb{R} \cup \infty$, which are the roots of the quadratic polynomial

$$Q_A(z,1) = cz^2 + (d-a)z - b = 0 \tag{0.1}$$

of discriminant $D = (a+d)^2 - 4 > 0$. One point, denoted by w, is attracting ($A'(w) = \frac{1}{(cw+d)^2} < 1$) and the other, denoted by u, is repelling ($A'(u) > 1$). An oriented geodesic in

---

1991 *Mathematics Subject Classification.* Primary 20H05, 58F17; Secondary 20H10, 11H55.

Key words and phrases. modular group, modular surface, closed geodesics, "–" continued fractions, reduction theory.

The work of the author was partially supported by NSF grant DMS-9404136.

Published in *Geometriae Dedicata*, **63:** 123–145, 1996.

$\mathcal{H}$ from u to w, called the *axis* of γ and denoted by $\mathcal{C}(\gamma)$, is clearly γ–invariant, and does not change if we choose the matrix $-A$ instead of A. If γ belongs to a Fuchsian group Γ, i.e. a discrete subgroup of $PSL(2,\mathbb{R})$ (for necessary information see e.g. [6]), its axis becomes an oriented *closed geodesic* in the quotient–space $\Gamma\backslash\mathcal{H}$. If two elements γ_1 and γ_2 are conjugate in Γ, i.e. $\gamma_1 = \gamma\gamma_2\gamma^{-1}$ for some $\gamma \in \Gamma$, then γ maps the axis of γ_2 to the axis of γ_1, hence they represent the same oriented closed geodesic in $\Gamma\backslash\mathcal{H}$. Conversely, every oriented closed geodesic in $\Gamma\backslash\mathcal{H}$ represents the conjugacy class of a primitive (i.e. not a power of another) hyperbolic transformation in Γ. In what follows closed geodesics will be always considered oriented.

Let Γ be a finitely generated Fuchsian group of the first kind. The idea of coding geodesics with respect to a given Dirichlet fundamental region $\mathcal{D}$ for Γ, goes back to Morse [8] (see also [9, p. 104]). A Dirichlet region for such a Γ always has an even number of sides which are identified by generators of Γ which we denote by $\{\gamma_i\}$. We label the sides of $\mathcal{D}$ by elements of the set $\{\gamma_i\}$ as follows: if a side s is identified in $\mathcal{D}$ with the side $\gamma_j(s)$, we label the side s by γ_j. By labeling all the images of s under Γ by the same generator γ_j we obtain the labeling of the whole net $\mathcal{N}$ of images of sides of $\mathcal{D}$, such that each side in $\mathcal{N}$ has two labels corresponding to the two images of $\mathcal{D}$ shared by this side. Any oriented geodesic in $\mathcal{H}$ may be coded by a (two-sided) sequence of generators of Γ which label the successive sides of $\mathcal{N}$ it crosses; at each crossing we choose the label corresponding to the image the geodesic *enters*. We describe the coding sequence of a geodesic in the assumption that it does not pass through the vertices of the net $\mathcal{N}$. We may assume that the geodesic intersects $\mathcal{D}$ and choose an initial point on it inside $\mathcal{D}$. After exiting $\mathcal{D}$, the geodesic enters a neighboring image of $\mathcal{D}$ through the side labeled, say, by γ_1 (see Fig. 1). Therefore this image is $\gamma_1(\mathcal{D})$, and first symbol in the code is γ_1. If it enters the second image of $\mathcal{D}$ through the side labeled by γ_2, the second image is $(\gamma_1\gamma_2\gamma_1^{-1})(\gamma_1(\mathcal{D})) = \gamma_1\gamma_2(\mathcal{D})$, and the second symbol in the code is γ_2, and so on. Thus we obtain a sequence of all images of $\mathcal{D}$ crossed by our geodesic in the direction of its orientation: $\mathcal{D}$, $\gamma_1(\mathcal{D})$, $\gamma_1\gamma_2(\mathcal{D}), \dots$. If a geodesic is the axis of a primitive hyperbolic element $\gamma \in \Gamma$, we have

$$\gamma = \gamma_1\gamma_2\dots\gamma_n \tag{0.2}$$

for some n. In this case the sequence is periodic with the least period $[\gamma_1, \gamma_2, \dots, \gamma_n]$.

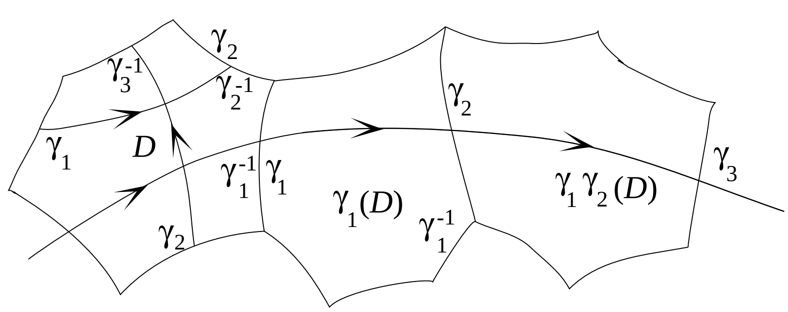

Figure 1

By mapping the oriented geodesic segments between every two consecutive crossings of the net $\mathcal{N}$ back to $\mathcal{D}$ (as shown in Fig. 1), we obtain a *geodesic in* $\mathcal{D}$. The coding sequence described above may be obtained by taking inverses of the generators labeling the sides of $\mathcal{D}$ the geodesic hits consequently. The axis of a primitive hyperbolic element, $\mathcal{C}(\gamma)$ becomes a *closed geodesic in* $\mathcal{D}$. If the geodesic passes through a vertex of $\mathcal{D}$, an ambiguity appears in assigning a code to it. We do not elaborate on this point since in the case of $SL(2,\mathbb{Z})$ the ambiguity can be removed, as is explained below.

If two hyperbolic elements are conjugate in Γ, their closed geodesics in $\mathcal{D}$ coincide, and hence the periods in their coding sequences differ by a cyclic permutation. Conversely, if two primitive hyperbolic elements have the periods in their coding sequences that differ by a cyclic permutation, by (0.2) they are conjugate in Γ, and hence their closed geodesics coincide. We call the period of the coding sequence of $\mathcal{C}(\gamma)$ with respect to a given Dirichlet region, up to a cyclic permutation, the *Morse code* of a closed geodesic associated to the conjugacy class of γ, and denote it by $[\gamma] = [\gamma_1, \gamma_2, \dots, \gamma_n]$. The axis of the inverse transformation, $\mathcal{C}(\gamma^{-1})$, is the same geodesic as $\mathcal{C}(\gamma)$, but oriented in the opposite direction. It is easy to see that its Morse code is given by $[\gamma^{-1}] = [\gamma_n^{-1}, \gamma_{n-1}^{-1}, \dots, \gamma_1^{-1}]$. The Morse code of a matrix A, denoted also by $[A]$, is the Morse code of the corresponding Möbius transformation.

The Morse code of a matrix $A \in SL(2,\mathbb{Z})$ with respect to the standard fundamental region $F = \{z \in \mathcal{H} \mid |z| \geq 1, \ |\operatorname{Re} z| \leq \frac{1}{2}\}$ can be described as a finite sequence of integers.

