DISCRETE AND CONTINUOUS
DYNAMICAL SYSTEMS
Volume 3, Number 1, January 1997 pp. 107–116

A BILLIARD IN THE HYPERBOLIC PLANE WITH DECAY OF CORRELATION OF TYPE n^{-2}

M. Bauer

Department of Mathematics
Campus de Beaulieu, Université de Rennes I
35042 Rennes - Cedex, France

A. Lopes

Instituto de Matemática
Universidade Federal RGS Av. Bento Gonçalves 9500
91500 Porto Alegre - RS, Brazil

(Communicated by M. Shub)

Abstract. We consider billiard trajectories on the geodesic triangle D in the hyperbolic half-plane with internal angles 0, 0 and $\pi/2$ at the vertices ∞, 1 and i. The billiard map of D sends a given intersection point (together with the angle of incidence) of a given billiard trajectory γ with the boundary ∂D of D to the next intersection point of γ with ∂D.

By choosing an appropriate cross section for the billiard map, we show that the decay of correlation of the first return map is slower than n^{-2}. As a by-product, we enumerate the billiard trajectories in terms of their cutting sequences and relate the boundary expansion of the billiard map to continued fractions with even partial quotients.

0. Introduction. In thepresent paper we consider the billiard that corresponds to the Theta Group (see section 1). Adapting some ideas of D. Fried ([1]) we enumerate conjugacy classes of geodesics on the billiard (Theorem 2 in section 3). The billiard map is the first return map of the geodesic flow on the billiard for a certain cross section (section 3). Using standard geometric techniques we represent the billiard map by a function F in two variables (Theorem 2 and Corollary 1 of section 3), the so called "boundary expansion". The projection to its first component is denoted by T. Following Series ([4]) we find in section 4 the invariant measure m of T.

The map T and the measure m were studied in a paper by Kraaikamp and Lopes ([2]) about the Theta Group. The map T is related to the Continued Fraction expansion with even partial quotients in the same way that the Gauss map is related to the usual continued fraction expansion (See remark 1 of section 4). So the present paper presents a geometric proof (or interpretation) of some of the results in [2]. In particular the map F is a geometric way to define a natural extension of the continued fraction expansion in even partial quotients.

Our main result is theorem 3 in section 5 which states that the billiard map presents for a certain function (interpreted as a first return map in the remark following the theorem) a decay of correlation of order larger than n^{-2}.

1991 *Mathematics Subject Classification.* 58F11, 11J70, 54H20.

It is a pleasure for the first author to acknowledge the hospitality of IMPA during February and the Universidade Federal do Rio Grande do Sul, Brazil during Spring and Summer of 1994.

This situation is quite different from the modular case or the case of compact hyperbolic Riemann surfaces where the decay of correlation is of exponential type.

In [3] the decay of correlation of the Maneville-Pomeau map is obtained by means of Renewal Theory considerations. The map considered here is similar to the only case ($\gamma = 2$) to which the methods of [3] did not apply.

1. Definition of the billiard. We analyse the hyperbolic billiard that is determined by the triangle D which is contained in the hyperbolic upper half-plane $\mathbb{H}^2 = \{x + iy \in \mathbb{C} : y > 0\}$ and whose (geodesic) sides are $A_0 = \{iy : y \geq 1\}$, $B_0 = \{1 + iy : y > 0\}$ and $C_0 = \{x + iy : x^2 + y^2 = 1, 0 \leq x < 1\}$. D is illustrated by the shaded region in figure 1. The other parts of the figure will be explained in the remainder of this section.

Roughly speaking, we consider the unit speed motion of a point mass in D that will rebound when hitting the boundary $\partial D = A_0 \cup B_0 \cup C_0$ of D according to the law of "incoming angle equals outgoing angle" and whose trajectory is a hyperbolic geodesic segment between any two consecutive boundary hits. In the notation of figure 1, the arcs $\alpha_0, \alpha_1, \alpha_2$ and α_3 are geodesic segments (i.e. intersections with D of circles whose centers are on the real axis) that combine to a connected subset of a billiard trajectory.

Note that the vertices 1 and ∞ are at infinity (in the hyperbolic metric) and can not be reached by a point mass in finite time that starts in D. The interior angle at these two vertices is of course 0.

