

GEOMETRICALLY MARKOV GEODESICS ON THE MODULAR SURFACE

SVETLANA KATOK AND ILIE UGARCOVICI

Dedicated to Yu. S. Ilyashenko on the occasion of his sixtieth birthday

ABSTRACT. The Morse method of coding geodesics on a surface of constant negative curvature consists of recording the sides of a given fundamental region cut by the geodesic. For the modular surface with the standard fundamental region each geodesic (which does not go to the cusp in either direction) is represented by a bi-infinite sequence of non-zero integers called its *geometric code*.

In this paper we show that the set of all geometric codes is not a finite-step Markov chain, and identify a maximal 1-step topological Markov chain of admissible geometric codes which we call, as well as the corresponding geodesics, *geometrically Markov*. We also show that the set of geometrically Markov codes is the maximal symmetric 1-step topological Markov chain of admissible geometric codes, and obtain an estimate from below for the topological entropy of the geodesic flow restricted to this set.

INTRODUCTION

Let $\mathcal{H} = \{z = x + iy : y > 0\}$ be the upper half-plane endowed with the hyperbolic metric, and Γ be a finitely generated Fuchsian group of the first kind. Oriented geodesics on $M = \Gamma \backslash \mathcal{H}$, with exception of those which go to a cusp of M in either direction, can be coded with respect to a fixed Dirichlet fundamental region F for Γ as follows. The fundamental region F is a polygon with an even number of sides which are paired by generators of Γ and their inverses, and labeled accordingly [K1, K2]. Let us denote this generating set by Γ_F . Then any oriented geodesic on M which does not pass through vertices of F —we call such *general position geodesics*— can be coded by a bi-infinite sequence of elements from Γ_F by taking inverses of the generators labeling the sides of the fundamental region F cut by the geodesic. This construction goes back to Morse [M] (see also [K2] and [S1, S3] for details), and we will refer to these sequences as *Morse coding sequences*. For general position geodesics, a coding sequence is periodic if and only if the geodesic is closed. An ambiguity in assigning a Morse code occurs if a geodesic passes through a vertex of F : such geodesics have more than one code, and closed geodesics have non-periodic codes along with periodic ones (see [BiS, GL] for relevant discussion).

For free groups Γ with properly chosen fundamental regions, all reduced (here it simply means that a generator g does not follow or precede g^{-1}) bi-infinite

Date: May 28, 2003 and, in revised form, February 10, 2005.

2000 Mathematics Subject Classification. Primary 37D40, 37B40; Secondary 11A55, 20H05.

Key words and phrases. Modular surface, geodesic flow, topological entropy, topological Markov chain.

Published in Moscow Mathematical Journal **5**, no.1 (2005), 135–155.

sequences of elements from Γ_F are realized as Morse coding sequences of geodesics on M [S3] but, in general, this is not the case. Even for the classical example of $\Gamma = PSL(2, \mathbb{Z})$ with the standard fundamental region

$$F = \{z \in \mathcal{H} : |z| \geq 1, |\operatorname{Re} z| \leq 1/2\},$$

no elegant description of admissible Morse coding sequences is known and probably does not exist. Important results in this direction were obtained in [GL], where the admissible coding sequences were described in terms of forbidden blocks. The set of generating forbidden blocks found by the authors has an intricate structure attesting the complexity of the Morse code.

Let us review the Morse coding procedure for this classical case. The boundary of F consists of four sides: left and right vertical, identified and labeled by T ($T(z) = z + 1$) and T^{-1} , respectively, and left and right circular identified and labeled by S ($S(z) = -\frac{1}{z}$) (see Figure 1). In this case, the Morse coding sequence of an oriented general position geodesic γ can be represented by a bi-infinite sequence of non-zero integers. We choose an initial point on the circular part of the boundary of F and move in the direction of geodesic, assigning a positive number to a block of consecutive T 's and a negative number to a block of consecutive T^{-1} 's separated by S 's. The bi-infinite sequence of non-zero integers

$$[\gamma] = [\dots, n_{-1}, n_0, n_1, n_2, \dots],$$

uniquely defined up to a shift, is called the *geometric code* of γ . Moving the initial point in either direction until its return to one of the circular sides of F corresponds to a shift of the geometric coding sequence $[\gamma]$.

An oriented geodesic with geometric code $[\gamma]$ can be lifted to the upper half-plane \mathcal{H} (by choosing the initial point appropriately) so that it intersects

$$T^{\pm 1}(F), \dots, T^{n_1}(F), T^{n_1}S(F), \dots, T^{n_1}ST^{n_2}S(F), \dots$$

in the positive direction (the sign in the first group of terms is chosen in accordance with the sign of n_1 , etc.) and

$$S(F), ST^{\mp 1}(F), \dots, ST^{-n_0}(F), \dots, ST^{-n_0}ST^{-n_{-1}}(F), \dots$$

in the negative direction [GK].

An oriented geodesic on \mathcal{H} (a semicircle or a ray orthogonal to the real axis) connects two points "at infinity" $\mathbb{R} \cup \{\infty\}$, which are called its *repelling* and *attracting* end points and will be denoted by u and w , respectively.

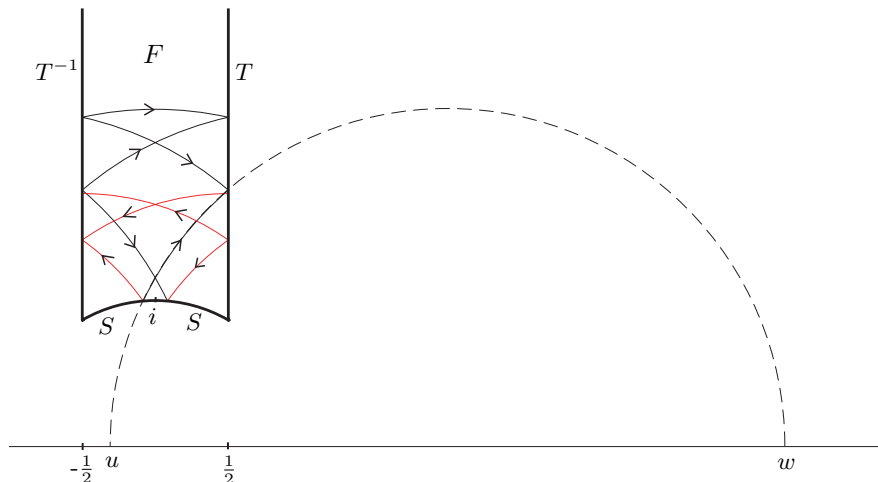
A geodesic γ on M is closed, if and only if it is the projection on M of the axis of a hyperbolic transformation $A \in PSL(2, \mathbb{Z})$. In this case, the geometric coding sequence of a general position geodesic is periodic, and we call the least period $[n_1, n_2, \dots, n_m]$, defined up to a cyclic permutation, its geometric code. Moreover, if we choose A such that its axis enters F through the circular boundary, then

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

For example, the Morse coding sequence of the closed geodesic shown on Figure 1 is

$$[T, T, T, T, S, T^{-1}, T^{-1}, T^{-1}, S],$$

hence the periodic geometric code is $[4, -3]$. The lift of the geodesic on \mathcal{H} shown with a dashed line is the axis of the transformation $G = T^4ST^{-3}S$, $G(z) = (13z + 4)/(3z + 1)$.


 FIGURE 1. The fundamental region and a geodesic on M

The case when a geodesic passes through the corner $\rho = \frac{1}{2} + i\frac{\sqrt{3}}{2}$ of F was described to a great extent in [GL, Section 7]. Such a geodesic has more than one code obtained by approximating it by general position geodesics which pass near the corner ρ slightly higher or slightly lower. If a geodesic hits the corner only once it has exactly two codes. If a geodesic hits the corner at least twice, it hits it infinitely many times and is closed; if it hits the corner n times in its period, it has exactly $2n + 2$ codes, i.e. shift-equivalent classes of coding sequences, some of which are not periodic. Canonical codes considered in [K2] were obtained by the convention that a geodesic passing through ρ in the clockwise direction exits F through the right vertical side of F labeled by T (this corresponds to the approximation by geodesics which pass near the corner ρ slightly higher). According to this convention, the geometric codes of the axes of transformations $A_4 = T^4S$, $A_{3,6} = T^3ST^6S$ and $A_{6,3} = T^6ST^3S$ are $[4]$, $[3, 6]$ and $[6, 3]$, respectively (see Figures 5, 4, and 7 below). However, all these geodesics have other codes. For example, the axis of A_4 has a code $[2, -1]$ obtained by approximation by geodesics which pass near the corner ρ slightly lower (see Figure 3), and two non-periodic codes for the same closed geodesic are

$$[\dots, 4, 4, 4, 3, -1, 2, -1, 2, -1, 2, \dots] \text{ and } [\dots, 2, -1, 2, -1, 2, -1, 3, 4, 4, \dots].$$

The main results of this paper are the following. We identify a class of bi-infinite sequences that are realized as geometric codes (Theorem 1.5) and describe it as a topological 1-step Markov chain. (Some preliminary results in identifying admissible geometric codes were obtained by J. Noel and M. Richter while working under the supervision of the authors of this paper during the REU program in the summer of 2002.) We also show that the set of all geometric codes is not a finite-step Markov chain (Theorem 2.6) and that the class of admissible geometric codes found in Theorem 1.5 is a maximal 1-step transitive topological Markov chain (Theorem 2.4), and the maximal symmetric 1-step transitive topological Markov chain (Theorem 2.5) in the set of all geometric codes. We call the set of geodesics with geometric codes identified in Theorem 1.5 *geometrically Markov*, and the restriction of the geodesic flow to this set *geometrically Markov geodesic flow*. In Section 3 we give a

lower bound for the topological entropy of the geometrically Markov geodesic flow (Theorem 3.1).

