

ARITHMETIC CODING AND ENTROPY FOR THE POSITIVE GEODESIC FLOW ON THE MODULAR SURFACE

BORIS GUREVICH AND SVETLANA KATOK

This article, whose authors had the privilege and good fortune of studying at Moscow University when I. G. Petrovsky was its Rector, is dedicated to his memory.

ABSTRACT. In this article we study geodesics on the modular surface by means of their arithmetic codes. Closed geodesics for which arithmetic and geometric codes coincide were identified in [9]. Here they are described as periodic orbits of a special flow over a topological Markov chain with countable alphabet, which we call the *positive geodesic flow*. We obtain an explicit formula for the ceiling function and two-sided estimates for the topological entropy of the positive geodesic flow, which turns out to be separated from one: the topological entropy of the geodesic flow on the modular surface.

2000 MATH. SUBJ. CLASS. 37D40, 37B40, 20H05.

KEY WORDS AND PHRASES. Geodesic flow, modular surface, Fuchsian group, entropy, topological entropy.

INTRODUCTION

Let $\mathcal{H} = \{z \in \mathbb{C} : \text{Im } z > 0\}$ be the upper half-plane endowed with the hyperbolic metric. Geodesics on the modular surface $M = \text{PSL}(2, \mathbb{Z}) \backslash \mathcal{H}$ can be coded in two different ways. Let

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

be the standard fundamental region for $\text{PSL}(2, \mathbb{Z})$ whose sides are identified by the generators of $\text{PSL}(2, \mathbb{Z})$, $T(z) = z + 1$ and $S(z) = -1/z$ (see Figure 1). In this article we will consider only oriented geodesics which do not go to the cusp of M in either direction (the corresponding geodesics on F contain no vertical segments), and often refer to them simply as geodesics. Notice that all closed geodesics belong to this set, and the lift of this set to SM , the unit tangent bundle of M , is an invariant set of the geodesic flow on SM . The geometric code with respect to F can be assigned to any geodesic γ on F and can be described by a double-infinite sequence of integers as follows. We choose an initial point on the circular part

Received July 1, 2001; in revised form September 26, 2001.

The first author was supported in part by RFBR Grants 99-01-00284 and 99-01-00314.

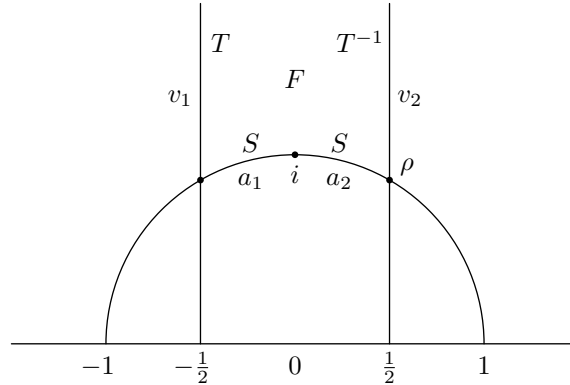


FIGURE 1.

$a_1 \cup a_2$ of the boundary of F and move in the direction of the geodesic, counting the number of times it hits the vertical sides of the boundary of F , so that a positive integer is assigned to each block of hits of v_2 , and a negative one, to each block of hits of v_1 . Moving the initial point in the opposite direction allows us to continue the sequence backwards. Thus we obtain a double-infinite sequence of integers

$$[\gamma] = [\dots, m_1, m_2, m_3, \dots]$$

called the *geometric code* of γ . Moving the initial point in either direction until its return to $a_1 \cup a_2$ corresponds to a shift of the geometric coding sequence $[\gamma]$.

A 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

$$T^{\pm 1}(F), \dots, T^{m_1}(F), T^{m_1}S(F), \dots, T^{m_1}ST^{m_2}S(F), \dots$$

in the positive direction (the sign in the first group of terms is chosen in accordance with the sign of m_1 , etc.) and

$$S(F), ST^{\mp 1}(F), \dots, ST^{-m_0}(F), \dots, ST^{-m_0}ST^{-m_{-1}}(F), \dots,$$

in the negative direction. It follows that if two geodesics on M have the same geometric code, they can be lifted to geodesics on \mathcal{H} which intersect the same infinite sequence of images of F in each direction, hence have the same end points and therefore coincide, i.e., the geometric code classifies geodesics on M . The geometric code is periodic if and only if the geodesic is closed, i.e., is the axis of a matrix in $SL(2, \mathbb{Z})$. 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 . This construction is universal for all finitely generated Fuchsian groups and goes back to Morse [14]; see [9] for details.

In [9] “-” continued fractions were used to produce another code classifying closed geodesics on the modular surface, precisely, the period (up to a cyclic permutation) of the “-” continued fraction expansion of the attracting fixed point

of the corresponding hyperbolic matrix $A \in \mathrm{SL}(2, \mathbb{Z})$, which is a quadratic irrationality. We call it the *arithmetic code* of the conjugacy class of A , or of the corresponding closed geodesic, and denote it by (A) .

A hyperbolic matrix in $\mathrm{SL}(2, \mathbb{Z})$ is called *reduced* if its attracting and repelling fixed points, denoted by w and u respectively, satisfy $w > 1$, and $0 < u < 1$. The set of reduced matrices conjugate to A is called the *A-cycle*.

A hyperbolic matrix A in $\mathrm{SL}(2, \mathbb{Z})$ is called *totally F-reduced* if the axes of all matrices in the A -cycle intersect F .

