A POINCARÉ MAP FOR THE HOROCYCLE FLOW ON $PSL(2,\mathbb{Z})\backslash\mathbb{H}$ AND THE STERN-BROCOT TREE

CLAUDIO BONANNO, ALESSIO DEL VIGNA, AND STEFANO ISOLA

ABSTRACT. We construct a Poincaré map $\mathcal{P}_h$ for the positive horocycle flow on the modular surface $PSL(2,\mathbb{Z})\backslash\mathbb{H}$, and begin a systematic study of its dynamical properties. In particular, we give a complete characterisation of the periodic orbits of $\mathcal{P}_h$, and show that they are equidistributed with respect to the invariant measure of $\mathcal{P}_h$ and can be organised in a tree by using the Stern-Brocot tree of rational numbers. In addition we introduce a time-reparameterisation of $\mathcal{P}_h$ which gives an insight into the dynamics of the non-periodic orbits.

1. INTRODUCTION

Hyperbolic geometry has been a crucial ground for the development of important ideas of ergodic theory since its inception. Geodesic flows on surfaces of constant negative curvature were the first non-trivial examples of ergodic flows. In addition, several geometric coding techniques have been developed for such flows, stemming from a 1898 work of J. Hadamard, later developed by M. Morse and G. Hedlund in the 1920s and 30s. Subsequently, interest also grew in the study of positive and negative horocycle flows, which move respectively along the stable and unstable manifolds for the geodesic flow. More concretely, these types of flow move tangent vectors sideways along the horocycle whose center is in the direction of the tangent vector.

The geodesic flow is also one of the first systems for which coding was used to study its dynamical properties. It was E. Artin in the 1920s who noticed that by coding a geodesic on the hyperbolic Poincaré half-plane $\mathbb{H}$ by means of the continued fraction expansions of its end points, it was possible to deduce the existence of periodic and everywhere dense geodesic on the modular surface $\mathcal{M} := PSL(2,\mathbb{Z})\backslash\mathbb{H}$. This observation was later developed in the 80s in papers by R. L. Adler and L. Flatto, and by C. Series (see [1, 27]). Nowadays it is a classical result that the geodesic flow on the unit tangent bundle of the modular surface $S\mathcal{M}$ has a Poincaré section for which the associated Poincaré map has a factor map on one of the coordinates which is the Gauss map. This connection has been very useful in the study of the rich dynamics of the geodesic flow, and has led to connections between different aspects of the two systems, such as S. G. Dani's correspondence ([10]) between Diophantine approximation properties and the behaviour of certain orbits of the geodesic flow on the modular surface, and the D. H. Mayer's correspondence ([20]) between the Selberg Zeta function and the dynamical zeta function of the Gauss map.

Many dynamical properties have been studied also for the (positive/negative) horocycle flow on surfaces of constant negative curvature (see e.g. [11]). In particular it is known that the dynamics is less rich than

that of the geodesic flow. In the case of the modular surface $\mathcal{M}$, the only invariant measures on $S\mathcal{M}$ are the hyperbolic volume form and the measures supported on the periodic orbits. Instead in the case of compact surfaces, the horocycle flow is uniquely ergodic and has no periodic orbits. Finer and quantitative ergodic properties are also known: the rigidity of the flow and its rate of mixing have been studied by M. Ratner ([22, 23]); in the case of the modular surface, the periodic orbits are known to be asymptotically equidistributed on $S\mathcal{M}$ ([34, 26, 17, 28, 32]); the Birkhoff averages of suitably regular observables have been proved to have a polynomial rate of convergence ([13, 29, 24]). However the beautiful results that we have recalled have been obtained by methods of harmonic analysis and representation theory. Notable exceptions are equidistribution results in [12] and [28], which are proved by dynamical methods but do not contain estimates for the rate of convergence. In a more general context, it is also worth mentioning the very general construction of O. Sarig and M. Schmoll which builds the horocycle flow as a suspension flow over adic transformations [25].

The main objective of this work is to construct a Poincaré map for the positive horocycle flow on the unit tangent bundle $S\mathcal{M}$, and begin a systematic study of its properties. We believe that this paper represents a first step in the study of the ergodic properties of the horocycle flow without using other structures of the modular surface. First of all, we work in the same framework used to approach the geodesic flow. We define a Poincaré map $\mathcal{P}_h$ for the positive horocycle flow using a cross-section with respect to which the geodesic flow has a Poincaré map whose factor map on the first coordinate is the extended Farey map, the slow version of the Gauss map (see Appendix A). We obtain a map $\mathcal{P}_h$ which is defined on a two-dimensional set $\mathbb{W}$ homeomorphic to $\mathbb{R}^+ \times \mathbb{R}^+$, whose action on the first coordinate is defined in terms of matrices in $SL(2, \mathbb{Z})$, and preserves an infinite measure ν which is absolutely continuous with respect to the Lebesgue measure (the same holds for the Poincaré map of the geodesic flow). The map $\mathcal{P}_h$ is defined in Section 4, in which we also give the explicit expression of the first return time function to the cross-section along the horocycle flow, thus constructing a suspension flow isomorphic to the horocycle flow on $S\mathcal{M}$.

Let us remark that as far as we know the only other Poincaré maps for the horocycle flow on the modular surface which have been constructed are those in [6] (see also [18]) and, more recently, in [19]. These maps use a cross-section defined in terms of the identification of $S\mathcal{M}$ with the space of unimodular lattices. The methods of [6] have been applied to construct sections for the horocycle flow on quotients of $S\mathbb{H}$ with respect to other groups than $PSL(2, \mathbb{Z})$, see [5, 15, 33, 31] and the review [4].

Looking at the previous constructions from the ergodic theoretical point-of-view, the main difference with our approach is that the Poincaré map on which the suspension flow is built preserves an infinite measure. This is a difficulty as we cannot use the standard methods of ergodic theory and we need to appeal at the results of infinite ergodic theory, but we are confident that this is the feature of the system responsible for example for its slow mixing behaviour[1]. We believe that our approach will turn useful in studying the horocycle flow by looking at the dynamical properties of its Poincaré map and its first return time function only without using any geometric or algebraic structure of the modular surface.

In Section 5 we study the periodic points of $\mathcal{P}_h$. Without appealing to known results, we give a complete characterisation of the periodic points. We prove that there exists a one-parameter family of periodic points which are characterised by having the first coordinate in $\mathbb{Q}^+$ (Theorem 5.8), and show that the periodic orbits are asymptotically equidistributed on $\mathbb{W}$ with respect to the invariant measure ν (Theorem 5.11). The behaviour of the periodic points may be studied also by looking at their rational coordinate. Using the Stern-Brocot tree, which is a binary tree containing all the rational numbers in $\mathbb{Q}^+$, we construct an algorithm to obtain all the points in a periodic orbit and to list them in the dynamical order (see Appendix B). It is immediate to realise that as the period of the orbit diverges, the rational coordinates of the points in the periodic orbit cover all the positive rational numbers.

---

[1]Recently, attention to ergodic properties of suspension flows based on nonuniformly hyperbolic maps has increased.

In Section 6 we study the behaviour of the non-periodic orbits. We show that there exists a time-reparameterization $\mathcal{T}_h$ of $\mathcal{P}_h$ which has factor map on the first coordinate isomorphic to the so-called backward continued fraction map, whereas the second coordinate of $\mathcal{T}_h$ is strictly increasing for non-periodic orbits. This can be interpreted as a proof of the existence of a subsequence of times along which the non-closed horocycles converge to the cusp of $\mathcal{M}$.

Lastly, in Section 7 we discuss the relations with the construction in [6].

Before getting into the construction of the Poincaré map $\mathcal{P}_h$ in Section 4, we collect what the reader needs to know about the properties of the Stern-Brocot tree in Section 2 and about the modular surface $\mathcal{M}$ in Section 3.

2. Stern-Brocot trees of positive rational numbers

In this section we recall the Stern-Brocot tree and its permuted versions, binary trees containing all the positive rational numbers in their reduced form exactly once. The *Farey sum* of two fractions is defined to be

$$\frac{a}{c} \oplus \frac{b}{d} := \frac{a+b}{c+d},$$

and is used to generate all the positive fractions. Let $1/0$ denote the "fraction" of infinity, and let $F_{-1} = \{0/1, 1/0\}$. Then for $k \geq 0$ we recursively define the *Stern-Brocot sets* F_k as the set of fractions containing F_{k-1} and the Farey sum of two consecutive fractions in F_{k-1}. In each set the fractions are arranged in increasing order. Recall that the Farey sum of two fractions a/c and b/d satisfies $a/c < a/c \oplus b/d < b/d$. Thus we find

$$F_0 = \left\{ \frac{0}{1}, \frac{1}{1}, \frac{1}{0} \right\}, \quad F_1 = \left\{ \frac{0}{1}, \frac{1}{2}, \frac{1}{1}, \frac{2}{1}, \frac{1}{0} \right\}, \quad F_3 = \left\{ \frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}, \frac{3}{2}, \frac{2}{1}, \frac{3}{1}, \frac{1}{0} \right\},$$

and so on. A *Farey pair* consists of two consecutive fractions in a Stern-Brocot set. In the following we use that a Farey pair $p/q, p'/q'$ with $p/q < p'/q'$ satisfies $p'q - q'p = 1$.

2.1. The Stern-Brocot tree.
The Stern-Brocot sets form a strictly increasing sequence of sets which eventually contain all positive rational numbers in reduced form. One can then arrange the fractions in a binary tree, the Stern-Brocot tree $\mathcal{T}$, according to the order they appear in the Stern-Brocot sets. Figure 1 shows the first few levels of the Stern-Brocot tree. The fractions $0/1$ and $1/0$ are the generators of all the fractions, hence they constitute the first level of the tree, $\mathcal{T}_{-1}$. Then the level $\mathcal{T}_0$ is given by $1/1$ the only fraction in $F_0 \setminus F_{-1}$. Recursively, the k-th level $\mathcal{T}_k$ of the tree is given by the fractions in $F_k \setminus F_{k-1}$. The fractions are arranged in increasing order, and each is connected to a fraction of the previous level. It is clear by definition of the Stern-Brocot sets that each fraction p/q in F_k which is not in F_{k-1} is the Farey sum of two fractions $a/c, b/d$ in F_{k-1}, of which exactly one, say a/c, is in the level $\mathcal{T}_{k-1}$ of the Stern-Brocot tree. In this case we say that a/c is a *parent* of p/q, which in turn is the *daughter* of a/c. In addition, all fractions in a level $\mathcal{T}_k$, excluding $0/1$ and $1/0$, have two daughters in $\mathcal{T}_{k+1}$. This induces a natural notion of *sisters fractions*, two fractions which have a parent in common.