Let us remark that there is another method of coding geodesics which uses the boundary expansions of the end points of the geodesic at infinity and a certain “reduction theory”. It was first applied by Artin [Ar] to the modular group (he used continued fractions for the boundary expansions), used by Hedlund [H], and developed by Bowen and Series in [BoS] and [S1, S3] for other Fuchsian groups. Notably, an elegant *arithmetic* code for geodesics on the modular surface was obtained in [K2, GK] using minus continued fraction expansions of the endpoints, and it was interpreted via a particular “cross-section” of SM in [GK]. The set of arithmetic coding sequences was identified in [GK]: in contrast with the set of geometric coding sequences, it is a symbolic Bernoulli system on the infinite alphabet $\mathcal{N}^+ = \{n \in \mathbb{Z} : n \geq 2\}$. This code is a “relative” of Artin’s code. The space of Artin’s coding sequences is a 1-step Markov chain, and can also be interpreted via another cross-section as described in [KU] (see also [AF1, AF2]).

We are very glad that this article appears in the collection honoring the achievements of Yulij Sergeevich Ilyashenko. An outstanding mathematician, he has made an invaluable contribution into preserving and developing the vitality of the Moscow mathematical school. The world mathematical community owes a great debt of gratitude to him.

1. ADMISSIBLE GEOMETRIC CODES

In this section, we give a sufficient condition for a bi-infinite sequence of integers to be realized as a geometric code of a geodesic on M . We start with a preparatory lemma about convergence of “generalized” minus continued fractions. Although the result might exist in the vast literature on continued fractions and their generalizations, we were not able to find it in this form, useful for our study.

Lemma 1.1. *Let $\{a_n\}$, $n = 0, 1, \dots$, be a sequence of nonzero integers such that $|a_i| = 1$ implies $a_i \cdot a_{i+1} < 0$ (i.e. the symbol 1 must be followed by a negative integer and the symbol -1 must be followed by a positive integer). Then the sequence of rational numbers*

$$r_n = (a_0, a_1, \dots, a_n) := a_0 - \frac{1}{a_1 - \frac{1}{a_2 - \frac{1}{\ddots - \frac{1}{a_n}}}}$$

converges to a real number denoted by (a_0, a_1, \dots) , and $a_0 - 1 \leq (a_0, a_1, \dots) \leq a_0 + 1$.

Proof. We first assume that $|a_n| \geq 2$ for all $n \geq 1$. The proof in this case follows the lines of the proof for the case of minus continued fractions [K3], where $a_n \geq 2$, for all $n \geq 1$. We define inductively two sequences of integers $\{p_n\}$ and $\{q_n\}$ for $n \geq -2$:

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

The following properties are proved by induction (for $n \geq 0$):

- (i) $1 = q_0 < |q_1| < |q_2| < \dots < |q_n| < \dots$;
- (ii) $(a_0, a_1, \dots, a_n, x) = \frac{xp_n - p_{n-1}}{xq_n - q_{n-1}}$, for any real number x , $|x| \geq 1$;
- (iii) $p_n q_{n+1} - p_{n+1} q_n = 1$;
- (iv) $(a_0, 2, \dots, 2) \leq (a_0, a_1, \dots, a_n) \leq (a_0, -2, \dots, -2)$.

Properties (i)–(iii) show that $r_n = p_n/q_n$ for $n \geq 0$. Moreover, the sequence r_n is a Cauchy sequence because

$$|r_{n+1} - r_n| = \frac{1}{|q_n q_{n+1}|} \leq \frac{1}{(n+1)^2}.$$

Hence r_n is convergent. Using property (iv), one has

$$(a_0, 2, 2, \dots) \leq (a_0, a_1, \dots) \leq (a_0, -2, -2, \dots)$$

which implies that $a_0 - 1 \leq (a_0, a_1, \dots) \leq a_0 + 1$. (We used here that $(2, 2, 2, \dots) = 1$ and $(-2, -2, -2, \dots) = -1$.)

We return to the general situation, and prove by induction on n that: any rational $r_n = (a_0, a_1, \dots, a_n)$ can be rewritten as (b_0, b_1, \dots, b_l) , with b_i depending on a_0, a_1, \dots, a_n and $|b_i| \geq 2$ for all $i \geq 1$; moreover, if $|b_0| = 1$, then $b_0 = a_0$ and $b_0 \cdot b_1 < 0$; also $l \geq [n/2]$ if $|b_0| > 1$, and $l \geq [(n+1)/2]$ if $|b_0| = 1$.

The proof for the case $n = 1$ ($r_1 = (a_0, a_1)$) is immediate: if $|a_1| \geq 2$ let $b_0 = a_0$ and $b_1 = a_1$, if $a_1 = 1$ then $b_0 = a_0 - 1 \leq -2$ and $(a_0, a_1) = (b_0)$; if $a_1 = -1$, then $b_0 = a_0 + 1 \geq 2$ and $(a_0, a_1) = (b_0)$. Now, suppose that the statement is true for some $n > 1$ and let $r_{n+1} = (a_0, a_1, \dots, a_n, a_{n+1})$ be a sequence satisfying the assumptions of the lemma. Let $(b_1, b_2, \dots, b_{l+1})$ be the reduction of $(a_1, a_2, \dots, a_{n+1})$, where $l \geq [n/2]$ if $|b_1| \geq 2$, and $l \geq [(n+1)/2]$ if $|b_1| = 1$ (from the induction hypothesis). Thus $(a_0, a_1, \dots, a_{n+1}) = (a_0, b_1, b_2, \dots, b_{l+1})$. If $|b_1| > 1$, then the reduction is complete, $b_0 = a_0$, and $l+1 \geq [n/2] + 1 \geq [(n+2)/2]$. Suppose $b_1 = -1$. From the induction hypothesis, $b_1 = a_1$ (hence $a_0 > 0$) and $b_2 > 0$. Let $x = (b_3, \dots, b_{l+1})$. Notice that

$$\begin{aligned} (a_0, b_1, b_2, \dots, b_{l+1}) &= (a_0, b_1, b_2, x) = a_0 - \frac{1}{-1 - \frac{1}{b_2 - \frac{1}{x}}} = a_0 + 1 - \frac{1}{(b_2 + 1) - \frac{1}{x}} \\ &= (a_0 + 1, b_2 + 1, b_3, \dots, b_{l+1}), \end{aligned}$$

where $a_0 + 1 \geq 2$, $b_2 + 1 \geq 2$. The reduction is now complete and $l \geq [(n+1)/2]$. An analog reduction can be done for $b_1 = 1$.

Notice that, from the way a sequence $r_n = (a_0, a_1, \dots, a_n)$ has been rewritten as (b_0, b_1, \dots, b_l) , one obtains $a_0 - 1 \leq r_n \leq a_0 + 1$ if $|a_0| \geq 2$, $r_n \geq 1$ if $a_0 = 1$, and $r_n \leq -1$ if $a_0 = -1$.

We can prove now that $r_n = (a_0, a_1, \dots, a_n)$ is a Cauchy sequence. Let $\epsilon > 0$ and N such that $1/([N/2])^2 < \epsilon$. Let $n > m \geq N$ and consider the sequence $r_n = (a_0, a_1, \dots, a_m, a_{m+1}, \dots, a_n)$. Rewrite (a_0, a_1, \dots, a_m) as (b_0, b_1, \dots, b_l) with $l \geq [m/2]$ and let $x = (a_{m+1}, \dots, a_n)$ ($|x| \geq 1$). Let $p_k/q_k = (b_0, b_1, \dots, b_k)$ ($1 \leq k \leq l$) be defined as in the first part of the proof. Using property (iii) one has

$$(b_0, b_1, \dots, b_l, x) = \frac{xp_l - p_{l-1}}{xq_l - q_{l-1}},$$

hence

$$|r_n - r_m| = \left| \frac{xp_l - pl_{-1}}{xq_l - ql_{-1}} - \frac{pl}{ql} \right| = \frac{1}{|q_l||xq_l - ql_{-1}|} \leq \frac{1}{|q_l|} \leq \frac{1}{(l+1)^2} < \frac{1}{[m/2]^2} < \epsilon.$$

This shows that $\{r_n\}$ is a Cauchy sequence, hence convergent. The limit (a_0, a_1, \dots) satisfies the required inequalities, since each r_n does. \square

Remark 1.2. If not all a_i 's, $i \geq 1$ are equal to 2 (or -2), then

$$(1.1) \quad a_0 - 1 < (a_0, a_1, \dots) < a_0 + 1.$$

Next we show how fixed points of hyperbolic transformations in $PSL(2, \mathbb{Z})$ are related to “generalized” minus continued fraction expansions. The expression $(\overline{a_0, a_1, \dots, a_k})$ is used to denote a periodic “generalized” minus continued fraction expansion, with period (a_0, a_1, \dots, a_k) .