Notice that while a geometric code may contain positive or negative numbers, all numbers in any arithmetic code are ≥ 2 . It was proved in [9] that a hyperbolic matrix $A \in \mathrm{SL}(2, \mathbb{Z})$ is totally F -reduced if and only if its arithmetic and geometric codes coincide, which is also equivalent to the fact that all segments comprising the closed geodesic in F corresponding to the conjugacy class of A are clockwise (positively) oriented. We shall call closed geodesics identified in Theorems 1 and 2 of [9] *positive closed geodesics*. It was also shown in [9] that their codes are subject to restrictions which mysteriously coincide with the Schläfli symbols of regular tessellations of the sphere, include the five Platonic bodies $\{3, 3\}$, $\{3, 4\}$, $\{4, 3\}$, $\{3, 5\}$, and $\{5, 3\}$, and two “degenerate” series $\{2, p\}$ and $\{p, 2\}$ for $p \geq 2$. We shall refer to these as *Platonic restrictions*.

A positive closed geodesic with the arithmetic code (n_1, \dots, n_m) is a union of m “coils”, each “closely imitating” the closed geodesic corresponding to $T^{n_i}S$ with the code (n_i) . Lifted to the unit tangent bundle of M , SM , positive closed geodesics are exactly those which are completely contained in the positive half-space $S^+M = \{(z, \zeta) \in SM : \mathrm{Re} \zeta > 0\}$ (here (z, ζ) are the natural local coordinates on SM : $z \in \mathcal{H}$, $\zeta \in \mathbb{C}$, $|\zeta| = \mathrm{Im} z$).

The paper is organized as follows. In Section 1 we extend the reduction theory to all oriented geodesics on M (with the trivial exception of geodesics going to the cusp). This allows us to assign to each geodesic a doubly-infinite sequence of integers ≥ 2 , defined up to a shift and called its *arithmetic code*, which we interpret here as a code with respect to a particular cross-section of SM . It is natural to consider all geodesics on M whose arithmetic codes are subject to the Platonic restrictions. We show that they are exactly the geodesics comprised of segments in F which are positively (clockwise) oriented, and hence we call them *positive geodesics*. The set of vectors in SM tangent to positive geodesics is a noncompact invariant set of the geodesic flow on SM . We call the restriction of the geodesic flow to this set the *positive geodesic flow*. Since the geodesic flow is Anosov, the problem of studying positive geodesics fits in a popular branch of hyperbolic dynamics and its applications to number theory (see [10], [11]). In Section 2 we describe a representation of the positive geodesic flow as a special flow over a countable state topological Markov chain (a subshift of finite type with countable alphabet). This representation allows us to obtain (in Section 3) two-sided estimates for the topological entropy of this flow. There are different ways to extend the notion of topological entropy to dynamical systems with noncompact phase space (see [3, 4, 16]), which may give different quantities. We adopt the definition identifying the topological entropy with the supremum of the measure-theoretic entropies over the set of all flow-invariant Borel probability measures.

Our estimate shows in particular that the topological entropy of the positive geodesic flow is strictly less than one. It is worthwhile to compare this quantity with the topological entropy of the full geodesic flow. It has been known from the late 60's [12, 13, 20] that the topological entropy of the geodesic flow on any compact Riemann surface is equal to one (see also [7], Ch. 20). The arguments, however, do not carry over directly to the non-compact case. The authors consulted several leading experts in hyperbolic dynamics and thermodynamic formalism, and to our mutual amazement we found that no written reference exists for a fact that the topological entropy of the geodesic flow on the modular surface is equal to one. In Section 4 we prove this fact for quotients of \mathcal{H} by all geometrically finite Fuchsian groups of the first kind by computing the entropy of the geodesic flow with respect to the normalized Riemannian volume and showing that it is a measure of maximal entropy. It follows that the geodesic flow on the modular surface is not Borel measurable isomorphic to the positive geodesic flow.

Acknowledgments. We would like to thank A. Katok for helpful discussions on the estimates of the entropy of the full geodesic flow and for drawing our attention to his papers [5] and [6], R. Spatzier for pointing out the reference [21] to us, and V. Oseledets for helpful discussions. The first author is grateful to the Shapiro Fund and to the Center for Dynamical Systems at Penn State for their hospitality and support during his visit to Penn State in the Fall of 1999 when this work was initiated. And finally, we would like to thank the referee for the comments which helped us to clarify the exposition of the paper.

1. ARITHMETIC CODE OF ORIENTED GEODESICS ON THE MODULAR SURFACE

In this section we explain how to assign the arithmetic code to a geodesic on M .

Definition 1. An oriented geodesic on \mathcal{H} is called *reduced* if its repelling and attracting fixed points, denoted by u and w , respectively, satisfy $0 < u < 1$ and $w > 1$.

1.1. Construction of the cross-section. Let $P \cup Q$ be a subset of the unit tangent bundle SM , where P consists of all tangent vectors with base points in the side a_2 of the boundary of F (see Figure 1) such that the corresponding geodesics \mathcal{H} go in the positive (clockwise) direction and have both end points positive, and Q consists of all tangent vectors with base points in the vertical side v_2 of F pointed inwards, i. e., in the negative direction.

Proposition 2. $P \cup Q$ is a cross-section for the geodesic flow on SM .

Proof. If a geodesic on F goes in the positive direction, after a number of hits of v_2 (which do not change the direction of the geodesic) it hits either a_1 or a_2 . If it hits a_1 and does not change direction after the identification of a_1 with a_2 via $z \rightarrow -1/z$, it crosses P on SM . If it hits a_1 and changes direction or hits a_2 , it crosses v_1 in the negative direction and after the identification of v_1 with v_2 via $z \rightarrow z + 1$, enters F through the side v_2 in the negative direction, i. e., it crosses Q on SM . Assume now that a geodesic on F goes in the negative direction. If it first

crosses v_1 we arrive to the case already considered. If it first crosses a_2 and does not change direction after the identification of a_2 with a_1 , it will cross v_1 next. If it changes direction after hitting a_1 or a_2 we also arrive to the case which was already considered. \square

1.2. Description of the arithmetic code. Thus, every geodesic γ , lifted to SM , can be represented as a doubly-infinite sequence of segments σ_i between successive returns to the cross-section $P \cup Q$.

