Ergod. Th. & Dynam. Sys. (2012), **32**, 755–777
doi:10.1017/S0143385711000460

Applications of (a, b)-continued fraction transformations

SVETLANA KATOK† and ILIE UGARCOVICI‡

† *Department of Mathematics, Pennsylvania State University, University Park, PA 16802, USA*
(e-mail: katok_s@math.psu.edu)
‡ *Department of Mathematical Sciences, DePaul University, Chicago, IL 60614, USA*
(e-mail: iugarcov@depaul.edu)

(*Received* 31 *January* 2011 *and accepted in revised form* 26 *May* 2011)

Dedicated to the memory of Dan Rudolph

Abstract. We describe a general method of arithmetic coding of geodesics on the modular surface based on the study of one-dimensional Gauss-like maps associated to a two-parameter family of continued fractions introduced in [Katok and Ugarcovici. Structure of attractors for (a, b)-continued fraction transformations. *J. Modern Dynamics* **4** (2010), 637–691]. The finite rectangular structure of the attractors of the natural extension maps and the corresponding 'reduction theory' play an essential role. In special cases, when an (a, b)-expansion admits a so-called 'dual', the coding sequences are obtained by juxtaposition of the boundary expansions of the fixed points, and the set of coding sequences is a countable sofic shift. We also prove that the natural extension maps are Bernoulli shifts and compute the density of the absolutely continuous invariant measure and the measure-theoretic entropy of the one-dimensional map.

1. *Introduction and background*

In [**16**], the authors studied a new two-parameter family of continued fraction transformations. These transformations can be defined using the standard generators $T(x) = x + 1$, $S(x) = -1/x$ of the modular group $\mathrm{SL}(2, \mathbb{Z})$ and considering $f_{a,b} : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ given by

$$f_{a,b}(x) = \begin{cases} x + 1 & \text{if } x < a, \\ -\dfrac{1}{x} & \text{if } a \le x < b, \\ x - 1 & \text{if } x \ge b. \end{cases} \tag{1.1}$$

Under the assumption that the parameters (a, b) belong to the set

$$\mathcal{P} = \{(a, b) \mid a \le 0 \le b,\ b - a \ge 1,\ -ab \le 1\},$$

one can introduce corresponding continued fraction algorithms by using the first return map of $f_{a,b}$ to the interval $[a, b)$. Equivalently, these so-called (a, b)-*continued fractions* can be defined using a generalized integral part function:

$$\lfloor x \rceil_{a,b} = \begin{cases} \lfloor x - a \rfloor & \text{if } x < a, \\ 0 & \text{if } a \le x < b, \\ \lceil x - b \rceil & \text{if } x \ge b, \end{cases} \tag{1.2}$$

where $\lfloor x \rfloor$ denotes the integer part of x and $\lceil x \rceil = \lfloor x \rfloor + 1$.

A starting point of the theory is the following result [**16**, Theorem 2.1]: if $(a, b) \in \mathcal{P}$, then any irrational number x can be expressed uniquely as an infinite continued fraction of the form

$$x = n_0 - \cfrac{1}{n_1 - \cfrac{1}{n_2 - \cfrac{1}{\ddots}}} = \lfloor n_0, n_1, \ldots \rceil_{a,b} \quad (n_k \ne 0 \text{ for } k \ge 1),$$

where $n_0 = \lfloor x \rceil_{a,b}$, $x_1 = -1/(x - n_0)$ and $n_{k+1} = \lfloor x_{k+1} \rceil_{a,b}$, $x_{k+1} = -1/(x_k - n_k)$, i.e. the sequence of partial fractions $r_k = \lfloor n_0, n_1, \ldots, n_k \rceil_{a,b}$ converges to x.

It is possible to construct (a, b)-continued fraction expansions for rational numbers, too. However, such expansions will terminate after finitely many steps if $b \ne 0$. If $b = 0$, the expansions of rational numbers will end with a tail of 2s, since $0 = \lfloor 1, 2, 2, \ldots \rceil_{a,0}$.

The above family of continued fraction transformations contains three classical examples: the case $a = -1$, $b = 0$ described in [**12, 22**] gives the 'minus' (backward) continued fractions, the case $a = -1/2$, $b = 1/2$ gives the 'closest-integer' continued fractions considered first by Hurwitz in [**9**], and the case $a = -1$, $b = 1$ was presented in [**14, 19**] in connection with a method of coding symbolically the geodesic flow on the modular surface following Artin's pioneering work [**6**] and corresponds to the regular 'plus' continued fractions with alternating signs of the digits.

The main object of study in [**16**] is a two-dimensional realization of the *natural extension map* of $f_{a,b}$, $F_{a,b} : \bar{\mathbb{R}}^2 \backslash \Delta \to \bar{\mathbb{R}}^2 \backslash \Delta$, $\Delta = \{(x, y) \in \bar{\mathbb{R}}^2 \mid x = y\}$, defined by

$$F_{a,b}(x, y) = \begin{cases} (x + 1, y + 1) & \text{if } y < a, \\ \left(-\frac{1}{x}, -\frac{1}{y}\right) & \text{if } a \le y < b, \\ (x - 1, y - 1) & \text{if } y \ge b. \end{cases} \tag{1.3}$$

Here is the main result of that paper.

THEOREM 1.1. [**16**] *There exists an explicit uncountable set $\mathcal{E}$ of one-dimensional Lebesgue measure zero that lies on the diagonal boundary $b = a + 1$ of $\mathcal{P}$ such that:*

(1) *for all $(a, b) \in \mathcal{P} \backslash \mathcal{E}$ the map $F_{a,b}$ has an attractor $D_{a,b} = \bigcap_{n=0}^{\infty} F_{a,b}^n(\bar{\mathbb{R}}^2 \backslash \Delta)$ on which $F_{a,b}$ is essentially bijective;*

(2) *the set $D_{a,b}$ consists of two (or one, in degenerate cases) connected components each having* finite rectangular structure, *i.e. bounded by non-decreasing step-functions with a finite number of steps;*

(3) *almost every point (x, y) of the plane ($x \neq y$) is mapped to $D_{a,b}$ after finitely many iterations of $F_{a,b}$.*

Figure 1 shows the attractor $D_{a,b}$ in the case $a = -4/5$ and $b = 2/5$.

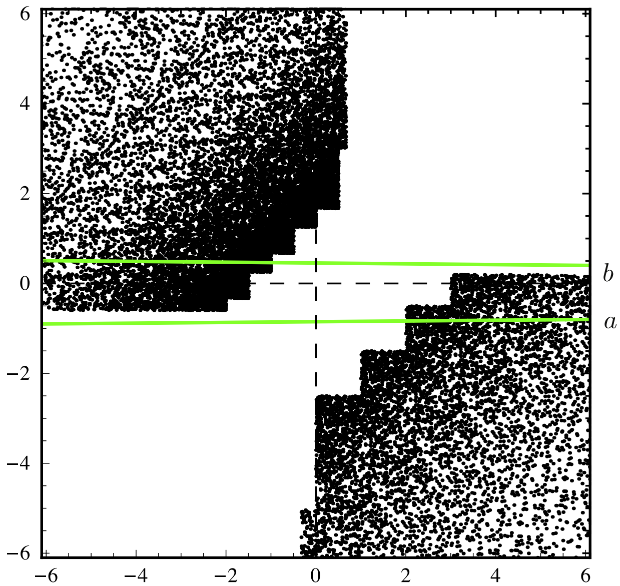

FIGURE 1. Attracting domain $D_{a,b}$ for $a = -\frac{4}{5}$, $b = \frac{2}{5}$.

An essential role in the argument is played by the forward orbits associated to a and b: to a, the *upper orbit* $\mathcal{O}_u(a)$ (i.e. the orbit of Sa) and the *lower orbit* $\mathcal{O}_\ell(a)$ (i.e. the orbit of Ta); and to b, the *upper orbit* $\mathcal{O}_u(b)$ (i.e. the orbit of $T^{-1}b$) and the *lower orbit* $\mathcal{O}_\ell(b)$ (i.e. the orbit of Sb). It was proved in [**16**] that if $(a, b) \in \mathcal{P} \backslash \mathcal{E}$, then $f_{a,b}$ satisfies the *finiteness condition*, i.e. for both a and b, their upper and lower orbits are either eventually periodic, or they satisfy the *cycle property*, i.e. they meet forming a cycle; more precisely, there exist $k_1, m_1, k_2, m_2 \geq 0$ such that

$$f_{a,b}^{m_1}(Sa) = f_{a,b}^{k_1}(Ta) = c_a \quad \text{(respectively, } f_{a,b}^{m_2}(T^{-1}b) = f_{a,b}^{k_2}(Sb) = c_b\text{)},$$

where c_a and c_b are the *ends of the cycles*. If the products of transformations over the upper and lower sides of the cycle are equal, the cycle property is *strong*; otherwise it is *weak*. In both cases the set $\mathcal{L}_{a,b}$ of the corresponding values is finite; ends of the cycles belong to the set $\mathcal{L}_{a,b}$ if and only if they are equal to 0, i.e. if the cycle is weak. The structure of the attractor $D_{a,b}$ is explicitly 'computed' from the finite set $\mathcal{L}_{a,b}$.

The paper is organized as follows. In §2 we give some background information about geodesic flows and their representations as special flows over symbolic dynamical systems, and define the coding map. In §3 we describe the reduction procedure for coding geodesics via (a, b)-continued fractions based on the study of the attractor of the associated natural extension map, define the corresponding cross-section set, and introduce the notion of *reduced geodesic*. In §4 we prove that the first return map to the cross-section corresponds to a shift of the coding sequence (Theorem 4.1) and, as a consequence, show that (a, b)-continued fractions satisfy the *tail property*, i.e. two $SL(2, \mathbb{Z})$-equivalent real numbers have the same tails in their (a, b)-continued fraction expansions. In §5 we introduce a notion of a *dual code* and show that if an (a, b)-expansion has a dual (a', b')-expansion, then the coding sequence of a reduced geodesic is obtained by juxtaposition of the (a, b)-expansion of its attracting endpoint w and the (a', b')-expansion of $1/u$, where u is its repelling endpoint. We also prove that if the (a, b)-expansion admits a dual, then the set of admissible coding sequences is a sofic shift (Theorem 5.8). In §6 we derive formulas for the density of the absolutely continuous invariant measure and the measure-theoretic entropy of the one-dimensional Gauss-type maps and their natural extensions. We also prove that the natural extension maps are Bernoulli shifts. Finally, in §7 we apply results of [**16**] to obtain explicit formulas for invariant measure for the one-dimensional maps for some regions of the parameter set $\mathcal{P}$.