Proposition 1.3. *Given $A \in PSL(2, \mathbb{Z})$, $A = T^{n_1} S T^{n_2} S \dots T^{n_k} S$ with $|n_i| \geq 1$ ($1 \leq i \leq k$), such that*

- if $|n_i| = 1$, then $n_i n_{i+1} < 0$;
- not all n_i 's are equal to 2 (or -2),

then A is hyperbolic and its fixed points are given by:

$$w = (\overline{n_1, n_2, \dots, n_k}), \quad u = \frac{1}{(\overline{n_k, n_{k-1}, \dots, n_1})}.$$

Proof. It is enough to prove that A fixes u and w and that $u \neq w$. This would imply in particular that A is a hyperbolic transformation. By a direct verification we obtain that $T^{n_k} S(w) = (n_k, \overline{n_1, \dots, n_k})$ and $A(w) = (n_1, n_2, \dots, n_k, \overline{n_1, n_2, \dots, n_k}) = (\overline{n_1, n_2, \dots, n_k})$. This proves that A fixes w , and similarly one can show that A fixes u . Since not all integers n_1, n_2, \dots, n_k are equal to 2 (or -2) we have that $n_0 - 1 < w < n_0 + 1$ and $-1 < u < 1$, hence $w \neq u$. Notice also that if $r_k = (n_1, n_2, \dots, n_k)$ (a finite expression as in Lemma 1.1), then $A^n(r_k)$ converges to w . Therefore, w is the attracting fixed point of A and u must be the repelling one. \square

The following remark will be used several times in the paper.

Remark 1.4. If the sequence $[\dots, n_{-1}, n_0, n_1, n_2, \dots]$ is a geometric code of a geodesic γ , then $[\dots, -n_1, -n_0, -n_{-1}, -n_{-2}, \dots]$ is a geometric code of the reversed geodesic, and $[\dots, -n_{-1}, -n_0, -n_1, -n_2, \dots]$ is a geometric code of the geodesic symmetric to γ with respect to the imaginary axis. This is so due to the fact that the fundamental region F is symmetric with respect to the imaginary axis.

Theorem 1.5. *Any bi-infinite sequence of nonzero integers $\{\dots, n_{-1}, n_0, n_1, n_2, \dots\}$ such that*

$$(1.2) \quad \left| \frac{1}{n_i} + \frac{1}{n_{i+1}} \right| \leq \frac{1}{2} \quad \text{for } i \in \mathbb{Z}$$

is realized as a geometric code of a geodesic on M .

Proof. Relation (1.2) is equivalent to saying that the pairs $\{2, p\}$, $\{-2, -p\}$, $\{p, 2\}$, $\{-p, -2\}$ ($p \geq 1$), $\{1, q\}$, $\{q, 1\}$, $\{-1, -q\}$, $\{-q, -1\}$ ($q \neq -1, -2$), $\{3, 3\}$, $\{-3, -3\}$, $\{3, 4\}$, $\{-3, -4\}$, $\{4, 3\}$, $\{-4, -3\}$, $\{3, 5\}$, $\{-3, -5\}$, $\{5, 3\}$, $\{-5, -3\}$ are forbidden.

Let $x = \{\dots, n_{-1}, n_0, n_1, n_2, \dots\}$ be a sequence satisfying (1.2) and assume without loss of generality that such a sequence is different from the periodic corner

sequences $[4], [-4], [3, 6], [-3, -6], [2, -1], [-2, 1]$, for which we already know that they are valid geometric codes. We will prove that for any such sequence x , the geodesic $\gamma(x)$ with the end points $w(x) = (n_1, n_2, \dots)$ and $u(x) = 1/(n_0, n_{-1}, \dots)$ (we use “generalized” minus continued fractions, here) is a general position geodesic and has the geometric code $[\gamma(x)] = [\dots, n_{-1}, n_0, n_1, n_2, \dots]$.

Since the first return of the geodesic to the circular base of F corresponds to a shift of the geometric coding sequence, it is enough to prove that $\gamma(x)$ enters F through its circular base and then intersects the regions $T(F), T^2(F), \dots, T^{n_1}F$, if $n_1 > 0$ (or $T^{-1}(F), T^{-2}(F), \dots, T^{n_1}(F)$, if $n_1 < 0$) before intersecting the circular base of $T^{n_1}(F)$.

First, notice that if a sequence $x = \{\dots, n_{-1}, n_0, n_1, n_2, \dots\}$ satisfies (1.2), then

- (i) if $n_1 = 2$, then $w(x) = (n_1, n_2, \dots) > 2$ (since $n_2 \leq -1$ and $(n_2, n_3, \dots) < 0$ by (1.1));
 - (ii) if $n_1 = -2$, then $w(x) = (n_1, n_2, \dots) < -2$ ($n_2 \geq 1$ and $(n_2, n_3, \dots) > 0$ by (1.1));
 - (iii) if $|n_1| \geq 3$, then $|n_0|, |n_1| \geq 2$, and using (i) and (ii)
- $$(1.3) \quad n_1 - \frac{1}{2} < w(x) = n_1 - \frac{1}{(n_2, n_3, \dots)} < n_1 + \frac{1}{2} \quad \text{and} \quad -\frac{1}{2} < u(x) < \frac{1}{2}.$$

The proof proceeds with a detailed analysis of the different values n_1 can take.

Case $n_1 = 1$. In this case n_2 and n_0 are either -1 or -2 . Notice that if $n_2 = -1$, then $n_3 = 1, 2$, and

$$(n_2, n_3, \dots) = (-1, n_3, \dots) > -2 > (\overline{-2, 1}) = -1 - \sqrt{3}.$$

Hence

$$w(x) = (1, -1, n_3, \dots) > 1 - \frac{1}{(\overline{-2, 1})} = (\overline{1, -2}) = \frac{1 + \sqrt{3}}{2}.$$

If $n_2 = -2$, then $n_3 \geq 1$. If $n_3 \geq 2$, then $(n_3, n_4, \dots) > 2 > (\overline{1, -2})$, and

$$w(x) = (1, -2, n_3, n_4, \dots) > 1 - \frac{1}{-2 - \frac{1}{(n_3, n_4, \dots)}} > 1 - \frac{1}{-2 - \frac{1}{(\overline{1, -2})}} = (\overline{1, -2}).$$

We are left with the case $n_3 = 1$ (and $w(x) = (1, -2, 1, n_4, \dots)$). Repeating inductively the procedure we described above, one obtains that for all sequences starting with $n_1 = 1$,

$$w(x) = (1, n_2, n_3, \dots) \geq (\overline{1, -2}) =: w_{1, -2}.$$

A similar argument shows that

$$w(x) = (1, n_2, n_3, \dots) \leq (\overline{1, -1}) = \frac{1 + \sqrt{5}}{2} =: w_{1, -1}.$$

With an analog procedure, one can also show that

$$u(x) = \frac{1}{(n_0, n_{-1}, \dots)} \leq \frac{1}{(\overline{-2, 1})} = \frac{1 - \sqrt{3}}{2} =: u_{-2, 1},$$

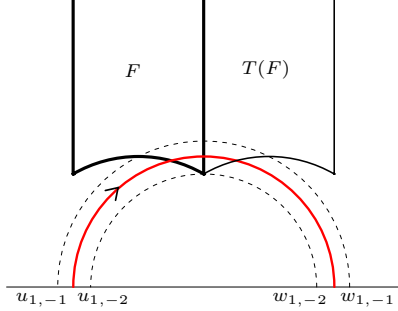
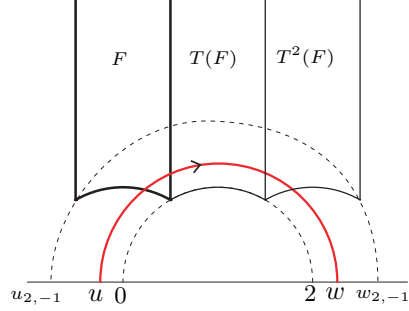
and

$$u(x) \geq \frac{1}{(\overline{-1, 1})} = \frac{1 - \sqrt{5}}{2} =: u_{1, -1}.$$

To summarize, we have found the following relations

$$w_{1, -2} \leq w(x) \leq w_{1, -1} \quad \text{and} \quad u_{1, -1} \leq u(x) \leq u_{-2, 1}.$$

This implies that the geodesic $\gamma(x)$ enters F through its circular base and then traverses $T(F)$ before intersecting the circular part of $T(F)$. Figure 2 illustrates the situation, with $\gamma(x)$ situated between the two bounding geodesics on \mathcal{H} : one from $u_{1,-2}$ to $w_{1,-2}$, and the other one from $u_{1,-1}$ to $w_{1,-1}$. Since $\gamma(x)$ is different from the corner geodesic $[1, -2]$, we notice that the segment of γ we study does not intersect any corners.