Each segment σ_i can be lifted to \mathcal{H} to a segment of a reduced geodesic which we denote γ_i . Indeed, if σ_i starts at $x \in P$, γ_i is the geodesic on \mathcal{H} determined by x and is reduced by the definition of P . If σ_i starts at $x \in Q$, γ_i is the geodesic on \mathcal{H} determined by $TS(x)$ which is also reduced. Thus we obtain a doubly-infinite sequence of reduced geodesics $\{\gamma_i\}_{i=-\infty}^{\infty}$ which represent the same geodesic γ on M . It follows that any reduced geodesic γ' on \mathcal{H} that is $PSL(2, \mathbb{Z})$ -equivalent to a lift of γ to \mathcal{H} , must coincide with one of the γ_i 's in this sequence.

For a reduced geodesic γ_i from u_i to w_i we write the “-” continued fraction expansions

$$w_i = n_1 - \frac{1}{n_2 - \frac{1}{n_3 - \frac{1}{\ddots}}}, \quad \frac{1}{u_i} = n_0 - \frac{1}{n_{-1} - \frac{1}{n_{-2} - \frac{1}{\ddots}}},$$

and assign to it a doubly-infinite sequence

$$(\gamma_i) = (\dots, n_{-3}, n_{-2}, n_{-1}, n_0, n_1, n_2, \dots), \quad n_i \geq 2.$$

Then γ_i intersects the circles $|z| = 1$ and $|z - n_1| = 1$ in the points z_1 and z'_1 , respectively (see Figure 2, where $n_1 = 5$). The corresponding segment σ_i projects

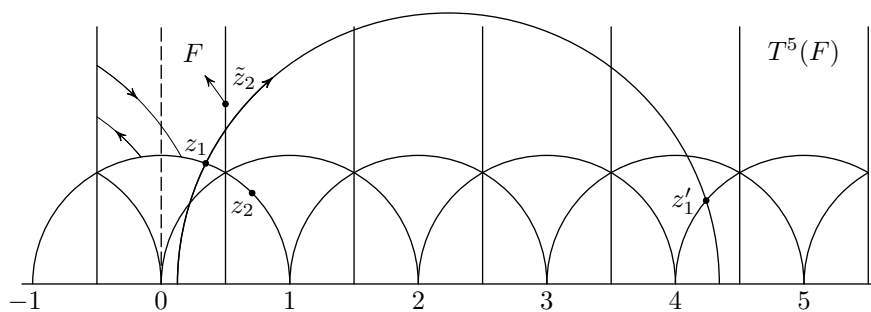


FIGURE 2.

to M to the geodesic segment $[z_1, \tilde{z}_2]$ where $z_2 = ST^{-n_1}(z'_1)$ belongs to the first quadrant of the circle $|z| = 1$ and $\tilde{z}_2 = (TS)^{-1}(z_2)$. Then $ST^{-n_1}(\gamma_i)$ is reduced, represents the next segment γ_{i+1} and produces the same coding sequence shifted one symbol to the left. Thus all reduced geodesics γ_i in the sequence produce the same, up to a shift, doubly-infinite coding sequence, which we call the *arithmetic code*

of γ . Conversely, each doubly-infinite sequence of integers ≥ 2 produces a doubly-infinite sequence of reduced geodesics on \mathcal{H} which project to the same geodesic on M . Thus we have,

Theorem 3. *Arithmetic code classifies oriented geodesics on M .*

Closed geodesics have periodic arithmetic codes. A geodesic γ on \mathcal{H} from u to w which becomes closed on M , has the arithmetic code $(\gamma) = (\overline{n_1, n_2, \dots, n_m})$ where the “-” continued fraction expansion of w has the period (n_1, n_2, \dots, n_m) and the “-” continued fraction expansion of $1/u$ has the period $(n_m, n_{m-1}, \dots, n_1)$. Thus this theory may be considered to be an extension of the reduction theory for closed geodesics. The coding sequence for a non-closed geodesic is obtained by the procedure described earlier in this section, and even for quadratic irrationalities is more complicated than just taking the purely periodic part of the “-” continued fraction, as the following example shows.

Example. Let γ be a geodesic on \mathcal{H} from $u = \sqrt{5}$ to $w = -\sqrt{3}$. The “-” continued fraction expansions are

$$w = (-1, 2, \overline{2, 3}), \quad \frac{1}{u} = (1, \overline{2, 6, 2, 2}).$$