To be more precise, we define the group Γ of hyperbolic isometries generated by the reflections $a(z) = -\overline{z}$, $b(z) = 2 - \overline{z}$ and $c(z) = 1/\overline{z}$, about the sides A_0, B_0, and C_0, respectively. Γ acts on $\mathbb{H}^2$ with fundamental domain D. Let $\pi : \mathbb{H}^2 \to D$ be the projection that sends a point $z \in \mathbb{H}^2$ to the unique point in D equivalent to z by an element of Γ. For instance, the regions in figure 1 that are denoted $a(D)$, $c(D)$, $ca(D)$, etc., are mapped bijectively to D under the map a, c, ca, etc., respectively, where ac is the map a followed by c. So we read compositions of maps from left to right.

We identify the quotient space $Q = \mathbb{H}^2/\Gamma$ with D. A *billiard trajectory* is by definition the projection of a geodesic γ of $\mathbb{H}^2$ under the map π to D. In the notation of figure 1, the arcs $\alpha_0, \beta_1, \beta_2$ and β_3 combine to a geodesic segment and project to $\alpha_0, \alpha_1, \alpha_2$ and α_3, respectively.

We remark in passing that D has the structure of an orbifold for which a billiard trajectory is a geodesic. Also, the index 2 subgroup Γ^+ of Γ that consists of an even number of compositions of isometries is the Theta group that acts on $\mathbb{H}^2$ with fundamental domain $D \cup a(D)$.

We will always assume that geodesics of $\mathbb{H}^2$ and billiard trajectories are oriented. Moreover, we ignore billiard trajectories that hit the point i or that converge to infinity or the point 1.

For each point z of a billiard trajectory $\tilde{\gamma}$ that is not on the boundary of D there is a well defined unit vector tangent to $\tilde{\gamma}$ at z. At a point $z \in \tilde{\gamma} \cap \partial D$ we have two choices, and we define the *(unit) tangent vector* v of $\tilde{\gamma}$ at z to be the one that points out of D.

Note that $\Gamma(\partial D) = \{g(X) : g \in \Gamma, X \in \{A_0, B_0, C_0\}\}$ forms a tessellation of $\mathbb{H}^2$ that we denote by $\mathcal{T}$. We attach to each image of the side A_0 (and B_0, C_0) under an element of the group Γ the same label A (and B, C, respectively). See figure 2 where we drew part of the tessellation $\mathcal{T}$ between $x = -1$ and $x = 1$. As the map ab is horizontal translation by 2 the tessellation between $2n-1$ and $2n+1$ is a horizontal translate of the part between -1 and 1, for all $n \in \mathbb{Z}$.

We next define the *cutting sequence* $\cdots X_{-2}X_{-1}X_0X_1X_2\cdots$ of a billiard trajectory $\tilde{\gamma}$ with respect to a point $z \in \tilde{\gamma} \cap C_0$ as follows: $X_0 = C$ and $X_1, X_2, \cdots$ ($X_{-1}, X_{-2}, \cdots$, respectively) are the labels of the sides of the triangle D that we encounter in succession when moving along $\tilde{\gamma}$ in the positive (negative) direction starting from z. For instance, the connected segment $\alpha_0 \cup \alpha_1 \cup \alpha_2 \cup \alpha_3$ shown in figure 1 contributes $BCABC$ to the cutting sequence of the billiard trajectory that contains the segment.

Alternatively, suppose that γ is the (unique) geodesic of $\mathbb{H}^2$ that passes through z with the same unit tangent vector as $\tilde{\gamma}$. As $\pi(\gamma) = \tilde{\gamma}$, the cutting sequence of $\tilde{\gamma}$ with respect to z agrees with the cutting sequence of γ with respect to z and the labels of the tessellation $\mathcal{T}$. For example the cutting sequence of $\alpha_0 \cup \beta_1 \cup \beta_2 \cup \beta_3$ is $BCABC$, which is what we found in the end of the last paragraph.

Note that it can not happen that the cutting sequence of γ (and hence of $\tilde{\gamma}$) contains one of the strings BA or CB an infinite number of times in succession. So a billiard trajectory can not "bounce off" to infinity.

2. A normal form for conjugacy classes in Γ. We represent in this section a given conjugacy class in Γ by a word in normal form.

Note first that Γ has presentation $\Gamma = \langle a, b, c | a^2 = b^2 = c^2 = 1; (ac)^2 = 1\rangle$, where we read letters of a word (as compositions of maps) from left to right. The map $ac = ca$ is the rotation by π about the point i, π being twice the angle of incidence of the sides A_0 and C_0. Note also that the elements of finite order in Γ are given by a, b, c and ac.