FIGURE 2. Case $n_1 = 1$ FIGURE 3. Case $n_1 = 2$

Case $n_1 = 2$. In this case $n_2 \leq -1$, and $n_0 \leq -1$. Notice that $w(x) > 2$ and $u(x) < 0$. We aim to prove that

$$w(x) \leq \overline{(2, -1)} = 1 + \sqrt{3} =: w_{2,-1} \quad \text{and} \quad u(x) \geq \frac{1}{(-1, 2)} = 1 - \sqrt{3} =: u_{2,-1}.$$

If $n_2 \leq -2$, then $(n_2, n_3, \dots) < -2 < \overline{(-1, 2)}$, and

$$w(x) = 2 - \frac{1}{(n_2, n_3, \dots)} < 2 - \frac{1}{(-1, 2)} = \overline{(2, -1)} = w_{2,-1}.$$

If $n_2 = -1$, then $w(x) = (2, -1, n_3, \dots)$, and $n_3 = 1, 2$. If $n_3 = 1$, then

$$(n_3, n_4, \dots) = (1, n_4, \dots) < 2 < \overline{(2, -1)},$$

hence

$$w(x) = (2, -1, 1, n_4, \dots) < 2 - \frac{1}{-1 - \frac{1}{(1, n_4, \dots)}} < 2 - \frac{1}{-1 - \frac{1}{\overline{(2, -1)}}} = \overline{(2, -1)}.$$

If $n_3 = 2$, then $w(x) = (2, -1, 2, n_4, \dots)$ and we can do inductively the same reasoning to conclude that for every sequence starting with $n_1 = 2$, one obtains $w(x) \geq w_{2,-1}$. A similar argument shows that

$$u(x) = \frac{1}{(n_0, n_{-1}, \dots)} \geq \frac{1}{(-1, 2)} = 1 - \sqrt{3} =: u_{2,-1}.$$

Hence, we proved that $2 < w(x) \leq w_{2,-1}$, and $u_{2,-1} \leq u(x) < 0$. This implies that $\gamma(x)$ satisfies the requirements specified above: it enters F through its base and then traverses $T(F)$ and $T^2(F)$ before intersecting the base of $T^2(F)$. Figure 3 illustrates the situation, with $\gamma(x)$ situated between two geodesics on \mathcal{H} : one from 0 to 2, and the other one from $u_{2,-1}$ to $w_{2,-1}$. Since $\gamma(x)$ is different from the

corner geodesic $[2, -1]$, we notice that the segment of γ we study does not intersect any corners.

Case $n_1 = 3$. In this case, $n_2 \geq 6$ or $n_2 \leq -2$, and $n_0 \geq 6$ or $n_0 \leq -2$. Notice that relation (1.3) implies $w(x) < 7/2$ and $u(x) > -1/2$. We aim to prove that

$$w(x) \geq (\overline{3,6}) = \frac{3 + \sqrt{7}}{2} \quad \text{and} \quad u(x) \leq \frac{1}{(\overline{6,3})} = \frac{3 - \sqrt{7}}{2}.$$

If $n_2 \leq -2$, then $w(x) > 3 > (\overline{3,6})$, and if $n_2 > 6$, then $(n_2, n_3, \dots) > 6 > (\overline{6,3})$, thus $w(x) > (\overline{3,6})$. If $n_2 = 6$, then $n_3 \geq 3$ or $n_3 \leq -2$. In the case $n_3 \leq -2$, $(n_2, n_3, \dots) = (6, n_3, \dots) > 6 > (\overline{6,3})$, hence $w(x) > (\overline{3,6})$. In the case $n_3 > 3$, $(n_3, n_4, \dots) > 3 > (\overline{3,6})$, hence $w(x) > (\overline{3,6})$. It remains to analyze the case $n_3 = 3$, i.e. $w(x) = (3, 6, 3, n_4, \dots)$. Repeating inductively the same steps, one concludes that for any sequence starting with $n_1 = 3$,

$$w(x) = (3, n_2, n_3, \dots) \geq (\overline{3,6}) =: w_{3,6}.$$

Similarly,

$$u(x) = \frac{1}{(n_0, n_{-1}, n_{-2}, \dots)} \leq \frac{1}{(\overline{6,3})} =: u_{3,6}.$$

Therefore,

$$w_{3,6} \leq w(x) < \frac{7}{2} \quad \text{and} \quad -\frac{1}{2} < u(x) \leq u_{3,6}.$$

Figure 4 illustrates the situation: $\gamma(x)$ enters F through its base and then it cuts $T(F)$, $T^2(F)$, $T^3(F)$ before intersecting the base of $T^3(F)$. Since $\gamma(x)$ is different from the corner geodesic $[3, 6]$, we notice that the segment of γ we study does not intersect any corners.

Case $n_1 = 4$. In this case $n_2 \geq 4$ or $n_2 \leq -2$, and $n_0 \geq 4$ or $n_0 \leq -2$. A similar argument as in the previous case, shows that

$$w_4 := 2 + \sqrt{3} = (\overline{4}) \leq w(x) < \frac{9}{2} \quad \text{and} \quad -\frac{1}{2} < u(x) \leq \frac{1}{(\overline{4})} = 2 - \sqrt{3} =: u_4.$$

The geometric situation is presented on Figure 5: $\gamma(x)$ encloses the geodesic from u_4 to w_4 whose geometric code is $[4]$ (the corner code corresponding to the axis of $A_4 = T^4S$).

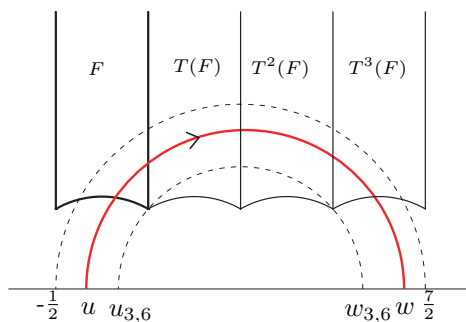


FIGURE 4. Case $n_1 = 3$

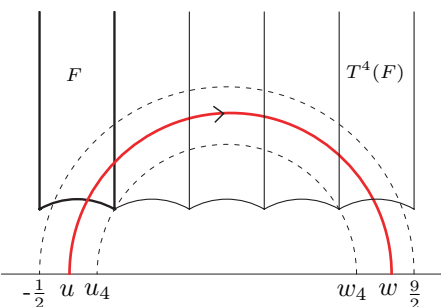


FIGURE 5. Case $n_1 = 4$

Case $n_1 = 5$. In this case $n_2 \geq 4$ or $n_2 \leq -2$, and $n_0 \geq 4$ or $n_0 \leq -2$. Hence

$$3 + \sqrt{3} = 5 - \frac{1}{(4)} \leq w(x) < \frac{11}{2} \text{ and } -\frac{1}{2} < u(x) \leq \frac{1}{(4)} = 2 - \sqrt{3}.$$

Figure 6 illustrates this situation: $\gamma(x)$ encloses the geodesic from $2 - \sqrt{3}$ to $3 + \sqrt{3}$, intersecting both F and $T^5(F)$, which can be easily verified. Therefore $\gamma(x)$ enters F through its circular base and then it traverses $T(F), T^2(F), T^3(F), T^4(F), T^5(F)$ before intersecting the base of $T^5(F)$.

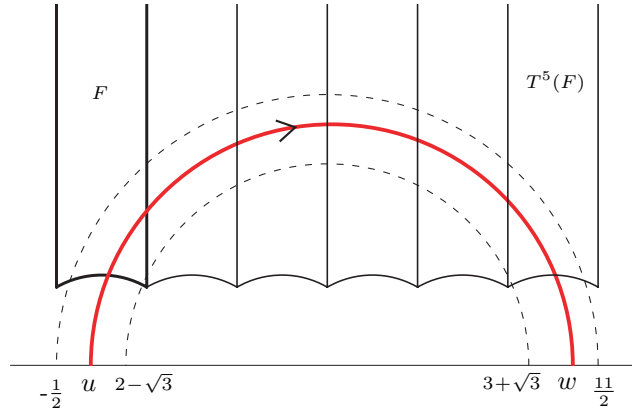


FIGURE 6. Case $n_1 = 5$

Case $n_1 = 6$. In this case $n_2 \geq 3$ or $n_2 \leq -2$, and $n_0 \geq 3$ or $n_0 \leq -2$. Hence

$$w_{6,3} := 3 + \sqrt{7} = (6, 3) \leq w(x) < \frac{13}{2} \text{ and } -\frac{1}{2} < u(x) \leq \frac{1}{(3, 6)} = 3 - \sqrt{7} =: u_{6,3}.$$

Therefore, $\gamma(x)$ satisfies the required geometric condition, since it encloses the geodesic from $u_{6,3}$ to $w_{6,3}$ whose geometric code is $[6, 3]$ (the corner code corresponding to the axis of $A_{6,3} = T^6 S T^3 S$). See Figure 7.