2. *Geodesic flow on the modular surface and its representation as a special flow over a symbolic dynamical system*

Let $\mathcal{H} = \{z = x + iy : y > 0\}$ be the upper half-plane endowed with the hyperbolic metric, $\mathcal{F} = \{z \in \mathcal{H} : |z| \geq 1, |\text{Re } z| \leq 1/2\}$ be the standard fundamental region of the modular group $PSL(2, \mathbb{Z}) = SL(2, \mathbb{Z})/\{\pm I\}$, and $M = PSL(2, \mathbb{Z})\backslash\mathcal{H}$ be the modular surface. Let $S\mathcal{H}$ denote the unit tangent bundle of $\mathcal{H}$. We will use the coordinates $v = (z, \zeta)$ on $S\mathcal{H}$, where $z \in \mathcal{H}$, $\zeta \in \mathbb{C}$, $|\zeta| = \text{Im}(z)$. The quotient space $PSL(2, \mathbb{Z})\backslash S\mathcal{H}$ can be identified with the unit tangent bundle of M, SM, although the structure of the fibered bundle has singularities at the elliptic fixed points (see [**11**, §3.6] for details). Recall that geodesics in this model are half-circles or vertical half-rays. The geodesic flow $\{\tilde{\varphi}^t\}$ on $\mathcal{H}$ is defined as an $\mathbb{R}$-action on the unit tangent bundle $S\mathcal{H}$ which moves a tangent vector along the geodesic defined by this vector with unit speed. The geodesic flow $\{\tilde{\varphi}^t\}$ on $\mathcal{H}$ descends to the *geodesic flow* $\{\varphi^t\}$ on the factor M via the canonical projection

$$\pi : S\mathcal{H} \to SM \tag{2.1}$$

of the unit tangent bundles. Geodesics on M are orbits of the geodesic flow $\{\varphi^t\}$.

A *cross-section* C for the geodesic flow is a subset of the unit tangent bundle SM visited by (almost) every geodesic infinitely often both in the future and in the past. In other words, every $v \in C$ defines an oriented geodesic $\gamma(v)$ on M which will return to C infinitely often. The 'ceiling' function $g : C \to \mathbb{R}$ giving the *time of the first return* to C is defined as follows: if $v \in C$ and t is the time of the first return of $\gamma(v)$ to C, then $g(v) = t$. The map $R : C \to C$ defined by $R(v) = \varphi^{g(v)}(v)$ is called the *first return map*. Thus $\{\varphi^t\}$ can be represented as a *special flow* on the space

$$C^g = \{(v, s) : v \in C, 0 \leq s \leq g(v)\},$$

given by the formula $\varphi^t(v, s) = (v, s + t)$ with the identification $(v, g(v)) = (R(v), 0)$ (see Figure 2).

Let $\mathcal{N}$ be a finite or countable alphabet, $\mathcal{N}^{\mathbb{Z}} = \{x = \{n_i\}_{i\in\mathbb{Z}} \mid n_i \in \mathcal{N}\}$ be the space of all bi-infinite sequences endowed with the Tikhonov (product) topology,

$$\sigma : \mathcal{N}^{\mathbb{Z}} \to \mathcal{N}^{\mathbb{Z}} \text{ defined by } (\sigma x)_i = n_{i+1}$$

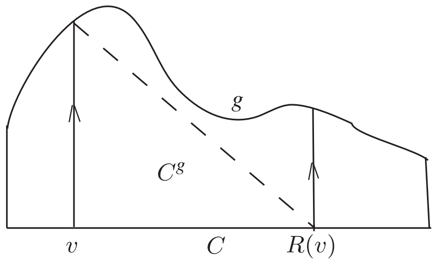

FIGURE 2. Geodesic flow is a special flow.

be the left shift map, and $\Lambda \subset \mathcal{N}^{\mathbb{Z}}$ be a closed σ-invariant subset. Then (Λ, σ) is called a *symbolic dynamical system*. There are some important classes of such dynamical systems. The space $(\mathcal{N}^{\mathbb{Z}}, \sigma)$ is called the *full shift* (or the *topological Bernoulli shift*). If the space Λ is given by a set of simple transition rules which can be described with the help of a matrix consisting of zeros and ones, we say that (Λ, σ) is a *one-step topological Markov chain* or simply a *topological Markov chain* (also called a *subshift of finite type*). A factor of a topological Markov chain is called a *sofic shift*. (See [**10**, §1.9] for the definitions.)

In order to represent the geodesic flow as a special flow over a symbolic dynamical system, one needs to choose an appropriate cross-section C and code it, i.e. to find an appropriate symbolic dynamical system (Λ, σ) and a continuous surjective map Cod : $\Lambda \to C$ (in some cases the actual domain of Cod is Λ except a finite or countable set of excluded sequences) defined such that the diagram

$$\begin{array}{ccc} \Lambda & \xrightarrow{\sigma} & \Lambda \\ {\scriptstyle \mathrm{Cod}}\downarrow & & \downarrow{\scriptstyle \mathrm{Cod}} \\ C & \xrightarrow{R} & C \end{array}$$

is commutative. We can then talk about *coding sequences* for geodesics defined up to a shift which corresponds to a return of the geodesic to the cross-section C. Notice that usually the coding map is not injective but only finite-to-one (see [**2**, §§3.2 and 5]).

There are two essentially different methods of coding geodesics on surfaces of constant negative curvature. The geometric code, with respect to a given fundamental region, is obtained by a construction universal for all Fuchsian groups. The second method is specific to the modular group and is of arithmetic nature: it uses continued fraction expansions of the end points of the geodesic at infinity and a so-called reduction theory (see [**14, 15**] for the three classical cases). Here we will describe a general method of arithmetic coding via (a, b)-continued fractions that is based on study of the attractor of the associated natural extension map. This approach, coupled with ideas of Bowen and Series [**7**], may be useful for coding of geodesics on quotients by general Fuchsian groups.

3. *The reduction procedure*

In what follows we will denote the end points of geodesics on $\mathcal{H}$ by u and w, and whenever we refer to such geodesics, we use (u, w) as their coordinates on $\bar{\mathbb{R}}^2$ $(u \neq w)$.

The reduction procedure for symbolically coding the geodesic flow on the modular surface via continued fraction expansions was presented for the three classical cases in [**14**]; for a survey on symbolic dynamics of the geodesic flow see also [**15**]. Here we describe the reduction procedure for (a, b)-continued fractions and explain how it can be used for coding purposes.

Let γ be an arbitrary geodesic on $\mathcal{H}$ from u to w (irrational end points), and $w = \lfloor n_0, n_1, \ldots \rceil_{a,b}$. We construct the sequence of real pairs $\{(u_k, w_k)\}$ $(k \geq 0)$ defined by

$$u_0 = u,\ w_0 = w \text{ and } w_{k+1} = ST^{-n_k} w_k, \quad u_{k+1} = ST^{-n_k} u_k. \tag{3.1}$$

Each geodesic γ_k from u_k to w_k is PSL$(2, \mathbb{Z})$-equivalent to γ by construction. It is convenient to describe this procedure using the *reduction map* that combines the appropriate iterate of the map $F_{a,b}$:

$$R_{a,b} : \mathbb{R}^2 \backslash \Delta \to \mathbb{R}^2 \backslash \Delta$$

given by the formula $R_{a,b}(u, w) = (ST^{-n}u, ST^{-n}w)$, where n is the first digit in the (a, b)-expansion of w. Notice that $(u_k, w_k) = R^k_{a,b}(u, w)$.

Definition 3.1. A geodesic in $\mathcal{H}$ from u to w is called (a, b)-*reduced* if $(u, w) \in \Lambda_{a,b}$, where

$$\Lambda_{a,b} = F_{a,b}(D_{a,b} \cap \{a \leq w \leq b\}) = S(D_{a,b} \cap \{a \leq w \leq b\}).$$

According to Theorem 1.1, for (almost) every geodesic γ from u to w in $\mathcal{H}$, the above algorithm produces in finitely many steps an (a, b)-reduced geodesic PSL$(2, \mathbb{Z})$-equivalent to γ, and an application of this algorithm to an (a, b)-reduced geodesic produces another (a, b)-reduced geodesic. In other words, there exists a positive integer ℓ such that $R^\ell_{a,b}(u, w) \in \Lambda_{a,b}$ and $R_{a,b} : \Lambda_{a,b} \to \Lambda_{a,b}$ is bijective (with the exception of some segments of the boundary of $\Lambda_{a,b}$ and their images).

Let γ be a reduced geodesic with the repelling point $u \neq 0$ and the attracting point

$$w = \lfloor n_0, n_1, \ldots \rceil_{a,b}. \tag{3.2}$$

Then, by successive applications of the map $R_{a,b}$, we obtain a sequence of real pairs $\{(u_k, w_k)\}$ $(k \geq 0)$ (see (3.1) above) such that each geodesic γ_k from u_k to w_k is (a, b)-reduced. Using the bijectivity of the map $R_{a,b}$, we extend the sequence (3.2) to the past to obtain a bi-infinite sequence of integers

$$\lfloor \gamma \rceil = \lfloor \ldots, n_{-2}, n_{-1}, n_0, n_1, n_2, \ldots \rceil, \tag{3.3}$$

called the *coding sequence* of γ, as follows. There exist an integer $n_{-1} \neq 0$ and a real pair $(u_{-1}, w_{-1}) \in \Lambda_{a,b}$ such that $ST^{-n_{-1}} w_{-1} = w = w_0$ and $ST^{-n_{-1}} u_{-1} = u = u_0$. Notice that $\lfloor w_{-1} \rceil_{a,b} = n_{-1}$. By uniqueness of the (a, b)-expansion, we conclude that $w_{-1} = \lfloor n_{-1}, n_0, n_1, \ldots \rceil_{a,b}$. Continuing inductively, we define the sequence of integers n_{-k} and the real pairs $(u_{-k}, w_{-k}) \in \Lambda_{a,b}$ $(k \geq 2)$, where

$$w_{-k} = \lfloor n_{-k}, n_{-k+1}, n_{-k+2}, \ldots \rceil_{a,b},$$

by $ST^{-n_{-k}} w_{-k} = w_{-(k-1)}$ and $ST^{-n_{-k}} u_{-k} = u_{-(k-1)}$. We also associate to γ a bi-infinite sequence of (a, b)-reduced geodesics

$$(\ldots, \gamma_{-2}, \gamma_{-1}, \gamma_0, \gamma_1, \gamma_2, \ldots), \tag{3.4}$$

where γ_k is the geodesic from u_k to w_k.

Remark 3.2. Notice that all 'intermediate' geodesics $T^{-s}\gamma_k$ $(1 \le s \le n_k)$ obtained from γ_k using the map $F_{a,b}$ are not (a, b)-reduced.