The structure of binary tree of $\mathcal{T}$ makes it possible to code each fraction $p/q \in \mathbb{Q}^+$ by the path one has to follow on $\mathcal{T}$ from the root $1/1$ to p/q by using L for left and R for right to denote the motion from a fraction to the left and to the right daughter respectively. Analogously, using the matrices in $SL(2, \mathbb{Z})$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and considering the set $\{L, R\}^*$ of all finite words with letters in the alphabet $\{L, R\}$, one can set

$$\frac{1}{1} \leftrightarrow I \quad \text{and} \quad \frac{p}{q} \leftrightarrow M\left(\frac{p}{q}\right) \in \{L, R\}^*$$
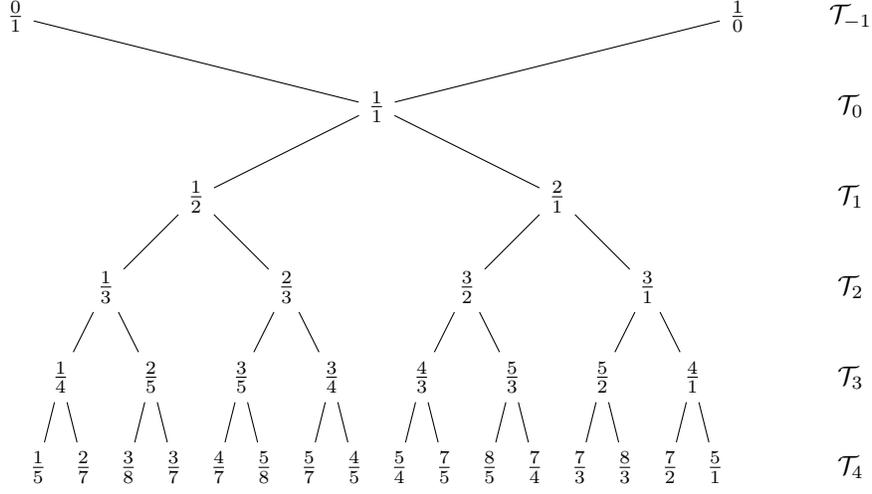
Figure 1. The first six levels of the Stern-Brocot tree.

where a motion on the tree corresponds to right multiplication by L if we move to the left and by R if we move to the right. In this way

$$\frac{p}{q} \leftrightarrow M\left(\frac{p}{q}\right) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \quad \text{with} \quad a,b,c,d \geq 0 \quad \text{and} \quad \frac{p}{q} = \frac{a}{c} \oplus \frac{b}{d},$$

so that $p = a + b$ and $q = c + d$.

Example 2.1. We have

$$\frac{1}{1} \leftrightarrow M\left(\frac{1}{1}\right) = I, \quad \frac{1}{2} \leftrightarrow M\left(\frac{1}{2}\right) = L, \quad \frac{2}{1} \leftrightarrow M\left(\frac{2}{1}\right) = R,$$

and thus for example

$$\frac{5}{8} \leftrightarrow M\left(\frac{5}{8}\right) = LRLR = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} \quad \text{and} \quad \frac{5}{8} = \frac{2}{3} \oplus \frac{3}{5}.$$

This coding can be extended to all $x \in \mathbb{R}^+$ by using the continued fraction expansion of an irrational number and its slow convergents $\{A_k/B_k\}_k$, so that

$$\mathbb{R} \setminus \mathbb{Q} \ni x \leftrightarrow M(x) \in \{L,R\}^{\mathbb{N}} \quad \text{and} \quad M(x) = \lim_{k \to \infty} M\left(\frac{A_k}{B_k}\right),$$

where now $\{L,R\}^{\mathbb{N}}$ denotes the set of infinite words with letters in the alphabet $\{L,R\}$.

Remark 2.2. Note that using this coding

$$x \in [0,1) \Leftrightarrow M(x) = L \cdots \quad \text{and} \quad x \in (1,\infty) \Leftrightarrow M(x) = R \cdots$$

2.2. **The permuted Stern-Brocot tree.** Starting from $\mathcal{T}$ and the $\{L,R\}$ coding of the fractions, the permuted Stern-Brocot tree $\hat{\mathcal{T}}$ (also known as Calkin-Wilf tree [9]) is obtained from the Stern-Brocot tree $\mathcal{T}$ by moving each fraction to the position reached by the motions of its coding on $\mathcal{T}$ as read from right to left. Figure 2 shows the first few levels of the the permuted Stern-Brocot tree. For example, since

$$\frac{2}{3} \leftrightarrow M\left(\frac{2}{3}\right) = LR$$

the fraction 2/3 can be reached in $\hat{\mathcal{T}}$ by the motions RL, that is starting from 1/1 we first go downward to the right and then downward to the left. This sequence of motions describes the new position of 2/3 in $\hat{\mathcal{T}}$.
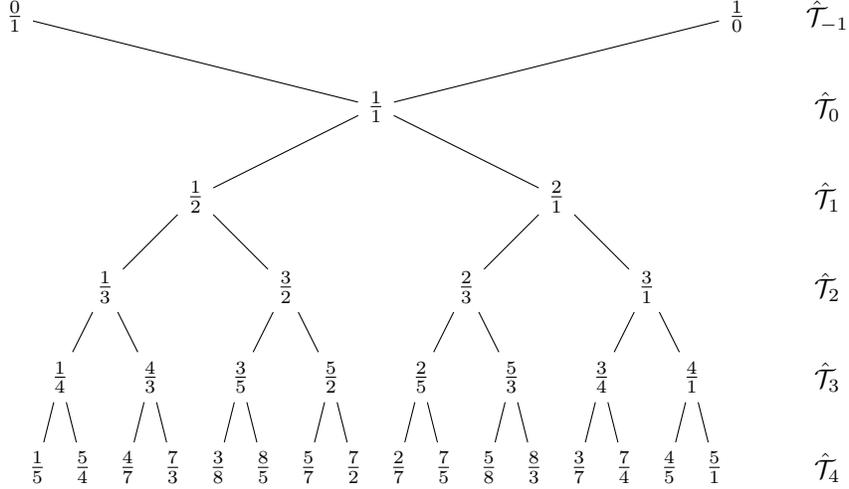
Figure 2. The first six levels of the permuted Stern-Brocot tree.

Given a finite word $w = (w_1 w_2 \cdots w_{n-1} w_n)$ let $\hat{w}$ denote its reverse, that is $\hat{w} = (w_n w_{n-1} \cdots w_2 w_1)$. Then $\hat{M}$ denotes the reverse sequence of motions of a rational number, that is $\hat{M}(p/q)$ is the reverse of the word $M(p/q)$, so that looking at the examples we have considered above we have

$$\hat{M}\left(\frac{2}{3}\right) = RL \quad \text{and} \quad \hat{M}\left(\frac{5}{8}\right) = RLRL.$$

Lemma 2.3. *The matrices $M(p/q)$ and its reverse $\hat{M}(p/q)$ satisfy the following relation*

$$M\left(\frac{p}{q}\right) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Leftrightarrow \quad \hat{M}\left(\frac{p}{q}\right) = \begin{pmatrix} d & b \\ c & a \end{pmatrix}.$$

Proof. Let J be the matrix

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

One easily verifies that $JL^n = R^n J$ and $JR^n = L^n J$ for all $n \geq 1$. In addition, the transpose L^t coincides with R. Hence, if

$$M\left(\frac{p}{q}\right) = L^{n_1} R^{m_1} L^{n_2} R^{m_2} \ldots L^{n_s} R^{m_s}$$

for any $s \geq 1$ and $n_i, m_i \geq 0$, then a standard computation gives

$$\hat{M}\left(\frac{p}{q}\right) = R^{m_s} L^{n_s} \ldots R^{m_2} L^{n_2} R^{m_1} L^{n_1} = J \left(M\left(\frac{p}{q}\right) \right)^t J.$$

The lemma immediately follows. $\qquad\square$

It is interesting to notice that $\hat{\mathcal{T}}$ is closely related to the *extended Farey map* $U : \mathbb{R}^+ \to \mathbb{R}^+$

$$U(x) := \begin{cases} \dfrac{x}{1-x}, & \text{if } 0 \leq x < 1, \\[2mm] x - 1, & \text{if } x \geq 1. \end{cases} \tag{2.1}$$

Indeed, given p/q its daughters in $\hat{\mathcal{T}}$ are $p/(p+q)$ and $(p+q)/q$ which correspond to the elements of $U^{-1}(p/q)$. In Appendix A the map U will pop up as the factor map on the first coordinate of the Poincaré map for the geodesic flow with respect to a section $\mathcal{C}$ (to be defined below) on the unit tangent $S\mathcal{M}$ of the modular surface. For more details about the Stern-Brocot trees and their relations with one-dimensional maps we refer to [8].

5

3. Preliminaries from hyperbolic geometry

We shall consider the *upper half-plane*

$$\mathbb{H} := \{z = x + iy \ : \ x, y \in \mathbb{R}, \ y > 0\}$$

with the hyperbolic metric $ds^2 = (dx^2 + dy^2)/y^2$, so that $\mathbb{H}$ becomes a Riemannian manifold with constant negative curvature -1. The boundary of the hyperbolic plane is $\partial\mathbb{H} = \mathbb{R}\cup\{\infty\}$, where $\mathbb{R} := \{x+iy \ : \ y = 0\}$. Given $\Gamma \in SL(2,\mathbb{R})$ we let $\Gamma : \mathbb{H} \to \mathbb{H}$ act as a Möbius transformation

$$\Gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \Gamma(z) = \frac{az + b}{cz + d}.$$

As the action of Γ and $-\Gamma$ coincide, the group of orientation preserving isometries of $\mathbb{H}$ is taken to be $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{\pm I\}$. An important quantity related to a matrix $\Gamma \in SL(2,\mathbb{R})$ is its *deformation factor* at a point z, which we denote by def $(\Gamma(z))$ and is given by

$$\Gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{def}\,(\Gamma(z)) := \frac{d}{dz}\Gamma(z) = (cz + d)^{-2}. \tag{3.1}$$

The deformation factor of a matrix is related to its lift as an action on the tangent bundle of $\mathbb{H}$.
Let us denote by $\zeta \in T_z\mathbb{H}$ the tangent vectors to $\mathbb{H}$ at the point z, then to each tangent vector ζ we associate the angle $\theta(\zeta) \in (-\pi, \pi]$, which is the angle between ζ and the positive y-axis measured counterclockwise from the y-axis. The unit tangent bundle $S\mathbb{H}$ can be identified with $PSL(2,\mathbb{R})$, to the effect that for each $(z, \zeta) \in S\mathbb{H}$ there exists a unique $\Gamma_{(z,\zeta)} \in PSL(2,\mathbb{R})$ such that

$$z = \Gamma_{(z,\zeta)}(i) \quad \text{and} \quad \zeta = d_i(\Gamma_{(z,\zeta)})(i), \tag{3.2}$$

where $i \in T_i\mathbb{H}$ is the unit vector based at $i \in \mathbb{H}$ with $\theta(i) = 0$. Note that using the lift $(\Gamma_{(z,\zeta)})_*$ of $\Gamma_{(z,\zeta)}$ to $S\mathbb{H}$, we can restate the above conditions as $(z, \zeta) = (\Gamma_{(z,\zeta)})_*(i, i)$.