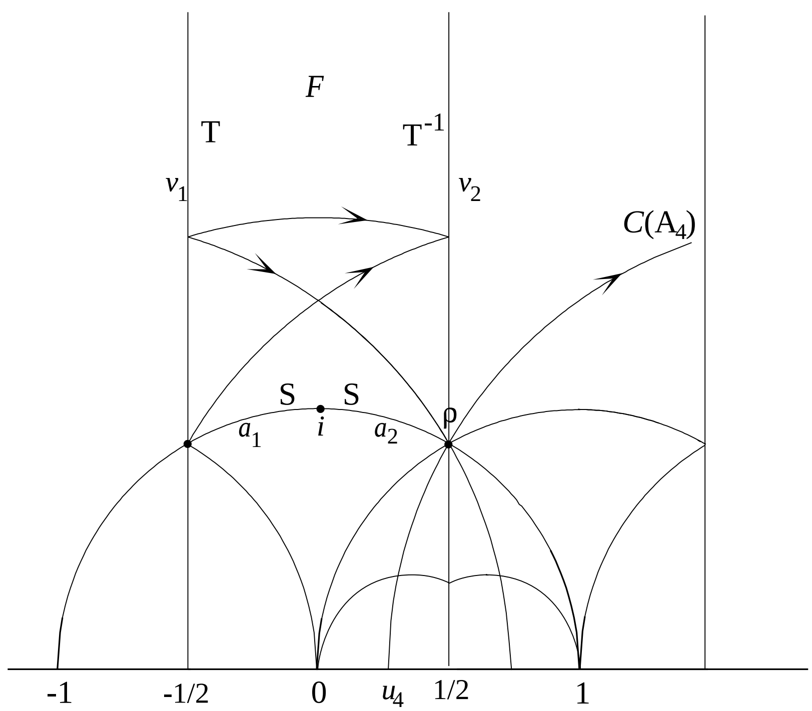

Figure 2

It is convenient to regard F (see Fig. 2) as a quadrilateral rather than a triangle, with the point i separating two sides, the arcs of the unit circle. The left vertical side v_1 is identified with the right one v_2 via the transformation $T(z) = z + 1$, and the two arcs of the unit circle, denoted respectively by a_1 and a_2, are interchanged by the transformation of order two $S(z) = -1/z$ which fixes i. According to our convention the side v_1 is labeled by T, v_2 is labeled by T^{-1}, and both a_1 and a_2 are labeled by S (see Fig. 2). In this particular situation the Morse code $[A]$ is a periodic sequence of symbols T, T^{-1}, and S. It is easy to see that it must contain at least one S, an S cannot be followed by another S, and a T cannot be followed by a T^{-1} and vice versa. Therefore $[A]$ is a finite sequence of blocks, defined up to a cyclic permutation, consisting of T's and T^{-1}'s that are separated by S's. Since it cannot both start and end by an S, we will always assume that it ends by an S. To each block of T's we assign a positive integer equal to its length, and to each block of T^{-1}'s we assign a negative integer whose absolute value is equal to its length. Thus we obtain a finite sequence of integers $[n_1, n_2, \dots, n_m]$, called the *geometric code* of A and also denoted by $[A]$, which classifies closed geodesic on the modular surface $PSL(2,\mathbb{Z})\backslash\mathcal{H}$. The coding sequence of a geodesic passing through the vertex ρ of F in the clockwise direction is obtained by the convention that it exits F through the side v_2. The axis of $A_4 = \begin{pmatrix} 4 & -1 \\ 1 & 0 \end{pmatrix}$, passing through the vertex ρ, and the corresponding closed geodesic in F are shown in Fig. 2. According to our convention, its Morse code is $[T, T, T, T, S]$, and its geometric code is $[4]$.

The idea of using continued fractions to study the modular group $PSL(2,\mathbb{Z})$ and geodesics on the modular surface occurred in [2] and [4], and was developed in [10]. Related ideas appear in [11] and [1]. In this paper we use "–" continued fractions to produce another code classifying closed geodesics on the modular surface (Proposition 2.2) which comes from the Gauss reduction theory. It is a finite sequence of integers $(n_1, \dots, n_m)$ with $n_i \geq 2$ also defined up to a cyclic permutation. We call it the *arithmetic code* of the conjugacy class of A and denote it by (A). The exact definitions are given in §2. Theorem 2 give a complete description of all closed geodesic for which the geometric and arithmetic codes coincide.

Suppose we have a set of elements with an equivalence relation. In the most general terms, a *reduction theory* is an algorithm for finding canonical representatives in each equivalence class. Such representatives are called "reduced" elements. Each equivalence class contains a canonical (finite and non–empty) set of reduced elements which form a cycle in a natural way, and following the reduction algorithm one can pass from a given element within its equivalence class to a reduced one in a finite number of steps. Applying the same algorithm to a reduced element, one obtains consequently all reduced elements in its cycle.

In the reduction algorithm for co-compact Fuchsian groups described in [5] all elements whose axes intersect a given fundamental region $\mathcal{D}$ are called "reduced". The cycle of Γ–conjugate reduced elements consists of all reduced elements with the same Morse code, and the intersections of their axes with $\mathcal{D}$ comprise the closed geodesic associated to this particular conjugacy class.

A complete account of the Gauss reduction theory for indefinite integral binary quadratic forms and its relation to the theory of "–" continued fractions is given in Zagier's book [12, Chapter 13]. It can be translated into the matrix language as follows. To each integral binary quadratic form of discriminant $D > 0$ corresponds a geodesic in $\mathcal{H}$ which connects the roots of the corresponding quadratic equation. Its image in $PSL(2,\mathbb{Z})\backslash\mathcal{H}$ is closed since there exists a hyperbolic matrix $A \in SL(2,\mathbb{Z})$ with the same axis (the set of integral matrices having this axis is a real quadratic field $\mathbb{Q}(\sqrt{D})$, where A corresponds to a non-trivial unit of norm 1). Conversely, we associate to a hyperbolic matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z})$ an integral binary quadratic form $Q_A(x,y) = cx^2 + (d-a)xy - by^2$ of discriminant $D = (a+d)^2 - 4 > 0$ which has already appeared in (0.1). Two matrices with the same trace are conjugate in $SL(2,\mathbb{Z})$ if and only if the corresponding quadratic forms (with the same discriminant) are equivalent in the narrow sense, i.e. via a matrix from $SL(2,\mathbb{Z})$. Thus the two theories are equivalent. The Gauss's notion of a "reduced" binary quadratic form translates into the following notion of a "reduced" matrix which is not connected with any particular fundamental region.

Definition. A hyperbolic matrix in $SL(2,\mathbb{Z})$ is called *reduced* if its attracting and repelling fixed points, denoted by w and u respectively, satisfy

$$w > 1, \quad 0 < u < 1. \tag{0.3}$$

Definition. The set of reduced matrices conjugate to a given matrix A is called the *A–cycle*.

In §2 we show that for any hyperbolic matrix $A \in SL(2,\mathbb{Z})$ the A–cycle consists of all reduced matrices B such that $(B) = (A)$, is finite and non–empty (Proposition 2.5), and that any m–tuple of integers $(n_1, \ldots, n_m)$ with $n_i \geq 2$ is an arithmetic code of some explicit A–cycle (Corollary 2.9)

Definition. Let $\mathcal{D}$ be any fundamental region for $SL(2,\mathbb{Z})$. A hyperbolic matrix in $SL(2,\mathbb{Z})$ is called *$\mathcal{D}$–reduced* if it is reduced and its axis intersects $\mathcal{D}$.

In what follows $\mathcal{D}$ will be either the standard fundamental region F (see Fig. 2) or one of its images under $SL(2,\mathbb{Z})$.

Definition. A hyperbolic matrix A in $SL(2,\mathbb{Z})$ is called *totally F–reduced* if all matrices in the A–cycle are F–reduced.

Since the direction of the axis of a hyperbolic transformation is not conjugacy invariant, some segments of a closed geodesic in F may be oriented clockwise, and the others – counter–clockwise (see an example in Fig. 3). A case when *all* segments are clockwise oriented is special and is identified in Theorem 1.