PROPOSITION 3.3. *A formal minus continued fraction consisting of the digits of the 'past' of (3.3),*

$$n_{-1} - \cfrac{1}{n_{-2} - \cfrac{1}{n_{-3} - \cfrac{1}{\ddots}}} = (n_{-1}, n_{-2}, n_{-3}, \ldots)$$

converges to $1/u$.

Proof. By [**13**, Lemma 1.1], it will be sufficient to check that $|n_{-k}| = 1$ implies $n_{-k} \cdot n_{-(k+1)} < 0$, i.e. the digit 1 must be followed by a negative integer and the digit -1 must be followed by a positive integer. We use the following properties of the set $\Lambda_{a,b}$ that can be derived from our knowledge of the shape of the set $D_{a,b}$ determined in [**16**, Lemmas 5.6, 5.10, 5.11]. The upper part of $\Lambda_{a,b}$ is contained in the region

$$\begin{array}{ll} [-1, 0] \times \left[-\dfrac{1}{a}, +\infty\right] \cup [0, 1] \times \left[-\dfrac{1}{b-1}, +\infty\right] & \text{if } b < 1, \\ [-1, 0] \times \left[-\dfrac{1}{a}, +\infty\right] & \text{if } b \ge 1. \end{array} \tag{3.5}$$

The lower part of $\Lambda_{a,b}$ is contained in the region

$$\begin{array}{ll} [-1, 0] \times \left[-\infty, -\dfrac{1}{a+1}\right] \cup [0, 1] \times \left[-\infty, -\dfrac{1}{b}\right] & \text{if } a > -1, \\ [0, 1] \times \left[-\infty, -\dfrac{1}{b}\right] & \text{if } a \le -1. \end{array} \tag{3.6}$$

Recall that $(u_{-(k+1)}, w_{-(k+1)}) = (T^{n_{-(k+1)}} S u_{-k}, T^{n_{-(k+1)}} S w_{-k})$ for an appropriate integer $n_{-(k+1)} \neq 0$. Suppose that $n_{-k} = 1$. Then $w_{-k} > 0$. If $u_{-k} < 0$, then $Su_{-k} > 0$ and $Sw_{-k} < 0$, and it takes a negative power of T to bring it back to (the lower component of) $\Lambda_{a,b}$, i.e. $n_{-(k+1)} < 0$. The case $u_{-k} > 0$, according to (3.5), can only occur if $b \le 1$. In this case, $-1/(b-1) \le w_{-k} < b + 1$, which is equivalent to $b > 1$, a contradiction. Therefore $n_{-k} = 1$ implies $n_{-(k+1)} < 0$. A similar argument shows that $n_{-k} = -1$ implies $n_{-(k+1)} > 0$. We conclude that the formal minus continued fraction converges. In order to prove that the limit is equal to $1/u$ we use the recursive definition of the digits $n_{-1}, n_{-2}, \ldots$, to write

$$\frac{1}{u} = n_{-1} - u_{-1} = n_{-1} - \frac{1}{n_{-2} - u_{-2}} = \cdots = (n_{-1}, n_{-2}, \ldots, n_{-k} - u_{-k}) = \cdots ,$$

and the conclusion follows since the formal minus continued fraction converges. □

Let

$$C = \{z \in \mathcal{H} \mid |z| = 1, -1 \le \operatorname{Re} z \le 1\}$$

be the upper-half of the unit circle, and

$$C^- = \{z \in \mathcal{H} \mid |z+1| = 1, -\tfrac{1}{2} \le \mathrm{Re}\, z \le 0\}$$

and

$$C^+ = \{z \in \mathcal{H} \mid |z-1| = 1, 0 \le \mathrm{Re}\, z \le \tfrac{1}{2}\}$$

be the images of the two vertical boundary components of the fundamental region $\mathcal{F}$ under S (see Figure 3).

PROPOSITION 3.4. *Every (a, b)-reduced geodesic intersects either C or both curves C^- and C^+.*

Proof. If a, b are such that $-1 \le a \le 0$ and $0 \le b \le 1$, then by properties (3.5) and (3.6) of the set $\Lambda_{a,b}$, if $(u, w) \in \Lambda_{a,b}$, then $-1 \le u \le 1$ and $w \ge -1/a$ or $w \le -1/b$, and hence all (a, b)-reduced geodesics intersect C. For the case $b > 1$ we have: if $-1 < u < 0$, then either $w > -1/a > b > 1$ or $w < -1/(a+1) < -1$, i.e. the geodesic intersects C; if $0 < u < 1$, then (3.5) implies that $w < -1/b < a < 0$, thus the corresponding geodesic intersects C if $w < -1$, and it intersects first C^+ and then C^- if $-1 < w < 0$. Similarly, for the case $a < -1$ we have: if $0 < u < 1$, then either $w < -1/b < a < -1$ or $w > -1/(b-1) > 1$, i.e. the geodesic intersects C; if $-1 < u < 0$, then (3.6) implies that $w > -1/a > b > 0$, therefore the corresponding geodesic intersects C if $w > 1$, and it intersects first C^- and then C^+ if $0 < w < 1$. □

Based on Proposition 3.4 we introduce the notion of the *cross-section point*. It is either the intersection of a reduced geodesic γ with C, or, if γ does not intersect C, its first intersection with $C^- \cup C^+$.

Now we can define a map

$$\varphi : \Lambda_{a,b} \to S\mathcal{H}, \quad \varphi(u, w) = (z, \zeta),$$

where $z \in \mathcal{H}$ is the cross-section point on the geodesic γ from u to w, and ζ is the unit vector tangent to γ at z. The map is clearly injective. Composed with the canonical projection π introduced in (2.1), we obtain a map

$$\pi \circ \varphi : \Lambda_{a,b} \to SM.$$

Let $C_{a,b} = \pi \circ \varphi(\Lambda_{a,b}) \subset SM$. This set can be described as follows: $C_{a,b} = P \cup Q_1 \cup Q_2$, where P consists of the unit vectors based on the circular boundary of the fundamental region $\mathcal{F}$ pointing inward such that the corresponding geodesic γ on the upper half-plane $\mathcal{H}$ is (a, b)-reduced, Q_1 consists of the unit vectors based on the right vertical boundary of $\mathcal{F}$ pointing inward such that either $S\gamma$ or $TS\gamma$ is (a, b)-reduced (notice that they cannot both be reduced), and Q_2 consists of the unit vectors based on the left vertical boundary of $\mathcal{F}$ pointing inward such that either $S\gamma$ or $T^{-1}S\gamma$ is (a, b)-reduced (see Figure 3). Then almost every orbit of $\{\varphi^t\}$ returns to $C_{a,b}$, i.e. $C_{a,b}$ is a *cross-section* for $\{\varphi^t\}$, and $\Lambda_{a,b}$ is a parametrization of $C_{a,b}$. The map $\pi \circ \varphi$ is injective, as follows from Remark 3.2: only one of the geodesics γ, $S\gamma$, $T^{-1}S\gamma$, and $TS\gamma$ can be reduced.

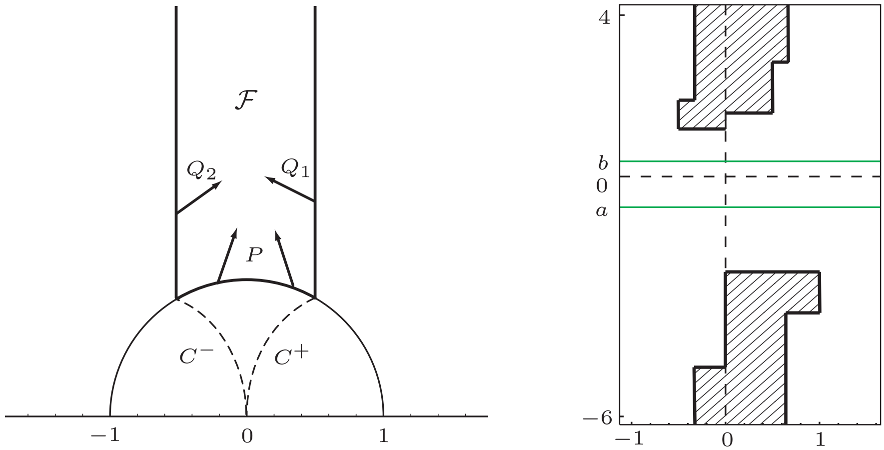

FIGURE 3. The cross-section (left) and its $\Lambda_{a,b}$ parametrization (right).

4. *Symbolic coding of the geodesic flow via (a, b)-continued fractions*

If γ is a geodesic on $\mathcal{H}$, we denote by $\bar{\gamma}$ the canonical projection of γ on M. For a given geodesic on M that can be reduced in finitely many steps, we can always choose its lift γ to $\mathcal{H}$ to be (a, b)-reduced.

The following theorem provides the basis for coding geodesics on the modular surface using (a, b)-coding sequences.

THEOREM 4.1. *Let γ be an (a, b)-reduced geodesic on $\mathcal{H}$ and $\bar{\gamma}$ its projection to M. Then:*

(1) *each geodesic segment of $\bar{\gamma}$ between successive returns to the cross-section $C_{a,b}$ produces an (a, b)-reduced geodesic on $\mathcal{H}$, and each reduced geodesic* $\mathrm{SL}(2, \mathbb{Z})$*-equivalent to γ is obtained in this way;*

(2) *the first return of $\bar{\gamma}$ to the cross-section $C_{a,b}$ corresponds to a left shift of the coding sequence of γ.*

Proof. (1) By lifting a geodesic segment on M starting on $C_{a,b}$ to $\mathcal{H}$, we obtain a segment of a geodesic γ on $\mathcal{H}$ that is reduced by the definition of the cross-section $C_{a,b}$. A coding sequence of $\gamma = \gamma_0$ that connects u_0 to $w_0 = \lfloor n_0, n_1, \ldots \rceil_{a,b}$,

$$\lfloor \gamma_0 \rceil = \lfloor \ldots, n_{-2}, n_{-1}, n_0, n_1, n_2, \ldots \rceil,$$

is obtained by extending the sequence of digits of w_0 to the past as explained in the previous section.

Let us assume that $w_0 > 0$, i.e. $n_0 \geq 1$. The case $w_0 < 0$ can be treated similarly. The geodesic $ST^{-n_0}\gamma_0 = \gamma_1$ is reduced by Theorem 1.1. Let z_0 and z_1 be the cross-section points on γ_0 and γ_1, respectively. Then $z_1' = T^{n_0}Sz_1 \in \gamma_0$; it is the intersection point of γ_0 with the circle $|z - n_0| = 1$. We will show that the geodesic segment of γ_0, $[z_0, z_1']$, projected to M is the segment between two successive returns to the cross-section $C_{a,b}$. Since $ST^{-n_0}(z_1') = z_1$ is the cross-section point on γ_1, the geodesic segment

$[z_0, z_1']$ projected to M is between two returns to $C_{a,b}$. Recall that a geodesic in $\mathcal{F}$ consists of countably many oriented geodesic segments between consecutive crossings of the boundary of $\mathcal{F}$ that are obtained by the canonical projection of γ_0 to $\mathcal{F}$.