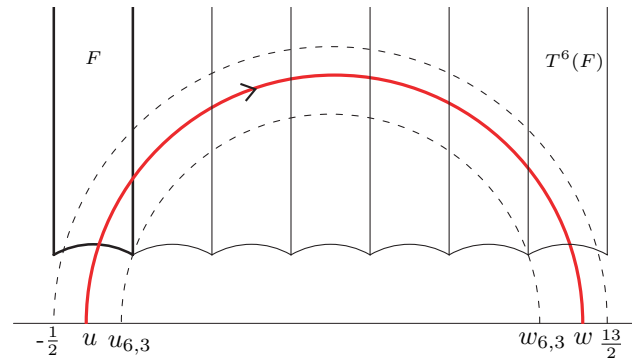


FIGURE 7. Case $n_1 = 6$

Case $n_1 > 6$. In this case $n_2 \geq 3$ or $n_2 \leq -2$, and $n_0 \geq 3$ or $n_0 \leq -2$. Hence

$$w(x) = (n_1, n_2, n_3, \dots) = n_1 - \frac{1}{(n_2, n_3, n_4, \dots)} \geq n_1 - \frac{1}{(3, 6)}$$

and

$$u(x) = \frac{1}{(n_0, n_{-1}, n_{-2}, \dots)} \leq \frac{1}{(3, 6)}.$$

Therefore,

$$n_1 - (3 - \sqrt{7}) = n_1 - \frac{1}{(3, 6)} \leq w(x) < n_1 + \frac{1}{2} \quad \text{and} \quad -\frac{1}{2} < u(x) \leq \frac{1}{(3, 6)} = 3 - \sqrt{7}.$$

The geodesic $\gamma(x)$ is enclosed by the geodesic from $-1/2$ to $n_1 + 1/2$ and encloses the geodesic from $\bar{u} = 3 - \sqrt{7}$ to $\bar{w} = n_1 - (3 - \sqrt{7})$. The equation of the semicircle from \bar{u} to \bar{w} is

$$(1.4) \quad x^2 + y^2 = (\bar{u} + \bar{w})x - \bar{u}\bar{w}$$

By direct calculations, $y > \sqrt{3}/2$ for $x = 1/2$, and $x = n_1 - 1/2$. Therefore the geodesic from \bar{u} to \bar{w} enters F through its circular base and then it cuts the regions $T(F)$, $T^2(F)$, \dots , $T^{n_1}F$ before intersecting the base of $T^{n_1}(F)$. The same will be true for $\gamma(x)$.

The situation when $n_1 < 0$, can be reduced to the previous discussion due to Remark 1.4. \square

Remark 1.6. The class of admissible geometric codes identified in Theorem 1.5 contains the so-called class of *positive* coding sequences found in [GK]: all bi-infinite sequences of positive integers $[\dots, n_{-1}, n_0, n_1, n_2, \dots]$ such that

$$(1.5) \quad \frac{1}{n_i} + \frac{1}{n_{i+1}} \leq \frac{1}{2} \quad \text{for all } i \in \mathbb{Z}.$$

Following [GK] we call oriented geodesics whose geometric coding sequences are positive, *positive geodesics*. It was proved in [GK] that the condition (1.5) is also necessary for a sequence of positive integers to be an admissible geometric code.

2. COUNTABLE TOPOLOGICAL MARKOV CHAINS

Let $\mathcal{N} = \{n \in \mathbb{Z} : |n| \geq 1\}$, $\mathcal{N}^{\mathbb{Z}}$ be the set of all bi-infinite sequences

$$\mathcal{N}^{\mathbb{Z}} = \{x = \{n_i\}_{i \in \mathbb{Z}} : n_i \in \mathcal{N}, i \in \mathbb{Z}\}$$

endowed with Tychonov product topology, and $\sigma : \mathcal{N}^{\mathbb{Z}} \rightarrow \mathcal{N}^{\mathbb{Z}}$ the left shift map given by $\{\sigma x\}_i = n_{i+1}$.

Let X_0 be the set of admissible geometric coding sequences for general position geodesics in M , and X be its closure in the Tychonov product topology. It was proved in [GL, Theorem 7.2] that every sequence in X is a geometric code of a unique oriented geodesic in M , and every geodesic in M has at least one and at most finitely many codes (see some examples in the Introduction). Thus X is a closed σ -invariant subspace of $\mathcal{N}^{\mathbb{Z}}$.

Let us recall the notion of a k -step topological Markov chain defined on the alphabet \mathcal{N} (see for example [KH]).

Definition 2.1. Given an integer $k \geq 1$ and a map $\tau : \mathcal{N}^{k+1} \rightarrow \{0, 1\}$, the set

$$X_\tau = \{x \in \mathcal{N}^{\mathbb{Z}} : \tau(n_i, n_{i+1}, \dots, n_{i+k}) = 1 \ \forall i \in \mathbb{Z}\}$$

with the restriction of σ to X_τ is called the k -step topological Markov chain with alphabet \mathcal{N} and transition map τ .

By definition X_τ is a closed σ -invariant subset of $\mathcal{N}^{\mathbb{Z}}$. Without loss of generality we will always assume that the map τ is *essential*, i.e., $\tau(n_1, n_2, \dots, n_{k+1}) = 1$ if and only if there exists a bi-infinite sequence in X_τ containing this $(k+1)$ -block $\{n_1, n_2, \dots, n_{k+1}\}$.

The simplest situation is the class of 1-step topological Markov chains which often are called just *topological Markov chains*. In this case the transition map τ is given by a matrix T called transition matrix.

Definition 2.2. A 1-step topological Markov chain X_T with a transition matrix T is called *symmetric* if $T(n, m) = T(m, n)$ for all $n, m \in \mathcal{N}$.

The set of all bi-infinite sequences satisfying relation (1.2) of Theorem 1.5 can be described as a symmetric 1-step countable topological Markov chain, with the alphabet \mathcal{N} and transition matrix A ,

$$(2.1) \quad A(n, m) = \begin{cases} 1 & \text{if } |1/n + 1/m| \leq 1/2, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that X_A is a closed σ -invariant subset of $\mathcal{N}^{\mathbb{Z}}$ and $X_A \subset X$. Moreover, X_A is a transitive topological Markov chain, i.e., for any $m, n \in \mathcal{N}$ there exists a bi-infinite sequence $x = \{n_i\}_{i \in \mathbb{Z}} \in X_A$ and $i < j$ such that $n_i = m$ and $n_j = n$.

Remark 2.3. The set X_A is strictly included in the set X . For example, $[5, 3, -2]$ is an admissible geometric code, obtained as the code of the closed geodesic corresponding to the axis of $T^5 S T^3 S T^{-2} S$ (see Figure 8).

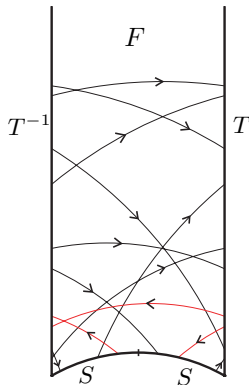


FIGURE 8. Geometric code $[5, 3, -2]$

Theorem 2.4. *The set X_A is a maximal, transitive 1-step countable topological Markov chain in the set of all geometric codes X .*

Proof. Consider a 1-step Markov chain $X_{\bar{A}} \subset X$, with \bar{A} its (essential) transition matrix, such that $X_A \subset X_{\bar{A}}$. We will show that, under this assumption, $A = \bar{A}$, hence X_A coincides with $X_{\bar{A}}$.

Assume that $\bar{A}(2, p) = 1$, with $p \geq 2$. Since $\bar{A}(p, q) = 1$ and $\bar{A}(q, 2) = 1$ for $q < -1$, this implies that $[p, q, 2]$ is a valid geometric code for any $q < -1$. Considering the axis of $T^p S T^q S T^2 S$ whose end points are $w = (\overline{p, q, 2})$ and $u = 1/(\overline{2, q, p})$, notice that, by choosing $|q|$ large enough, the geodesic from u to w does not intersect the fundamental region F , and moreover, it intersects the regions $TS(F)$, $T(F)$, $T^2(F), \dots, T^{p-1}(F)$, $T^{p-1}S(F)$, in this precise order. Hence its geometric code must contain the entry $p-1$, which contradicts our assumption. Using Remark 1.4 and the fact that the matrix A is symmetric, we also have $\bar{A}(p, 2) = 0$, $\bar{A}(-p, -2) = 0$, $\bar{A}(-2, -p) = 0$ (with $p \geq 2$). Indeed, let us show, for example, that $\bar{A}(-p, -2) = 0$ for $p \geq 2$. If $\bar{A}(-p, -2) = 1$, then $[-2, q, -p]$ would be an admissible code for $q > 1$ (since $\bar{A}(-2, q) = A(-2, q) = 1$ and $\bar{A}(q, -p) = A(q, -p) = 1$). Then, Remark 1.4 implies that $[p, -q, 2]$ would be also admissible, hence $\bar{A}(2, p) = 1$. But we have already proved that $\bar{A}(2, p) = 0$, hence the contradiction.