We say that $w \in \Gamma$, represented as a word in a, b, c, is in *normal form* if it is shortest in its conjugacy class (i.e. among all the representation of w or a conjugate of w as a word in the letters a, b, c, we take one that has the fewest letters), if it is of infinite order, if its first letter is c and if it does not contain the string ca. Note in particular that w will not contain any of the strings aa, bb, cc, aca and cac.

Given $w \in \Gamma$ that is not of finite order and that is not conjugate to $(ab)^n$. We represent w by a word that is shortest in its conjugacy class and replace each occurrence of the string ca by ac to get a representation w' of w that will again be shortest in its conjugacy class.

After possibly cyclically permuting the letters of w', we get a word whose first letter is c. If by doing so we create a string ca, we replace it by ac. If the first letter is now a we conjugate the word by a to get a word in normal form that represents a conjugate of w.

We showed that we can represent the conjugacy class of w by a word in normal form.

Suppose now that w_0 is a word in normal form. We write $w_0 = r_1w_1$, where r_1 is the longest initial string of letters of w_0 that contains exactly one letter c. This initial letter c is followed by the letter b, and each letter a and b of w_0 can only be followed by the letter b and a, respectively. So by defining

$$f(n,e) = \begin{cases} c(ba)^n, & \text{if } e = 1; \\ c(ba)^{n-1}b, & \text{if } e = -1, \end{cases}$$

we see that r_1 is equal to $f(n_1, e_1)$, for some $n_1 \geq 1$ and $e_1 \in \{\pm 1\}$.

w_1 will now be either the empty word or a word whose first letter is c. We write $w_1 = r_2 w_2$ where r_2 is the longest initial substring of w_1 that contains exactly one c. Note that as w_0 is shortest in its conjugacy class this c can not be the only letter of w_1. We see as before that $r_2 = f(n_2, e_2)$, for some $n_2 \geq 1$ and $e_2 \in \{\pm 1\}$.

Theorem 1. *If $w \in \Gamma$ is not of finite order, then w is conjugate to either $(ba)^n$ or an element $w' \in \Gamma$ in normal form. In the latter case we have*

$$w' = f(n_1, e_1) \cdots f(n_k, e_k),$$

where $n_i \geq 1$ and $e_i \in \{\pm 1\}$, for $i = 1, \dots, k$. Moreover, this normal form representation of the conjugacy class of w is unique up to a cyclic permutation of the factors $f(n_i, e_i)$.

We only need to prove the uniqueness part which is done in the following

Lemma 1. *If two conjugate elements w and w' of Γ are in normal form and represented as a product of factors $f(n,e)$ as in theorem 1, then the two representations can only differ by a cyclic reordering of the factors.*

Proof. (The argument follows the lines of a proof of a lemma by Fried in [1].)

Step 1. By taking squares we see that it suffices to prove the lemma for elements in the subgroup Γ^+ of Γ that consists of words of even length.

Using the Reidemeister-Schreier method for presenting subgroups, one sees that Γ^+ has presentation $\langle x, y | x^2 = 1 \rangle$, where $x = ca$ and $y = ab$, hence is the free product of $\mathbb{Z}_2$ and $\mathbb{Z}$. It follows that each Γ^+ conjugacy class of an element in Γ^+ can be written as $w = xy^{n_1} \cdots xy^{n_k}$, where $n_i \in J = \mathbb{Z} \setminus \{0\}$, for $i = 1, \dots, k$. We say that $(n_1, \dots, n_k)$ is the J-sequence associated to w. Note that the J-sequence of a conjugacy class is unique up to cyclic permutation.

As we are interested in Γ conjugacy classes of elements in Γ^+, we need in addition to consider conjugation by the element a. To that effect we note that $axa = x^{-1} = x$ and $aya = y^{-1}$, hence $axy^n a = xy^{-n}$.

If follows that if $w, w' \in \Gamma^+$ are conjugate in Γ, then their J-sequences can only differ by a cyclic rearrangement of their factors or an overall sign.