If z_0 is the intersection of γ_0 with C, there are two possibilities: first, when γ_0 intersects $\mathcal{F}$ or γ_0 does not intersect $\mathcal{F}$ and $ST^{-1}\gamma_0$ exits $\mathcal{F}$ through its circular boundary; and second, when γ_0 does not intersect $\mathcal{F}$ and $ST^{-1}\gamma_0$ exits $\mathcal{F}$ through its left vertical boundary. In the first case the segments in $\mathcal{F}$ are represented by the intersection with $\mathcal{F}$ of the following geodesics in $\mathcal{H}$: $T^{-1}\gamma_0, T^{-2}\gamma_0, \dots, T^{-n_0+1}\gamma_0$, either $ST^{-n_0+1}\gamma_0$ or $T^{-n_0}\gamma_0$, and either γ_0 or $ST^{-1}\gamma_0$.

Suppose that for some intermediate point $z \in \gamma_0$, $z \in [z_0, z_1']$ the unit vector tangent to γ_0 at z, (z, ζ), is projected to $C_{a,b}$. By tracing the geodesic γ_0 inside $\mathcal{F}$, we see that (z, ζ) must be projected to $(\bar{z}, \bar{\zeta})$ with $\bar{z}$ on the boundary of $\mathcal{F}$ and $\bar{\zeta}$ directed inward. Then the geodesic through $(\bar{z}, \bar{\zeta})$:

(a) enters $\mathcal{F}$ through its vertical boundary and exits it also through the vertical boundary;
(b) enters $\mathcal{F}$ through its vertical boundary and exits through its circular boundary; or
(c) enters $\mathcal{F}$ through its circular boundary and exits through its vertical boundary.

The following assertions are implied by the analysis of the attractor $D_{a,b}$. In case (a), $T^{-1}ST^{-s}\gamma_0$ is not reduced for $1 \le s < n_0$ since $s < n_0$, $T^{-s}w_0 > b$, hence $ST^{-s}w_0 > -1/b$, i.e. $(ST^{-s}u_0, ST^{-s}w_0) \notin D_{a,b}$, therefore

$$(T^{-1}ST^{-s}u_0, T^{-1}ST^{-s}w_0) \notin \Lambda_{a,b}.$$

In case (b), if the segment $T^{-n_0}\gamma_0$ exits through the circular boundary of $\mathcal{F}$, $ST^{-n_0}\gamma_0 = \gamma_1$ is reduced, and we reached the point z_1 on the cross-section. If the segment $T^{-n_0+1}\gamma_0$ intersects the circular boundary of $\mathcal{F}$, $ST^{-n_0+1}\gamma_0$ is not reduced. In case (c), ST^{-n_0+1} is not reduced.

In the second case the first digit of w_0 is $n_0 = 2$. This is because $n_0 = 1$ would imply that $b + 1 < w < -1/(b - 1)$ which is impossible. Thus $ST^{-2}\gamma_0 = \gamma_1$ is reduced. In this case the geodesic in $\mathcal{F}$ consists of the intersection with $\mathcal{F}$ of a single geodesic $ST^{-1}\gamma_0$ that enters $\mathcal{F}$ through its right vertical and leaves it through its left vertical boundary, since $(TS)T(ST^{-1}\gamma_0) = ST^{-2}\gamma_0 = \gamma_1$ is reduced. In all cases the geodesic segment $[z_0, z_1']$ projected to M is between two consecutive returns to $C_{a,b}$.

If $z_0 \notin C$, by Proposition 3.4, since $w_0 > 0$, $z_0 \in C^-$. Notice that this implies that $a < -1$ and $n_0 = 1$, and $\gamma_1 = ST^{-1}\gamma_0$ is reduced. In this case the geodesic in $\mathcal{F}$ also consists of the intersection with $\mathcal{F}$ of a single geodesic $S\gamma_0$ that enters $\mathcal{F}$ through its right vertical and leaves it through its left vertical boundary, since $(TS)T(S\gamma_0) = ST^{-1}\gamma_0 = \gamma_1$ is reduced, and hence the geodesic segment $[z_0, z_1']$ projected to M is between two consecutive returns to $C_{a,b}$. Continuing this argument by induction in both the positive and negative direction, we obtain a bi-infinite sequence of points

$$(\dots, z_{-2}, z_{-1}, z_0, z_1, z_2, \dots),$$

where z_k is the cross-section point of the reduced geodesic γ_k in the sequence of γ_0, that represents the sequence of all successive returns of the geodesic γ_0 in M to the cross-section $C_{a,b}$.

If $\tilde{\gamma}_0$ is a reduced geodesic in $\mathcal{H}$, $SL(2, \mathbb{Z})$-equivalent to γ_0, then both project to the same geodesic on M. Therefore, the cross-section point $\tilde{z}_0$ of $\tilde{\gamma}_0$ projects on $C_{a,b}$ to a cross-section point z_k of γ_k for some k. This completes the proof of (1).

(2) Since $\gamma_1 = ST^{-n_0}\gamma_0$, $w_1 = ST^{-n_0}w_0 = \lfloor n_1, n_2, \ldots \rceil_{a,b}$. The first digit of the past is evidently n_0, and the remaining digits are the same as for γ_0. Thus (2) follows. □

The following corollary is immediate.

COROLLARY 4.2. *If γ' is* $SL(2, \mathbb{Z})$*-equivalent to γ, and both geodesics can be reduced in finitely many steps, then the coding sequences of γ and γ' differ by a shift.*

This implies a very important property of (a, b)-continued fractions that escapes a direct proof.

COROLLARY 4.3. (The tail property) *For almost every pair of real numbers that are* $SL(2, \mathbb{Z})$*-equivalent, the 'tails' of their (a, b)-continued fraction expansions coincide.*

Remark 4.4. The set of exceptions in Corollary 4.3 is the same as the set described in Theorem 1.1(3).

Thus we can talk about *coding sequences of geodesics on M*. To any geodesic γ that can be reduced in finitely many steps we associate the coding sequence (3.3) of a reduced geodesic $SL(2, \mathbb{Z})$-equivalent to it. Corollary 4.2 implies that this definition does not depend on the choice of a particular representative: sequences for equivalent reduced geodesics differ by a shift.

Let $X_{a,b}$ be the closure of the set of admissible sequences and σ be the left shift map. The coding map $\text{Cod} : X_{a,b} \to C_{a,b}$ is defined by

$$\text{Cod}(\lfloor \ldots, n_{-2}, n_{-1}, n_0, n_1, \ldots \rceil) = (1/(n_{-1}, n_{-2}, \ldots), \lfloor n_0, n_1, \ldots \rceil_{a,b}). \tag{4.1}$$

This map is essentially bijective.

The symbolic system $(X_{a,b}, \sigma) \subset (\mathcal{N}^{\mathbb{Z}}, \sigma)$ is defined on the infinite alphabet $\mathcal{N} \subset \mathbb{Z}\backslash\{0\}$. The product topology on $\mathcal{N}^{\mathbb{Z}}$ is induced by the distance function

$$d(x, x') = \frac{1}{m},$$

where $x = (n_i)$, $x' = (n'_i) \in \mathcal{N}^{\mathbb{Z}}$, and $m = \max\{k \mid n_i = n'_i \text{ for } |i| \le k\}$.

PROPOSITION 4.5. *The map* Cod *is continuous.*

Proof. If $d(x, x') < 1/m$, then the (a, b)-expansions of the attracting end points $w(x)$ and $w(x')$ of the corresponding geodesics given by (3.2) have the same first m digits. Hence the first m convergents of their (a, b)-expansions are the same, and, using the properties of (a, b)-continued fractions and the rate of convergence of [**16**, Theorem 2.1], we obtain $|w(x) - w(x')| < 2/m$. Similarly, the first m digits in the convergent formal minus continued fraction of $1/u(x)$ and $1/u(x')$ are the same, and hence $|u(x) - u(x')| < 2|u(x)u'(x)|/m < 2/m$. Therefore the geodesics are uniformly $2/m$-close. But the tangent vectors $v(x)$, $v(x') \in C_{a,b}$ are determined by the intersection of the corresponding geodesic with the unit circle or the curves C^+ and C^-. Hence, by making m large enough we can make $v(x')$ as close to $v(x)$ as we wish. □

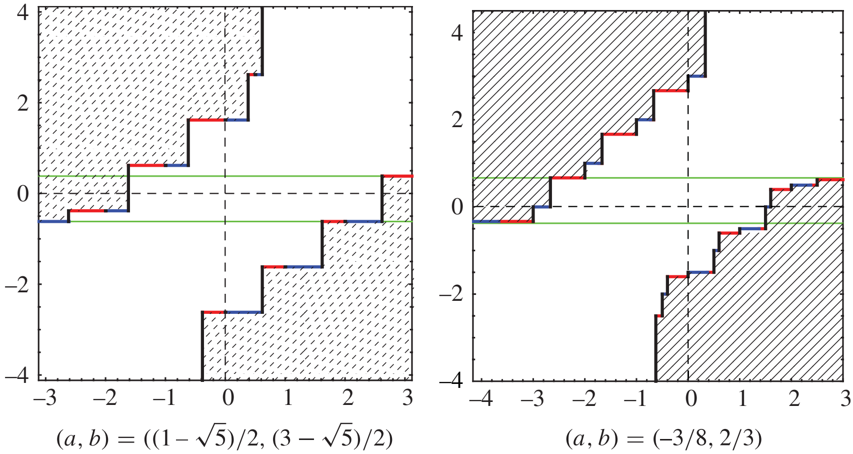

FIGURE 4. Domains of self-dual expansions.