Theorem 1. *Let $A \in SL(2,\mathbb{Z})$ be a hyperbolic matrix. The following statements are equivalent:*

(1) *A is totally F–reduced.*

(2) *The arithmetic and geometric codes of A coincide.*
(3) *All segments comprising the closed geodesic in F corresponding to the conjugacy class of A are clockwise oriented.*

Fig. 3 gives an illustration to Theorem 1. It shows the closed geodesic in F for the matrix $A = \begin{pmatrix} 15 & -8 \\ 2 & -1 \end{pmatrix}$. Its attracting fixed point $w = 4 + 2\sqrt{3}$ has a pure periodic "–" continuous fraction expansion $w = (\overline{8, 2})$, hence $(A) = (8, 2)$. The matrix A itself is not F–reduced since $u > \frac{1}{2}$, and hence its axis does not intersects F, therefore A is not totally F–reduced. The closed geodesic in F corresponding to the conjugacy class of A, shown in Fig. 3, consists of 10 oriented geodesic segments: segments numbered 1–7 are clockwise oriented while segments numbered 8–10 are oriented counter–clockwise. Following the closed geodesic in F we obtain its geometric code $[A] = [6, -2]$, which is different from its arithmetic code $(A) = (8, 2)$.

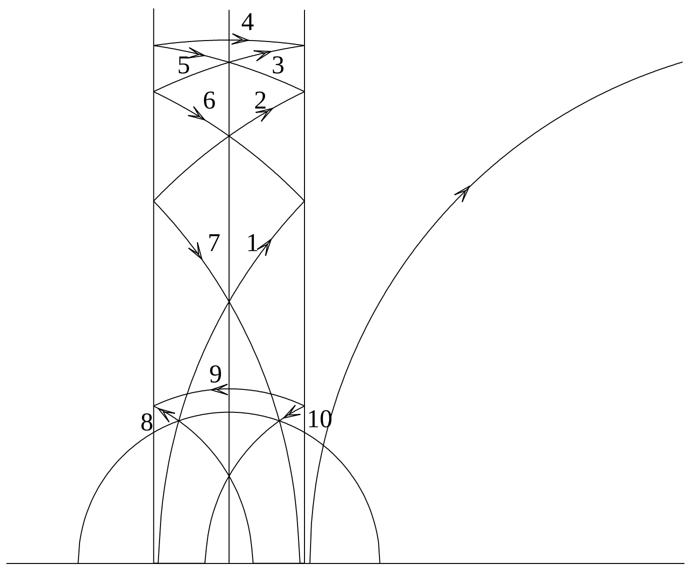

Figure 3

Here are two more examples of reduced matrices which are not totally F–reduced:

for $A = \begin{pmatrix} 14 & -3 \\ 5 & -1 \end{pmatrix}$, $(A) = (3, 5)$, but $[A] = [-1, 1, -1, 3]$;

for $A = \begin{pmatrix} 65 & -17 \\ 23 & -6 \end{pmatrix}$, $(A) = (3, 6, 4)$, but $[A] = [-1, 1, -1, 5, 3]$.

The following theorem gives a necessary and sufficient condition for a matrix to be totally F–reduced in terms of its arithmetic code.

Theorem 2. *Let $(A) = (n_1, \ldots, n_m)$. A is totally F–reduced if and only if $\frac{1}{n_i} + \frac{1}{n_{i+1}} \le \frac{1}{2}$ for all $i \pmod m$, i.e. the arithmetic code (A) does not contain 2 and the following pairs: $\{3, 3\}, \{3, 4\}, \{4, 3\}, \{3, 5\}$, and $\{5, 3\}$.*

An outline of the paper is as follows. In §1 we review the theory of "–" continued fractions [12]. In §2 we give a geometric interpretation of the Gauss reduction theory applied directly to matrices and prove that any arithmetic code is realized on a reduced matrix. Finally, in §3 we prove Theorems 1 and 2.

This paper was written during my visit to the Mathematics Research Centre at the University of Warwick in July 1994. I would like to thank the institute for their hospitality, financial support and excellent working conditions. The codes and graphics were produced using *Matlab* package.

1. The theory of "–" continued fractions. Let $n_0, n_1, n_2, \dots$ be a sequence of integers satisfying $n_1, n_2, \dots \geq 2$. Following Zagier [12] we denote by $(n_0, n_1, \dots, n_s)$ the finite "–" continued fraction

$$(n_0, n_1, \dots, n_s) = n_0 - \cfrac{1}{n_1 - \cfrac{1}{n_2 - \cfrac{1}{\ddots - \cfrac{1}{n_s}}}}$$

and by $(n_0, n_1, n_2 \dots)$ the limit $\lim_{s\to\infty}(n_0, n_1, \dots, n_s)$ whose existence is easily established. Conversely, every real number α has a unique "–" continued fractions expansion $\alpha = (n_0, n_1, n_2, \dots)$ with $n_i \in \mathbb{Z}$ and $n_1, n_2, \dots \geq 2$, by setting $n_0 = [\alpha] + 1$, and, inductively, $n_i = [\alpha_i] + 1$, $\alpha_{i+1} = \frac{1}{n_i - \alpha_i}$. This gives a one–to–one correspondence between the set of real numbers α and the set of infinite sequences $(n_0, n_1, n_2 \dots)$ with $n_i \in \mathbb{Z}$ and $n_1, n_2, \dots \geq 2$. Under this correspondence the following statements are true:

(1.1) α is rational if and only if from some point on all the n_i's are equal to 2;

(1.2) α is a quadratic irrationality, i.e. a root of a quadratic polynomial with coefficients in $\mathbb{Z}$, if and only if its "–" continued fraction expansion is eventually periodic, i.e. from some point on the n_i's repeat periodically: $\alpha = (n_0, n_1, \dots, n_k, \overline{n_{k+1}, \dots, n_{k+m}})$ (the line over $n_{k+1}, \dots, n_{k+m}$ signifies that those numbers repeat periodically and that m is the least period);

(1.3) α has a purely periodic "–" continued fraction expansion if and only if $\alpha > 1$, and $0 < \alpha' < 1$, where α' is conjugate to α, i.e. it is the other root of the same quadratic polynomial as α is;

(1.4) if $\alpha = (\overline{n_1, \dots, n_m})$, then $\frac{1}{\alpha'} = (\overline{n_m, \dots, n_1})$.

Property (1.3) is extremely important. It gives an equivalent definition of the reduced matrix: a matrix is reduced if and only if its attracting fixed point has a purely periodic "–" continued fraction expansion.

The following property is crucial for our purposes; it shows that the period of a "–" continued fraction expansion is a complete system of $SL(2,\mathbb{Z})$–invariants (just as the period of an ordinary continued fraction expansion is a complete system of $GL(2,\mathbb{Z})$–invariants, a standard fact, see e.g. [3, p. 142]).

Proposition 1.1. *Two quadratic irrationalities are obtained from one another by an application of a transformation from $SL(2,\mathbb{Z})$ if and only if the periods in their "–" continued fraction expansions are cyclic permutations of one another.*

Proof. If two quadratic irrationalities have periods in their "–" continued fraction expansions, which are cyclic permutations of one another, one can be obtained from another by consequent applications of transformations $T(z)=z+1$, $T^{-1}(z)=z-1$, and $S(z)=-1/z$. Since those transformations are in $SL(2,\mathbb{Z})$, the claim in this direction follows. Since the transformations S and T generate $SL(2,\mathbb{Z})$ (see e.g. [6, p. 74]), it is sufficient to prove the converse only for these particular transformations. Let w be a quadratic irrationality:

$$w=(n_0,n_1,\dots,n_k,\overline{n_{k+1},\dots,n_{k+m}}).$$

This representation is not unique: we can extend the part before the period by adding a period to it if we need to. Then obviously

$$T^{\pm 1}(w)=(n_0\pm 1,n_1,\dots,n_k,\overline{n_{k+1},\dots,n_{k+m}}).$$

In order to deal with S we first notice that if $n_0\geq 2$, then

$$S(w)=(0,n_0,n_1,\dots,n_k,\overline{n_{k+1},\dots,n_{k+m}})$$

which is a legitimate "–" continued fraction expansion. We remind that the following relations between S and T hold (in fact, these relations define $SL(2,\mathbb{Z})$, but we do not use it here):

$$S^2=\mathrm{Id},\quad STSTST=\mathrm{Id},$$

where Id denotes the identity transformation. In the next argument we use the following consequences of the second relation:

$$STS=T^{-1}ST^{-1},\quad ST^{-1}S=TST,$$

and for $p\geq 2$

$$ST^{-p}S=TS\underbrace{T^2S\dots T^2S}_{p-1\text{ times}}T.$$

If $n_0\leq -1$ we obtain

$$S(w)=(1,\underbrace{2,\dots,2}_{-n_0-1\text{ times}},n_1+1,n_2,\dots,n_k,\overline{n_{k+1},\dots,n_{k+m}}).$$