Step 2. Suppose that $w' \in \Gamma^+$ is as in theorem 1. Using the relation $cb = xy$, one shows that $f(n,1) = xy^n a$ and $f(n,-1) = xy^n$. So $w' = f(n_1, e_1) \cdots f(n_k, e_k)$ can be written as $xy^{n_1} a^{\epsilon_1} \cdots xy^{n_k} a^{\epsilon_k}$, where $n_i \geq 1$ and $\epsilon_i \in \{0, 1\}$, for $i = 1, \dots, k$. Using again the relation $axy^n a = xy^{-n}$ and the fact that $\epsilon_1 + \cdots + \epsilon_k$ is even, the J-sequence of w is of the form $(n_1, \pm n_2, \dots, \pm n_k)$, where the signs of the n_i are a function of the ϵ_j. Note that if we cyclically permute the factors of w' then we get the J-sequence of the permuted word by applying the same permutation to the factors of the J-sequence of w' plus a possible overall change of sign.

Step 3. Given now two elements w and w' in Γ^+ that are conjugate in Γ and in normal form. We write them as a product of factors $f(n,e)$ as in theorem 1. Construct the J-sequence S and S' of w and w', respectively, as described in Step 2. S and S' agree up to an eventual cyclic reordering and overall sign.

By cyclically reordering the factors of w', we can assume that S and S' agree up to maybe an overall sign. But the first element in the J-sequence is always positive, so $S = S'$. It follows that w and the (cyclically permuted) w' agree. □

3. The billiard map. We fix for this section a billiard trajectory $\tilde{\gamma}^0$ and a point $z_0 \in \tilde{\gamma}^0 \cap C_0$.

Suppose that $\cdots, z_{-1}, z_0, z_1, \cdots$ are the successive points where $\tilde{\gamma}^0$ hits the side C_0. The vector tangent to $\tilde{\gamma}$ at z_i is denoted by v_i, for $i \in \mathbb{Z}$.

Of course the pair (z_i, v_i) determines $\tilde{\gamma}^0$, hence in particular (z_{i+1}, v_{i+1}), for each $i \in \mathbb{Z}$. The application that maps (z_i, v_i) to (z_{i+1}, v_{i+1}), for each $i \in \mathbb{Z}$ is called the *billiard map* of the billiard D with respect to the side C_0.

The goal of this section is to describe the billiard map in some convenient representation using only the cutting sequence of $\tilde{\gamma}^0$. We proceed in two steps, the second being a modification of the first.

Step 1. Each (z_i, v_i) determines a unique oriented geodesic γ^i of $\mathbb{H}^2$. Suppose temporarily that $\alpha^i(t)$, $t \in \mathbb{R}$, is a unit speed (with respect to the hyperbolic metric) parametrisation of γ^i that respects the orientation of γ^i. Then $\gamma^i_+ = \lim_{t\to\infty} \alpha^i(t)$ and $\gamma^i_- = \lim_{t\to-\infty} \alpha^i(t)$ are points "at infinity", i.e. points on the boundary $\{(x,0) : x \in \mathbb{R}\} \cup \{\infty\}$ of the hyperbolic half-plane. We refer to γ^i_+ and γ^i_- as the positive and negative, respectively, endpoint of γ^i.

The following facts are crucial, well known and easy to verify (see [5] for example):

We fix i. Suppose that α is the connected segment of $\tilde{\gamma}^0$ that has initial point z_i and endpoint z_{i+1}. Then α is the union of smooth arcs $\alpha_1, \dots, \alpha_n$, for some n, where the endpoint of α_j is contained in a side of the triangle D, say with label X_j, and agrees with the initial point of α_{j+1}, for $j = 1, \dots, n-1$. Moreover, the initial point of α_1 and the endpoint of α_n are in C_0.

This means that the cutting sequence l of $\tilde{\gamma}^0$ with respect to z_i is of the form $\cdots \underline{C} X_1 \cdots X_{n-1} C \cdots$. By underlining a letter of a cutting sequence we indicate where we start to read it.

Let β be the initial segment of γ^i that starts at z_i and cuts through exactly n consecutive regions $D_1, \dots, D_n$, each of which is a copy of D by an element of Γ. Each D_j cuts out a segment β_j from β that projects to α_j, for $j = 1, \dots, n$.

Note that the endpoint of β_j is in a side of the tessellation $\mathcal{T}$ with label X_j, for $j = 1, \dots, n-1$, and β_n has its endpoints in a side with label C.

We use for the rest of this section the following convention: If the capital letter X denotes a label A, B or C, then x denotes the map a, b or c, respectively.

One shows inductively that $D_j = x_{j-1} \cdots x_1 c(D)$, and $\beta_j = x_{j-1} \cdots x_1 c(\alpha_j)$, hence $\alpha_j = c x_1 \cdots x_{j-1}(\beta_j)$, for $j = 1, \dots, n$.