3.1. Geodesic and horocycle flow on $S\mathbb{H}$.
As it is well known, the *geodesics* of $\mathbb{H}$ are the half lines orthogonal to $\partial\mathbb{H}$ and the half circles with center on $\partial\mathbb{H}$. Given (z, ζ) in the unit tangent bundle $S\mathbb{H}$, we denote by $\gamma_{z,\zeta}(t)$ the (unique) geodesic tangent to ζ at z and set

$$\gamma_{z,\zeta}^{\pm} := \lim_{t\to\pm\infty} \gamma_{z,\zeta}(t) \in \partial\mathbb{H}.$$

Note that $\gamma_{z,\zeta}^+ = \infty$ if and only if $\theta(\zeta) = 0$, and $\gamma_{z,\zeta}^- = \infty$ if and only if $\theta(\zeta) = \pi$. In these two cases the geodesic is a vertical half line with the other limit point on $\mathbb{R}$. The *geodesic flow* is defined as the one-parameter group g_t acting on the unit tangent bundle $S\mathbb{H}$ as

$$g_t : S\mathbb{H} \to S\mathbb{H}, \qquad g_t(z, \zeta) = (\gamma_{z,\zeta}(t), \gamma_{z,\zeta}'(t)),$$

where $\gamma_{z,\zeta}'(t)$ is the unit vector tangent to the geodesic at $\gamma_{z,\zeta}(t)$. Thanks to the identification of $S\mathbb{H}$ with $PSL(2,\mathbb{R})$, we can read the geodesic flow on $S\mathbb{H}$ as the action by right multiplication of matrices in $PSL(2,\mathbb{R})$. In particular, for $t \in \mathbb{R}$

$$\Gamma_{g_t(z,\zeta)} = \Gamma_{(z,\zeta)} \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}. \tag{3.3}$$

One defines the horocycle flow analogously. Let $W^+(z, \zeta)$ denote the *(positive) horocycle* at (z, ζ). If $\theta(\zeta) \neq 0$, the horocycle $W^+(z, \zeta)$ is the circle tangent to $\mathbb{R}$ at $\gamma_{z,\zeta}^+$, passing at z and orthogonal to ζ, with ζ pointing inward. In this case we denote by $r^+(z, \zeta)$ its radius. If instead $\theta(\zeta) = 0$, then $W^+(z, \zeta)$ is a horizontal line passing at z with ζ pointing upward. The *(positive) horocycle flow* is defined as the one-parameter group

h_s^+ on $S\mathbb{H}$ which moves vectors orthogonal to $W^+(z,\zeta)$ rightward on $W^+(z,\zeta)$ at unit speed. In terms of $PSL(2,\mathbb{R})$ matrices, for all $s \in \mathbb{R}$

$$\Gamma_{h_s^+(z,\zeta)} = \Gamma_{(z,\zeta)} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}. \tag{3.4}$$

Thanks to the reinterpretation of the flow action by right multiplication of matrices, it is straightforward to verify the following commutation rules:

$$g_t \circ g_s = g_s \circ g_t \quad \text{and} \quad g_t \circ h_s^+ = h_{se^{-t}}^+ \circ g_t \quad , \quad \forall t, s \in \mathbb{R}.$$

3.2. **The modular surface.** We now consider the *modular group* $PSL(2,\mathbb{Z})$, that is the subgroup of $PSL(2,\mathbb{R})$ generated by the two matrices

$$S := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad R := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

which act on $\mathbb{H}$ as

$$S(z) = -\frac{1}{z} \quad \text{and} \quad R(z) = z + 1$$

Since $S^2 = (SR)^3 = I$ the modular group is not free.

The *modular surface* is then defined to be the quotient $\mathcal{M} := PSL(2,\mathbb{Z})\backslash\mathbb{H}$, with the quotient topology. The standard fundamental domain $\mathcal{F}$ for $\mathcal{M}$ is the geodesic polygon

$$\mathcal{F} = \left\{ z = x + iy \in \mathbb{H} \ : \ |x| \le \frac{1}{2}, \ |z| \ge 1 \right\}.$$

The identification of $S\mathbb{H}$ with $PSL(2,\mathbb{R})$ implies that $S\mathcal{M}$ is identified with $PSL(2,\mathbb{Z})\backslash PSL(2,\mathbb{R})$. Let $\pi : \mathbb{H} \to \mathcal{M}$ be the projection on the modular surface and let $\pi_* : S\mathbb{H} \to S\mathcal{M}$ be its lift to the unit tangent bundles. We let $\tilde{g}_t : S\mathcal{M} \to S\mathcal{M}$ and $\tilde{h}_s^+ : S\mathcal{M} \to S\mathcal{M}$ be the projection of the geodesic and horocycle flows on the modular surface, that is

$$\tilde{g}_t \circ \pi_* = \pi_* \circ g_t \quad \text{and} \quad \tilde{h}_s^+ \circ \pi_* = \pi_* \circ h_s^+.$$

Motion on a geodesic or on a horocycle on $\mathcal{M}$ corresponds to the motion on one of the equivalent geodesics or horocycles on $\mathbb{H}$, up to identification of equivalent points.

4. The Poincaré map for the horocycle flow and its suspension

Given $n \in \mathbb{Z}$, let $\mathcal{I}_n$ denote the vertical line in $\mathbb{H}$ given by

$$\mathcal{I}_n = \{x + iy \in \mathbb{H} \ : \ x = n\}$$

and let

$$\mathcal{I}^+ := \mathcal{I}_0 \cap \mathcal{F} = \{z \in \mathbb{H} \ : \ x = 0, \ y \ge 1\} \quad \text{and} \quad \mathcal{I}^- := \mathcal{I}_0 \setminus \mathcal{I}^+ = \{z \in \mathbb{H} \ : \ x = 0, \ 0 < y < 1\}.$$

Our Poincaré section for the geodesic and the horocycle flows on $S\mathcal{M}$ will be the set

$$\mathcal{C} := \left\{ (z,\zeta) \in S\mathcal{M} \ : \ z \in \mathcal{I}^+, \ \theta(\zeta) \neq 0, \pi \right\}.$$

Set moreover $C := \{(z,\zeta) \in S\mathbb{H} \ : \ \pi_*(z,\zeta) \in \mathcal{C}\}$. Using $S(i) = i$ and $S(\mathcal{I}^+ \setminus \{i\}) = \mathcal{I}^-$, and the fact that the lift $S_* : S\mathbb{H} \to S\mathbb{H}$ satisfies

$$\theta(S_*(z,\zeta)) = \theta(\zeta) + \pi \pmod{2\pi} \qquad \forall\, (z,\zeta) \in S\mathbb{H} \text{ with } z \in \mathcal{I}_0,$$

for each $(z,\zeta) \in \mathcal{C}$ there exists $(z',\zeta') \in S\mathbb{H}$ with $z' \in \mathcal{I}_0$ and $\theta(\zeta') \in (-\pi, 0)$ such that $\pi_*(z',\zeta') = (z,\zeta)$. Hence in the following we identify $\mathcal{C}$ with the set of points on $\mathcal{I}_0$ with tangent unit vector ζ pointing towards the half-plane of points with strictly positive real part, in accordance with

$$\mathcal{C} = \pi_* \left(\{(z,\zeta) \in S\mathbb{H} \ : \ z \in \mathcal{I}_0\,, \ \theta(\zeta) \in (-\pi, 0)\} \right). \tag{4.1}$$

Stated otherwise, when we write $(z, \zeta) \in \mathcal{C}$ we think of it as a couple with $z \in \mathcal{I}_0$ and $\theta(\zeta) \in (-\pi, 0)$. Analogously we set

$$C = \{(z, \zeta) \in S\mathbb{H} \; : \; \exists \, \Gamma \in PSL(2, \mathbb{Z}) \text{ such that } \Gamma(z) \in \mathcal{I}_0 \, , \, \theta(\Gamma_*(z, \zeta)) \in (-\pi, 0)\} \, . \qquad (4.2)$$

Subsets of C which play an important role in the rest of the paper are points with base on the lines $\mathcal{I}_n$ and on the half-circles

$$\mathcal{J}_n := \left\{ x + iy \in \mathbb{H} \; : \; \left(x - n - \frac{1}{2} \right)^2 + y^2 = \frac{1}{4} \right\} \, .$$

Note that $\mathcal{I}_n = R^n(\mathcal{I}_0)$ and $\mathcal{J}_n = R^n L(\mathcal{I}_0)$, where

$$L := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \in PSL(2, \mathbb{Z}).$$

Note moreover that for each $n \in \mathbb{Z}$ the lines $\mathcal{I}_n$, $\mathcal{J}_n$ and $\mathcal{I}_{n+1}$ are the sides of the hyperbolic triangle

$$\Delta_n := \left\{ x + iy \in \mathbb{H} \; : \; n \le x \le n+1, \; \left(x - n - \frac{1}{2} \right)^2 + y^2 \ge \frac{1}{4} \right\}$$

with vertices in n, $n+1 \in \mathbb{R}$ and ∞ (each Δ_n contains three copies of the fundamental domain $\mathcal{F}$).

We now proceed with the construction of the Poincaré map on $\mathcal{C}$ for the horocycle flow $\tilde{h}_s^+$ for negative times s (the Poincaré map on $\mathcal{C}$ for the geodesic flow $\tilde{g}_t$ is already known, but we recall its construction in Appendix A). Let's start by defining the set

$$\mathbb{W} := \left\{ (\gamma, r, \varepsilon) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \{-1, 0, +1\} \; : \; r \ge \gamma, \; \varepsilon = 0 \text{ if and only if } r = \gamma \right\}$$

which is homeomorphic to $\mathbb{R}^+ \times \mathbb{R}^+$. Now, for each $(z, \zeta) \in \mathcal{C}$ the positive horocycle $W^+(z, \zeta)$ is tangent to $\mathbb{R}$ at $\gamma_{z, \zeta}^+$ and intersects $\mathcal{I}_0$. Its radius thus satisfies $r^+(z, \zeta) \ge \gamma_{z, \zeta}^+$, with strict inequality if and only if $W^+(z, \zeta)$ intersects $\mathcal{I}_0$ in two different points. We can therefore code $(z, \zeta) \in \mathcal{C}$ with a point in $\mathbb{W}$ using its horocycle $W^+(z, \zeta)$. We set

$$\gamma := \gamma_{z, \zeta}^+, \quad r := r^+(z, \zeta), \quad \text{and} \quad \varepsilon = \begin{cases} -1 \, , & \text{if } \theta(\zeta) \in (-\pi/2, 0); \\ 0 \, , & \text{if } \theta(\zeta) = -\pi/2; \\ +1 \, , & \text{if } \theta(\zeta) \in (-\pi, -\pi/2). \end{cases}$$

The variable ε is equal to 0 if the horocycle $W^+(z, \zeta)$ is tangent to $\mathcal{I}_0$ at z, whereas is equal to $+1$ or to -1 if z is the point of intersection of $W^+(z, \zeta)$ with $\mathcal{I}_0$ of highest, respectively lowest, variable y. We have thus defined a map

$$\mathcal{W} : \mathcal{C} \to \mathbb{W}, \quad (z, \zeta) \mapsto \mathcal{W}(z, \zeta) = (\gamma_{z, \zeta}^+, \, r^+(z, \zeta), \, \varepsilon). \qquad (4.3)$$

Given $(z, \zeta) \in \mathcal{C}$ let us consider the horocycle flow $\tilde{h}_s^+(z, \zeta)$ for $s \in (-\infty, 0)$, and let (z', ζ') be its first return to $\mathcal{C}$. We can thereby view the Poincaré map of $\tilde{h}_s^+$ on $\mathcal{C}$ as a map

$$\mathcal{P}_h : \mathbb{W} \to \mathbb{W}$$

$$(\gamma, r, \varepsilon) = \mathcal{W}(z, \zeta) \mapsto \mathcal{P}_h(\gamma, r, \varepsilon) = (\gamma', r', \varepsilon') = \mathcal{W}(z', \zeta') \, .$$

It is clear that (z', ζ'), viewed as a point of the set C (with slight abuse of notation), depends on the configuration of $W^+(z, \zeta)$ and on the value of ε (see Figures 3 and 4 below). In particular the first return to C for $\tilde{h}_s^+$ with negative s occurs on one of the following sets: $\mathcal{J}_{-1}$, $\mathcal{I}_{-1}$ or $\mathcal{I}_0$ if $\varepsilon = -1$; $\mathcal{I}_1$, $\mathcal{I}_2$ or $\mathcal{J}_0$ if $\varepsilon = 0, +1$ (here we are using that $\theta(\zeta') \ne -\pi$). Then $W^+(z', \zeta')$ will be the positive horocycle defined by (z', ζ') as a point in $\mathcal{C}$. If $\Gamma \in PSL(2, \mathbb{Z})$ sends $z' \in C$ to $\mathcal{I}_0$, that is $\Gamma(z') \in \mathcal{I}_0$ is the representative of z' in $\mathcal{C}$, then $W^+(z', \zeta') = \Gamma(W^+(z, \zeta))$.