If $n_0=0$, we have

$$S(w)=(n_1,\dots,n_k,\overline{n_{k+1},\dots,n_{k+m}}).$$

Now let $n_0=1$. If $n_1\geq 3$, we have

$$S(w)=(-1,n_1-1,n_2,\dots,n_k,\overline{n_{k+1},\dots,n_{k+m}}).$$

Since w is irrational, there is a n_i in the period that is greater than 2, so we suppose that $n_s\geq 3$, and $n_i=2$ for all $1\leq i\leq s-1$. Then

$$S(w)=(-s,n_s-1,\dots,n_k,\overline{n_{k+1},\dots,n_{k+m}}).\quad\square$$

2. Reduction Theory for $\mathbf{SL}(2,\mathbb{Z})$ and factorization of reduced matrices. In order to understand the connection between the geometric and arithmetic codes we present here an equivalent of Gauss reduction theory for matrices in $SL(2,\mathbb{Z})$.

We will need the following standard fact, the proof of which we include for the sake of completeness.

Lemma 2.1. *Let Γ be a Fuchsian group, $\gamma_1, \gamma_2 \in \Gamma$ hyperbolic elements having a common fixed point. Then their second fixed points also coincide, hence they have the same axis, and both are powers of a primitive matrix with the same axis.*

Proof. By a standard conjugation we may assume that both γ_1 and γ_2 fix the ∞, so

$$\gamma_1(z) = \lambda z \ (\lambda > 1), \quad \text{and} \quad \gamma_2(z) = \mu z + k \ (\mu \neq 1,\ k \neq 0).$$

Then

$$\gamma_1^{-n}\gamma_2\gamma_1^n(z) = \lambda^{-n}(\mu(\lambda^n z) + k) = \mu z + \lambda^{-n}k.$$

Hence $\|\gamma_1^{-n}\gamma_2\gamma_1^n\| = \sqrt{\mu^2 + \lambda^{-2n}k^2 + 1}$ is bounded as $n \to \infty$, and hence the sequence $\{\gamma_1^{-n}\gamma_2\gamma_1^n\}$ contains a converging subsequence of distinct terms, a contradiction with discreteness of Γ. Therefore $k = 0$, and hence both γ_1 and γ_2 fix 0. $\square$

Now we can characterize the conjugate matrices in $SL(2,\mathbb{Z})$.

Proposition 2.2. *Two hyperbolic matrices A and B in $SL(2,\mathbb{Z})$ with the same trace are conjugate in $SL(2,\mathbb{Z})$ if and only if their attracting (repelling) fixed points have periods in their "–" continued fraction expansions that are cyclic permutations of one another.*

Proof. Let w_A and w_B be attracting fixed points of A and B respectively, such that their periods in "–" continued fraction expansion differ by a cyclic permutation. Then by Proposition 1.1 there exists $C \in SL(2,\mathbb{Z})$ such that $w_A = Cw_B$. Then matrices CBC^{-1} and A have the same fixed point w_A, and by Lemma 2.1, since they have the same trace, either $CBC^{-1} = A$ or $CBC^{-1} = A^{-1}$. Since both w_A and w_B are attracting, w_A is attracting for both, A and CBC^{-1}, and therefore $CBC^{-1} = A$. Conversely, suppose two matrices in $SL(2,\mathbb{Z})$ are conjugate. Then their attracting fixed points w_A and w_B are obtained from each other by an application of a matrix $C \in SL(2,\mathbb{Z})$. Then by Proposition 1.1 the periods in "–" continued fraction expansions of w_A and w_B differ by a cyclic permutation. $\square$

Thus, in addition to the geometric code described in the Introduction, we obtain two other invariants of a closed geodesic, also defined up to a cyclic permutation: the periods of the "–" continued fraction expansions of its attracting and repelling fixed points. The first invariant, which we call the *arithmetic code* of A and denote by (A), coincides with the geometric code for a large class of closed geodesics identified in Theorem 2. The second is the arithmetic code of the inverse matrix (A^{-1}) which corresponds to the same closed geodesic with the opposite orientation. Notice that if A is reduced, the relation (1.4) holds.

The following lemma describes reduced matrices in terms of their entries.

Lemma 2.3. *Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a matrix in $SL(2,\mathbb{Z})$ with $a+d>2$. A is reduced if and only if $c>0$, $a+b-c-d>0$, and $b<0$. For a reduced A with the attracting fixed point w we also have*

(i) $a>0$, $c+d>0$, $d\le 0$;
(ii) $\frac{a+b}{c+d} < w < \frac{a}{c}$.

Proof. Suppose A is reduced. Since $a+d>0$, the attractive and repelling fixed points w and u are given by

$$w = \frac{a-d+\sqrt{D}}{2c} \text{ and } u = \frac{a-d-\sqrt{D}}{2c}.$$

Since for a hyperbolic matrix in $SL(2,\mathbb{Z})$ $c \neq 0$, (0.3) implies $c>0$, $|a-d-2c| < \sqrt{D}$, which in its turn implies $a+b-c-d>0$, and $b<0$. Conversely, the first two conditions imply $|a-d-2c| < \sqrt{D}$, hence $w>1$, and $u<1$. The first and the third imply $|a-d| > \sqrt{D}$, and since $a-d > c-b > 0$, we obtain $u>0$.

Since $a+d > \sqrt{D}$, $\frac{a}{c} > w > 1$, hence $a>0$. We have

$$Q_A(\frac{a+b}{c+d},1) = \frac{c+d-a-b}{(c+d)^2} < 0$$

since the expression in the numerator is negative, hence by (0.3) $\frac{a+b}{c+d} < w$, which proves (ii). Moreover, $\frac{1}{c+d} = \frac{a}{c} - \frac{a+b}{c+d} > 0$, hence $c+d>0$. Since $bc \le -1$, $ad = 1+bc \le 0$, hence $d \le 0$, and (i) follows. □

Corollary 2.4. *Let A be a reduced matrix with the attracting fixed point w. Then, considered as a function on $\mathbb{R}\cup\infty$, $A(x)$ is increasing for $x>1$. For any fixed number $x>1$ the sequence $\{A^n(x)\}$ (converging to w) is decreasing if $x>w$ and is increasing if $1<x<w$.*

Proof. Let $A(x) = \frac{ax+b}{cx+d}$. For $x > -\frac{d}{c}$, $A(x)$ is an increasing, concave–up function with a horizontal asymptote $y = \frac{a}{c}$. Notice that $A(x) < x$ for $x>w$ and $A(x) > x$ for $1<x<w$. The assertion follows since, by Lemma 2.3 (i), $-\frac{d}{c} < 1$. □

Proposition 2.5 (Reduction Algorithm). *There is a finite number of reduced matrices in $SL(2,\mathbb{Z})$ with a given trace t, $|t|>2$. Any hyperbolic matrix in $SL(2,\mathbb{Z})$ with a trace t can be reduced by a finite number of standard conjugations. Applied to a reduced matrix A, this conjugation gives another reduced one. Any reduced matrix conjugate to A is obtained from A by a number of standard conjugations. Thereby the set of reduced matrices is decomposed into disjoint cycles of conjugate matrices.*

Proof. The proof of the first assertion is an adaptation of the proof in [12] for matrices. We give it here for the sake of completeness. Suppose $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is reduced. Let

$k = a - d - 2c$. It was shown in the proof of Lemma 2.3 that $|k| < \sqrt{D}$, and hence k can take only finitely many values for a given $D = t^2 - 4$. We have $D - k^2 = 4c(a+b-c-d) > 0$, hence $c \,|\, \frac{D-k^2}{4}$ and also can take only finitely many values. We express a, b, and d in terms of c and k as follows:

$$\begin{aligned} a &= \frac{t+k+2c}{2}, \\ b &= \frac{D-k^2}{4c} - (k+c), \\ d &= \frac{t-k-2c}{2}. \end{aligned}$$

and thus obtain the finiteness of the number of reduced matrices with a given trace t.