As γ^{i+1} is a continuation of α_n, it follows that $\gamma^{i+1} = c x_1 \cdots x_{n-1}(\gamma^i)$, hence in particular $\gamma^{i+1}_\pm = c x_1 \cdots x_{n-1}(\gamma^i_\pm)$. Note that the cutting sequence of γ^{i+1} with respect to z^{i+1} is $\cdots C X_1 \cdots X_{n-1} \underline{C} \cdots$. This means that we get the cutting sequence of γ^{i+1} (or $\tilde{\gamma}^0$) with respect to z_{i+1} by underlining the first C that occurs after the C that is underlined in the cutting sequence of γ^i (or $\tilde{\gamma}^0$) with respect to z_i.

Step 2. This is almost what we want, but not quite. It is clear that the cutting sequence l of $\tilde{\gamma}^0$ does not contain any of the strings AA, BB, CC, ACA, or CAC. We already remarked that l does not contain the string BA an infinite number of times in succession. l might however contain the string CA.

We therefore define the *modified cutting sequence* l_m of the cutting sequence l to be the sequence we get from l by replacing each occurrence of CA by AC. If by doing that we replace the string CA in l, with C underlined, by AC, then we continue to underline the same C in l_m.

It follows by an argument similar to the one we used in section 2 to characterize words in normal form that $l_m = \cdots F(n_{-1}, e_{-1})F(n_0, e_0)F(n_1, e_1)\cdots$, where

$$F(n,e) = \begin{cases} C(BA)^n, & \text{if } e = 1; \\ C(BA)^{n-1}B, & \text{if } e = -1. \end{cases}$$

Recall the sequences $(z_i, v_i)_{i=-\infty}^{i=\infty}$ and $(\gamma^i)_{i=-\infty}^{i=\infty}$ from step 1. The cutting sequence l^i of γ^i with respect to z_i is either of the form $l^i = \cdots \underline{C}B \cdots$ or $l^i = \cdots \underline{C}A \cdots$. In the first case we define $\delta^i = \gamma^i$ and take the cutting sequence of δ^i with respect to z_i. In the second case we define $\delta^i = a(\gamma^i)$ and take the cutting sequence of δ^i with respect to $a(z_i)$. Note that the cutting sequences of δ^i and γ^i agree in either case.

As a geodesic in $\mathbb{H}^2$ can not cut the same line of the tessellation $\mathcal{T}$ twice, we can convince ourselves by a quick consultation of figure 2 that $\delta^i_+ \in I = (0,1)$ and $\delta^i_- \in J = (-\infty, -1) \cup (1, \infty)$. Indeed, if l^i contains $\underline{C}B$, then $\delta^i = \gamma^i$ crosses first C_0, and then the side of $c(D)$ whose label is B. So $\delta^i_+ \in I$ and $\delta^i_- \in J$. If however l^i contains $\underline{C}A$, then $\delta^i = a(\gamma^i)$ crosses $a(C_0)$ and then the side of $ca(D)$ whose label is A. So $\delta^i_+ \in I$ and $\delta^i_- \in (-\infty, -1) \subset J$.

Using the notation of steps 1 and 2 above, we determine in the next theorem the billiard map acting on pairs (δ^i_+, δ^i_-).

Theorem 2. *Let l be the cutting sequence of the billiard trajectory $\tilde{\gamma}^0$ with respect to the point $z_0 \in C_0$.*

(1) *Then the modified cutting sequence l_m of l is of the form*

$$l_m = \cdots F(n_{-1}, e_{-1})F(n_0, e_0)F(n_1, e_1)\cdots,$$

where $n_i \geq 1$ and $e_i \in \{\pm 1\}$, for $i \in \mathbb{Z}$.

(2) *We assume the notation in part 1 to be chosen such that the initial letter C of $F(n_0, e_0)$ is underlined (i.e. is the label that corresponds to z_0). The billiard map is then given by*

$$\delta^{i+1}_\pm = f(n_i, e_i)(\delta^i_\pm), \quad \text{for } i \in \mathbb{Z}.$$

Moreover, the cutting sequence l^k of $\tilde{\gamma}^0$ with respect to z_k equals l, except that the kth C that follows the underlined C of l is now underlined, for $k \in \mathbb{N}$. Similarly, the modified cutting sequence l^k_m of l^k equals l_m except that the initial C of $F(n_k, e_k)$ is now underlined.