In conclusion, the geodesic flow becomes a special flow over a symbolic dynamical system $(X_{a,b}, \sigma)$ on the infinite alphabet $\mathcal{N} \subset \mathbb{Z}\backslash\{0\}$. The ceiling function $g_{a,b}(x)$ on $X_{a,b}$ coincides with the time of the first return of the associated geodesic $\gamma(x)$ to the cross-section $C_{a,b}$. One can establish an explicit formula for $g_{a,b}(x)$ as the function of the end points of the corresponding geodesic $\gamma(x)$, $u(x)$, $w(x)$, following the ideas explained in [**8**]. If $-1 \le a \le 0$ and $0 \le b \le 1$, then $g_{a,b}(x)$ is cohomologous to $2 \log |w(x)|$; more precisely,

$$g_{a,b}(x) = 2 \log |w(x)| + \log h(x) - \log h(\sigma x)$$

where

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

5. *Dual codes*

We have seen that a coding sequence for a reduced geodesic from u to w (see (3.3)) is comprised of the sequence of digits in the (a, b)-expansion of w and the 'past', an infinite sequence of non-zero integers, each digit of which depends on w and u. In some special cases the 'past' only depends on u, and, in fact, will coincide with the sequence of digits of $1/u$ by using a so-called *dual expansion* to (a, b).

Let $\psi(x, y) = (-y, -x)$ be the reflection of the plane about the line $y = -x$.

Definition 5.1. If $\psi(D_{a,b})$ coincides with the attractor set $D_{a',b'}$ for some $(a', b') \in \mathcal{P}$, then the (a', b')-expansion is called the *dual* expansion to (a, b). If $(a', b') = (a, b)$, then the (a, b)-expansion is called *self-dual*.

Example 5.2. The classical situations of $(-1, 0)$- and $(-1, 1)$-expansions are self-dual. Two more sophisticated examples, $((1 - \sqrt{5})/2, (3 - \sqrt{5})/2)$ and $(-3/8, 2/3)$, are shown in Figure 4.

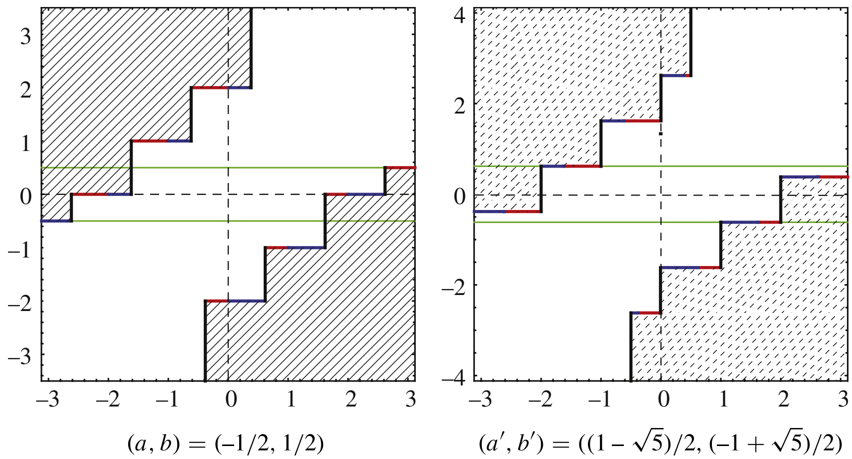

FIGURE 5. Dual expansions.

Example 5.3. The expansions $(-1/n, 1-(1/n))$, $n \geq 1$, satisfy a weak cycle property and have dual expansions that are periodic. A classical example in this series is the Hurwitz case $(-1/2, 1/2)$ whose dual is $((1-\sqrt{5})/2, (-1+\sqrt{5})/2)$ (see [**9, 14**]). Their domains are shown in Figure 5.

The following result gives equivalent characterizations for an expansion to admit a dual.

PROPOSITION 5.4. *The following are equivalent:*

(i) *the (a, b)-expansion has a dual;*

(ii) *the boundary of the lower part of the set $D_{a,b}$ does not have y-levels with $a < y < 0$, and the boundary of the upper part of the set $D_{a,b}$ does not have y-levels with $0 < y < b$;*

(iii) *a and b do not have the strong cycle property.*

Proof. If the (a, b)-expansion has a dual (a', b')-expansion, then the parameters a', b' are obtained from the boundary of $D_{a,b}$ as follows: the right vertical boundary of the upper part of $D_{a,b}$ is the ray $x = 1 - b'$, and the left vertical boundary of the lower part of $D_{a,b}$ is the ray $x = -1 - a'$. Now assume that (ii) does not hold. Then at least one of the parameters a, b has the strong cycle property, and either the left boundary of the upper part of $\Lambda_{a,b}$ or the right boundary of the lower part of $\Lambda_{a,b}$ is not a straight line. Assume the former. Then the reflection of $D_{a,b}$ with respect to the line $y = -x$ is not $D_{a',b'}$ since the map $F_{a',b'}$ is not bijective on it: the black rectangle in Figure 6 belongs to it, but its image under T^{-1}, colored in grey, does not. Thus (i) ⇒ (ii).

Conversely, let the vertical line $x = 1 - b'$ be the right boundary of the upper part of $D_{a,b}$ and the vertical line $x = -1 - a'$ be the left boundary of the lower part of $D_{a,b}$. Let $[x_a, \infty] \times \{a\}$ be the intersection of $D_{a,b}$ with the horizontal line at the level a, and $[-\infty, x_b] \times \{b\}$ be the intersection of $D_{a,b}$ with the horizontal line at the level b. Then $a' = 1/x_b$ and $b' = 1/x_a$. We also see that $1 - b' = -1/t$, where $t = x_b$ or $t < x_b$ if

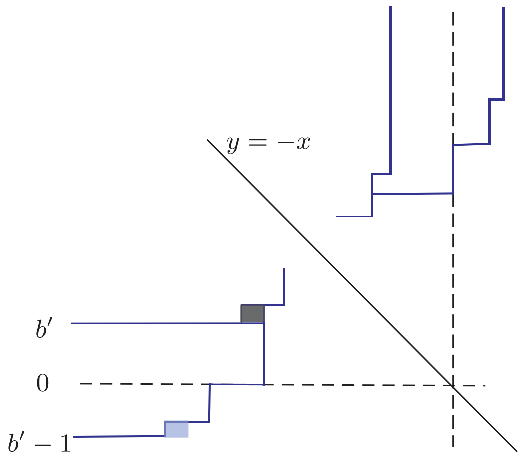

FIGURE 6. Dual expansions and $D_{a,b}$.

$[t, x_b] \times \{0\}$ is a segment of the boundary of $D_{a,b}$. Then $-b' + 1 = -1/t \le a'$, which implies that $b' - a' \ge 1$. By [**16**, Lemma 5.6], $x_b \le -1$ and $x_a \ge 1$, therefore

$$-1 \le a' \le 0 \le b' \le 1, \tag{5.1}$$

and

$$\Lambda_{a,b} = D_{a,b} \cap \{(u, w) \in \bar{\mathbb{R}}^2 : -b' \le u \le -a'\}. \tag{5.2}$$

We now show that $\psi(D_{a,b}) = D_{a',b'}$ is the attractor for $F_{a',b'}$, where

$$F_{a',b'} = \psi \circ F_{a,b}^{-1} \circ \psi^{-1}. \tag{5.3}$$

For $(u, w) \in D_{a',b'}$ with $a' < w < b'$, we have $\psi^{-1}(u, w) = (-w, -u)$ with $-b' < u < -a'$, so $\psi^{-1}(u, w) \in \Lambda_{a,b}$ by (5.2), hence $F_{a,b}^{-1}(-w, -u) = (1/w, 1/u)$, and $F_{a',b'}(u, w) = (-1/u, -1/w)$. For $(u, w) \in D_{a',b'}$ with $w > b'$, we have $\psi^{-1}(u, w) = (-w, -u)$ with $u < -b'$, so $F_{a,b}^{-1}(-w, -u) = (-w + 1, -u + 1)$, and $F_{a',b'}(u, w) = (u - 1, w - 1)$. Similarly, for $(u, w) \in D_{a',b'}$ with $w < a'$, we have $\psi^{-1}(u, w) = (-w, -u)$ with $u > -a'$, so $F_{a,b}^{-1}(-w, -u) = (-w - 1, -u - 1)$, and $F_{a',b'}(u, w) = (u + 1, w + 1)$. This proves that (ii) ⇒ (i).

Notice that (ii) and (iii) are equivalent by [**16**, Theorems 4.2 and 4.5]. □

Remark 5.5. Notice that if an (a, b)-expansion has a dual, then $-1 \le a \le 0 \le b \le 1$. This follows from (5.1) and the fact that the relation of duality is symmetric.

THEOREM 5.6. *If an (a, b)-expansion admits a dual expansion (a', b'), and γ_0 is an (a, b)-reduced geodesic, then its coding sequence,*

$$\lfloor \gamma_0 \rceil = \lfloor \ldots, n_{-2}, n_{-1}, n_0, n_1, n_2, \ldots \rceil, \tag{5.4}$$

is obtained by juxtaposing the (a, b)-expansion of $w_0 = \lfloor n_0, n_1, n_2, \ldots \rceil_{a,b}$ and the (a', b')-expansion of $1/u_0 = \lfloor n_{-1}, n_{-2}, \ldots \rceil_{a',b'}$. This property is preserved under the left shift of the sequence.

Proof. We will show that the digits in the (a', b')-expansion of $1/u_0$ coincide with the digits of the 'past' of (5.4). By (5.3), the diagram

$$\begin{array}{ccc} \Lambda_{a,b} & \xrightarrow{S\psi} & \Lambda_{a',b'} \\ {\scriptstyle R_{a,b}^{-1}}\big\downarrow & & \big\downarrow{\scriptstyle R_{a',b'}} \\ \Lambda_{a,b} & \xrightarrow{S\psi} & \Lambda_{a',b'} \end{array}$$

is commutative. The pair $(u_0, w_0) \in \Lambda_{a,b}$, therefore $(Su_0, Sw_0) \in S\Lambda_{a,b} \subset D_{a,b}$, and $(1/w_0, 1/u_0) \in \Lambda_{a',b'}$. The first digit of the (a', b')-expansion of $1/u_0$ is n_{-1}, so

$$R_{a',b'}(1/w_0, 1/u_0) = (ST^{-n_{-1}}(1/w_0), ST^{-n_{-1}}(1/u_0))$$

maps $\Lambda_{a',b'}$ to itself. Then

$$(u_{-1}, w_{-1}) := R_{a,b}^{-1}(u_0, w_0) = (T^{n_{-1}}Su_0, T^{n_{-1}}Sw_0) \in \Lambda_{a,b}$$

and

$$(ST^{-n_{-1}}u_{-1}, ST^{-n_{-1}}w_{-1}) = (u_0, w_0).$$

Also $w_{-1} = \lfloor n_{-1}, n_0, n_1, \ldots \rceil_{a,b}$ and $ST^{-n_{-1}}(1/u_0) = 1/u_{-1} = \lfloor n_{-2}, \ldots \rceil_{a',b'}$.

Continuing by induction, one proves that all digits of the 'past' of the sequence (5.4) are the digits of the (a', b')-expansion of $1/u_0$.