Our aim is now to describe the map $\mathcal{P}_h$ on the points of $\mathbb{W}$, also in relation with the $\{L, R\}$ coding of $\gamma \in \mathbb{R}^+$ introduced in Section 2. The following lemma is a key result in order to characterize the map $\mathcal{P}_h$ in dependence on the subset of C on which the first return occurs.

Lemma 4.1. *For $(z,\zeta) \in \mathcal{C}$, let $(\gamma, r, \varepsilon) = \mathcal{W}(z,\zeta)$. If the first return to C for $\tilde{h}_s^+$ with negative s along $W^+(z,\zeta)$ occurs on $\Gamma^{-1}(\mathcal{I}_0)$ for some $\Gamma \in PSL(2,\mathbb{Z})$, then $(\gamma', r', \varepsilon') = \mathcal{P}_h(\gamma, r, \varepsilon)$ satisfies*

$$\gamma' = \Gamma(\gamma) \quad and \quad r' = \mathrm{def}\,(\Gamma(\gamma))\, r\,,$$

where $\mathrm{def}\,(\Gamma(\cdot))$ *is the deformation factor defined in* (3.1).

Proof. The statement for γ' simply follows from $W^+(z',\zeta') = \Gamma(W^+(z,\zeta))$ whenever the first return occurs on $\Gamma^{-1}(\mathcal{I}_0)$. Let us now see how the radius of the positive horocycle $W^+(z,\zeta)$ changes when a matrix in $PSL(2,\mathbb{Z})$ acts on it. We first consider the action of the generators R and S. If $\Gamma \in \{R, R^{-1}\}$ then the horocycle $\Gamma(W^+(z,\zeta))$ is obtained by just translating $W^+(z,\zeta)$ to the right or to the left. Therefore the radius does not change, and $\mathrm{def}\,(\Gamma(z)) = 1$ for all z. If, on the other hand, $\Gamma = S = S^{-1}$, one easily checks that $x + iy \in W^+(z,\zeta)$, a circle of radius r and center at $\gamma + ir$, if and only if $(x - \gamma)^2 + y^2 - 2ry = 0$, and $S(x+iy) = X+iY$ satisfies $(X - \gamma') + Y^2 - 2r'Y = 0$ with $\gamma' = -1/\gamma$ and $r' = r/\gamma^2$. Since $\mathrm{def}\,(S(\gamma)) = \gamma^{-2}$ the result is proven also in this case. The general case now follows by noting that

$$\mathrm{def}\,(\Gamma_1\Gamma_2(z)) = \mathrm{def}\,(\Gamma_1(\Gamma_2(z)))\,\mathrm{def}\,(\Gamma_2(z))$$

for all z and all $\Gamma_1, \Gamma_2 \in PSL(2,\mathbb{Z})$. $\qquad\square$

The following tables describe the different cases that have to be considered to define $\mathcal{P}_h$ on $\mathbb{W}$, providing the following information: the subset of C on which the first return occurs; the matrix Γ as in Lemma 4.1; the image $(\gamma', r', \varepsilon') = \mathcal{P}_h(\gamma, r, \varepsilon)$ computed using the lemma along with simple arguments for ε'; the finite word $M(\gamma') \in \{L, R\}^*$ as a function of $M(\gamma)$ as defined in Section 2, using the notation $\sigma : \{L, R\}^* \to \{L, R\}^*$ for the shift map.

We start with the case $\varepsilon = +1$, for which we refer to Figure 3. We need to distinguish between the cases $\gamma > 1$ and $\gamma < 1$. In the first case the first return occurs on $\mathcal{I}_1$, but in the last case it may occur on $\mathcal{I}_1$ or $\mathcal{J}_0$ depending on the value of the radius r of the horocycle. The case $\gamma = 1$ is special because the case $\theta(\zeta') = 0$ is not admissible for points in $\mathcal{C}$. The different possibilities are described in Table 1. We remark that in some circumstances it is necessary to add the action of S to have $\theta(\zeta')$ in the interval $(-\pi, 0)$. Moreover the action on $M(\gamma)$ has been computed by concatenation of the letters describing Γ and $M(\gamma)$, using some relations among the matrices such as $SR^{-1} = RL^{-1}$.

Conditions on γ and r	First return	Γ	$\mathcal{P}_h(\gamma, r, \varepsilon)$	$M(\gamma')$
$\gamma > 1$	$\mathcal{I}_1$	R^{-1}	$(\gamma - 1, r, +1)$	$\sigma(M(\gamma))$
$\gamma = 1$	$\mathcal{I}_2$	SR^{-2}	$(1, r, -1)$	$M(\gamma)$
$\gamma < 1,\ r > 1 - \gamma$	$\mathcal{I}_1$	SR^{-1}	$\left(\frac{1}{1-\gamma}, \frac{r}{(1-\gamma)^2}, -1\right)$	$R\sigma(M(\gamma))$
$\gamma < 1,\ r = 1 - \gamma$	$\mathcal{I}_1$	SR^{-1}	$\left(\frac{1}{1-\gamma}, \frac{1}{1-\gamma}, 0\right)$	$R\sigma(M(\gamma))$
$\gamma < 1,\ r < 1 - \gamma$	$\mathcal{J}_0$	L^{-1}	$\left(\frac{\gamma}{1-\gamma}, \frac{r}{(1-\gamma)^2}, +1\right)$	$\sigma(M(\gamma))$

Table 1. Definition of the map $\mathcal{P}_h$ for $\varepsilon = +1$.

In the case $\varepsilon = -1$, following $\tilde{h}_s^+$ with s negative, the points slide on the negative abscissa half-plane, so the first return may occur on $\mathcal{J}_{-1}$ or on $\mathcal{I}_{-1}$, or back on $\mathcal{I}_0$ (see Figure 4). The different possibilities are described in Table 2. In this case note that $r \in (1 + \gamma, \gamma(1 + \gamma))$ is possible only when $\gamma > 1$, and it corresponds to the only situation for which the return occurs on $\mathcal{I}_{-1}$, since the horocycle does not intersect $\mathcal{J}_{-1}$ but does intersect $\mathcal{I}_{-1}$.

It remains to describe the case $\varepsilon = 0$. In this case we necessarily have $r = \gamma$ and everything works analogously to the case $\varepsilon = +1$. See Table 3.
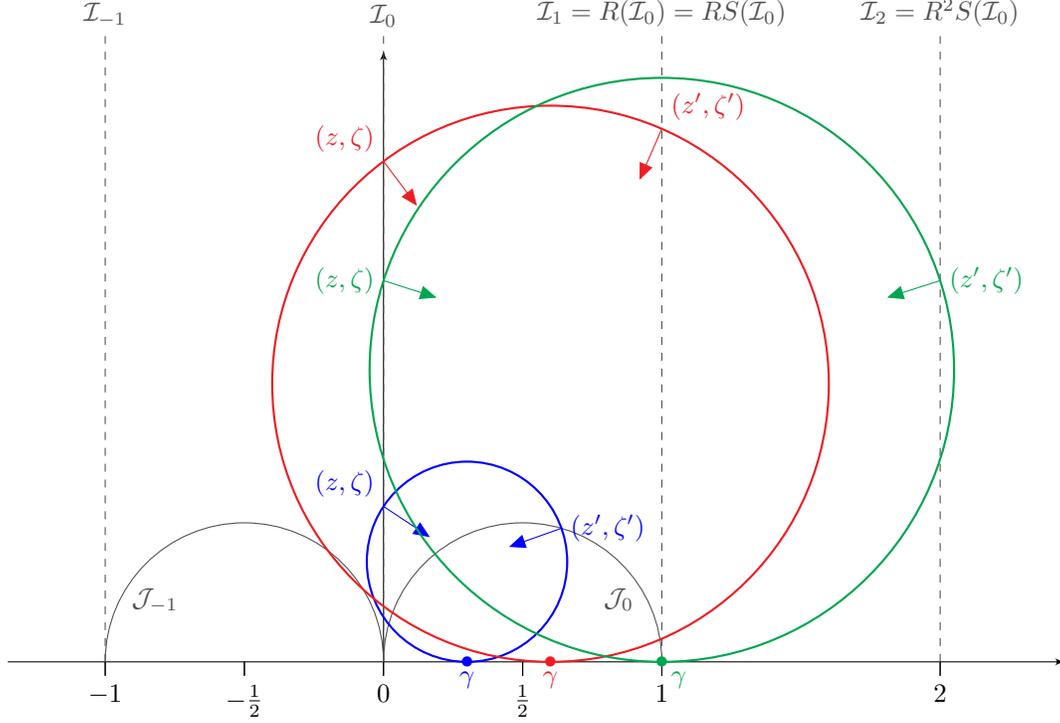
Figure 3. Action of the first return map $\mathcal{P}_h$ for $\varepsilon = +1$. The red horocycle has $\gamma < 1$ and $r > 1 - \gamma$, so that the first return occurs on $\mathcal{I}_1$; the blue horocycle has $\gamma < 1$ and $r < 1 - \gamma$, so that the first return occurs on $\mathcal{J}_0$; the green horocycle has $\gamma = 1$ so that the first return occurs on $\mathcal{I}_2$.

Conditions on γ and r	First return	Γ	$\mathcal{P}_h(\gamma, r, \varepsilon)$	$M(\gamma')$
$r > \gamma(1+\gamma)$	$\mathcal{J}_{-1}$	L	$\left(\frac{\gamma}{1+\gamma}, \frac{r}{(1+\gamma)^2}, -1\right)$	$L\,M(\gamma)$
$r = \gamma(1+\gamma)$	$\mathcal{J}_{-1}$	L	$\left(\frac{\gamma}{1+\gamma}, \frac{\gamma}{1+\gamma}, 0\right)$	$L\,M(\gamma)$
$\gamma > 1,\, 1+\gamma < r < \gamma(1+\gamma)$	$\mathcal{I}_{-1}$	R	$(\gamma+1, r, -1)$	$R\,M(\gamma)$
$\gamma > 1,\, r = 1+\gamma$	$\mathcal{I}_{-1}$	R	$(\gamma+1, \gamma+1, 0)$	$R\,M(\gamma)$
$r < \min\{1+\gamma, \gamma(1+\gamma)\}$	$\mathcal{I}_0$	I	$(\gamma, r, +1)$	$M(\gamma)$

Table 2. Definition of the map $\mathcal{P}_h$ for $\varepsilon = -1$.

Conditions on γ and r	First return	Γ	$\mathcal{P}_h(\gamma, r, \varepsilon)$	$M(\gamma')$
$\gamma = r > 1$	$\mathcal{I}_1$	R^{-1}	$(\gamma - 1, \gamma, +1)$	$\sigma(M(\gamma))$
$\gamma = r = 1$	$\mathcal{I}_2$	SR^{-2}	$(1, 1, 0)$	$M(\gamma)$
$\frac{1}{2} < \gamma = r < 1$	$\mathcal{I}_1$	SR^{-1}	$\left(\frac{1}{1-\gamma}, \frac{\gamma}{(1-\gamma)^2}, -1\right)$	$R\sigma(M(\gamma))$
$\gamma = r = \frac{1}{2}$	$\mathcal{I}_1$	SR^{-1}	$(2, 2, 0)$	$R\sigma(M(\gamma))$
$\gamma = r < \frac{1}{2}$	$\mathcal{J}_0$	L^{-1}	$\left(\frac{\gamma}{1-\gamma}, \frac{\gamma}{(1-\gamma)^2}, +1\right)$	$\sigma(M(\gamma))$

Table 3. Definition of the map $\mathcal{P}_h$ for $\varepsilon = 0$.