The segment $\gamma \cap F$ lifted to SM represents a segment beginning and ending on Q , hence $TS(\gamma) = \gamma'$ from $TS(u) = u'$ to $TS(w) = w'$ is reduced. Indeed,

$$w' = (\overline{2, 3}), \quad \frac{1}{u'} = (\overline{6, 2, 2, 2}),$$

and the arithmetic code of γ is

$$(\dots, 2, 2, 2, 6, 2, 2, 2, 6, 2, 3, 2, 3, 2, 3, \dots).$$

Thus the set of all oriented geodesics on M can be described symbolically as the set

$$X := \mathcal{N}^{\mathbb{Z}} = \{x = (n_i)_{i=-\infty}^{\infty} : n_i \in \mathcal{N}, i \in \mathbb{Z}\}.$$

of doubly-infinite sequences on the infinite alphabet $\mathcal{N} = \{n \in \mathbb{Z}, n \geq 2\}$ with the Tychonov product topology. Each oriented geodesic γ corresponds to its arithmetic code $(\gamma) \in X$, and each $x \in X$ produces a geodesic on \mathcal{H} from $w(x)$ to $u(x)$, where

$$w(x) = (n_1, n_2, \dots) \quad \text{and} \quad \frac{1}{u(x)} = (n_0, n_{-1}, \dots) \tag{1}$$

are respective “-” continued fraction expansions, which projects to a geodesic on M . The left shift $\sigma: X \rightarrow X$ given by $(\sigma x)_i = n_{i+1}$ transforms $\gamma(x)$ to $ST^{-n_1}\gamma(x)$. Periodic sequences, corresponding to closed geodesics, are periodic orbits of σ and are dense in X .

1.3. Calculation of the return time. We define a function $f(x)$ on X to be the length of the segment between successive returns of the geodesic $\gamma(x)$ to the cross-section $P \cup Q$ of SM , i. e., the time of the first return to $P \cup Q$ if moving with the unit speed along the geodesic $\gamma(x)$. We have found an explicit formula for the function $f(x)$.

Theorem 4. *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 $f(x) = 2 \log w(x) + \log g(x) - \log g(\sigma x)$ with*

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

Proof. Using the same notations as in Theorem 3, we denote the intersection points of $\gamma(x)$ with the circles $|z| = 1$ and $|z - n_1| = 1$ by z_1 and z_2 , respectively. Projected to SM , $[z_1, z'_1]$ is exactly a segment between two successive returns to the cross-section $P \cup Q$, and the value $f(x)$ is equal to its hyperbolic length (see Figure 2). The next segment is given by $ST^{n_1}(\gamma(x))$, and we write $z_j = x_j + iy_j$ for $j = 1, 2$, $z'_1 = x'_1 + iy_2$, $w_1 = w(x)$, $u_1 = u(x)$, and $w_2 = ST^{-n_1}w_1$, $u_2 = ST^{-n_1}u_1$. A simple calculation shows that

$$f(x) = \ell(z_1, z'_1) = \log \frac{y_2|w_1 - z_1|^2}{|w_1 - z'_1|^2 y_1}.$$

The expression under the logarithm can be simplified

$$\begin{aligned} \frac{y_2|w_1 - z_1|^2}{|w_1 - z'_1|^2 y_1} &= \frac{y_2((w_1 - x_1)^2 + y_1^2)}{((w_1 - x'_1)^2 + y_2^2)y_1} = \frac{y_2(w_1^2 - 2w_1x_1 + x_1^2 + y_1^2)}{((x_2 - \frac{1}{w_2})^2 + y_2^2)y_1} \\ &= \frac{(w_1^2 - 2w_1x_1 + 1)w_2^2 y_2}{(w_2^2 - 2w_2x_2 + 1)y_1}. \end{aligned}$$

Thus

$$f(x) - 2 \log w_1 = \log \frac{\frac{w_1^2 - 2w_1x_1 + 1}{y_1 w_1^2}}{\frac{w_2^2 - 2w_2x_2 + 1}{y_2 w_2^2}}.$$

Now we notice that for $j = 1, 2$

$$x_j = \frac{1 + u_j w_j}{u_j + w_j} \quad \text{and} \quad y_j = \frac{\sqrt{(w_j^2 - 1)(1 - u_j^2)}}{u_j + w_j}.$$

Thus

$$\frac{w_j^2 - 2w_j x_j + 1}{y_j w_j^2} = \frac{(w_j - u_j)\sqrt{w_j^2 - 1}}{w_j^2 \sqrt{1 - u_j^2}},$$

and therefore the error term can be written in the form

$$\log g(x) - \log g(\sigma x),$$

where

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

and

$$f(x) = 2 \log w(x) + \log g(x) - \log g(\sigma x). \tag{2}$$

□

Corollary 5. *The length of a closed geodesic with the arithmetic code (n_1, \dots, n_m) is equal to*

$$2 \log \prod_{i=1}^m w_i,$$

where w_1, \dots, w_m are the attractive fixed points of all reduced matrices corresponding to this closed geodesic.

2. SPECIAL REPRESENTATION OF THE POSITIVE GEODESIC FLOW

2.1. Characterization of positive geodesics.

Definition 6. We call a geodesic γ *positive* if all segments of the double-infinite sequence comprising it begin and end in the set P .

Since a positive geodesic avoids the set Q , its arithmetic code counts the number of times it hits the side v_2 of F in the positive direction. The following result gives a characterization of positive geodesics in terms of their arithmetic code and is a generalization of Theorems 1 and 2 of [9].

Theorem 7. *The following statements are equivalent:*

- (i) γ is positive;
- (ii) $(\gamma) = (\dots, n_{-1}, n_0, n_1, n_2, \dots)$ is subject to the Platonic restrictions, i. e., $\frac{1}{n_i} + \frac{1}{n_{i+1}} \leq \frac{1}{2}$ for all i ;
- (iii) the geometric and arithmetic codes of γ coincide, i. e., $(\gamma) = [\gamma]$;
- (iv) all segments comprising the geodesic γ in F are positively (clockwise) oriented.

Proof. It is clear from the definition and the discussion preceding Theorem 3 that (i), (ii) and (iv) are equivalent. It remains to prove that (i) and (ii) are equivalent.