In order to see what happens under a left shift, we reverse the diagram to obtain

$$\begin{array}{ccc} \Lambda_{a,b} & \xrightarrow{S\psi} & \Lambda_{a',b'} \\ {\scriptstyle R_{a,b}}\big\downarrow & & \big\downarrow{\scriptstyle R_{a',b'}^{-1}} \\ \Lambda_{a,b} & \xrightarrow{S\psi} & \Lambda_{a',b'} \end{array}$$

Since the first digit of (a, b)-expansion of w_0 is n_0,

$$R_{a,b}(u_0, w_0) = (ST^{-n_0}u_0, ST^{-n_0}w_0)$$

maps $\Lambda_{a,b}$ to itself. Then $(u_1, w_1) := (ST^{-n_0}u_0, ST^{-n_0}w_0)$ and $w_1 = \lfloor n_1, n_2, \ldots \rceil_{a,b}$. Also

$$(1/w_1, 1/u_1) = R_{a',b'}^{-1}(1/w_0, 1/u_0) = (T_0^n S(1/w_0), T_0^n S(1/u_0)),$$

hence $1/u_1 = \lfloor n_0, n_{-1}, n_{-2}, \ldots \rceil_{a',b'}$. □

Remark 5.7. Under conditions of Theorem 5.6, if γ_0 projects to a closed geodesic on M, then its coding sequence is periodic, and $w_0 = \lfloor \overline{n_0, n_1, \ldots, n_m} \rceil_{a,b}$, $1/u_0 = \lfloor \overline{n_m, \ldots, n_1, n_0} \rceil_{a',b'}$.

THEOREM 5.8. *If an* (a, b)*-expansion admits a dual expansion, then the symbolic space* $(X_{a,b}, \sigma)$ *is a sofic shift.*

Proof. The 'natural' (topological) partition of the set $\Lambda_{a,b}$ related to the alphabet $\mathcal{N}$ is $\Lambda_{a,b} = \bigcup_{n\in\mathcal{N}} \Lambda_n$, where Λ_n are labeled by the symbols of the alphabet $\mathcal{N}$ and are defined by the condition $\Lambda_n = \{(u, w) \in \Lambda_{a,b} \mid n_0(u, w) = n_0(w) = n\}$. In order to prove that the

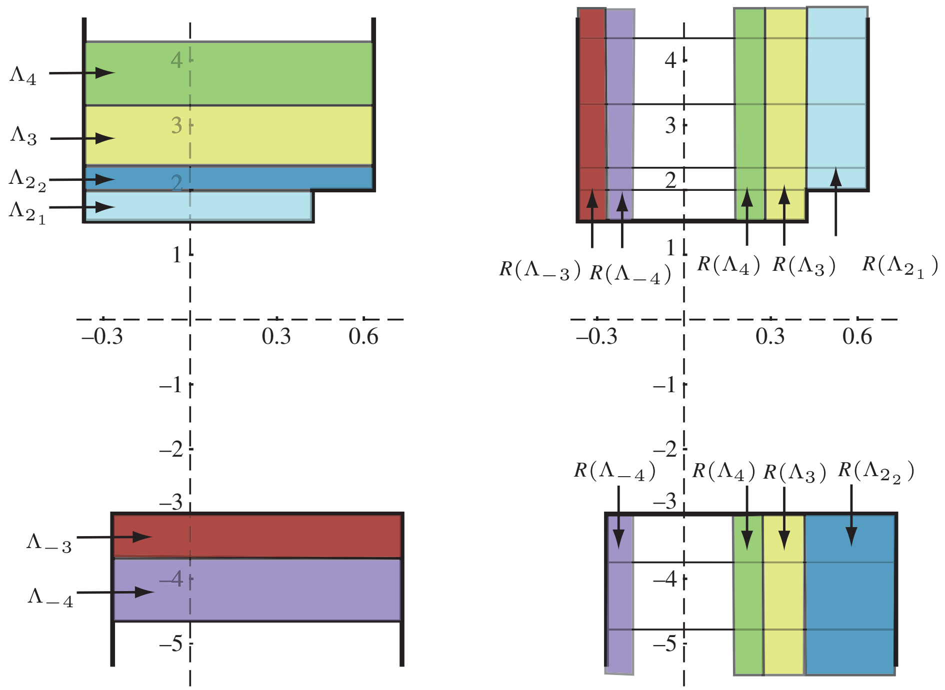

FIGURE 7. The partition of $\Lambda_{a,b}$ and its image through $R_{a,b}$.

space $(X_{a,b}, \sigma)$ is sofic one needs to find a topological Markov chain $(M_{a,b}, \tau)$ and a surjective continuous map $h : M_{a,b} \to X_{a,b}$ such that $h \circ \tau = \sigma \circ h$.

Notice that the elements Λ_n are rectangles for large n; in fact, at most two elements in the upper part and at most two elements in the lower part of $\Lambda_{a,b}$ are incomplete rectangles (see Figure 7).

Since $\Lambda_{a,b}$ has finite rectangular structure, we can subdivide these incomplete rectangles horizontally into rectangles, and extend the alphabet $\mathcal{N}$ by adding subscripts to the corresponding elements of $\mathcal{N}$. For example, if Λ_2 is subdivided into two rectangles, $\Lambda_2 = \bigcup_{i=1}^{2} \Lambda_{2_i}$, the 'digit' 2 will give rise to two digits, 2_1, 2_2 in the extended alphabet $\mathcal{N}'$ (see Figure 7). We denote the new partition of $\Lambda_{a,b}$ by $\bigcup_{n\in\mathcal{N}'} M_n$. Notice that it consists of rectangles with horizontal and vertical sides. Since the first return R to $\Lambda_{a,b}$ corresponds to the left shift of the coding sequence x associated to the geodesic (u, w), we see that $x = \{n_k\}_{-\infty}^{\infty}$, where n_k is defined by $R^k(u, w) \in \Lambda_{n_k}$. Now we define the symbolic space $M_{a,b}$ as follows: to each sequence $x \in X_{a,b}$ we associate a geodesic (u, w) by (4.1), and define a new coding sequence $y = \{m_k\}_{-\infty}^{\infty}$, where m_k is defined by $R^k(u, w) \in M_{m_k}$, and τ is the left shift.

We will prove that $(M_{a,b}, \tau)$ is a topological Markov chain. For this, in accordance to [**2**, Theorem 7.9], it is sufficient to prove that for any pair of distinct symbols $n, m \in \mathcal{N}'$, $R(M_n)$ and M_m either do not intersect or intersect 'transversally', i.e. their intersection is a rectangle with two horizontal sides belonging to the horizontal boundary of M_m

and two vertical sides belonging to the vertical boundary of $R(M_n)$. Let us recall that $-1 \le a \le 0 \le b \le 1$ (see Remark 5.5). Therefore, if $M_n = \Lambda_n$ is a complete rectangle, it is, in fact, a 1×1 square, and its image under R is an infinite vertical rectangle intersecting all M_m transversally. If M_n is obtained by subdivision of some Λ_k and belongs to the lower part of $\Lambda_{a,b}$, its horizontal boundaries are the levels of the step-function defining the lower component of $D_{a,b}$, and by Proposition 5.4, since the lower boundary of $D_{a,b}$ does not have y-levels with $a < y < 0$, its image is a vertical rectangle intersecting only the lower component of $D_{a,b}$ whose horizontal boundaries are the levels of the step-function defining the lower component of $D_{a,b}$. Therefore, all possible intersections with M_m are transversal. A similar argument applies to the case where M_n belongs to the upper part of $\Lambda_{a,b}$. The map $h : M_{a,b} \to X_{a,b}$ is obviously continuous, surjective, and, in addition, $h \circ \tau = \sigma \circ h$. □

6. *Invariant measures and ergodic properties*

Based on the finite rectangular geometric structure of the domain $D_{a,b}$ and the connections with the geodesic flow on the modular surface, we study some of the measure-theoretic properties of the Gauss-type map $\hat{f}_{a,b} : [a, b) \to [a, b)$,

$$\hat{f}_{a,b}(x) = -\frac{1}{x} - \left\lfloor -\frac{1}{x} \right\rceil_{a,b}, \quad \hat{f}_{a,b}(0) = 0. \tag{6.1}$$

Notice that the associated natural extension map $\hat{F}_{a,b}$,

$$\hat{F}_{a,b}(x, y) = \left(\hat{f}_{a,b}(x), -\frac{1}{y - \lfloor -1/x \rceil_{a,b}} \right), \tag{6.2}$$

is obtained from the map $F_{a,b}$ induced on the set $\Lambda_{a,b}$ by the change of coordinates

$$x = -1/w, \quad y = u \tag{6.3}$$

(or, equivalently, on the set $D_{a,b} \cap \{(u, w) \mid a \le w < b\}$ by the change of coordinates $x = w$, $y = -1/u$). Therefore the domain $\hat{\Lambda}_{a,b}$ of $\hat{F}_{a,b}$ is easily identified knowing $\Lambda_{a,b}$ and may be considered as its ‘compactification’.

Many of the measure-theoretic properties of $\hat{f}_{a,b}$ and $\hat{F}_{a,b}$ (existence of an absolutely continuous invariant measure, ergodicity) follow from the fact that the geodesic flow φ^t on the modular surface M can be represented as a special flow $(R_{a,b}, \Lambda_{a,b}, g_{a,b})$ on the space

$$\Lambda_{a,b}^{g_{a,b}} = \{(u, w, t) : (u, w) \in \Lambda_{a,b}, 0 \le t \le g_{a,b}(u, w)\}$$

(see §2). We recall that $R_{a,b} = F_{a,b}|_{\Lambda_{a,b}}$ and $g_{a,b}$ is the ceiling function (the time of the first return to the cross-section $C_{a,b}$) parametrized by $(u, w) \in \Lambda_{a,b}$.

We start with the fact that the geodesic flow $\{\varphi^t\}$ preserves the smooth (Liouville) measure $dm = du\, dw\, dt/(w - u)^2$ (see [**3**]), hence $R_{a,b}$ preserves the absolutely continuous measure $d\rho = du\, dw/(w - u)^2$. Using the change of coordinates (6.3), the map $\hat{F}_{a,b}$ preserves the absolutely continuous measure $d\nu = dx\, dy/(1 + xy)^2$.