Having constructed the map $(\mathbb{W}, \mathcal{P}_h)$, we now proceed to construct the associated suspension flow. Let $(z, \zeta) \in \mathcal{C}$ and $(\gamma, r, \varepsilon) := \mathcal{W}(z, \zeta)$. The first return of the horocycle flow $\tilde{h}_s^+$ along $W^+(z, \zeta)$ for negative
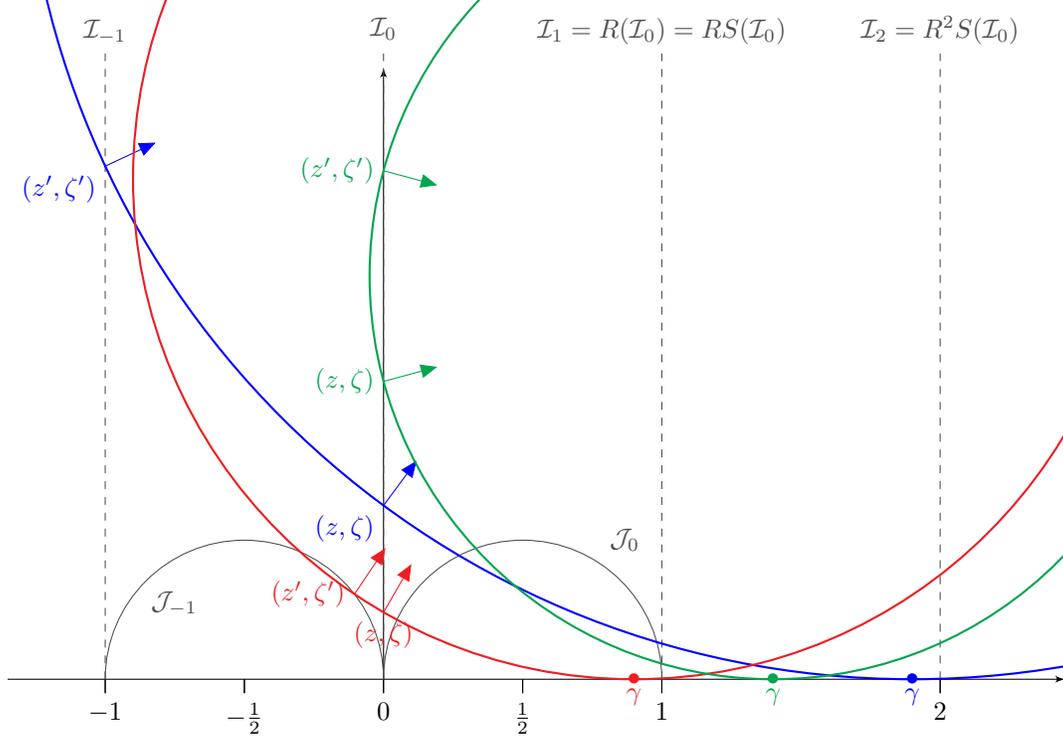
Figure 4. Action of the first return map $\mathcal{P}_h$ for $\varepsilon = -1$. The red horocycle has $\gamma < 1$ and $r > \gamma(1 + \gamma)$, so that the first return occurs on $\mathcal{J}_{-1}$; the blue horocycle has $\gamma > 1$ and $1 + \gamma < r < \gamma(1 + \gamma)$, so that it does not intersect $\mathcal{J}_{-1}$ and thus the first return occurs on $\mathcal{I}_{-1}$; the green horocycle has $\gamma < 1$ and $r < \min\{1 + \gamma, \gamma(1 + \gamma)\}$, so that the first return occurs on $\mathcal{I}_0$.

times s occurs at a time $s_h(z, \zeta) < 0$, which can accordingly be written as a function $s_h(\gamma, r, \varepsilon)$. Set

$$\Sigma_h := \{(\gamma, r, \varepsilon, \xi) \in \mathbb{W} \times \mathbb{R} \ : \ s_h(\gamma, r, \varepsilon) \leq \xi \leq 0\}$$

and use the map $\mathcal{P}_h$ to define the flow $\phi_s : \Sigma_h \to \Sigma_h$ by

$$\phi_s(\gamma, r, \varepsilon, \xi) = \begin{cases} (\gamma, r, \varepsilon, \xi + s) & \text{if } s_h(\gamma, r, \varepsilon) < \xi + s \leq 0 \\ (\mathcal{P}_h(\gamma, r, \varepsilon), 0) & \text{if } \xi + s = s_h(\gamma, r, \varepsilon) \end{cases}. \tag{4.4}$$

We made explicit the flow for s negative, the case s positive is defined analogously as $\mathcal{P}_h$ is an invertible map. The next Proposition summarizes some basic ergodic properties of the dynamical systems $(\mathbb{W}, \mathcal{P}_h)$ and (Σ_h, ϕ_s).

Proposition 4.2. (i) *The map $\mathcal{P}_h : \mathbb{W} \to \mathbb{W}$ preserves the infinite measure ν which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{W}$ with density $k(\gamma, r, \varepsilon) = r^{-2}$.*

(ii) *The flow $\phi_s : \Sigma_h \to \Sigma_h$ preserves the measure $\tilde{\nu}$ which is absolutely continuous with respect to the Lebesgue measure on Σ_h with density $\tilde{k}(\gamma, r, \varepsilon, \xi) = r^{-2}$. Moreover $\tilde{\nu}(\Sigma_h) = 4\pi^2/3$.*

(iii) *The dynamical system $(\Sigma_h, \tilde{\nu}, \phi_s)$ is isomorphic to the system $(S\mathcal{M}, \tilde{m}, \tilde{h}_s^+)$, where $\tilde{m}$ is the projection on $S\mathcal{M}$ of the Liouville measure $dm(x, y, \theta) = y^{-2} \, dx \, dy \, d\theta$ on $S\mathbb{H}$, with $z = x + iy$ and $\theta = \theta(\zeta)$.*

Proof. Given $(z, \zeta) \in S\mathbb{H}$ with $\theta(\zeta) \notin \{0, \pi\}$, consider its coordinates (γ, r, ξ) defined as $\gamma = \gamma_{z,\zeta}^+$, $r = r^+(z, \zeta)$ being the radius of the horocycle $W^+(z, \zeta)$, and $\xi \in \mathbb{R}$ such that $(z, \zeta) = h_\xi^+(z', \zeta')$ where z' is the point in

$W^+(z,\zeta)$ with maximum value of the y component. We obtain the following expressions for the coordinates (x, y, θ) of (z, ζ), where $z = x + iy$ and $\theta = \theta(\zeta)$:

$$x(\gamma, r, \xi) = \gamma - \frac{2r\xi}{\xi^2 + 1}, \quad y(\gamma, r, \xi) = \frac{2r}{\xi^2 + 1}, \quad \theta(\gamma, r, \xi) = -2 \arctan \frac{1}{\xi}. \tag{4.5}$$

From (4.5) it follows that the Liouville measure $dm(x, y, \theta) = y^{-2} \, dx \, dy \, d\theta$ becomes

$$dm(\gamma, r, \xi) = \frac{1}{2 \, r^2} \, d\gamma \, dr \, d\xi$$

in the coordinates (γ, r, ξ), and we set $d\tilde{\nu} = r^{-2} d\gamma dr d\xi$. Moreover by the definition of ξ, the horocycle flow h_s^+ acts in the coordinates (γ, r, ξ) simply by translation on ξ, that is

$$h_s^+(\gamma, r, \xi) = (\gamma, r, \xi + s).$$

These facts, together with $\tilde{m}(S\mathcal{M}) = 2\pi^2/3$, prove (ii) and (iii). Finally (i) follows from (ii). □

Let us calculate an explicit expression for the return time $s_h(\gamma, r, \varepsilon)$. From Lemma 4.1 it follows that $s_h(\gamma, r, \varepsilon)$ is the time spent by flowing along the horocycle until its intersection with the right representative of $\mathcal{I}_0$. Now, by the identification of $S\mathbb{H}$ with $PSL(2, \mathbb{R})$ (see (3.2)), the map $\mathcal{W}$ can be defined also as a map from $PSL(2, \mathbb{R})$ to $\mathbb{W}$. In particular

$$(z, \zeta) = \mathcal{W}^{-1}(\gamma, r, \varepsilon) \quad \Rightarrow \quad \Gamma_{(z,\zeta)} = \begin{pmatrix} \frac{\gamma}{\sqrt{2r}} & -\frac{r + \varepsilon\sqrt{r^2 - \gamma^2}}{\sqrt{2r}} \\ \frac{1}{\sqrt{2r}} & \frac{r - \varepsilon\sqrt{r^2 - \gamma^2}}{\gamma\sqrt{2r}} \end{pmatrix}. \tag{4.6}$$

Proposition 4.3. *For* $(z, \zeta) = \mathcal{W}^{-1}(\gamma, r, \varepsilon) \in \mathcal{C}$ *with* $z \in \mathcal{I}_0$ *and* $\theta(\zeta) \in (-\pi, 0)$, *let* $\Gamma_{(z,\zeta)} \in PSL(2, \mathbb{R})$ *be as in* (4.6). *Assume that the first return to* C *for* $\tilde{h}_s^+$ *with negative* s *along* $W^+(z, \zeta)$ *occurs on* $\Gamma^{-1}(\mathcal{I}_0)$ *with* $\Gamma \in PSL(2, \mathbb{Z})$, *and set* $(\gamma', r', \varepsilon') = \mathcal{P}_h(\gamma, r, \varepsilon)$ *and* $(z', \zeta') = \mathcal{W}^{-1}(\gamma', r', \varepsilon')$. *Write also the corresponding matrix* $\Gamma_{(z', \zeta')}$ *in the form* (4.6). *Then* $s_h(\gamma, r, \varepsilon)$ *satisfies the equation*

$$\begin{pmatrix} 1 & s_h(\gamma, r, \varepsilon) \\ 0 & 1 \end{pmatrix} = \Gamma_{(z,\zeta)}^{-1} \, \Gamma^{-1} \, \Gamma_{(z', \zeta')}.$$

Proof. The horocycle flow is represented by the right action of matrices in $PSL(2, \mathbb{R})$ as in (3.4). Hence

$$\Gamma_{h_{s_h(\gamma, r, \varepsilon)}^+(z, \zeta)} = \Gamma_{(z, \zeta)} \begin{pmatrix} 1 & s_h(\gamma, r, \varepsilon) \\ 0 & 1 \end{pmatrix}.$$

In addition, if the first return to $\mathcal{C}$ occurs on $\Gamma^{-1}(\mathcal{I}_0)$ at the point $\Gamma^{-1}(z', \zeta')$, we have

$$\Gamma_{h_{s_h(\gamma, r, \varepsilon)}^+(z, \zeta)} = \Gamma^{-1} \Gamma_{(z', \zeta')}.$$

This concludes the proof. □

Using the above Proposition along with Tables 1, 2, 3, and Equation (4.6), we can make the form of $s_h(\gamma, r, \varepsilon)$ explicit for all $(\gamma, r, \varepsilon) \in \mathbb{W}$. By means of the functions

$$u(\gamma, r, \varepsilon) := -\varepsilon \frac{\sqrt{r^2 - \gamma^2}}{\gamma}, \quad v(\gamma, r, \varepsilon) := \varepsilon \frac{r}{\gamma}, \quad \delta(\gamma, r, \varepsilon) := \delta_0(\varepsilon), \tag{4.7}$$

where $\delta_0(\cdot)$ takes the value 1 at 0, and vanishes elsewhere, we get the expressions collected in the following two tables. One may readily check that $s_h(\gamma, r, \varepsilon) < 0$ for all $(\gamma, r, \varepsilon) \in \mathbb{W}$ and also $s_h \in L^1(\mathbb{W}, \nu)$, so that the set $(\Sigma_h, \tilde{\nu})$ as defined in Proposition 4.2 is a finite measure space.