(ii) \Rightarrow (i). Assume that (ii) holds, but γ is not positive. This means that it crosses Q . The geodesic γ can be approximated by a sequence of closed geodesics whose codes are subject to Platonic restrictions. By Theorems 1 and 2 of [9], these closed geodesics are positive, i. e., avoid Q . Thus γ belongs to the closure of the set of positive geodesics. But the only geodesics in the closure which are not positive are those which contain infinite vertical segment. Since such geodesics have been excluded from coding from the beginning, γ is positive.

(i) \Rightarrow (ii). Assume that γ is positive. First let us show that (γ) does not contain 2. Indeed, if $n_i = 2$, then the reduced geodesic γ_i associated to (γ) must intersect the circles $|z| = 1$ and $|z - 2| = 1$ at the points $z_1 = x_1 + iy_1$ and $z'_1 = x'_1 + iy'_1$ with $0 < x_1 < 1/2$ and $3/2 < x'_1 < 2$. Since the geodesic γ' passing through the points $\frac{1}{2} + i\frac{\sqrt{3}}{2}$ and $\frac{3}{2} + i\frac{\sqrt{3}}{2}$ also passes through 0 and 2, this is impossible since this would imply that two different geodesics, γ and γ' , have two intersection points.

Now let us assume that (γ) contains a forbidden pair $\{n_i, n_{i+1}\}$, so that

$$\frac{1}{n_i} + \frac{1}{n_{i+1}} > \frac{1}{2} \tag{3}$$

and $n_i, n_{i+1} \geq 3$. We shall show that the geodesic γ_i with the end points $w = (n_{i+1}, n_{i+2}, \dots)$ and $u = \frac{1}{(n_i, n_{i-1}, \dots)}$ is not positive by showing that it intersects

the unit circle $|z| = 1$ at a point $z_0 = x_0 + iy_0$ with $x_0 > 1/2$, which can be rewritten in terms of the end points of γ , u and w as

$$2(uw + 1) - (u + w) > 0. \tag{4}$$

We write $w = n_{i+1} - 1/\alpha$ and $1/u = n_i - 1/\beta$, where $\alpha = (n_{i+2}, n_{i+3}, \dots)$ and $\beta = (n_{i-1}, n_{i-2}, \dots)$. Since $n_i \geq 3$, $\alpha > 2$ and $\beta > 2$. Now we can rewrite (4) as

$$\frac{\alpha\beta(2(n_{i+1} + n_i) - n_{i+1}n_i - 1) + (n_{i+1} - 2)\alpha + (n_i - 2)\beta - 1}{\alpha(n_i\beta - 1)}.$$

Since (3) is equivalent to $2(n_{i+1} + n_i) - n_{i+1}n_i - 1 \geq 0$ and n_i and n_{i+1} are at least 3, we see that the numerator of the last expression is positive. The denominator is also positive, which completes the proof. \square

2.2. Symbolic dynamics for the positive geodesic flow. The set of sequences $(\dots, n_0, n_1, n_2, \dots)$ which satisfy $\frac{1}{n_i} + \frac{1}{n_{i+1}} \leq \frac{1}{2}$ for all i , i.e., are subject to the Platonic restrictions, can be described with the help of an infinite matrix A of zeros and ones where $A(i, j) = 0$ precisely when the pair $\{i, j\}$ is prohibited:

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 1 & \dots \\ 0 & 0 & 1 & 1 & 1 & 1 & \dots \\ 0 & 0 & 1 & 1 & 1 & 1 & \dots \\ 0 & 1 & 1 & 1 & 1 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

In compliance with Theorem 7 we shall call such sequences *positive*. The set of positive sequences can be described as

$$X_A = \{x \in X : A(n_i, n_{i+1}) = 1\},$$

and the restriction $\sigma|_{X_A} = \sigma_A$ is a countable *topological Markov chain*. Positive periodic sequences correspond to the periodic orbits of σ_A and are dense in X_A . By [9], Cor. 2.9 they correspond to the positive closed geodesics on the modular surface.

By Theorem 7, the behavior of a positive geodesic $\gamma(x)$ in F is pretty regular. It enters the fundamental region F through the side a_2 (see Figure 1), and after a number of hits of the vertical side v_2 hits the side a_1 (we call this a “coil”), thus returning to the side a_2 after the identification in F . The function $f(x)$ restricted to X_A is the length of this coil, or the time of the first return to the set P . Using the function $f(x)$ as the ceiling function we define the usual 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 $\varphi_t(x, y) = (x, y + t)$ using the identification $(x, f(x)) = (\sigma_A x, 0)$. This is a symbolic representation of the *positive geodesic flow* $\{\varphi_t^+\}$ defined as a restriction of the geodesic flow $\{\varphi_t\}$ on SM to the subset $\Sigma^+ \subset S^+M$ consisting of all vectors in SM tangent to positive geodesics (see Introduction). The set Σ^+ is $\{\varphi_t\}$ -invariant and noncompact. It is the maximal $\{\varphi_t\}$ -invariant subset of Σ^+ .

There is a one-to-one correspondence between the periodic orbits of σ_A and the closed orbits of the special flow ϕ_t so that the lengths of the latter are equal to the lengths of the corresponding positive closed geodesics.

Since $f(x)$ is cohomologous to a simple function $2 \log w(x)$, and the special flows corresponding to those functions are topologically conjugate [15], it is sufficient to consider the special flow with the ceiling function $f(x) = 2 \log w(x)$.

3. ENTROPY OF THE POSITIVE GEODESIC FLOW

In this section we obtain a two-sided estimate for the topological entropy of the flow $\{\varphi_t^+\}$. Under the definition adopted in this paper (see Introduction), the topological entropy $h(\cdot)$ turns out to be invariant with respect to a continuous conjugacy (and even a Borel measurable conjugacy) of dynamical systems. Therefore $h(\{\varphi_t^+\}) = h(\{\phi_t\})$, where $\{\phi_t\}$ is the special flow over X_A with the ceiling function $f(x) = 2 \log w(x)$, described above.