The attracting fixed point of A has a "–" continued fraction expansion

$$(n_0, n_1, \dots, n_k, \overline{n_{k+1}, \dots, n_{k+m}})$$

where the notations are as in (1.2). Conjugating A by $S^{-1}T^{-n_0}$ we obtain a matrix $A_0 = S^{-1}T^{-n_0}AT^{n_0}S$, and inductively, $A_i = S^{-1}T^{-n_i}A_{i-1}T^{n_i}S$ for $i = 1, 2, \dots$. The attracting fixed point of the matrix

$$A_k = (S^{-1}T^{-n_k}S^{-1} \dots T^{-n_1}S^{-1}T^{-n_0})A(ST^{-n_k}S \dots T^{-n_1}ST^{-n_0})^{-1},$$

w, has a purely periodic "–" continued fraction expansion $w = (\overline{n_{k+1}, \dots, n_{k+m}})$, and according to (1.3), $w > 1$, $0 < u < 1$, i.e. A_k is reduced. Applying the same procedure to A_k we obtain m reduced matrices in a sequence corresponding to the period of w. Conversely, if two reduced matrices are conjugate, their attracting fixed points have pure periodic "–" continued fraction expansions whose periods, by Proposition 2.2, differ by a cyclic permutation. Hence they belong to the same cycle and are obtained from one another by a number of standard conjugations. □

Remark. Proposition 2.5 asserts the finiteness of the number of conjugacy classes of matrices in $SL(2,\mathbb{Z})$ with a given trace t, which corresponds to the class number of the real quadratic field $\mathbb{Q}(\sqrt{t^2-4})$ (in the narrow sense), a standard and very important fact in number theory. This fact, however, is much more general, it is valid for all Fuchsian groups of the first kind. For the co–compact Fuchsian groups it follows from the expansiveness of the geodesic flow and can be found e.g. in [7, p.p. 212, 549, 569–70].

Our next goal is to show that any arithmetic code is realized on some reduced matrix in $SL(2,\mathbb{Z})$ giving thus a variety of examples. The main step is Proposition 2.6, whose proof is an adaptation of the proof in [12] for quadratic forms.

Proposition 2.6. *Let $n_1, \dots, n_m \geq 2$ be integers, and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a primitive hyperbolic matrix with a positive trace and the attractive fixed point*

$$w = (\overline{n_1, \dots, n_m}). \tag{2.1}$$

Then it can be represented in the form

$$A = T^{n_1} S T^{n_2} S \dots T^{n_m} S,$$

where $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *and* $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Proof. First we need a technical lemma.

Lemma 2.7. *Let $n \geq 2$ be an integer, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$, and w a real number such the following conditions are satisfied:*
(2.2) $c, c + d > 0$,
(2.3) $n - 1 < w < n$,
(2.4) $\frac{a+b}{c+d} < w < \frac{a}{c}$.
Then $n - 1 < \frac{a}{c} \leq n$.

Proof. The left inequality follows immediately. Suppose $n < \frac{a}{c}$. Then

$$\frac{a+b}{c+d} < n < \frac{a}{c}.$$

But then $ac + bc < nc(c + d) < ac + ad$ in contradiction with $ad - bc = 1$. $\square$

Let $A_0 = A$, $w_0 = w$, $A_s = S^{-1}T^{-n_s} \dots S^{-1}T^{-n_1}A = \begin{pmatrix} a_s & b_s \\ c_s & d_s \end{pmatrix}$, and

$$w_s = S^{-1}T^{-n_s} \dots S^{-1}T^{-n_1} w_0 = (\overline{n_{s+1}, \dots, n_m, n_1, \dots, n_s}) \tag{2.5}$$

We are going to prove by induction that A_s, w_s, n_{s+1} satisfy (2.2)–(2.4) for $0 \leq s < m$, and $A_m = 1_2$, the identity matrix. Since

$$A_s A (A_s)^{-1} = (S^{-1}T^{-n_s}S^{-1} \dots S^{-1}T^{-n_1}) A (S^{-1}T^{-n_s}S^{-1} \dots S^{-1}T^{-n_1})^{-1},$$

we see that w_s is the attractive fixed point of $A_s A (A_s)^{-1}$. It has a pure periodic "–" continued fraction expansion. For $s < m$ it is different from w_0 since m is the least period of w_0, which implies that $A_s \neq 1_2$ for $s < m$. It follows that for all $0 \leq s < m$, $c_s \neq 0$. Otherwise we would have $a_s d_s = 1$, which implies $A_s(z) = z + b$, $b \in \mathbb{Z}$, and the second fixed point, $w_s' = w_0' + b$ stays between 0 and 1 only if $b = 0$.

The induction hypothesis holds for $A_0 = A$, $w_0 = w$, and n_1: $n_1 - 1 < w < n_1$ gives (2.3), and (2.2) and (2.4) follow from in Lemma 2.3. Suppose the induction hypothesis holds for $0 < s - 1 < m$. Then by Lemma 2.7 for A_{s-1} we have $c_s = n_s c_{s-1} - a_{s-1} > 0$ and

$$c_s < c_{s-1}. \tag{2.6}$$

Also $c_s + d_s = n_s(c_{s-1} + d_{s-1}) - (a_{s-1} + b_{s-1}) > 0$ by (2.3) and (2.4) for A_{s-1}, thus condition (2.2) for A_s follows. Condition (2.3) for A_s follows from (2.5). We have $\frac{a_s}{c_s} = S^{-1}T^{-n_s}(\frac{a_{s-1}}{c_{s-1}})$, and $\frac{a_s+b_s}{c_s+d_s} = S^{-1}T^{-n_s}(\frac{a_{s-1}+b_{s-1}}{c_{s-1}+d_{s-1}})$, and condition (2.4) for A_s follows from (2.4) for A_{s-1} and the fact that the function $S^{-1}T^{-n_s}(x) = \frac{1}{-x+n_s}$ is increasing for $x < n_s$.

By (2.6) we see that the coefficient c_s (an integer!) is decreasing monotonically with s, until it attains 0. It remains to show that $c_m = 0$. Suppose $c_m > 0$; then by (2.6) we still have $c_m < c_{m-1} < c$. On the other hand, $A_m A$ fixes w_0, hence by Lemma 2.1 it is equal to A^n for some $n \in \mathbb{Z}$. The following lemma delivers a contradiction, hence $A_m A = 1_2$.