Conditions on γ and r	$s_h(\gamma, r, \varepsilon)$
$r < 1 - \gamma$	$((u - v) \circ \mathcal{P}_h)(\gamma, r, \varepsilon) - (u - v)(\gamma, r, \varepsilon) + (\delta \circ \mathcal{P}_h + \delta)(\gamma, r, \varepsilon)$
$r \geq 1 - \gamma$	$((u + v) \circ \mathcal{P}_h)(\gamma, r, \varepsilon) - (u + v)(\gamma, r, \varepsilon) - (\delta \circ \mathcal{P}_h + \delta)(\gamma, r, \varepsilon)$

Table 4. The return time s_h for $\varepsilon \in \{0, +1\}$ in terms of the functions defined in (4.7).

Conditions on γ and r	$s_h(\gamma, r, \varepsilon)$
$r < \min\{1 + \gamma, \gamma(1 + \gamma)\}$	$(u \circ \mathcal{P}_h - u)(\gamma, r, \varepsilon)$
$1 + \gamma \leq r < \gamma(1 + \gamma)$	$((u - v) \circ \mathcal{P}_h)(\gamma, r, \varepsilon) - (u - v)(\gamma, r, \varepsilon) + (\delta \circ \mathcal{P}_h + \delta)(\gamma, r, \varepsilon)$
$r \geq \gamma(1 + \gamma)$	$((u + v) \circ \mathcal{P}_h)(\gamma, r, \varepsilon) - (u + v)(\gamma, r, \varepsilon) - (\delta \circ \mathcal{P}_h + \delta)(\gamma, r, \varepsilon)$

Table 5. The return time s_h for $\varepsilon = -1$ in terms of the functions defined in (4.7).

Remark 4.4. The expressions for the return time $s_h(\gamma, r, \varepsilon)$ displayed in Table 4 and 5 show that the suspension flow $\phi_s : \Sigma_h \to \Sigma_h$ over $(\mathbb{W}, \mathcal{P}_h)$, defined in (4.4), has roof function which is not cohomologous to a constant. If it were, there should be a function $\Psi : \mathbb{W} \to \mathbb{R}$ so that

$$s_h - \int_{\mathbb{W}} s_h \, d\nu = \Psi \circ \mathcal{P}_h - \Psi \,.$$

This does not happen because the sets on which we have defined the return time are not $\mathcal{P}_h$-invariant. For example, choosing (γ, r, ε) with $\varepsilon = +1$ and $r \in ((1 - \gamma)(1 - 2\gamma), 1 - \gamma)$, we see that $(\gamma', r', \varepsilon') = \mathcal{P}_h(\gamma, r, \varepsilon)$ satisfies $\varepsilon' = +1$ and $r' = r/(1 - \gamma)^2 > 1 - \gamma' = \gamma/(1 - \gamma)$. On the other hand, the reader will have noticed that the function v changes sign when passing from the first to the second row in Table 4.

5. Closed horocycles and the permuted Stern-Brocot trees

We have shown in Proposition 4.2 that the suspension flow ϕ_s over the system $(\mathbb{W}, \mathcal{P}_h)$ is isomorphic to the horocycle flow on $S\mathcal{M}$ with respective invariant measures. This implies that there is a one-to-one correspondence between the closed horocycles on $S\mathcal{M}$ which intersect $\mathcal{C}$ and the periodic points of the map $\mathcal{P}_h$. This correspondence may be stated by saying that for a point $(z, \zeta) \in \mathcal{C}$ with $(\gamma, r, \varepsilon) = \mathcal{W}(z, \zeta) \in \mathbb{W}$, there exist $\Gamma \in SL(2, \mathbb{Z})$ and $s \in (-\infty, 0)$ such that $(z, \zeta) = \Gamma_*(\tilde{h}_s^+(z, \zeta))$ if and only if there exists $n \in \mathbb{N}$ such that $\mathcal{P}_h^n(\gamma, r, \varepsilon) = (\gamma, r, \varepsilon)$.

It is well-known (see e.g. [11, chap. 11]) that the closed horocycles are generated by the points $(iy, i) \in S\mathcal{M}$. More precisely, they are the projections on $S\mathcal{M}$ of the horocycles $W^+(iy, i)$ in $S\mathbb{H}$ given by the set of points $(x + iy, i)$ with $x \in \mathbb{R}$. Since $h_s^+ : (x + iy, i) \to (x + sy + iy, i)$, they have length $1/y$ when projected on $S\mathcal{M}$.

Proposition 5.1. *If $y \leq \frac{1}{2}$, the projection of the horocycle $W^+(iy, i)$ on $S\mathcal{M}$ intersects $\mathcal{C}$.*

Proof. By applying S to $W^+(iy, i)$ we see that $S(W^+(iy, i))$ is the circle tangent to $\mathbb{R}$ at 0 and of radius $1/(2y) \geq 1$. Then $R(S(W^+(iy, i)))$, the circle tangent to $\mathbb{R}$ at 1 and of radius $1/(2y)$, is equal to $W^+(z, \zeta)$ with $(z, \zeta) \in \mathcal{C}$ and $z \in \mathcal{I}_0$. If $y < 1/2$, we can choose (z, ζ) to be the determined by the intersection of $R(S(W^+(iy, i)))$ with $\mathcal{I}_0$ with smaller value of the variable y, therefore it follows that $\mathcal{W}(z, \zeta) = (1, 1/(2y), -1)$. If $y = 1/2$, the intersection of $R(S(W^+(iy, i)))$ with $\mathcal{I}_0$ consists of a single point, and therefore $\mathcal{W}(z, \zeta) = (1, 1, 0)$. $\square$

Remark 5.2. In the previous Proposition, one can use the identification between $S\mathbb{H}$ and $PSL(2, \mathbb{R})$ recalled in Section 3, to prove that for the point $(z, \zeta) \in \mathcal{C}$ in the intersection between $R(S(W^+(iy, i)))$ and $\mathcal{I}_0$ we have

$$\Gamma_{(z,\zeta)} = \begin{pmatrix} \sqrt{y} & \frac{-1 + \sqrt{1 - 4y^2}}{2\sqrt{y}} \\ \sqrt{y} & \frac{1 + \sqrt{1 - 4y^2}}{2\sqrt{y}} \end{pmatrix}.$$

13

Moreover $\tilde{h}_s^+(z, \zeta) = (z, \zeta)$ for $s = -1/y$ since

$$SR^{-2}\begin{pmatrix} \sqrt{y} & \frac{-1+\sqrt{1-4y^2}}{2\sqrt{y}} \\ \sqrt{y} & \frac{1+\sqrt{1-4y^2}}{2\sqrt{y}} \end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{y} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{y} & \frac{-1+\sqrt{1-4y^2}}{2\sqrt{y}} \\ \sqrt{y} & \frac{1+\sqrt{1-4y^2}}{2\sqrt{y}} \end{pmatrix}.$$

A consequence of Proposition 5.1 is that closed horocycles of length $\ell \geq 2$ correspond to periodic points of the map $\mathcal{P}_h$, and in particular the horocycle of length ℓ corresponds to the point $(1, \ell/2, \varepsilon) \in \mathbb{W}$ with $\varepsilon = \pm 1$ if $\ell > 2$, and $\varepsilon = 0$ if $\ell = 2$. In the following we characterise the periodic points of $\mathcal{P}_h$, their orbits, and their periods.

Example 5.3. Let us follow the orbits of two points of the form $(1, r, \varepsilon)$ with $r \in \mathbb{N}$ under the action of $\mathcal{P}_h$. First, setting $r = 1$, the point $(1, 1, 0)$ is a fixed point of $\mathcal{P}_h$ as shown in Table 3.
Setting instead $r = 2$, we can consider the point $(1, 2, -1)$. Applying $\mathcal{P}_h$ according to Table 2 for $r = \gamma(1+\gamma)$, we find $\mathcal{P}_h(1, 2, -1) = (1/2, 1/2, 0)$. Then $\mathcal{P}_h(1/2, 1/2, 0) = (2, 2, 0)$ and $\mathcal{P}_h(2, 2, 0) = (1, 2, +1)$ by Table 3. Finally we use Table 1 to get $\mathcal{P}_h(1, 2, +1) = (1, 2, -1)$. Hence $(1, 2, -1)$ is periodic of period 4.

First we characterise the periodic points of $\mathcal{P}_h$ without using the known results on closed horocycles.

Proposition 5.4. *If a point* $(\gamma, r, \varepsilon) \in \mathbb{W}$ *is periodic for* $\mathcal{P}_h$ *then* $\gamma \in \mathbb{Q}$.

Proof. By using the map $\mathcal{W}$ given in (4.3), one realises that a closed horocycle on $S\mathcal{M}$ which intersects $\mathcal{C}$ can be associated to a point (γ, r, ε) in $\mathbb{W}$ with $\varepsilon \in \{-1, 0\}$. Indeed, if $(\gamma, r, +1) = \mathcal{W}(z, \zeta)$ is periodic, then also $(\gamma, r, -1)$ must be periodic, since they define the same closed horocycle.
Then, looking at Table 3 it follows that if $(\gamma, r, 0)$ is periodic then either $(\gamma, r) = (1, 1)$ or $(2, 2)$, and we are dealing with the fixed point or the period 4 orbit of Example 5.3, or else it can be reduced to a periodic point with $\varepsilon = -1$.
We are thus left to study a periodic point of the form $(\gamma, r, -1)$. Studying its orbit and the structure of the correspondent closed horocycle (see Figure 4), one readily checks that there exists $k \geq 1$ such that $\mathcal{P}_h^k(\gamma, r, -1) = (\gamma, r, +1)$. Hence by Lemma 4.1 there exists $\Gamma_1 \in SL(2, \mathbb{Z})$ which yields $\gamma = \Gamma_1(\gamma)$ and $r = \text{def}\,(\Gamma_1(\gamma))\,r$. Moreover, since the orbit of $(\gamma, r, -1)$ is periodic, there exists $m \geq 1$ such that $\mathcal{P}_h^m(\gamma, r, +1) = (\gamma, r, -1)$, so that $n = m + k$ is the period of $(\gamma, r, -1)$. Again this implies that there exists $\Gamma_2 \in SL(2, \mathbb{Z})$ which yields $\gamma = \Gamma_2(\gamma)$ and $r = \text{def}\,(\Gamma_2(\gamma))\,r$.
From Tables 1 and 2 it follows that Γ_1 could also be the identity, but Γ_2 cannot be such since $\Gamma_2^{-1}(\mathcal{I}_0)$ is the set on which one of the return along the horocycle occurs, after the return corresponding to $(\gamma, r, +1)$. This means that γ satisfies

$$\frac{a\gamma + b}{c\gamma + d} = \gamma \quad \text{and} \quad |c\gamma + d| = 1$$

for $a, b, c, d \in \mathbb{Z}$. Therefore $\gamma \in \mathbb{Q}$. □

The following analysis of the relationship between the periodic orbits and the permuted Stern-Brocot tree $\hat{\mathcal{T}}$ provides the opposite implication, that if $\gamma \in \mathbb{Q}$ then (γ, r, ε) is periodic for all r, ε.