The set $\Lambda_{a,b}$ has finite measure $d\rho$ if $a \ne 0$ and $b \ne 0$, since it is uniformly bounded away from the line $\Delta = \{(u, w) : u = w\} \subset \mathbb{R}^2$ (see relations (3.5) and (3.6)). In this situation, we can normalize the measure $d\rho$ to obtain the smooth probability measure

$$d\rho_{a,b} = \frac{d\rho}{K_{a,b}} = \frac{du\,dw}{K_{a,b}(w-u)^2} \tag{6.4}$$

where $K_{a,b} = \rho(\Lambda_{a,b})$. Similarly, if $a \neq 0$ and $b \neq 0$, the map $\hat{F}_{a,b}$ preserves the smooth probability measure

$$d\nu_{a,b} = \frac{dx\,dy}{K_{a,b}(1+xy)^2} \tag{6.5}$$

and $K_{a,b} = \rho(\Lambda_{a,b}) = \nu(\hat{\Lambda}_{a,b})$.

Returning to the Gauss-type map, $\hat{f}_{a,b}$, one can obtain explicitly a Lebesgue-equivalent invariant probability measure $\mu_{a,b}$ by projecting the measure $\nu_{a,b}$ onto the x-coordinate (pushforward); this is equivalent to integrating $\nu_{a,b}$ over $\hat{\Lambda}_{a,b}$ with respect to the y-coordinate as explained in [**4**].

We can immediately conclude that the systems $(\hat{F}_{a,b}, \nu_{a,b})$ and $(\hat{f}_{a,b}, \mu_{a,b})$ are ergodic from the fact that the geodesic flow $\{\varphi^t\}$ is ergodic with respect to dm. By using some well-known results about one-dimensional maps that are piecewise monotone and expanding, and the implications for their natural extension maps, we can establish stronger measure-theoretic properties: $(\hat{f}_{a,b}, \mu_{a,b})$ is exact and $(\hat{F}_{a,b}, \nu_{a,b})$ is a Bernoulli shift. Here we follow the presentation from [**23**] based on [**18, 21**].

THEOREM 6.1. *For any $a \neq 0$ and $b \neq 0$, the system $(\hat{f}_{a,b}, \mu_{a,b})$ is exact and its natural extension $(\hat{F}_{a,b}, \nu_{a,b})$ is a Bernoulli shift.*

Proof. Let us consider first the case $-1 < a < 0 < b < 1$. The interval (a, b) admits a countable partition $\xi = \{X_i\}_{i\in\mathbb{Z}\setminus\{0\}}$ of open intervals and the map $\hat{f}_{a,b}$ satisfies conditions (A), (F), (U) listed in [**23**]. Condition (A) is Adler's distortion estimate:

$$\hat{f}''_{a,b}/(\hat{f}'_{a,b})^2 \text{ is bounded on } X = \bigcup_{i\in\mathbb{Z}\setminus\{0\}} X_i, \tag{A}$$

condition (F) requires the finite image property of the partition ξ,

$$\hat{f}_{a,b}(\xi) = \{\hat{f}_{a,b}(X_i)\}_{i\in\mathbb{Z}\setminus\{0\}} \text{ is finite,} \tag{F}$$

while condition (U) is a uniformly expanding condition

$$|\hat{f}'_{a,b}| \geq \tau > 1 \text{ on } X. \tag{U}$$

Let $m \geq 0$ and $n \geq 0$ be such that $a - m - 1 \leq -1/b < a - m$ and $b + n \leq -1/a < b + n + 1$. Consider the open intervals

$$X_1 = \left(-\frac{1}{a-m-1}, b\right), \quad X_i = \left(-\frac{1}{a-m-i}, -\frac{1}{a-m-i+1}\right) \quad \text{for } i \geq 2$$

and

$$X_{-1} = \left(a, -\frac{1}{b+n+1}\right), \quad X_{-i} = \left(-\frac{1}{b+n+i-1}, -\frac{1}{b+n+i}\right) \quad \text{for } i \geq 2.$$

The map $\hat{f}_{a,b}$ satisfies conditions (A), (F), (U) with respect to the partition $\xi = \{X_i\}_{i\in\mathbb{Z}\setminus\{0\}}$. Indeed, $|\hat{f}''_{a,b}/(\hat{f}'_{a,b})^2| \leq 2$ on X, the collection of images $\hat{f}_{a,b}(\xi)$ consists of four sets $\hat{f}_{a,b}(X_1)$, $\hat{f}_{a,b}(X_{-1})$, $(b-1, b)$, $(a, a+1)$, and $|\hat{f}'_{a,b}| \geq \min\{1/a^2, 1/b^2\} > 1$ on X. Zweimüller [**23**] showed that any one-dimensional map for which conditions (A), (F), (U) hold is exact and satisfies Rychlik's conditions described in [**18**], hence its natural extension map is Bernoulli.

We now analyze the case $b \geq 1$. Let $K > 0$ be the smallest integer such that $b(a+1)^K < 1$. We will show that there exists $\gamma > 1$ such that, for every $x \in \bigcap_{i=0}^{K} \hat{f}_{a,b}^{-i}(X)$, some iterate $\hat{f}_{a,b}^n(x)$ with $n \leq K+1$ is expanding, i.e. $|(\hat{f}_{a,b}^n)'(x)| \geq \gamma$. (For the rest of the proof, we simplify the notation and let $\hat{f}$ denote the map $\hat{f}_{a,b}$.) Notice that if $x \in \bigcap_{i=0}^{n-1} \hat{f}^{-i}(X)$, then $\hat{f}^n$ is differentiable at x and

$$\frac{d}{dx}\hat{f}^n(x) = \frac{1}{(x\hat{f}(x)\cdots\hat{f}^{n-1}(x))^2}.$$

Assume that $ab > -1$. We look at the following cases:

(i) If $a < x < 0$, then $b-1 \leq \hat{f}(x) \leq b$, and $|x\hat{f}(x)| \leq |ab| < 1$.

(ii) If $0 < x < b$, then $a \leq \hat{f}(x) \leq a+1$. Let K be such that $b(a+1)^K < 1$. Then either there exists $1 \leq n \leq K$ such that $0 < \hat{f}^i(x) < a+1$ for $i = 1, 2, \ldots, n-1$ and $a < \hat{f}^n(x) < 0$, or $0 < \hat{f}^i(x) < a+1$ for $i = 1, 2, \ldots, K$. In the former case we have that

$$|x\hat{f}(x)\cdots\hat{f}^n(x)| \leq |ab(a+1)^{n-1}| < 1, \tag{6.6}$$

while in the latter case

$$|x\hat{f}(x)\cdots\hat{f}^K(x)| \leq |b(a+1)^K| < 1. \tag{6.7}$$

In the case $ab = -1$, let $\tau, \epsilon > 0$ be sufficiently small such that

$$b < -1/(a+\tau) < b+1 \quad \text{and} \quad a-1 < -1/(b-\epsilon) < a.$$

We again consider two cases:

(i) If $a < x < a+\tau$, then $b-1 < \hat{f}(x) < -1/(a+\tau)$, and $|x\hat{f}(x)| \leq |a/(a+\tau)| < 1$. If $a+\tau \leq x < 0$, then $|x\hat{f}(x)| \leq |b(a+\tau)| < 1$.

(ii) If $b-\epsilon < x < b$, then $0 < \hat{f}(x) < a+1$ and we have either (6.6) with $n \geq 2$ or (6.7). If $0 < x \leq b-\epsilon$, then we have (6.6) or (6.7) where b is replaced by $b-\epsilon$.

In conclusion, there exists a constant $\gamma > 1$ such that for every $x \in \bigcap_{i=0}^{K} \hat{f}_{a,b}^{-i}(X)$ some iterate $\hat{f}_{a,b}^n(x)$ with $n \leq K+1$ satisfies the condition $|(\hat{f}_{a,b}^n)'(x)| \geq \gamma$. This implies that the iterate $\hat{f}_{a,b}^N$, with $N = (K+1)!$, is uniformly expanding, i.e. it satisfies property (U). Since properties (A) and (F) are automatically satisfied by any iterate of $\hat{f}_{a,b}$ (see [**23**]), we have that $\hat{F}_{a,b}^N$ is Bernoulli. Using one of Ornstein's results [**17**, Theorem 4, p. 39], it follows that $\hat{F}_{a,b}$ is Bernoulli. □

The next result gives a formula for the measure-theoretic entropy of $(\hat{F}_{a,b}, \nu_{a,b})$.

THEOREM 6.2. *The measure-theoretic entropy of $(\hat{F}_{a,b}, \nu_{a,b})$ is given by*

$$h_{\nu_{a,b}}(\hat{F}_{a,b}) = \frac{1}{K_{a,b}}\frac{\pi^2}{3}. \tag{6.8}$$

Proof. To compute the entropy of this two-dimensional map, we use Abramov's formula [**1**],

$$h_{\tilde m}(\{\phi^t\}) = \frac{h_{\rho_{a,b}}(R_{a,b})}{\int_{\Lambda_{a,b}} g_{a,b}\, d\rho_{a,b}},$$

where $\tilde m$ is the normalized Liouville measure $d\tilde m = dm/m(SM)$. It is well known that $m(SM) = \pi^2/3$ (see [**3**]) and $h_{\tilde m}(\{\phi^t\}) = 1$ (see [**20**]). The measure $d\tilde m$ can be represented by the Ambrose–Kakutani theorem [**5**] as a smooth probability measure on the space $\Lambda_{a,b}^{g_{a,b}}$,

$$d\tilde m = \frac{d\rho_{a,b}\, dt}{\int_{\Lambda_{a,b}} g_{a,b}\, d\rho_{a,b}}, \tag{6.9}$$

where $d\rho_{a,b}$ is the probability measure on the cross-section $\Lambda_{a,b}$ given by (6.4). This implies that

$$d\tilde m = \frac{d\rho\, dt}{K_{a,b}\int_{\Lambda_{a,b}} g_{a,b}\, d\rho_{a,b}} = \frac{dm}{K_{a,b}\int_{\Lambda_{a,b}} g_{a,b}\, d\rho_{a,b}}.$$

Therefore $K_{a,b}\int_{\Lambda_{a,b}} g_{a,b}\, d\rho_{a,b} = m(SM) = \pi^2/3$ and

$$h_{\nu_{a,b}}(\hat F_{a,b}) = h_{\rho_{a,b}}(R_{a,b}) = \int_{\Lambda_{a,b}} g_{a,b}\, d\rho_{a,b} = \frac{1}{K_{a,b}}\frac{\pi^2}{3}.$$ □

Since $(\hat F_{a,b}, \nu_{a,b})$ is the natural extension of $(\hat f_{a,b}, \mu_{a,b})$, the measure-theoretic entropies of the two systems coincide, hence

$$h_{\mu_{a,b}}(\hat f_{a,b}) = \frac{1}{K_{a,b}}\frac{\pi^2}{3}\,. \tag{6.10}$$

As an immediate consequence of the above entropy formula we derive a growth rate relation for the denominators of the partial quotients p_n/q_n of (a, b)-continued fraction expansions, similar to the classical cases.