Proof. We only need to prove the second part of the theorem. To that effect we start by noting that $l^0 = l$ will be of the form $l = \cdots \underline{C}A^{\epsilon_1}(BA)^n BA^{\epsilon_2}CA^{\epsilon_3}\cdots$, where $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_2 + \epsilon_3 \in \{0,1\}$. It follows from step 1 that the cutting sequence l^1 of γ^1 with respect to z_1 is l^0 except that we underline the first letter C that follows the one underlined in l^0. Moreover, $\gamma^1 = ca^{\epsilon_1}(ba)^n ba^{\epsilon_2}(\gamma^0)$. Note that $\delta^0 = a^{\epsilon_1}(\gamma^0)$ and $\delta^1 = a^{\epsilon_3}(\gamma^1)$. It follows that $\delta^1 = c(ba)^n ba^{\epsilon_2+\epsilon_3}(\delta^0)$. (Recall that we read maps from left to right and that $aca = c$.)

Now $l^0_m = \cdots A^{\epsilon_1}\underline{C}(BA)^n BA^{\epsilon_2+\epsilon_3}C\cdots$, hence $F(n_0, e_0) = C(BA)^n BA^{\epsilon_2+\epsilon_3}$. But then $f(n_0, e_0) = c(ba)^n ba^{\epsilon_2+\epsilon_3}$ which proves the claim in case $i = 0$.

Note that the modified cutting sequence l^1_m of l^1 is the same as l^0_m except that we underline the first C that follows the underlined C of l^0_m. An obvious induction argument proves the claim. □

We remark in passing that the first part of theorem 2 can be used to enumerate billiard trajectories. The only restriction on the modified cutting sequence of a billiard trajectory is that it can not contain an infinite number of times the string $F(1,-1) = CB$ in succession. To construct a geodesic with a given modified cutting sequence one can use the second part of remark 1 is section 4 below.

For $n \geq 1$ and $e \in \{\pm 1\}$, we compute

$$f(n,e)(x) = e\left[\frac{1}{x} - 2n\right],$$

where $x \in \mathbb{R} \setminus \{0\}$, and define

$$B(n,e) = \begin{cases} \left(\frac{1}{2n+1}, \frac{1}{2n}\right), & \text{if } e = 1; \\ \left(\frac{1}{2n}, \frac{1}{2n-1}\right), & \text{if } e = -1. \end{cases}$$

Recall that $\delta_+^i \in I = (0,1)$ and $\delta_-^i \in J = (-\infty,-1) \cup (1,\infty)$. One readily verifies that $f(n,e)$ maps $B(n,e)$ bijectively onto $(0,1)$, for $e = \pm 1$ and $n \in \mathbb{N}$. Moreover, $f(n,e)$ maps J onto $(-2n-1,-2n) \cup (-2n,-2n+1)$, if $e = 1$, and it maps J onto $(2n-1,2n) \cup (2n,2n+1)$, if $e = -1$. Thus, we proved

Corollary 1. *The billiard map, acting on (δ_+^i, δ_-^i) as described in Theorem 2 is given by*

$$\tilde{F} : \begin{array}{ccc} I \times J & \to & I \times J \\ (x,y) & \mapsto & (f(n,e)(x), f(n,e)(y)), \end{array}$$

if $x \in B(n,e)$, for $n \geq 1$ and $e \in \{\pm 1\}$. $\tilde{F}$ is a bijection between the subset of $I \times J$ where it is defined and its image.

4. The invariant measure of the billiard map. It is convenient to use the change of coordinates $h : J \to (-1,1)$ defined by $h(y) = -1/y$. For $n \in \mathbb{N}$ and $e \in \{\pm 1\}$, we define $g(n,e) = h \circ f(n,e) \circ h$ and compute

$$g(n,e)(y) = \frac{e}{(y+2n)}.$$

The billiard map then becomes

$$F : \begin{array}{ccc} (0,1) \times (-1,1) & \to & (0,1) \times (-1,1) \\ (x,y) & \mapsto & (f(n,e)(x), g(n,e)(y)), \end{array}$$

if $x \in B(n,e)$, for $n \geq 1$ and $e \in \{\pm 1\}$.

Denote by $T : (0,1) \to (0,1)$ the projection and restriction of F (and $\tilde{F}$) to its first factor, namely

$$T(x) = f(n,e)(x), \text{ if } x \in B(n,e),$$

for $e \in \{\pm 1\}$ and $n \in \mathbb{N}$. It will be convenient to add the right endpoint to the intervals $(0,1)$ and $B(n,e)$, for $n \geq 1$ and $e \in \{\pm 1\}$. See figure 3 for part of the graph of T. The map F that we defined in an essentially geometric way (up to the renormalization) agrees with the natural extension of the map T considered in [2].