5.1. **Periodic orbits of $\mathcal{P}_h$ and the permuted Stern-Brocot tree.** We start by studying the orbit of the points $(1, r, -1) \in \mathbb{W}$ with $r > 1$. The case $r = 1$ is easy since the corresponding point in $\mathbb{W}$ is $(1, 1, 0)$ and has been already discussed in Example 5.3.
From the argument in the proof of Proposition 5.4 it follows that there exists $k \geq 1$ such that $\mathcal{P}_h^k(1, r, -1) = (1, r, +1)$. We then apply Table 1 to get $\mathcal{P}_h(1, r, +1) = (1, r, -1)$. Hence the points $(1, r, -1)$ are periodic for all $r > 1$, and all the points in their orbits correspond to returns to $\mathcal{C}$ occurring on sets $\Gamma(\mathcal{I}_0)$, for $\Gamma \in SL(2, \mathbb{Z})$, in $\{z \in \mathbb{H} : \text{Re}(z) \leq 0\}$. In particular, let W_r^+ be the horocycle which is tangent to $\mathbb{R}$ at 1 and has radius $r > 1$. We need to determine all the sets $\Gamma^{-1}(\mathcal{I}_0)$ with $\Gamma \in SL(2, \mathbb{Z})$ which intersect W_r^+ in $\{z \in \mathbb{H} : \text{Re}(z) \leq 0\}$.

The easy case is given by the lines $\mathcal{I}_n = R^n(\mathcal{I}_0)$ with $n \leq 0$. It is clear that $W_r^+ \cap \mathcal{I}_n$ is not empty if and only if $r \geq 1 - n$. It turns out that it is more useful to study the intersections with the geodesics with rational end points in the intervals $[n - 1, n]$ for $n \leq 0$.

For a Farey pair $p/q, p'/q' \in \mathbb{Q}^+$ with $p/q < p'/q'$ (see Section 2), let

$$\mathfrak{g}\left(\frac{p}{q}, \frac{p'}{q'}\right) := \left\{ x + iy \in \mathbb{H} : \left(x + \frac{1}{2}\frac{p}{q} + \frac{1}{2}\frac{p'}{q'} \right)^2 + y^2 = \frac{1}{(2qq')^2} \right\} \tag{5.1}$$

be the geodesic with end points in $-p/q$ and $-p'/q'$. For example $\mathfrak{g}(0/1, 1/1) = \mathcal{J}_{-1}$.

Lemma 5.5. *For a Farey pair $p/q, p'/q' \in \mathbb{Q}^+$ we have*

$$\mathfrak{g}\left(\frac{p}{q}, \frac{p'}{q'}\right) = \left(\hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'}\right) \right)^{-1} (\mathcal{I}_0),$$

where $\hat{M}(a/b) \in \{L, R\}^*$ is the reverse word associated to the rational number a/b as in Section 2.2, and every symbol in $\hat{M}(a/b)$ is interpreted as a matrix in $SL(2, \mathbb{Z})$ acting on $\mathbb{H}$.

Proof. Let $p/q < p'/q'$, then by the discussion in Section 2.1 and Lemma 2.3

$$M\left(\frac{p}{q} \oplus \frac{p'}{q'}\right) = \begin{pmatrix} p' & p \\ q' & q \end{pmatrix} \quad \text{and} \quad \hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'}\right) = \begin{pmatrix} q & p \\ q' & p' \end{pmatrix}.$$

Then

$$\left(\hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'}\right) \right)^{-1} = \begin{pmatrix} p' & -p \\ -q' & q \end{pmatrix}$$

which as a matrix in $SL(2, \mathbb{Z})$ acting on $\mathbb{H}$ maps the end points $0, \infty$ of the geodesic $\mathcal{I}_0$ into the points $-p/q$ and $-p'/q'$ respectively, which are the end points of the geodesic $\mathfrak{g}(p/q, p'/q')$. Hence the lemma is proved. $\square$

From the previous lemma it follows that the geodesics $\mathfrak{g}(p/q, p'/q')$ may be arranged in a tree which corresponds to the Stern-Brocot tree $\mathcal{T}$ by the bijection

$$\frac{m}{s} = \frac{p}{q} \oplus \frac{p'}{q'} \quad \leftrightarrow \quad \mathfrak{g}\left(\frac{p}{q}, \frac{p'}{q'}\right)$$

with the convention that $1/1$ corresponds to $\mathcal{I}_0 = \mathfrak{g}(0/1, 1/0)$. If m/s is a daughter of m'/s' in $\mathcal{T}$, then the two corresponding geodesics have one end point in common, given by the common parent of m/s and m'/s', and the geodesic corresponding to m/s lies in $\mathbb{H}$ below that of m'/s'.

At the same time, the tree of geodesics is in bijection with the permuted Stern-Brocot tree $\hat{\mathcal{T}}$, since for $\mathfrak{g}(p/q, p'/q')$ the word $\hat{M}(p/q \oplus p'/q')$ gives the position of $m/s = p/q \oplus p'/q'$ in $\hat{\mathcal{T}}$. This relation plays a fundamental role when studying the action of the map $\mathcal{P}_h$ thanks to Lemma 4.1.

We now give an elementary result on the intersection of $\mathfrak{g}(p/q, p'/q')$ with the horocycle W_r^+ tangent to $\mathbb{R}$ at 1 and of radius $r > 1$.

Lemma 5.6. *For a Farey pair $p/q, p'/q' \in \mathbb{Q}^+$, the horocycle tangent to $\mathbb{R}$ at $\rho > 0$ and of radius $r \geq \rho$ intersects $\mathfrak{g}(p/q, p'/q')$ if and only if $r \geq (q\rho + p)(q'\rho + p')$.*

Proof. We look for an equivalent condition to the existence of solutions to the system given by the polynomial equation in (5.1) describing $\mathfrak{g}(p/q, p'/q')$ and the polynomial equation $(x - \rho)^2 + y^2 - 2ry = 0$ describing the horocycle tangent to $\mathbb{R}$ at $\rho > 0$ and of radius $r \geq \rho$. The two circles have a tangency point if and only if

$$\left(r + \frac{1}{2qq'} \right)^2 = r^2 + \left(\rho + \frac{1}{2}\left(\frac{p}{q} + \frac{p'}{q'} \right) \right)^2$$

15

which, using that $p'q - pq' = 1$ and elementary computations, is equivalent to $r = (q\rho + p)(q'\rho + p')$. Then the two circles intersect if and only if the radius r is greater than or equal to the values for which they are tangent. $\qquad\square$

We are now ready to state the main result of this section, which provides a precise characterisation of the orbits under $\mathcal{P}_h$ of the periodic points $(1, r, -1)$ for all $r \geq 1$ and of all the periodic points of $\mathcal{P}_h$.

Definition 5.7. Given $p/q \in \mathbb{Q}^+$ with p, q coprime, we call *energy* of p/q the number $\xi(p/q) := pq$.

Theorem 5.8. *For all $r \geq 1$, the point $(1, r, -1) \in \mathbb{W}$ is periodic for $\mathcal{P}_h$ and has period $\mathrm{per}(r)$ given by*

$$\mathrm{per}(r) = 2 \cdot \# \left\{ (a,b) \in \mathbb{N} \times \mathbb{N} : (a,b) = 1, ab < r \right\} + \# \left\{ (a,b) \in \mathbb{N} \times \mathbb{N} : (a,b) = 1, \ ab = r \right\}.$$

In addition, a point $(\gamma, r, \varepsilon) \in \mathbb{W}$ is periodic if and only if $\gamma \in \mathbb{Q}$, and in this case if $\gamma = a/b$ with a, b coprime, it is in the orbit of the point $(1, rb^2, -1)$.

Proof. Let us consider a point $(1, r, -1) \in \mathbb{W}$ with $r \geq 1$. Note that the case $r = 1$ has been already discussed in Example 5.3, from which $\mathrm{per}(1) = 1$. Then let $r > 1$. We have seen that $(1, r, -1)$ is periodic and that $(1, r, \pm 1)$ are in the same orbit. Moreover, all the other points in the orbit of $(1, r, -1)$ are given by the intersections of W_r^+, the horocycle tangent to $\mathbb{R}$ at 1 and of radius r, with the sets $\Gamma^{-1}(\mathcal{I}_0)$ for some $\Gamma \in SL(2, \mathbb{Z})$. In particular, by Lemma 4.1, if W_r^+ intersects $\Gamma^{-1}(\mathcal{I}_0)$ then the point $(\gamma', r', \varepsilon')$, with $\gamma' = \Gamma(1)$ and $r' = \mathrm{def}\,(\Gamma(1))r$, is in the orbit of $(1, r, -1)$ for $\mathcal{P}_h$. From this it follows first of all that a first class of points in the orbit of $(1, r, -1)$ are given by the intersections of W_r^+ with $\mathcal{I}_n = R^n(\mathcal{I}_0)$ for $1 - r \leq n < 0$. If $n = 1 - r$ there is only one intersection, which gives $\varepsilon' = 0$ for the corresponding point, and if $n > 1 - r$ there are two intersections giving $\varepsilon' = \pm 1$ for the point in the orbit. By Lemma 4.1 the sets

$$\left\{ (1 + |n|, r, \pm 1) : n \in \mathbb{Z}^- , 1 \leq |n| < r - 1 \right\} \cup \left\{ (1 + |n|, r, 0) : n \in \mathbb{Z}^- , |n| = r - 1 \right\} \tag{5.2}$$

are made of points in the orbit of $(1, r, -1)$. In the formula for the period of $(1, r, -1)$, together with $(1, r, \pm 1)$ they correspond to all the points with $p \in \mathbb{N}$ and $q = 1$.

The other class of points in the orbit of $(1, r, -1)$ is given by the intersections of W_r^+ with the geodesics $\mathfrak{g}(p/q, p'/q')$ defined in (5.1). It is an elementary observation that, together with $\mathcal{I}_n$, these geodesics cover all the sets $\Gamma^{-1}(\mathcal{I}_0)$ for some $\Gamma \in SL(2, \mathbb{Z})$ in $\mathbb{H} \cap \{\mathrm{Re}(z) < 0\}$. By Lemma 5.6 the horocycle W_r^+ intersects $\mathfrak{g}(p/q, p'/q')$ if and only if $r \geq (p + q)(p' + q')$, and by Lemma 5.5 the intersection corresponds to the point $(\gamma', r', \varepsilon)$ with

$$\gamma' = \hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'} \right)(1), \quad r' = \mathrm{def}\left(\hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'} \right)(1) \right) r,$$

with $\varepsilon' = 0$ if the two circles are tangent, $\varepsilon' = \pm 1$ if there are two intersections. Let $m/s = p/q \oplus p'/q'$. Then $\gamma' = \hat{M}(m/s)(1)$ is the fraction that in $\hat{\mathcal{T}}$ occupies the place of m/s in $\mathcal{T}$.