Lemma 2.8. *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *be a hyperbolic matrix with* $c > 0$*, and for* $n \in \mathbb{Z}$*,* $A^n = A = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$*. Then for* $n \neq 0$ *either* $c_n > c$*, or* $c_n < 0$*.*

Proof. Since all transformations in $\{A^n \mid n \in \mathbb{Z}\}$ except for the identity have the same axes, the corresponding quadratic polynomials Q_{A^n} must be multiples of Q_A, and therefore for some λ_n, $c_n = c\lambda_n$, $b_n = b\lambda_n$, and $a_n - d_n = (a-d)\lambda_n$. Let us denote $\operatorname{tr} A = t$, and $\operatorname{tr} A^n = t_n$. Comparing discriminants of Q_{A^n} and Q_A we obtain

$$\lambda_n^2 = \frac{t_n^2 - 4}{t^2 - 4}.$$

Let $\mu > 1$ and $\frac{1}{\mu} < 1$ be the eigenvalues of A. Then $t = \mu + \frac{1}{\mu}$ and $t_n = \mu^n + \frac{1}{\mu^n}$, and hence

$$t_n - t = \mu^{-n}(\mu^{n-1} - 1)(\mu^{n+1} - 1) > 0,$$

which implies $|\lambda_n| > 1$, and $c_n > c$ or $c_n < 0$. □

This completes the proof of Proposition 2.6. □

Corollary 2.9. *Let* $n_1, \ldots, n_m \geq 2$ *be integers,* $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *and* $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$*. Then the matrix*

$$A = T^{n_1} S T^{n_2} S \ldots T^{n_m} S$$

is hyperbolic with a positive trace, reduced, and $(A) = (n_1, \ldots, n_m)$*.*

Proof. It can be easily proved by induction on m using Lemma 2.3 that $\operatorname{tr} A > 2$. A simple calculation shows that the matrix $A = T^{n_1} S T^{n_2} S \ldots T^{n_m} S$ fixes the point $w = (\overline{n_1, \ldots, n_m})$. We need to show that w is attracting. Suppose it is not. Then w is the attracting fixed point of A^{-1}, which is also hyperbolic with a positive trace. Applying Proposition 2.6 to A^{-1} we obtain that $A = A^{-1}$, which contradicts the hyperbolicity of A. □

3. Geometric and arithmetic codes of closed geodesics on the modular surface. Let us recall that a matrix A is called totally F–reduced if all matrices in the A–cycle are F–reduced.

Theorem 1. *The following statements are equivalent:*

1. *A matrix A is totally F–reduced.*
2. *The arithmetic and geometric codes of A coincide.*
3. *All segments comprising the closed geodesic in F corresponding to the conjugacy class of A are clockwise oriented.*

Proof. (1) $\Rightarrow$ (2) Suppose A is totally F–reduced. Then the axes of all matrices in A–cycle must enter F through the side a_2, and since they are clockwise oriented, they must leave F through the side v_2. Let the "–" continued fraction expansion of the attracting point of A, w, be $(\overline{n_1, \ldots, n_m})$. Then the axes of $T^{-1}AT$ will enter F through v_1 and exit F through v_2. Suppose the axes of $T^{-i}AT^i$ have the same property for all $1 \leq i < k$, and the axis of $T^{-k}AT^k$ enters F through v_1 and leaves F either through a_1 or a_2. We want to show that $k = n_1$. If $k < n_1$ then the axis of $T^{-n_1}AT^{n_1}$ does not intersect F, which, in its turn contradicts the fact that $S^{-1}T^{-n_1}AT^{n_1}S$ is F–reduced. On the other hand, the axis of $T^{-n_1+1}AT^{n_1+1}$ does not intersect F, hence $k = n_1$, the axis of $T^{-n_1}AT^{n_1}$ must exit F through a_1, and the first number in $[A]$ is n_1. (The axis of a reduced matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ cannot pass through the point i. For, if it does, plugging $z = i$ into its equation

$$c|z|^2 + (d-a)x - b = 0$$

we obtain $c = b$, which by Lemma 2.3 contradicts to the fact that A is reduced. The argument is still valid if the axis of A passes through the vertex ρ of F. In this case the axis of $T^{-n_1}AT^{n_1}$ passes through the vertex $\rho - 1$. We shall see later (see Corollary 3.3) that there are only three primitive F–reduced matrices whose axes pass through the vertex ρ.) Continuing this process and counting the number of T's between the S's we obtain the geometric code of A in the form $[n_1, n_2, \ldots, n_m]$ i.e. it is equal to its arithmetic code.

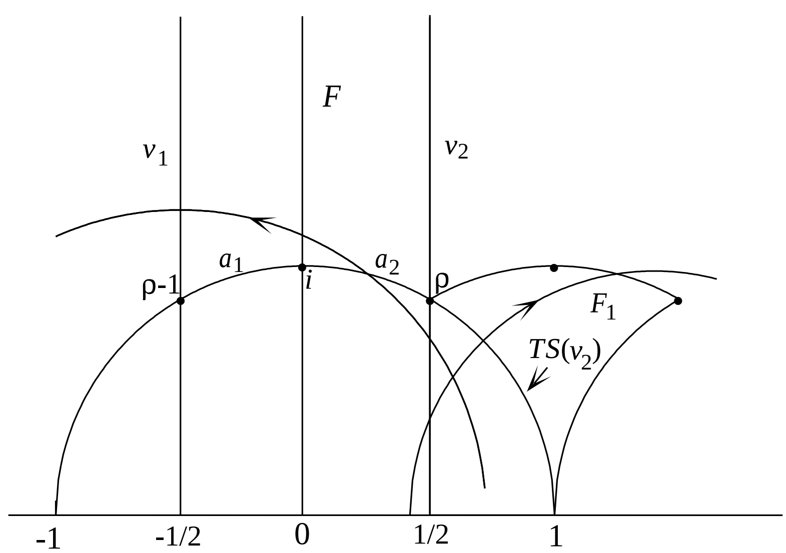

Figure 4

Notice that the intersection of the axes of the matrices in A–cycle with F constitute only a part of the closed geodesic in F representing the conjugacy class of A, its "reduced" part; the remaining pieces represent the axis of the shifts of the reduced geodesics. We have seen that all geodesic segments obtained in the process turned out to be clockwise oriented. Thus (1) $\Rightarrow$ (3).

(2) $\Rightarrow$ (1) Now suppose that A is not totally F–reduced. Then at least one matrix in A–cycle is not F–reduced. We assume that A itself is reduced but not F–reduced. Then it is F_1–reduced, where $F_1 = TS(F)$, and its axis must enter F_1 through the side $TS(v_2)$ (see Fig. 4). Hence the axis of its conjugate $T^{-1}S^{-1}T^{-1}ATST$ intersects F in the counter–clockwise direction and exits the side v_1 of F. This means that the geometric code of A contains at least one T^{-1} or, in other words, at least one number in it is negative, and hence it cannot coincide with any arithmetic code.

(3) $\Rightarrow$ (1) If A is not totally F–reduced, by the previous argument, one of the segments comprising the closed geodesic associated to the conjugacy class of A is counter–clockwisely oriented, a contradiction. $\square$

The following theorem identifies all totally F–reduced matrices in terms of their arithmetic code.

Theorem 2. *Let $(A) = (n_1, \dots, n_m)$. A is totally F–reduced if and only if $\frac{1}{n_i} + \frac{1}{n_{i+1}} \le \frac{1}{2}$ for all $i \pmod m$, i.e the arithmetic code (A) does not contain 2 and the following pairs: $\{3,3\}, \{3,4\}, \{4,3\}, \{3,5\}$, and $\{5,3\}$.*

Proof. Let $(A) = (n_1, \dots, n_m)$. First we prove that if (A) contains a forbidden combination, then A is not totally F–reduced. Suppose (A) contains a pair $\{k, n\}$ such that

$$\frac{1}{k} + \frac{1}{n} > \frac{1}{2}. \tag{3.1}$$

Passing, if necessary, to a matrix in A–cycle, we may assume that $n_1 = n$, $n_m = k$. We shall show that such an A is not F–reduced. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, its axis has the equation

$$c|z|^2 + (d-a)x - b = 0.$$

Since both codes are invariant under $SL(2,\mathbb{Z})$–conjugacy, we may assume that A is reduced. Then its axis intersects the circle $|z| = 1$. It intersects F or not depending whether for the intersection point $z = x + iy$, $x \le \frac{1}{2}$ or $x > \frac{1}{2}$. Hence we obtain the following condition for a reduced matrix A not to be F–reduced: $\frac{a-d}{c-b} - 2 < 0$.