PROPOSITION 6.3. *Let $\{q_n(x)\}$ be the sequence of the denominators of the partial quotients p_n/q_n associated to the (a, b)-continued fraction expansion of $x \in [a, b)$. Then*

$$\lim_{n\to\infty}\frac{\log q_n(x)}{n} = \frac{1}{2}h_{\mu_{a,b}}(\hat f_{a,b}) = \frac{1}{K_{a,b}}\frac{\pi^2}{6} \quad \textit{for a.e. } x. \tag{6.11}$$

Proof. The proof is similar to the classical case: using Birkhoff's ergodic theorem, we have

$$\lim_{n\to\infty}\frac{\log q_n(x)}{n} = -\int_a^b \log|x|\, d\mu_{a,b}.$$

At the same time, Rokhlin's formula tells us that

$$h_{\mu_{a,b}}(\hat f_{a,b}) = \int_a^b \log|\hat f'_{a,b}|\, d\mu_{a,b} = -2\int_a^b \log|x|\, d\mu_{a,b},$$

hence the conclusion. □

7. *Some explicit formulas for the invariant measure $\mu_{a,b}$*

In order to obtain explicit formulas for $\mu_{a,b}$ and $h_{\mu_{a,b}}(\hat f_{a,b})$, one obviously needs an explicit description of the domain $D_{a,b}$. In [**16**] we describe an algorithmic approach for finding the boundaries of $D_{a,b}$ for all parameter pairs (a, b) outside of a negligible exceptional parameter set $\mathcal{E}$. Let us point out that the set $D_{a,b}$ may have an arbitrary large number of horizontal boundary segments. The qualitative structure of $D_{a,b}$ is given by the cycle properties of a and b. This structure remains unchanged for all pairs (a, b) having cycles with similar combinatorial complexity. For a large part of the parameter set the cycle descriptions are relatively simple (see [**16**, §4]) and we discuss them here.

In what follows, we focus our attention on the situation $-1 \le a \le 0 \le b \le 1$, and due to the symmetry of the parameter set with respect to the parameter line $a = -b$ we assume that $a \le -b$.

We treat the case $1 \le -1/a \le b + 1$ and $a \le -1/b + m \le a + 1$ (for some $m \ge 1$). The coordinates of the corners of the boundary segments in the upper region $D_{a,b} \cap \{(u, w) \mid u < 0,\ a \le w \le b\}$ are given by

$$(-2, b-1), \left(-\frac{3}{2}, T^{-2}S(b-1)\right), \ldots, \left(-\frac{m+1}{m}, (T^{-2}S)^{(m-1)}(b-1)\right),$$
$$\left(-1, -\frac{1}{a} - 1\right),$$

while the corners of the boundary segments in the lower region $D_{a,b} \cap \{(u, w) \mid u > 0,\ a \le w \le b\}$ are given by

$$\left(m, -\frac{1}{b} + m\right), (m+1, a+1).$$

Therefore the set $\hat{\Lambda}_{a,b}$ is given by

$$\begin{aligned}\hat{\Lambda}_{a,b} = &\bigcup_{p=1}^{m-1} [(T^{-2}S)^{p-1}(b-1), (T^{-2}S)^{p}(b-1)] \times \left[0, \frac{p}{p+1}\right] \\ &\cup \left[(T^{-2}S)^{m-1}(b-1), -\frac{1}{a} - 1\right] \times \left[0, \frac{m}{m+1}\right] \cup \left[-\frac{1}{a} - 1, b\right] \times [0, 1] \\ &\cup \left[a, -\frac{1}{b} + m\right] \times \left[-\frac{1}{m}, 0\right] \cup \left[-\frac{1}{b} + m, a+1\right] \times \left[-\frac{1}{m+1}, 0\right]. \qquad (7.1)\end{aligned}$$

Figure 8 shows a typical set $\hat{\Lambda}_{a,b}$ for this case.

THEOREM 7.1. *If $1 \le -1/a \le b + 1$ and $a \le -(1/b) + m \le a + 1$, then*

$$\mu_{a,b} = \frac{1}{K_{a,b}} h_{a,b}(x)\, dx,$$

where $K_{a,b} = \log[(m-a)(1+b)^{2-m}]$ and $h_{a,b}(x) = h^{+}_{a,b}(x) + h^{-}_{a,b}(x)$ with

$$h^{+}_{a,b}(x) = \begin{cases} \dfrac{1}{x + ((p+1)/p)} & \text{if } (T^{-2}S)^{p-1}(b-1) \le x < (T^{-2}S)^{p}(b-1), \\ & p = 1, \ldots, m-1, \\ \dfrac{1}{x + ((m+1)/m)} & \text{if } (T^{-2}S)^{m-1}(b-1) \le x < -\dfrac{1}{a} - 1, \\ \dfrac{1}{x+1} & \text{if } -\dfrac{1}{a} - 1 \le x < b \end{cases}$$

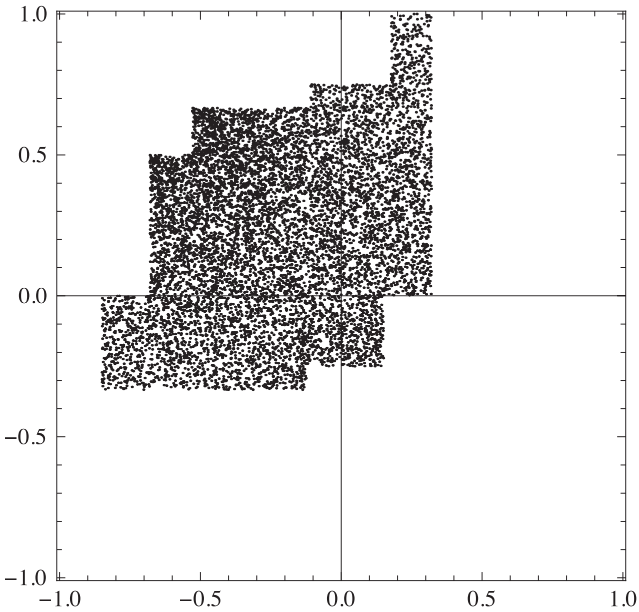

FIGURE 8. Typical domain $\hat{\Lambda}_{a,b}$ for the case studied.

and

$$h^-_{a,b}(x) = \begin{cases} \dfrac{1}{m-x} & \text{if } a \le x < -\dfrac{1}{b} + m, \\ \dfrac{1}{m+1-x} & \text{if } -\dfrac{1}{b} + m \le x < a+1 \,. \end{cases}$$

Proof. The density formulas are obtained from the simple integration result

$$\int_c^d \frac{1}{(1+xy)^2}\, dy = -\frac{1}{x}\left(\frac{1}{1+dx} - \frac{1}{1+cx}\right) = \frac{d}{1+dx} - \frac{c}{1+cx}. \tag{7.2}$$

For the density in the upper part of $\hat{\Lambda}_{a,b}$, $y \ge 0$, all integrals have the lower boundary $c = 0$, hence the result of (7.2) becomes $1/(x + 1/d)$. This gives the description of $h^+_{a,b}(x)$. For the density in the lower part of $\hat{\Lambda}_{a,b}$, $y \le 0$, all integrals have the upper boundary $d = 0$, hence the result $-1/(-1/c - x)$ and the description of $h^-_{a,b}(x)$. By a somewhat tedious computation, we get

$$K_{a,b} = \int_{\Lambda_{a,b}} h_{a,b}(x)\, dx = \log[(m-a)(1+b)^{2-m}],$$

and this completes the proof. □

Acknowledgement. The second author is partially supported by NSF grant DMS-0703421.

REFERENCES

[1] L. M. Abramov. On the entropy of a flow. *Sov. Math. Dokl.* **128**(5) (1959), 873–875.

[2] R. Adler. Symbolic dynamics and Markov partitions. *Bull. Amer. Math. Soc.* **35**(1) (1998), 1–56.

[3] R. Adler and L. Flatto. *Cross Section Maps for Geodesic Flows, I (The Modular Surface) (Progress in Mathematics, 21)*. Ed. A. Katok. Birkhäuser, Boston, MA, 1982, pp. 103–161.

[4] R. Adler and L. Flatto. Geodesic flows, interval maps, and symbolic dynamics. *Bull. Amer. Math. Soc.* **25**(2) (1991), 229–334.

[5] W. Ambrose and S. Kakutani. Structure and continuity of measurable flows. *Duke Math. J.* **9** (1942), 25–42.

[6] E. Artin. Ein mechanisches System mit quasiergodischen Bahnen. *Abh. Math. Semin. Univ. Hambg.* **3** (1924), 170–175.

[7] R. Bowen and C. Series. Markov maps associated with Fuchsian groups. *Publ. Math. Inst. Hautes Études Sci.* **50** (1979), 153–170.

[8] B. Gurevich and S. Katok. Arithmetic coding and entropy for the positive geodesic flow on the modular surface. *Moscow Math. J.* **1**(4) (2001), 569–582.

[9] A. Hurwitz. Über eine besondere Art der Kettenbruch-Entwicklung reeler Grossen. *Acta Math.* **12** (1889), 367–405.

[10] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, Cambridge, 1995.

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

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

[13] S. Katok and I. Ugarcovici. Geometrically Markov geodesics on the modular surface. *Moscow Math. J.* **5** (2005), 135–151.

[14] S. Katok and I. Ugarcovici. *Arithmetic Coding of Geodesics on the Modular Surface via Continued Fractions (CWI Tract 135)*. Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam, 2005, pp. 59–77.

[15] S. Katok and I. Ugarcovici. Symbolic dynamics for the modular surface and beyond. *Bull. Amer. Math. Soc.* **44** (2007), 87–132.

[16] S. Katok and I. Ugarcovici. Structure of attractors for (a, b)-continued fraction transformations. *J. Mod. Dyn.* **4** (2010), 637–691.

[17] D. Ornstein. *Ergodic Theory, Randomness, and Dynamical Systems*. Yale University Press, New Haven, CT, 1973.

[18] M. Rychlik. Bounded variation and invariant measures. *Studia Math.* **76** (1983), 69–80.

[19] C. Series. On coding geodesics with continued fractions. *Enseign. Math.* **29** (1980), 67–76.

[20] D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Math.* **153** (1984), 259–277.

[21] P. Walters. Invariant measures and equilibrium states for some mappings which expand distances. *Trans. Amer. Math. Soc.* **236** (1978), 121–153.

[22] D. Zagier. *Zetafunkionen und quadratische Körper: eine Einführung in die höhere Zahlentheorie*. Springer, Berlin–New York, 1982.

[23] R. Zweimüller. Ergodic structure and invariant densities of non-Markovian interval maps with indifferent fixed points. *Nonlinearity* **11** (1998), 1263–1276.