We next determine the invariant measure of the maps F and T following an argument of Series in [4].

The invariant measure of the geodesic flow induces an invariant measure for the first return map of a given cross section. By representing a geodesic by its endpoints (x,y), this measure becomes $\frac{1}{(x-y)^2}dxdy$, defined in our case on $I \times J$.

By applying the change of coordinates h to the variable y we find $\frac{1}{(1+xy)^2}dxdy$ as the invariant measure for F.

The invariant measure m for T on $(0,1)$ can then be found by integrating the invariant measure of F over the interval $(-1,1)$. We find that m is given by

$$dm = \left(\frac{1}{1-x} + \frac{1}{1+x}\right) dx.$$

Remark 1. (1) Suppose that we are in the situation of Theorem 2 of section 3. The formula for the billiard map given there implies that

$$\delta_+^i = \frac{1}{2n_i + e\delta_+^{i+1}},$$

for $i \geq 0$. We conclude that

$$\delta_+^0 = \cfrac{1}{2n_0 + \cfrac{e_0}{2n_1 + \cfrac{e_1}{2n_2 + \cfrac{e_2}{2n_3 + \ddots}}}}.$$

But this is just the continued fraction expansion with even partial quotients of δ^0 (see [2]). This means that we determined the continued fraction expansion with even partial quotients of the positive endpoint δ_+^0 of the geodesic δ^0 using the (modified) cutting sequence l_m of δ^0.

(2) If one applies the method of this paper to the group Γ_1 of hyperbolic isometries generated by the reflections $a_1(z) = -\overline{z}$, $b_1(z) = 1 - \overline{z}$ and $c_1(z) = 1/\overline{z}$, then instead of T one obtains the Gauss map $T_1(x) = 1/x - [1/x]$. Further, the map F_1 that corresponds to F is the natural extension of T_1, and the measure m_1 that is invariant under T_1, when normalized to be a probability measure, is the Gauss measure $\frac{1}{\log 2}\frac{1}{1+x}dx$. Finally, the endpoints of a (normalized) geodesic can be expanded as a (usual) continued fraction using the cutting sequence of the geodesic. See [4] and especially [1]. Note that the fundamental domain for Γ_1 is half of the fundamental domain for the modular group.

We close this section by checking directly that m is invariant under T.

Recall that $dm = g(x)dx$, where

$$g(x) = \left(\frac{1}{1-x} + \frac{1}{1+x}\right).$$

If $y \in (0,1)$, then the collection of inverse images of y under T is given by $\{a_n(y) = 1/(y+2n) : n \geq 1\}$ and $\{b_n(y) = 1/(2n-y) : n \geq 1\}$. To show that m is invariant under T it suffices to verify that g satisfies the functional equation

$$\sum_{n=1}^{\infty} [g(a_n(y))|a_n'(y)| + g(b_n(y))|b_n'(y)|] = g(y).$$

To that end we compute $a_n' = -1/(y+2n)^2$ and $b_n' = 1/(2n-y)^2$ and find that

$$\begin{aligned} g(a_n(y))|a_n'(y)| &= \{1/(1-\frac{1}{y+2n}) + 1/(1+\frac{1}{y+2n})\}\frac{1}{(y+2n)^2} \\ &= \{\frac{1}{y+2n-1} + \frac{1}{y+2n+1}\}\frac{1}{y+2n} \\ &= \frac{2}{(y+2n-1)(y+2n+1)} \\ &= \frac{1}{y+2n-1} - \frac{1}{y+2n+1}. \end{aligned}$$

By summing over n we find $1/(1+y)$.

Similarly, one shows that by performing the above computation for b_n instead of a_n one obtains $1/(1-y)$. g therefore satisfies the above functional equation.

5. The decay of correlation. Recall from the last section that the map T is given by $T(x) = 2 - \frac{1}{x}$, if $x \in [.5, 1]$. Note that for $k \geq 1$ the point $x_k = \frac{k-1}{k}$ satisfies $T(x_{k+1}) = x_k$, and therefore T is a bijective map from $(x_{k+1}, x_{k+2}]$ onto $(x_k, x_{k+1}]$.

If we set $M_k = (x_{k+1}, x_{k+2}]$, for $k \geq 0$, then we conclude from the above considerations that $T^k(M_k) = M_0 = (0, .5]$.