In order to explain how to obtain the estimates of $h(\{\varphi_t^+\})$, we consider a more general situation. Let Λ be a σ -invariant Borel subset of the sequence space $X = \mathcal{N}^{\mathbb{Z}}$. Given a Borel measurable function $g: \Lambda \rightarrow \mathbb{R}$ such that $\inf_{x \in \Lambda} g(x) > 0$, one can define a special flow $\{\psi_t\} = (\Lambda, g)$ on the space

$$\Lambda^g = \{(x, y) : x \in \Lambda, 0 \leq y \leq g(x)\}$$

much as the flow $\{\phi_t\}$ was defined above. Let $\tilde{\mu}$ be an arbitrary $\{\psi_t\}$ -invariant Borel probability measure on Λ^g and μ' its projection onto Λ . The sets $C_x = \{x\} \times \Delta_x$, $x \in \Lambda$, where $\Delta_x = \{y : 0 \leq y \leq f(x)\}$, constitute a measurable partition of Λ^g . The $\{\psi_t\}$ -invariance of $\tilde{\mu}$ implies that the conditional measure on C_x induced by $\tilde{\mu}$ is the normalized Lebesgue measure on Δ_x for μ' -almost all x (here we identify C_x and Δ_x). By definition the function $x \mapsto 1/g(x)$ is bounded and hence μ' -integrable. So we can introduce a measure μ on Λ by

$$\mu(dx) = K \frac{\mu'(dx)}{g(x)}, \quad \text{where} \quad K = \left[\int_{\Lambda} (1/g(x)) \mu'(dx) \right]^{-1}.$$

It is easy to check that $K = \int_{\Lambda} g d\mu$, μ is a probability measure, and that $\tilde{\mu}$ is the restriction to Λ^g of the direct product $\mu \times \ell$, divided by K , where ℓ is the Lebesgue measure on \mathbb{R} . Moreover, μ is σ -invariant.

Conversely, given a σ -invariant probability measure μ on Λ such that $\int_{\Lambda} g d\mu < \infty$, one can define $\tilde{\mu}$ as above and make sure that $\tilde{\mu}$ is a $\{\psi_t\}$ -invariant Borel probability measure on Λ^g . Thus we have a one-to-one correspondence between the set $I_g(\Lambda)$ of σ -invariant probability measures on Λ under which g is integrable and the set $I(\Lambda^g)$ of all $\{\psi_t\}$ -invariant probability measures on Λ^g .

For each measure $\mu \in I_g(\Lambda)$ we denote by h_{μ} the measure-theoretic entropy of σ with respect to μ . The entropy of the flow $\{\psi_t\}$ with respect to the measure $\tilde{\mu}$ will be denoted by $h_{\tilde{\mu}}(\{\psi_t\})$. Recall that by definition $h_{\tilde{\mu}}(\{\psi_t\}) = h_{\tilde{\mu}}(\psi_1)$ and that by Abramov's formula [1] $h_{\tilde{\mu}}(\{\psi_t\}) = h_{\mu} / \int_{\Lambda} g d\mu$. Hence

$$h(\{\psi_t\}) = \sup_{\mu \in I_g(\Lambda)} h_{\mu} \left(\int_{\Lambda} g d\mu \right)^{-1}. \tag{5}$$

Let $g \geq g'$ on Λ , and let $\{\psi'_t\} = (\Lambda, g')$. Then by (5) $h(\{\psi_t\}) \leq h(\{\psi'_t\})$.

Theorem 8. $0.7771 < h(\{\varphi_t^+\}) < 0.8161$.

Proof. The special flow $\{\phi_t\}$ is of the type considered above with $\Lambda = X_A$ and $g = f$, i. e., $\{\phi_t\} = (X_A, f)$. The function $f(x) = 2 \log w(x)$ can be extended to the whole space X since the formulae (1) make sense for any $x \in X$. For every $x = (n_i)_{i \in \mathbb{Z}} \in X$ we write $n_i = n_i(x)$. It is easy to show that since for $x \in X_A$ $n_i(x) \geq 3$, we have

$$2 \log cn_1(x) \leq f(x) \leq 2 \log n_1(x), \quad \text{where } c = (3 + \sqrt{5})/6 \approx 0.8726. \quad (6)$$

Thus the ceiling function is estimated by two functions which depend only on the first coordinate $n_1(x)$. Using a formula for the topological entropy by Polyakov [17] based on a result of Savchenko [19] for (X_A, g_δ) , with $g_\delta(x) = 2 \log \delta n_1(x)$, for $\delta = 1$ and $\delta = c$ we obtain the estimates

$$h_1 = 0.7771 < h(\{\varphi_t^+\}) < 0.8161 = h_c.$$

The estimated values h_δ are solutions of the equation $\Psi_\delta(s) = 1$, where

$$\Psi_\delta(s) = \frac{G(s)(1 + (3\delta)^{-2s} - (12\delta)^{-2s} - (15\delta)^{-2s})}{1 - (4\delta)^{-2s} - (5\delta)^{-2s}},$$

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

$$G(s) = \delta^{-2s} \left(\zeta(-2s) - \sum_{n=1}^5 n^{-2s} \right).$$

They were obtained with the help of the computer package Pari-GP. □

4. ENTROPY OF THE GEODESIC FLOW FOR CONSTANT CURVATURE

We conclude this paper by showing that the topological entropy of the geodesic flow on a quotient of \mathcal{H} by any geometrically finite Fuchsian group of the first kind is equal to one. This confirms that the topological entropy of the positive geodesic flow on the modular surface is strictly less than that of the full geodesic flow.