According to Corollary 2.9, $A = A_n A_{n_2} \dots A_k$, where $A_i = T^i S = \begin{pmatrix} i & -1 \\ 1 & 0 \end{pmatrix}$. We write $B = A_{n_2} \dots A_{n_{m-1}} = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$, where B is either reduced (by Corollary 2.9) or is the identity matrix 1_2. Then

$$A = \begin{pmatrix} n & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \begin{pmatrix} k & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a'nk - c'k + b'n - d' & -a'n + c' \\ a'k + b' & -a' \end{pmatrix}.$$

Then

$$\frac{a-d}{c-b}-2=\frac{a'nk-c'k+b'n-d'+a'}{a'k+b'+a'n-c'}-2$$

$$=\frac{(nk-2(n+k)+1)a'-(k-3)c'+(n-2)b'-(d'+c')}{a'k+b'+a'n-c'}. \tag{3.2}$$

By Corollary 2.9 $a+d>0$, by Lemma 2.3 we have $c-b>0$, i.e. the denominator of (3.2) is positive. If B is reduced, by Corollary 2.9 it has positive trace, hence by Lemma 2.3 $c'>0$, $b'<0$, $c'+d'>0$; if $B=1_2$ $c'=b'=0$, $d'=1$. In any case, since by (3.1)

$$nk-2(n+k)+1\le 0, \tag{3.3}$$

the numerator of (3.2) is negative, hence A is not F–reduced. The solutions of (3.1) in integers ≥ 2 give us exactly the forbidden pairs $\{2,q\},\{p,2\},\{3,3\},\{3,4\},\{4,3\},\{3,5\}$, and $\{5,3\}$.

Now we show that any code not containing forbidden pairs gives a totally F–reduced matrix. First we need to prove the "monotonicity" of the property to be totally F–reduced.

Lemma 3.1. *If $A=A_{n_1}\dots A_{n_i}\dots A_{n_m}$ is totally F–reduced and $n>n_i$, then $A(n)=A_{n_1}\dots A_n\dots A_{n_m}$ is also totally F–reduced.*

Proof. We write

$$A(n)=\begin{pmatrix}a_1 & b_1\\ c_1 & d_1\end{pmatrix}\begin{pmatrix}n & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}a_2 & b_2\\ c_2 & d_2\end{pmatrix}=\begin{pmatrix}a_1a_2n+b_1a_2-a_1c_2 & a_1b_2n+b_1b_2-a_1d_2\\ c_1a_2n+d_1a_2-c_1c_2 & c_1b_2n+d_1b_2-c_1d_2\end{pmatrix},$$

where $\begin{pmatrix}a_1 & b_1\\ c_1 & d_1\end{pmatrix}$ or $\begin{pmatrix}a_2 & b_2\\ c_2 & d_2\end{pmatrix}$ can be equal to the identity matrix 1_2. Let

$$f(n)=\frac{a_1a_2n+b_1a_2-a_1c_2-c_1b_2n-d_1b_2+c_1d_2}{c_1a_2n+d_1a_2-c_1c_2-a_1b_2n-b_1b_2+a_1d_2}-2.$$

Since $A=A(n_i)$ is totally F–reduced, $f(n_i)\ge 0$. The lemma will follow if we prove that $f'(n)>0$. For, notice that since $f(n)$ is a fractional linear transformation up to an additive constant, $f'(n)>0$ if and only if the determinant of the corresponding matrix is positive:

$$(a_1a_2-c_1b_2)(d_1a_2-c_1c_2-b_1b_2+a_1d_2)-(b_1a_2-a_1c_2-d_1b_2+c_1d_2)(c_1a_2-a_1b_2)=a_1^2-c_1^2+a_2^2-b_2^2.$$

If $\begin{pmatrix}a_1 & b_1\\ c_1 & d_1\end{pmatrix}\ne 1_2$ it is reduced, and by Lemma 2.7 $\frac{a_1}{c_1}>n_1-1\ge 3-1=2$. Hence $a_1^2>4c_1^2>c_1^2$, i.e $a_1^2-c_1^2>0$. If $\begin{pmatrix}a_1 & b_1\\ c_1 & d_1\end{pmatrix}=1_2$, the above equality also holds. Then

$$a_1^2-c_1^2+a_2^2-b_2^2>a_2^2-b_2^2>0.$$

The last inequality follows from Lemma 2.3: if $\begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$ is reduced: $a_2+b_2 > c_2+d_2 > 0$ and is trivial if $\begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = 1_2$. $\square$

To finish the proof of Theorem 2 we consider the following three matrices: $A_4 = \begin{pmatrix} 4 & -1 \\ 1 & 0 \end{pmatrix}$, $A_3A_6 = \begin{pmatrix} 17 & -3 \\ 6 & -1 \end{pmatrix}$, and $A_6A_3 = \begin{pmatrix} 17 & -6 \\ 3 & -1 \end{pmatrix}$. All these matrices are reduced with arithmetic codes (4), $(3,6)$, and $(6,3)$, respectively. The attracting fixed points of the respective transformations are $w_4 = 2+\sqrt{3} \approx 3.7321$, $w_{36} = \frac{3+\sqrt{7}}{2} \approx 2.8229$, and $w_{63} = 3+\sqrt{7} \approx 5.6458$ Their repelling fixed points are $u_4 = 2-\sqrt{3} \approx 0.2679$, $u_{36} = \frac{3-\sqrt{7}}{2} \approx 0.1771$, and $u_{63} = 3-\sqrt{7} \approx 0.3542$. Their axes pass through the vertex of F, $\rho = \frac{1}{2} + \frac{\sqrt{3}}{2}i$ (see Fig. 5), and hence all three matrices are totally F–reduced. They play a special role in the proof.

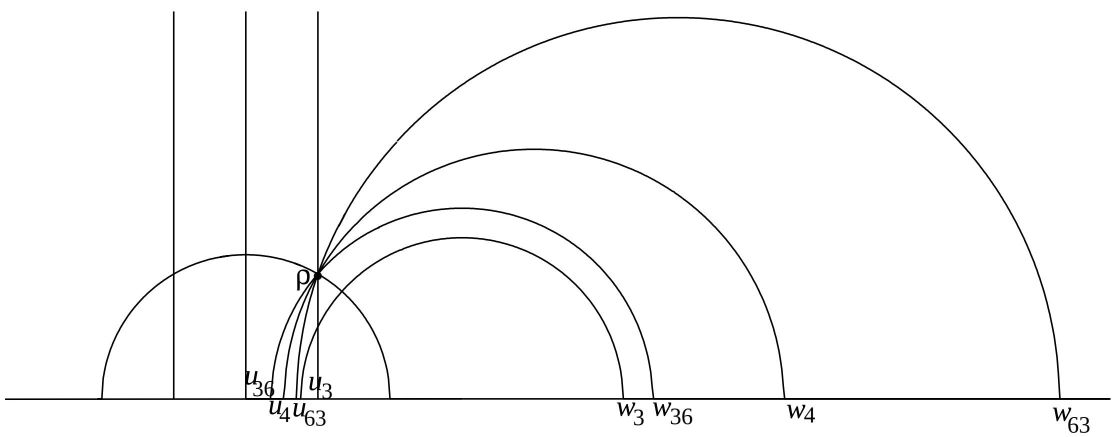

Figure 5

Let us denote the set of all arithmetic codes not containing forbidden pairs by $\mathcal{A}$. Those are exactly the codes satisfying the inequality

$$\frac{1}{n_i} + \frac{1}{n_{i+1}} \leq \frac{1}{2}$$

for all $i \pmod m$ by $\mathcal{A}$. Let $(n_1, \ldots, n_m) \in \mathcal{A}$. If $n_i > 6$ is adjacent to a 3, we can decrease it by 1 and still obtain a code in $\mathcal{A}$. Similarly, by decreasing an $n_i \geq 5$ which is not adjacent to a 3 by 1 we also obtain a code in $\mathcal{A}$. Clearly, starting with a particular code, we can successfully perform this procedure at most a finite number of times until we obtain a code for which it cannot be performed anymore. We call such codes *corner codes* of $\mathcal{A}$. Notice that the above procedure does not change the number and positions of 3's in the code. Conversely, each code in $\mathcal{A}$ can be obtain form a corner code by a finite number of reverse procedures. For example, every code in $\mathcal{A}$ not containing a 3 can be thus obtained from the code (4). If we show that all corner codes in $\mathcal{A}$ are totally reduced, the Theorem will follow from Lemma 3.1.