Now, assume that $\bar{A}(p, 1) = 1$, with $p \geq 1$. This means that there exists a geodesic whose geometric code contains the pair $\{p, 1\}$: starting from the circular boundary of F , it intersects the domains $T(F)$, $T^2(F)$, \dots , $T^p(F)$ and then hits the circular boundary of $T^p(F)$. It is impossible for this geodesic to traverse the region $T^p S T(F)$ and then to enter the region $T^p S T S(F)$ (see Figure 9). Hence the pair $\{p, 1\}$ is not admissible. Using Remark 1.4 and the symmetry of A , one also has $\bar{A}(1, p) = 0$, $\bar{A}(-1, -p) = 0$ and $\bar{A}(-p, -1) = 0$ with $p \geq 1$.

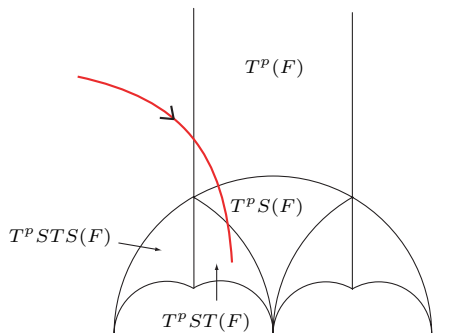


FIGURE 9. The pair $\{p, 1\}$ is not admissible

Assume now that $\bar{A}(p, -1) = 1$, with $p \geq 3$. Since $\bar{A}(-1, 2) = \bar{A}(2, -2) = \bar{A}(-2, p) = 1$, then the code $[p, -1, 2, -2]$ is admissible and can be obtained by coding the axis of the matrix $T^p S T^{-1} S T^2 S T^{-2} S$ whose endpoints are given by $u = \frac{1}{(-2, 2, -1, p)}$ and $w = (\overline{p, -1, 2, -2})$. Following the geodesic from u to w , one notices that, starting on the circular boundary of F , the geodesic traverses $p+1$ images of F consecutively, hence $[p, -1, 2, -2]$ is not a valid geometric code. Hence $\bar{A}(p, -1) = 0$ with $p \geq 3$. Using Remark 1.4 and the symmetry of A , we also have $\bar{A}(-1, p) = 0$, $\bar{A}(-p, 1) = 0$, and $\bar{A}(1, -p) = 0$ if $p \geq 3$.

Assume now that $\bar{A}(p, q) = 1$, with $\{p, q\}$ being one of the pairs $\{3, 4\}$, $\{3, 3\}$, $\{3, 5\}$, $\{4, 3\}$, $\{5, 3\}$. Since $\bar{A}(q, 7) = 1$ and $\bar{A}(7, p) = 1$, the periodic code $[p, q, 7]$

would be admissible. But this contradicts the result of [K2] about positive geodesics (see also Remark 1.6). Thus $\bar{A}(p, q) = 0$, and also $\bar{A}(-p, -q) = 0$, by using Remark 1.4.

We have just proved that there is no pair $\{p, q\}$ such that $\bar{A}(p, q) = 1$ and $A(p, q) = 0$. Therefore, $\bar{A} = A$, and $X_A = X_{\bar{A}}$, which implies that X_A is a maximal Markov chain in X . \square

Theorem 2.5. *The set X_A is the maximal symmetric 1-step countable topological Markov chain in the set of all geometric codes X .*

Proof. Let $X_{\bar{A}}$ be a symmetric 1-step Markov chain $X_{\bar{A}} \subset X$, with \bar{A} its (essential) transition matrix. We will show that, under this assumption, $A(n, m) = 0 \Rightarrow \bar{A}(n, m) = 0$. If $\bar{A}(n, m) = 1 = \bar{A}(m, n)$, then the periodic sequence $[n, m] \in X_{\bar{A}} \subset X$ is a valid geometric code. It was proved in [K2] that $[2, p]$ ($p \geq 2$), $[3]$, $[3, 4]$, $[3, 5]$ are not valid geometric codes. We proved in Theorem 2.4 that the periodic codes $[p, 1]$ ($|p| \geq 3$) are not valid, either. Using Remark 1.4, one obtains that $\bar{A}(n, m) = 0$ if $|1/n + 1/m| > 1/2$, hence $X_{\bar{A}} \subset X_A$, i.e., X_A is the maximal symmetric Markov chain in X . \square

Theorem 2.6. *The shift space X of geometric codes is not a finite-step topological Markov chain.*

*Proof.*¹ Suppose that X can be represented as a k -step topological Markov chain with transition map τ . Since any k -step Markov chain is obviously $(k + 1)$ -step Markov, we may assume without loss of generality that k is an odd number. Let

$$\bar{u} = \frac{1}{(-2, 2)} = 1 - \sqrt{2},$$

$$\bar{w} = (3, \overline{4, -8}) = 3 - \frac{1}{(4, -8)} = 3 - (-4 + 3\sqrt{2}) = 7 - 3\sqrt{2}.$$

One can check that the geodesic from \bar{u} to \bar{w} passes through the left corner of $T^3(F)$.

Let $k \geq 1$ and consider two periodic sequences of periods given by

$$A = [3, (4, -8)^k, 5, -2, (2, -2)^{lk}] \text{ and } B = [3, (4, -8)^k, 3, -2, (2, -2)^{lk}],$$

where l is a positive integer (to be determined later in the proof) and $(4, -8)^k$ and $(2, -2)^{lk}$ denote the fact that the pairs $\{4, -8\}$ and $\{2, -2\}$ are repeated k times and lk times, respectively. We will show that A is an admissible geometric code, i.e., the periodic geodesic γ_A from u_A to w_A , where

$$u_A = \frac{1}{((-2, 2)^{lk}, -2, 5, (-8, 4)^k, 3)} \text{ and } w_A = (3, \overline{(4, -8)^k, 5, -2, (2, -2)^{lk}}),$$

is in general position and its geometric code coincides with A . Indeed, the first entry in the geometric code of γ_A is 3 because u_A is close to \bar{u} and w_A is close to \bar{w} , with $u_A < \bar{u}$, $w_A > \bar{w}$ (hence γ_A enters the fundamental domain F through its circular boundary, traverses F , $T(F)$, $T^2(F)$ and $T^3(F)$ and then it hits the circular boundary of $T^3(F)$). For the proof to be complete one needs to consider

¹This is a correction of the published proof.

all geodesic segments in the period of A between two consecutive returns to the circular base of F . The next shift of γ_A has end points

$$u = \frac{1}{(3, (-2, 2)^{lk}, -2, 5, (-8, 4)^k)} \quad \text{and} \quad w = \overline{((4, -8)^k, 5, -2, (2, -2)^{lk}, 3)}.$$

Since

$$0 < u < \frac{1}{(3, -2, 2)} = \frac{2 - \sqrt{2}}{2}, \quad \frac{4 + 3\sqrt{2}}{2} = \overline{(4, -8)} < w < 4 + \frac{1}{2},$$

it encloses the geodesic from $\frac{2-\sqrt{2}}{2}$ to $\frac{4+3\sqrt{2}}{2}$ which passes through the right corner of F , and the corresponding symbol in the coding sequence of γ_A is 4. All other shift geodesics have end points satisfying $-1/2 < u < 0$ and $n < w < n + 1/2$ (if the first entry n of w is positive), and $0 < u < 1/2$, $n - 1/2 < w < n$ (if the first entry n of w is negative). Since $|n| \geq 2$, one can easily see that such a geodesic enters F through its circular side and then traverses $T(F)$, $T^2(F)$, \dots , $T^n(F)$ if $n > 0$ (or $T^{-1}(F)$, $T^{-2}(F)$, \dots , $T^n(F)$ if $n < 0$) before intersecting the circular side of $T^n(F)$. Therefore, the corresponding symbol in the coding sequence of γ_A is n , as required.

For the closed geodesic γ_B from u_B to w_B , where

$$u_B = \frac{1}{((-2, 2)^{lk}, -2, 3, (-8, 4)^k, 3)} \quad \text{and} \quad w_B = \overline{(3, (4, -8)^k, 3, -2, (2, -2)^{lk})},$$

the end point u_B is close to \bar{u} with $u_B < \bar{u}$, and w_B is close to \bar{w} with $w_B < \bar{w}$. Choose l large enough (depending on k) such that u_B is closer to \bar{u} than w_B to \bar{w} . A direct computation shows that the first entry in the geometric code of γ_B will be 2 and not 3. Therefore B is not an admissible geometric code. Since we assumed that the space X of geometric codes is k -step Markov, this implies the existence of a $(k+1)$ -tuple in the infinite sequence given by B such that $\tau(n_i, n_{i+1}, \dots, n_{i+k}) = 0$. Notice that such a $(k+1)$ -tuple must contain the symbol “3” from the beginning of the sequence B . Otherwise, by using Theorem 1.5, the periodic sequence $[n_i, n_{i+1}, \dots, n_{i+k}]$ is a valid geometric code ($k+1$ is even), so $\tau(n_i, n_{i+1}, \dots, n_{i+k})$ must be 1. But any $(k+1)$ -tuple containing the initial “3” appears in the sequence A , contradicting the fact that A is an admissible code. \square

Remark 2.7. The idea of the proof was inspired by the proof of [GL, Theorem 7.3], where the authors showed that the space of all Morse codes considered over the alphabet $\{T, T^{-1}, S\}$ is not a sofic shift space. Their result does not imply directly that the space we consider—of geometric codes over the infinite alphabet \mathcal{N} —is not a k -step countable Markov chain.