Let Γ be a geometrically finite Fuchsian group of the first kind, $M = \Gamma \backslash \mathcal{H}$, SM its unit tangent bundle, and m be the Riemannian volume on SM , i. e., the measure induced by the hyperbolic measure on the upper half-plane. It is standard that $m(SM) < \infty$ and that m is an invariant measure for the geodesic flow $\{\varphi_t\}$ on SM . A theorem of Sullivan for geometrically finite Kleinian groups [21] states that the entropy of the geodesic flow relative to the (normalized) canonical measure obtained from the unique geometric measure on the limit set $\Lambda(\Gamma)$ (on the set of non-wandering geodesics) is equal to the Hausdorff dimension of $\Lambda(\Gamma)$. In the case when Γ is a geometrically finite Fuchsian group of the first kind, the canonical measure is the Riemannian volume on SM and $\Lambda = \mathbb{R} \cup \infty$, so its Hausdorff dimension is equal to one. Therefore, the entropy of the geodesic flow on SM relative to the normalized Riemannian volume is equal to one.

In [21] Sullivan also states that this measure must be a measure of maximal entropy, as in the compact case. The argument we present below is based on an adaptation of arguments of A. Katok [5] and [6] to a non-compact case, and the

Prime Geodesic Theorem of Sarnak for constant curvature ([18], Ch. 3) obtained with the use of the Selberg Trace Formula.

Let μ be any Borel probability measure on SM invariant with respect to the geodesic flow $\{\varphi_t\}$, and $P(s)$ be the number of periodic orbits of $\{\varphi_t\}$ of period $\leq s$.

Theorem 9.
$$\varliminf_{s \rightarrow \infty} \frac{\log P(s)}{s} \geq h_\mu(\{\varphi_t\}).$$

Proof. Theorem 4.1 of [6] establishes the corresponding estimate for any $C^{1+\delta}$ ($\delta > 0$) flow without fixed points on any 3-dimensional compact manifold (see the original argument for diffeomorphisms in [5]). The argument works without compactness assumption but with certain restrictions on the flow, in particular, in our case.

Let d be a distance function on SM generated by the hyperbolic metric on M . For every $s > 0$ we define a metric d_s on SM by

$$d_s(x, y) = \max_{0 \leq t \leq s} d(\varphi_t x, \varphi_t y).$$

Its discrete counterpart d'_s is defined by the same formula where s and t are assumed to be nonnegative integers. For $s, \epsilon, \delta > 0$ we denote by $N(s, \epsilon, \delta)$ (respectively, $N_c(s, \epsilon, \delta)$) the minimal number of ϵ -balls in the metric d_s that cover a set (respectively, a compact set) of measure $> 1 - \delta$. Let $N'(s, \epsilon, \delta)$ and $N'_c(s, \epsilon, \delta)$ be corresponding numbers for the metric d'_s . It is easy to verify that

$$N_c(s, \epsilon, \delta) \geq N'_c([s], \epsilon, \delta) \geq N'([s], \epsilon, \delta). \tag{7}$$

We now show that

$$h_\mu(\varphi_1) \leq \lim_{\epsilon \rightarrow 0} \varliminf_{s \rightarrow \infty} \frac{\log N'([s], \epsilon, \delta)}{s}. \tag{8}$$

Theorem 1.1 in [5] contains such an estimate for an arbitrary homeomorphism f of a compact metric space (X, d) and any f -invariant Borel probability measure μ on X . The compactness assumption is not satisfied in our case, i. e., for $f = \varphi_1$. But by analyzing the argument in [5], one can make sure that it goes through if X is σ -compact and f is continuous, and even Borel measurable, and is based only on the existence of a partition described in the following lemma.

Lemma 10. *For any $\alpha > 0$ there is a finite measurable partition ξ of X such that*

$$h_\mu(f, \xi) \geq h_\mu(f) - \alpha$$

and $\mu(\partial c) = 0$ for every element $c \in \xi$, where ∂c is the boundary of c .

By assumption there is a sequence of compact subsets $X_n \subset X$ such that $X_{n+1} \supset X_n$ and $\bigcup_n X_n = X$. In our case a partition ξ will consist of a finite number of elements of sufficiently small diameter covering X_n for sufficiently large n , and one element of sufficiently small measure. It is easy to make the measures of the boundaries of all elements equal to zero. The estimate

$$h_\mu(\{\varphi_t\}) \leq \lim_{\epsilon \rightarrow 0} \varliminf_{s \rightarrow \infty} \frac{\log N(s, \epsilon, \delta)}{s} \tag{9}$$

now follows from (8), (7).

For every compact set $K \subset SM$ with $\mu(K) > 1 - \delta$ we denote by $P(s, \epsilon, \delta, K)$ the maximal number of points in K which are ϵ -separated in the metric d_s . Let

$$P(s, \epsilon, \delta) = \inf_K P(s, \epsilon, \delta, K).$$

Since $N(s, \epsilon, \delta) \leq P(s, \epsilon, \delta)$, one obtains from (9) a similar estimate of $h_\mu(\{\varphi_t\})$ in terms of the exponential growth of $P(s, \epsilon, \delta)$ which is the same as the exponential growth of ϵ -separated orbit segments which return much closer than ϵ to their initial conditions (say, ϵ^2 -close) in a compact subset of SM of measure $\geq 1 - \delta$ in the time between $(1 - \epsilon)s$ and s . The last conclusion is derived from the flow version of the argument in the proof of Theorem 4.3 of [5] and uses the Birkhoff ergodic theorem as in the compact case.