We want to estimate

$$\int I_{M_0}(x) I_{M_0}(T^k(x))dm(x).$$

Note that from the graph of T it is easy to see that

$$I_{M_0}(x) I_{M_0}(T^k(x)) = I_{(M_0 \cap T^{-1}M_{k-1}) \cup N_k},$$

for some set N_k that is not very important for us.

Therefore

$$\int I_{M_0}(x) I_{M_0}(T^k(x))dm(x) \geq m(M_0 \cap T^{-1}M_{k-1}).$$

We will show next that $m(M_0 \cap T^{-1}M_{n-1})$ decreases to zero as n^{-2}.

We calculate the measure $m(M_n)$ to be $\log\frac{2n+3}{2n+1}$. This last term is of the order n^{-1} by L'Hôpital's rule.

Now from the invariance of the measure m we find

$$m(M_n) = m(M_{n+1}) + m(M_0 \cap T^{-1}(M_n)).$$

Therefore $m(M_0 \cap T^{-1}(M_{n+1}))$ is of the order $n^{-1} - (n+1)^{-1}$ which in turn is of the order n^{-2}.

The final conclusion is:

Theorem 3. *The decay of correlation of the function I_{M_0} for the billiard map T (or F) is greater than or equal to a constant times n^{-2}.*

Remark 2. We adopt the notation of theorem 2 of section 3 and represent a billiard trajectory $\tilde{\gamma}^0$ by its modified cutting sequence l_m. The theorem then states that the billiard map relative to our choice of cross section (the side C_0 of D) acts by the "shift" map that consists in underlining the C that follows the previously underlined C of l_m. As C is always followed by B we look for the substring CB of l_m.

In theorem 3 however we restrict attention to M_0 which can be interpreted as choosing a smaller cross section. Namely, $\delta^0 \in (0,1)$ implies that the cutting sequence of $\tilde{\gamma}^0$ contains either $\underline{C}BA$ or $\underline{C}ABA$. This means that the modified cutting sequence contains $\underline{C}BA$. So by representing a billiard trajectory by its modified cutting sequence we choose the cross section CBA rather than only CB. The shift map consists now in underlining the initial C of the first string CBA that follows the previously underlined C (of a string CBA). This action is captured by the random variable I_{M_0} whose correlation is shown to have polynomial decay in theorem 3 above.

References.

[1] D. Fried, Billiard on triangular groups, to appear in Invent Math.

[2] C. Kraaikamp and A. Lopes, The Theta group and the continued fraction with even partial quotients, Geometricæ Dedicatæ, Vol 59, (1996).

[3] A. Lopes, The Zeta function, non-differentiability of pressure and the critical exponent of transition, Adv. in Math., Vol 101, (2) (1993).

[4] C. Series, The modular surface and continued fractions, J. London Math. Soc., (2) 31 (1985).

[5] C. Series, Geometrical methods of symbolic coding, in "Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces" (eds. T. Bedford, M. Keane, C. Series), Oxford University Press, Oxford, (1991).

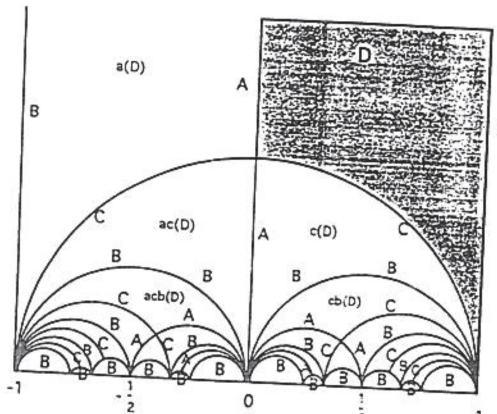

Figure 2

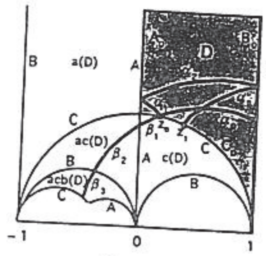

Figure 1

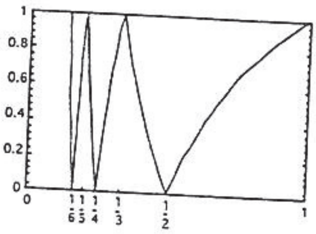

Figure 3

Received in revised form August 1996.

E-mail address: `mbauer@univ-rennes1.fr`, `alopes@if.ufrgs.br`