The lifts of oriented geodesics on M to its unit tangent bundle SM are the orbits of the geodesic flow $\{g^t\}$ on SM . We will conclude this section by providing a symbolic representation of $\{g^t\}$ as a special flow over the space of all geometric codes X .

A cross-section is a subset of SM which each geodesic visits infinitely often both in the future and in the past, therefore it can be identified with the space of all admissible geometric codes X .

The following is a particular cross-section B which captures the geometric code. It consists of all unit vectors in SM with base points on the circular sides of F pointing inside F (see Figure 10), and can be parameterized by (ϕ, θ) , where $\phi \in [-\pi/6, \pi/6]$ parameterizes the circle arc and $\theta \in [-\phi, \pi - \phi]$ is the angle the

unit vector makes with the positive horizontal axis in the clockwise direction. The partition of B corresponding to the geometric code is shown on Figure 11. Its

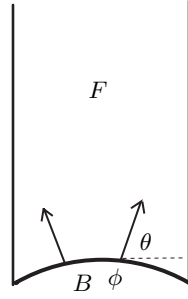


FIGURE 10. Cross section B

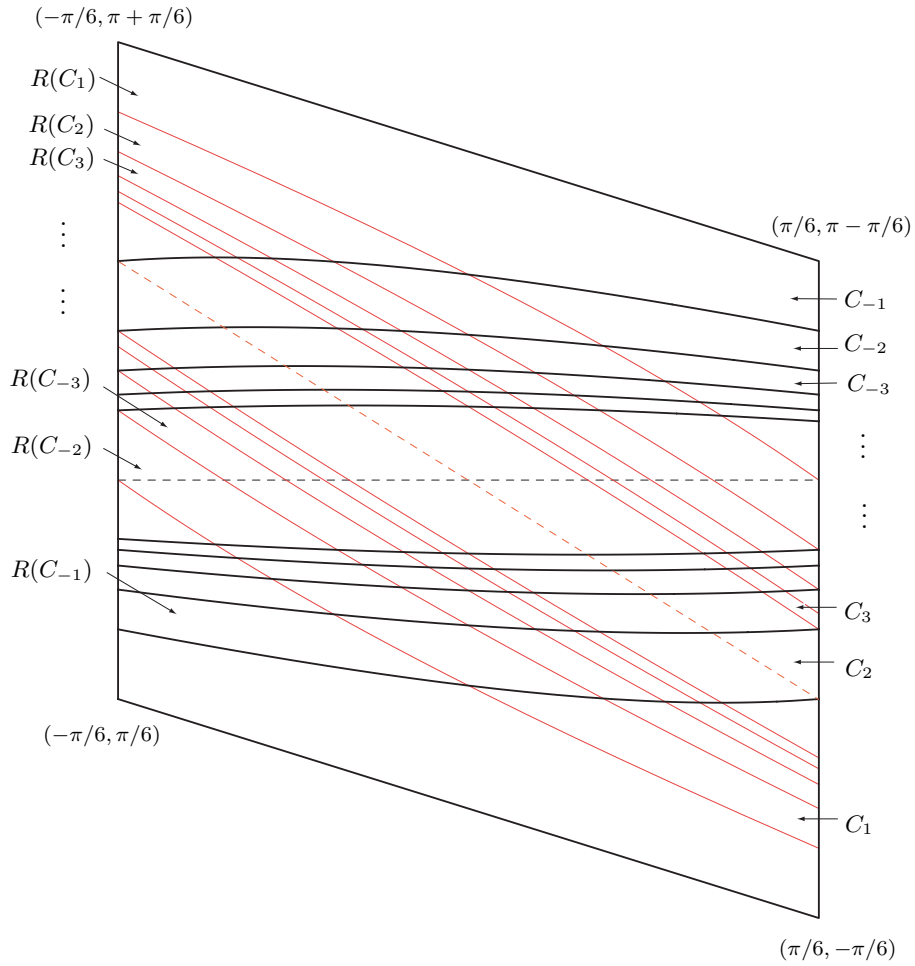


FIGURE 11. The infinite geometric partition and its image under the return map

elements are labeled by the symbols of the alphabet \mathcal{N} , $B = \sqcup_{n \in \mathcal{N}} C_n$, and are defined by the following condition: $C_n = \{v \in B : n_1(v) = n\}$, i.e., it consists of all tangent vectors v in B such that the coding sequence $x \in X$ of the corresponding geodesic with this initial vector has its first symbol in the geometric code $n_1(x) = n$.

Boundaries between the elements of the partition shown on Figure 11 correspond to geodesics going into the corner; the two vertical boundaries of the cross-section B are identified and correspond to geodesics emanating from the corner. They have more than one code. For example, the codes $[\dots, 2, -1, 2, -1, 2, -1, 3, 4, 4, 4, \dots]$ and $[4]$ correspond to the point on the right boundary of B between C_4 and C_3 , and the codes $[2, -1]$ and $[\dots, 4, 4, 4, 3, -1, 2, -1, 2, -1, 2, \dots]$ correspond to the point on the left boundary between C_2 and C_3 which are identified and are the four codes of the axis of A_4 .

Let $R : B \rightarrow B$ be the first return map. Then $n_1(R(v)) = n_2(v)$, i.e., the first return to the cross-section exactly corresponds to the left shift of the coding sequence x associated to v . This provides a symbolic representation of the geodesic flow $\{g^t\}$ on SM as a special flow over (X, σ) with the ceiling function f being the time of the first return to B (see [GK], §2.2 for a similar construction). The explicit formula for the first return time $f(x)$ is obtained similarly to that in [GK], Theorem 4:

Theorem 2.8. *Let $x \in X$, $x = [\dots, n_0, n_1, n_2, \dots]$, and $w(x)$ and $u(x)$ be the end points of the corresponding geodesic $\gamma(x)$. Then*

$$f(x) = 2 \log |w(x)| + \log g(x) - \log g(\sigma x)$$

where

$$g(x) = \frac{|w(x) - u(x)| \sqrt{w(x)^2 - 1}}{w(x)^2 \sqrt{1 - u(x)^2}}.$$

Some results of this section can be illustrated geometrically since the Markov property of the partition (for exact definition see [Ad], Section 6) is equivalent to the Markov property of the shift space.

The elements C_n and their forward iterates $R(C_n)$ are shown on Figure 11. Each C_n is a curvilinear quadrilateral with two vertical and two “horizontal” sides, and each $R(C_n)$ is a curvilinear quadrilateral with two vertical and two “slanted” sides. The horizontal sides of C_n are mapped to vertical sides of $R(C_n)$, and the vertical sides of C_n are stretched across the parallelogram representing B and mapped to the “slanted” sides of $R(C_n)$.

If $n_1(v) = n$ and $n_2(v) = m$ for some vector $v \in B$, then $R(C_n) \cap C_m \neq \emptyset$. Therefore, as follows from Figure 11, in the geometric code 2 cannot be followed by 1, 2, 3, 4 and 5.

We say that C_m and $R(C_n)$ intersect “transversally” if their intersection is a curvilinear parallelogram with two “horizontal” sides belonging to the horizontal boundary of C_m and two “slanted” sides belonging to the slanted boundary of $R(C_n)$. Notice that for each transverse intersection $R(C_n) \cap C_m$ its forward iterate under R stretches to a strip inside $R(C_m)$ between its two vertical sides.

We also observe that the elements C_m and $R(C_n)$ intersect transversally if and only if $|n| \geq 2$, $|m| \geq 2$, and

$$|1/n + 1/m| \leq 1/2,$$

so the restriction of our partition to the flow-invariant subset of X_A identified in Theorem 1.5 is indeed Markov in accordance with Theorem 7.9 of [Ad].

3. TOPOLOGICAL ENTROPY OF THE GEOMETRICALLY MARKOV GEODESIC FLOW

Let us denote by Σ the subset of the unit tangent bundle SM , consisting of vectors in SM tangent to geometrically Markov geodesics, i.e., geodesics whose codes are in X_A . The set Σ is flow invariant and noncompact. Let $\{g^t\}$ be the geodesic flow on SM , and $\{g^t|_\Sigma\}$ the restriction of the geodesic flow to Σ . The definition of topological entropy for a dynamical system with noncompact phase space adopted in this paper is the supremum of the measure-theoretic entropies over the set of all Borel invariant probability measures. The following theorem gives a lower bound estimate for $h(\{g^t|_\Sigma\})$ —the topological entropy of the flow $\{g^t|_\Sigma\}$. Let us recall that $h(\{g^t\}) = 1$, see [GK].

Theorem 3.1. $0.8417 < h(\{g^t|_\Sigma\})$.

Proof. The geodesic flow $\{g^t|_\Sigma\}$ has $B \cap \Sigma$ as a cross section. To an initial vector $v \in B \cap \Sigma$ we associate its corresponding geometric code $x \in X_A$, and the first return time of v to $B \cap \Sigma$ can be expressed in terms of the end points of the associated geodesic $\gamma(x)$ from $u(x) = 1/(n_0, n_{-1}, \dots)$ to $w(x) = (n_1, n_2, \dots)$.