Let (A) be a corner code different from (4), $(3,6)$ and $(6,3)$. Then (A) contains at least one 3 and, up to a cyclic permutation, can be represented as a series of several blocks of the form $\{6,3,6,\dots,3,6,4,\dots,4\}$. Notice that some of the blocks may not contain 4's at all, but they still begin and end with 6, and if they are next to each other, then we will have a pair $\{6,6\}$ in the code. In the following lemma we collect all necessary simple facts that we are going to use in the rest of the proof.

Lemma 3.2.

(i) *For* $x>1$ *and* $m\ge 3$, $A_m(x)>1$.
(ii) *For* $x>1$, $A_6A_6(x)>A_6A_3A_6(x)$.
(iii) *For* $x>1$ *and* $n>0$, $A_4^n(x)>A_3(x)$.
(iv) *For* $x>1$ *and* $n>0$, $(A_6A_3)^n(x)>A_4(x)$.

Proof. (i) and (ii) are obtained by a simple calculation; (iii) follows from $A_4(x)>A_3(x)$, the fact that $w_4=2+\sqrt{3}>A_3(x)$ and Corollary 2.4. Similarly, (iv) follows from $A_6A_3(x)>A_4(x)$, the fact that $w_{63}=3+\sqrt{7}>A_4(x)$ and Corollary 2.4. □

In order to prove that A is totally F–reduced, we have to consider six essential cases:

(1) $A=A_6A_3A_6\dots A_4^{n_1}\dots\dots A_4^{n_2}\dots A_6A_3A_6A_4^{n_m}$, where $n_i\ge 0$ and $n_m>0$. In this case we show that $w_A>w_{63}$ and $u_A<u_{63}$ where w and u denote attractive and repelling fixed points of the corresponding transformations. By Lemma 2.4, the Möbius transformation corresponding to any reduced matrix represents an increasing function for real $x>1$. Using Lemma 3.2 (i)–(iii) we may substitute A_3 for each $A_4^{n_i}$ with $n_i>0$ and insert a A_3 between any two neighboring A_6's to obtain for some integer $N>0$

$$w_A=A_6A_3A_6\dots A_4^{n_1}\dots A_4^{n_2}\dots A_6A_3A_6A_4^{n_m}w_A>(A_6A_3)^Nw_A.$$

If $w_A\le w_{63}$, by Corollary 2.4, we would have $(A_6A_3)^Nw_A\ge w_A$, a contradiction, hence $w_A>w_{63}$. By property (1.4) $\frac{1}{u_A}$ has the period reversed to that of w_A. Hence

$$\frac{1}{u_A}=A_4^{n_m}A_6A_3A_6\dots A_4^{n_1}\dots A_6A_3A_6\frac{1}{u_A}>(A_3A_6)^N\frac{1}{u_A},$$

and it follows that $\frac{1}{u_A}>w_{36}=\frac{1}{u_{63}}$. Thus we obtain the second inequality $u_A<u_{63}$. The axis of A encloses the axis of A_6A_3, and hence intersects F properly.

(2) $A=A_6A_3A_6\dots A_4^{n_1}\dots A_6A_3$. The same argument as above shows that in this case also $w_A>w_{63}$ and $u_A<u_{63}$. As in case (1), the axis of A encloses the axis of A_6A_3, and hence intersects F properly.

(3) $A=A_6A_3A_6\dots A_4^{n_1}\dots\dots A_4^{n_2}\dots A_6A_3A_6$. The same argument as in (1) shows that for some $M>0$

$$w_A=A_6A_3A_6\dots A_4^{n_1}\dots A_4^{n_2}\dots A_6A_3A_6w_A\ge(A_6A_3)^MA_6w_A.$$

Using Lemma 3.2 (iv) and (iii) we obtain $w_A > A_4A_6w_A > A_3A_6w_A$, hence $w_A > w_{36}$. Then $A_6w_A = 6 - \frac{1}{w_A} > 6 - \frac{1}{w_{36}} = w_{63}$, and therefore $w_A > w_{63}$. For $\frac{1}{u_A}$, the inequality $\frac{1}{u_A} > w_{36}$ already gives $u_A < u_{63}$. As in case (1), the axis of A encloses the axis of A_6A_3, and hence intersects F properly.

(4) $A = A_3A_6A_3A_6 \dots A_4^{n_1} \dots A_6A_3A_6$. The same argument as in case (1) shows that in this case $w_A > w_{36}$ and $u_A < u_{36}$. Hence the axis of A encloses the axis of A_3A_6 and hence intersects F properly.

(5) $A = A_4^{n_1}A_6A_3A_6A_3A_6 \dots A_6A_3A_6A_4^{n_m}$, where $n_1 > 0$, $n_m > 0$. By Lemma 3.2 (ii)–(iv) we obtain

$$w_A = A_4^{n_1}A_6A_3A_6A_3A_6 \dots A_6A_3A_6A_4^{n_m}w_A > A_4^{n_1}(A_6A_3)^K w_A > A_4^{n_1+1}w_A.$$

Therefore $w_A > w_4$ and similarly, $\frac{1}{u_A} > w_4$ which implies $u_A < \frac{1}{w_4} = u_4$. In this case the axis of A encloses the axis of A_4 , and hence intersects F properly.

(6) $A = A_4^{n_1}A_6A_3A_6A_3A_6 \dots A_6A_3A_6$, where $n_1 > 0$. Lemma 3.2 (iii) and the same argument as in case (4) shows that $w_A > w_{36}$ and $u_A < u_{36}$, hence the axis of A encloses the axis of A_3A_6 and hence intersects F properly.

□

Corollary 3.3. *The only primitive totally F–reduced matrices whose axes pass through the vertex ρ of the fundamental region F are A_4, A_3A_6, and A_6A_3.*

Proof. We have seen in the proof of Theorem 2, that the axis of any totally F–reduced matrix with a corner code different from a power of A_4, A_3A_6, or A_6A_3 encloses the axis of one of the three transformations A_4, A_3A_6, or A_6A_3, and hence intersects F properly and does not pass through the vertex ρ. By Lemma 3.1 the same is true for any totally F–reduced matrix which does not have a corner code. □

References

1. R. Adler and L. Flatto, *Cross section maps for geodesic flows, I (The Modular surface)*, Birkhäuser, Progress in Mathematics (ed. A. Katok) (1982), 103–161.
2. E. Artin, *Ein Mechanisches System mit quasiergodischen Bahnen*, Addison Wesley, Collected Papers, (1965), 499–501.
3. G.H. Hardy and E.M. Wright, *Introduction to the theory of numbers*, Oxford University Press, 1979.
4. G.A. Hedlund, *A metrically transitive group defined by the modular group*, Amer. J. Math. **57** (1935), 668–678.
5. S. Katok, *Reduction theory for Fuchsian groups*, Math. Ann. **273** (1986), 461–470.
6. S. Katok, *Fuchsian groups*, The University of Chicago Press, Chicago and London, 1992.
7. A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, New York, 1995.
8. M. Morse, *Symbolic dynamics*, (unpublished), Institute for Advanced Study Notes, Princeton (1966).
9. C. Series, *Symbolic dynamics for geodesic flows*, Acta Math. **146** (1981), 103–128.
10. C. Series, *The modular surface and continued fractions,*, J.London Math.Soc. (2) **31** (1985), 69-80.
11. C. Series, *Geometrical Markov coding of geodesics on surfaces of constant negative curvature*, Ergod. Th. & Dynam. Sys. **6** (1986), 601-625.
12. D. Zagier, *Zetafunktionen und quadratische Körper: eine Einführung in die höhere Zahlentheorie*, Hochschultext, Springer, Berlin, Heidelberg, New York, 1982.