Finally, the periodic orbits are produced exactly as in the compact case by the choice of a particular compact subset of SM of measure $\geq 1 - \delta$ and with the help of the Anosov Closing Lemma [2] which holds in our case (see the argument for geodesic flow on quotients by any Fuchsian group of the first kind in Appendix to [8]) instead of the main closing type lemma required in [5], §3. ϵ^2 -closeness of ϵ -separated orbit segments assures us that the obtained periodic orbits are different for different ϵ^2 -close orbits. \square

The same inequality now holds for the topological entropy.

Corollary 11.

$$\varliminf_{s \rightarrow \infty} \frac{\log P(s)}{s} \geq h(\{\varphi_t\}).$$

Theorem 12. *The topological entropy of the geodesic flow on a quotient of \mathcal{H} by a geometrically finite Fuchsian group of the first kind is equal to one.*

Proof. We know that $h(\{\varphi_t\}) \geq 1$. Assume that $h(\{\varphi_t\}) = h > 1$. Then

$$\varliminf_{s \rightarrow \infty} \frac{\log P(s)}{s} \geq h > 1,$$

while by Sarnak's Prime Geodesic Theorem ([18], Ch. 3) $P(s) \sim e^s/s$, a contradiction. \square

REFERENCES

- [1] L. M. Abramov, *On the entropy of a flow*, Dokl. Akad. Nauk SSSR **128** (1959), 873–875 (Russian). MR [22 #4816](#)
- [2] D. V. Anosov, *Geodesic flows on closed Riemannian manifolds of negative curvature*, Trudy Mat. Inst. Steklov. **90** (1967), 209 (Russian). MR [36 #7157](#)
- [3] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Trans. Amer. Math. Soc. **153** (1971), 401–414. MR [43 #469](#)
- [4] R. Bowen, *Topological entropy for noncompact sets*, Trans. Amer. Math. Soc. **184** (1973), 125–136. MR [49 #3082](#)
- [5] A. Katok, *Lyapunov exponents, entropy and periodic orbits for diffeomorphisms*, Inst. Hautes Études Sci. Publ. Math. (1980), no. 51, 137–173. MR [81i:28022](#)
- [6] A. Katok, *Entropy and closed geodesics*, Ergodic Theory Dynam. Systems **2** (1982), no. 3–4, 339–365 (1983). MR [85b:53047](#)

- [7] A. Katok, B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, no. 54, Cambridge University Press, Cambridge, 1995. MR [96c:58055](#)
- [8] S. Katok, *Modular forms associated to closed geodesics and arithmetic applications*, Ph.D. thesis, University of Maryland, 1983.
- [9] S. Katok, *Coding of closed geodesics after Gauss and Morse*, *Geom. Dedicata* **63** (1996), no. 2, 123–145. MR [97j:20045](#)
- [10] D. Y. Kleinbock, G. A. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces*, *Sinai's Moscow Seminar on Dynamical Systems*, Amer. Math. Soc., Providence, RI, 1996, pp. 141–172. MR [96k:22022](#)
- [11] C. Liverani, *Decay of correlations*, *Ann. of Math. (2)* **142** (1995), no. 2, 239–301. MR [96e:58090](#)
- [12] G. A. Margulis, *Certain applications of ergodic theory to the investigation of manifolds of negative curvature*, *Funktional. Anal. i Prilozhen.* **3** (1969), no. 4, 89–90 (Russian). MR [41#2582](#). English translation: *Funct. Anal. Appl.* **3** (1969), 335–336.
- [13] G. A. Margulis, *Certain measures that are connected with U-flows on compact manifolds*, *Funktional. Anal. i Prilozhen.* **4** (1970), no. 1, 62–76 (Russian). MR [42#7865](#). English translation: *Funct. Anal. Appl.* **4** (1970), 55–67.
- [14] M. Morse, *Symbolic dynamics*, Institute of Advanced Study Notes, Princeton, 1996, unpublished.
- [15] W. Parry, M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, *Astérisque* (1990), no. 187–188, 268. MR [92f:58141](#)
- [16] Ya. B. Pesin, B. S. Pitskel, *Topological pressure and the variational principle for noncompact sets*, *Funktional. Anal. i Prilozhen.* **18** (1984), no. 4, 50–63, 96 (Russian). MR [86i:28031](#). English translation: *Funct. Anal. Appl.* **18** (1984), 50–63.
- [17] 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). MR [186i:374](#)
- [18] P. Sarnak, *Prime geodesic theorem*, Ph.D. thesis, Stanford, 1980.
- [19] S. V. Savchenko, *Special flows constructed from countable topological Markov chains*, *Funktional. Anal. i Prilozhen.* **32** (1998), no. 1, 40–53, 96 (Russian). MR [99m:28040](#). English translation: *Funct. Anal. Appl.* **32** (1998), no. 1, 259–277.
- [20] Ja. G. Sinai, *Gibbs measures in ergodic theory*, *Uspehi Mat. Nauk* **27** (1972), no. 4(166), 21–64 (Russian). MR [53#3265](#). English translation: *Russian Math. Surveys* **27** (1972), no. 4, 21–69.
- [21] D. Sullivan, *Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups*, *Acta Math.* **153** (1984), no. 3–4, 259–277. MR [86c:58093](#)

FACULTY OF MECHANICS & MATHEMATICS, MOSCOW STATE UNIVERSITY, VOROBYOVY GORY,
MOSCOW 119899, RUSSIA

E-mail address: bmg@gibbs.math.msu.su

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

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