Using the first return time function $f(x)$ (Theorem 2.8), one defines the special flow ϕ^t on the space

$$X_A^f = \{(x, y) : x \in X_A, 0 \leq y \leq f(x)\}$$

by the formula $\phi^t(x, y) = (x, y + t)$, using the identification $(x, f(x)) = (\sigma x, 0)$. The special flow $\{\phi^t\}$ and the geodesic flow $\{g^t|_\Sigma\}$, are conjugate, hence they have the same topological entropy. We proceed by finding a lower bound estimate for $h(\{\phi^t\})$.

Notice that $f(x)$ is cohomologous to $2 \log |w(x)|$, therefore we can assume that $f(x) = 2 \log |w(x)|$, since any two special flows with cohomologous ceiling functions are topologically conjugate.

The estimates of Theorem 1.5 show that $|w(x)| \geq (1 + \sqrt{3})/2 > 1.3$. Hence $f(x)$ is bounded away from zero, and using Abramov's formula:

$$(3.1) \quad h(\{\phi^t\}) = \sup_{\mu \in \mathcal{I}_f(X_A)} \frac{h_\mu}{\int_{X_A} f d\mu},$$

where $\mathcal{I}_f(X_A)$ is the set of all σ -invariant probability measures on X_A under which f is integrable, and h_μ is the measure-theoretic entropy of the shift map σ on X_A with respect to μ .

In order to estimate $h(\{\phi^t\})$, we will use a method developed by Polyakov [P], based on [Sa]. The method requires the countable Markov chain to be a local perturbation of the full Bernoulli shift (i.e., the number of forbidden transitions must be finite), and the first return time function $f(x)$ to depend only on the first coordinate $n_1(x)$. In order to have these conditions satisfied, we restrict X_A to the alphabet $\tilde{\mathcal{N}} = \{n \in \mathbb{Z} : |n| \geq 3\}$. Let \tilde{A} be the restriction of matrix A to $\tilde{\mathcal{N}}$. Hence $X_{\tilde{A}}$ is a σ -invariant subspace of X_A , and a local perturbation of the full Bernoulli shift on the alphabet $\tilde{\mathcal{N}}$. Let $\{\tilde{\phi}^t\}$ denote the special flow over $(X_{\tilde{A}}, f)$. If x is in $X_{\tilde{A}}$, then

$$|w(x)| \leq |n_1(x)| + \left| \frac{1}{(n_2, n_3, \dots)} \right| \leq |n_1(x)| + \frac{1}{(3, 6)} = |n_1(x)| + 3 - \sqrt{7} \leq c|n_1(x)|,$$

where $c = 1 + (3 - \sqrt{7})/3 \approx 1.11808$. Let h_c be the topological entropy of the special flow over $(X_{\tilde{A}}, g_c)$, where $g_c(x) = 2 \log c |n_1(x)|$. Since $f(x) \leq g_c(x)$, and using (3.1), we have:

$$h(\{g^t|_{\Sigma}\}) = h(\{\phi^t\}) \geq h(\{\tilde{\phi}^t\}) \geq h_c.$$

From [P], Theorems 1 and 2, the value h_c is the unique positive solution of the equation $\Psi_c(s) = 1$, where

$$\Psi_c(s) = \psi_1(s) + \frac{(3c)^{-2s} G(s) \psi_2(s) \psi_3(s)}{1 - G(s) \psi_4(s)}.$$

$G(s)$ is related with the Riemann ζ -function by the formula

$$G(s) = 2 \cdot c^{-2s} \sum_{n=6}^{\infty} n^{-2s} = 2 \cdot c^{-2s} \left(\zeta(2s) - \sum_{n=1}^{n=5} n^{-2s} \right)$$

and functions $\psi_1, \psi_2, \psi_3, \psi_4$ are given by

$$\begin{aligned} \psi_1(s) &= (c^{-8s}(225^{-2s} + 2 \cdot 180^{-2s} + 144^{-2s}) \\ &\quad - c^{-6s}(75^{-2s} + 48^{-2s} + 2 \cdot 60^{-2s} + 2 \cdot 45^{-2s} + 2 \cdot 36^{-2s}) \\ &\quad + c^{-4s}(12^{-2s} + 9^{-2s})/\Delta(s) \\ \psi_2(s) = \psi_3(s) &= (c^{-6s}(75^{-2s} + 2 \cdot 60^{-2s} + 48^{-2s}) \\ &\quad - 2 \cdot c^{-4s}(15^{-2s} + 12^{-2s}) \\ &\quad - c^{-2s}(5^{-2s} + 4^{-2s} - 3^{-2s}) + 1)/\Delta(s) \\ \psi_4(s) &= (-2 \cdot c^{-4s}(15^{-2s} + 12^{-2s}) + 3^{-2s} \cdot c^{-2s} + 1)/\Delta(s). \end{aligned}$$

The denominator $\Delta(s)$ satisfies the relation

$$\begin{aligned} \Delta(s) &= c^{-6s}(75^{-2s} + 2 \cdot 60^{-2s} + 48^{-2s}) \\ &\quad - c^{-4s}(15^{-2s} + 12^{-2s}) - 2 \cdot c^{-2s}(5^{-2s} + 4^{-2s}) + 1. \end{aligned}$$

We used the computer algebra software Maple to perform these computations and obtained $h_c \approx 0.84171$. \square

Acknowledgment. We would like to thank Dan Genin who wrote the initial version of the Mathematica program for plotting closed geodesics on the modular surface.

REFERENCES

- [Ad] R. Adler, *Symbolic dynamics and Markov partitions*, Bull. Amer. Math. Soc. **35** (1998), no. 1, 1–56.
- [Ar] E. Artin, *Ein Mechanisches System mit quasiergodischen Bahnen*, Abh. Math. Sem. Univ. Hamburg **3** (1924), 170–175.
- [AF1] R. 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), 103–161, Progr. Math., 21, Birkhäuser, Boston, 1982.
- [AF2] R. Adler and L. Flatto, *Cross section map for the geodesic flow on the modular surface*, Contemp. Math. **26** (1984), 9–24.
- [BiS] J. Birman, C. Series, *Dehn's algorithm revisited, with applications to simple curves on surfaces*, in Combinatorial group theory and topology (Alta, Utah, 1984), 451–478, Ann. of Math. Stud., 111, Princeton Univ. Press, Princeton, NJ, 1987.
- [BoS] R. Bowen and C. Series, *Markov maps associated with Fuchsian groups*, Inst. Hautes Études Sci. Publ. Math. No. 50 (1979), 153–170.

- [GK] B. Gurevich and S. Katok, *Arithmetic coding and entropy for the positive geodesic flow on the modular surface*, Mosc. Math. J. **1** (2001), no. 4, 569–582.
- [GL] D. J. Grabiner and J. C. Lagarias, *Cutting sequences for geodesic flow on the modular surface and continued fractions*, Monatsh. Math. **133** (2001), no. 4, 295–339.
- [H] G. A. Hedlund, *A metrically transitive group defined by the modular group*, Amer. J. Math. **57** (1935), 668–678.
- [KH] A. Katok, B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.
- [K1] S. Katok, *Fuchsian Groups*, University of Chicago Press, 1992.
- [K2] S. Katok, *Coding of closed geodesics after Gauss and Morse*, Geom. Dedicata **63** (1996), 123–145.
- [K3] S. Katok, *Continued fractions, hyperbolic geometry and quadratic forms*, MASS Selecta, Amer. Math. Soc., 121–160, 2003.
- [KU] S. Katok, I. Ugarcovici, *Arithmetic coding of geodesics on the modular surface via continued fractions*, to appear, CWI Tracts **135**, Centrum voor Wiskunde en Informatica, Amsterdam, 2005.
- [M] M. Morse, *A one-to-one representation of geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. **22** (1921), 33–51.
- [P] A. B. Polyakov, *On a measure with maximal entropy for a special flow over a local perturbation of a countable topological Bernoulli scheme*, Mat. Sb. **192** (2001), no. 7, 73–96 (Russian). English translation in Sb. Math. **192** (2001), no. 7-8, 1001–1024.
- [Sa] S. V. Savchenko, *Special flows constructed from countable topological Markov chains*, Funktsional. Anal. i Prilozhen. **32** (1998), no. 1, 40–53, 96 (Russian). English translation in Funct. Anal. Appl. **32** (1998), no. 1, 32–41.
- [S1] C. Series, *Symbolic dynamics for geodesic flows*, Acta Math. **146** (1981), 103–128.
- [S2] C. Series, *The modular surface and continued fractions*, J. London Math. Soc. (2) **31** (1985), 69–80.
- [S3] C. Series, *Geometrical Markov coding of geodesics on surfaces of constant negative curvature*, Ergodic Theory Dynam. Systems **6** (1986), 601–625.

DEPARTMENT OF MATHEMATICS, THE PENNSYLVANIA STATE UNIVERSITY, UNIVERSITY PARK, PA 16802

E-mail address: `katok_s@math.psu.edu`

DEPARTMENT OF MATHEMATICS, THE PENNSYLVANIA STATE UNIVERSITY, UNIVERSITY PARK, PA 16802

Current address: Department of Mathematics, Rice University, Houston, TX 77005

E-mail address: `idu@rice.edu`