Using the notations introduced in Section 2.1 and Lemma 2.3 we have

$$M\left(\frac{p}{q} \oplus \frac{p'}{q'} \right) = \begin{pmatrix} p' & p \\ q' & q \end{pmatrix} \quad \text{and} \quad \hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'} \right) = \begin{pmatrix} q & p \\ q' & p' \end{pmatrix}.$$

In particular, setting $\gamma' = m'/s'$, it follows that

$$\gamma' = \frac{m'}{s'} = \frac{p + q}{p' + q'}.$$

Moreover, using Definition 5.7 of the energy of a positive rational number, we obtain

$$\xi(\gamma') = (p + q)(p' + q') \quad \text{and} \quad \mathrm{def}\left(\hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'} \right)(1) \right) = (p' + q')^{-2} = (s')^{-2}. \tag{5.3}$$

16

We can sum up the consequences of (5.2) and (5.3) by stating that a rational number a/b, with $(a, b) = 1$, is the first component of a point in the orbit of $(1, r, -1)$ under $\mathcal{P}_h$ if and only if r is greater or equal than the energy $\xi(a/b)$, and more precisely the orbit of $(1, r, -1)$ is given by

$$\left\{ \left(\frac{a}{b}, \frac{r}{b^2}, \varepsilon \right) : (a, b) = 1, \, ab < r, \, \varepsilon \in \{-1, +1\} \right\} \cup \left\{ \left(\frac{a}{b}, \frac{r}{b^2}, 0 \right) : (a, b) = 1, \, ab = r \right\}.$$

The first part of the statement of the theorem is proved. Proposition 5.4 and the characterisation of the orbit of $(1, r, -1)$ readily imply the second part of the statement. $\square$

The result of Theorem 5.8 may be refined by studying the dynamical ordering of the points in the orbit of $(1, r, -1)$. By looking at the intersections of a horocycle W_r^+ with the geodesics $\mathfrak{g}(p/q, p'/q')$ defined in (5.1), and at the ordering of these intersections described by the relative geometrical positions of the geodesics in $\mathbb{H}$, we can produce an algorithm to determine the points of the periodic orbit of $(1, r, -1)$ in their dynamical order up to $(1, r, +1)$ by using the permuted Stern-Brocot tree $\hat{\mathcal{T}}$. A description of this algorithm is given in Appendix B.

Remark 5.9. Adopting an alternative point of view, we can look at the period $\text{per}(r)$ of the $\mathcal{P}_h$-periodic orbit containing $(1, r, -1)$ as the number of intersections of the horocycle $W^+(iy, i) = \{x + iy : x \in \mathbb{R}\}$, with $y = 1/(2r)$, with the (geodesic) sides of the triangles forming the Farey tassellation of the domain $\{x + iy \in \mathbb{H} : 0 \le x \le 1\}$ (see Figure 5 below).
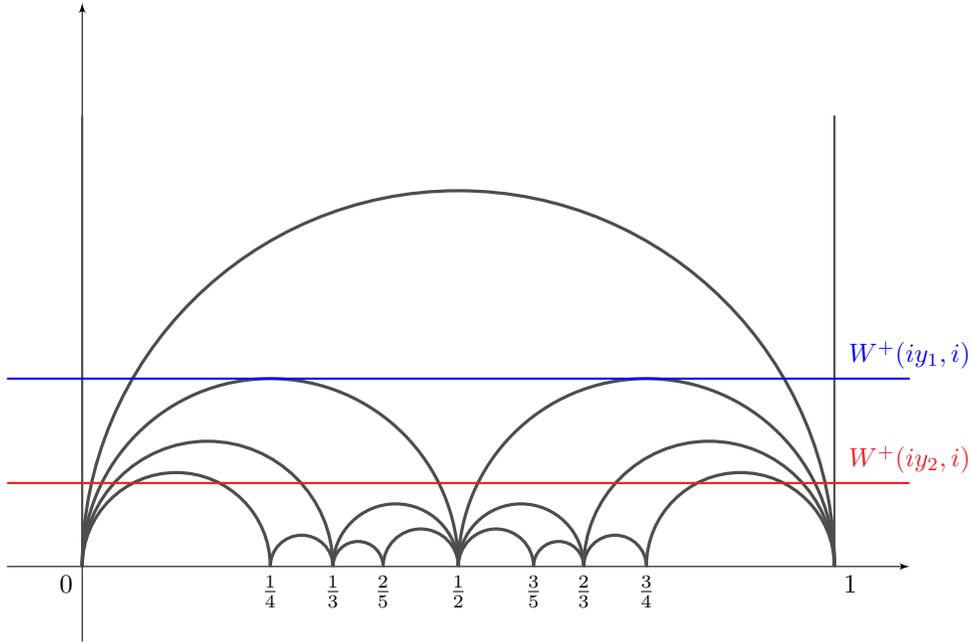

Figure 5. Farey tassellation of the domain $\{x + iy \in \mathbb{H} : 0 \le x \le 1\}$ up to level 4 as defined in Section 2. We also show two horizontal horocycles $W^+(iy_k, i)$ with $y = 1/(2r_k)$, $k = 1, 2$: the blue horocycle has $r_1 = 2$ and is tangent to some of the geodesic sides of the triangles of the Farey tassellation; the red horocycle has $4 < r_2 < 5$.

In particular, the horocycle $W^+(iy, i)$ intersects the geodesic joining a Farey pair $p/q, p'/q' \in [0, 1]$, with $p/q < p'/q'$, if and only if $y \le 1/(2qq')$. Note moreover that given a Farey pair $p/q, p'/q' \in [0, 1]$, the symmetric pair $(q - p)/q, (q' - p')/q'$ is in $[0, 1]$ and yields a geodesic side of equal "height" $(= 1/(2qq'))$, and there are no other pairs with this property. By the way, this symmetry accounts for the corresponding symmetry of the walks associated to the dynamics of the periodic points of $\mathcal{P}_h$ on the permuted Stern-Brocot

tree (see Appendix B). The formula for the period can thus be rewritten as

$$\text{per}(r) = 2 \cdot \# \left\{ \frac{p}{q}, \frac{p'}{q'} \text{ Farey pairs in } [0,1] \ : \ qq' < r \right\} + \# \left\{ \frac{p}{q}, \frac{p'}{q'} \text{ Farey pairs in } [0,1] \ : \ qq' = r \right\}.$$

To further clarify the connection with the construction made in the proof of Theorem 5.8, we observe that the projection on SM of the geodesic joining a Farey pair $p/q, p'/q'$, via RS, is the geodesic joining $P/Q, P'/Q'$ with $P = q - p$ and $Q = p$ (and similar primed relations), so that $(P + Q)(P' + Q') = qq'$.

Finally, this picture would also provide a sort of duality between periodic horocycles and scattering geodesics (i.e. vertical geodesics landing at rational points), in the sense that, in order to know the corresponding paths on the tree, in the first case we have to look at the intersections (of the horocycle) with Farey triangles (arcs of geodesics), while in the second case we have to look at the intersections (of the scattering geodesics) with Ford circles (horocycles).

5.2. **Distribution of periodic orbits.** According to Theorem 5.8, the periodic orbits of $\mathcal{P}_h$ are nothing more than the orbits of the one-parameter family of points $\{(1, R, -1)\}_{R \in (1,\infty)}$ in $\mathbb{W}$. Denoting by $\mathcal{O}(1, R, -1)$ the orbit of $(1, R, -1)$, which consists of $\text{per}(R)$ points, we first prove the following proposition.

Proposition 5.10. *The period* $\text{per}(R)$ *of the* $\mathcal{P}_h$*-periodic orbits* $\mathcal{O}(1, R, -1)$ *satisfies*

$$\text{per}(R) = \frac{12}{\pi^2} R \log R + O(R), \quad R \to \infty.$$

Proof. By the formula for the period $\text{per}(R)$ of the point $(1, R, -1)$ given in Theorem 5.8, we have to estimate the asymptotic behaviour of the cardinality of the set $\Pi(R) := \{(a, b) \in \mathbb{N} \times \mathbb{N} \ : \ (a, b) = 1, \ ab \leq R\}$, that is

$$\#\Pi(R) = \sum_{n=1}^{R} \sum_{(a,b)=1, \ ab=n} 1$$

Now, the inner sum counts the number of ways in which one can split the set of prime numbers dividing n into two disjoint sets. This in turn is equal to the number of square-free divisors of n. It is a standard result in number theory (see e.g. [3]) that is obtained by summing over the divisors on n, denoted $d|n$, the square of the Möbius function $\mu(\cdot)$, that is $\mu^2(d) = 1$ if $d|n$ and it is square-free, and $\mu^2(d) = 0$ if $d|n$ but is not square-free. Therefore,

$$\#\Pi(R) = \sum_{n=1}^{R} \sum_{(a,b)=1, \ ab=n} 1 = \sum_{n=1}^{R} \sum_{d|n} \mu^2(d).$$

The last sum may be rewritten by letting d vary between 1 and R and writing $n = qd \leq R$, hence we get

$$\sum_{n=1}^{R} \sum_{d|n} \mu^2(d) = \sum_{d=1}^{R} \mu^2(d) \sum_{q \leq \frac{R}{d}} 1 = \sum_{d=1}^{R} \mu^2(d) \left\lfloor \frac{R}{d} \right\rfloor = R \sum_{d=1}^{R} \frac{\mu^2(d)}{d} + O(R),$$

where in the last equality we have used that $\mu^2(d) \leq 1$ for all d. At this point, using $\sum_{d \leq R} \mu^2(d) = 6R/\pi^2 + O(\sqrt{R})$ (see e.g. [14, Theorem 333]), we can write by means of Abel's identity (see e.g. [3, Theorem 4.2])

$$\sum_{d=1}^{R} \frac{\mu^2(d)}{d} = \frac{1}{R} \sum_{d=1}^{R} \mu^2(d) + \int_{1}^{R} \frac{1}{t^2} \left(\sum_{d \leq t} \mu^2(d) \right) dt - 1 =$$

$$= \frac{6}{\pi^2} + O(R^{-\frac{1}{2}}) + \frac{6}{\pi^2} \int_{1}^{R} \frac{1}{t} dt + O\left(\int_{1}^{R} t^{-\frac{3}{2}} dt \right) - 1 =$$

$$= \frac{6}{\pi^2} \log R + O(1).$$

Finally

$$\mathrm{per}(R) = 2\,\#\Pi(R) + O(R) = 2R \sum_{d=1}^{R} \frac{\mu^2(d)}{d} + O(R) = \frac{12}{\pi^2}\,\log R + O(R).$$

and the proof is finished. □

Furthermore, a more refined number-theoretic argument allows us to conclude that the periodic orbits are equidistributed with respect to the $\mathcal{P}_h$-invariant measure ν. Although this is not surprising after the results in [34, 26, 28, 6], we believe it is interesting as a further case of equidistribution of the periodic orbits for a dynamical system with infinite invariant measure (see [16]).

Theorem 5.11. *Let ν be the $\mathcal{P}_h$-invariant measure on $\mathbb{W}$ with density $k(\gamma, r, \varepsilon) = r^{-2}$. Then for all $A \subseteq \mathbb{W}$ measurable sets with $0 < \nu(A) < \infty$, we have*

$$\sum_{(\gamma,r,\varepsilon)\in\mathcal{O}(1,R,-1)} \chi_A(\gamma, r, \varepsilon) = \frac{3}{\pi^2}\,R\,\nu(A) + O\left(\sqrt{R}\,\log R\right).$$

In particular, for $A, B \subseteq \mathbb{W}$ measurable sets with $0 < \nu(A), \nu(B) < \infty$, we have

$$\lim_{R\to\infty} \frac{\sum_{(\gamma,r,\varepsilon)\in\mathcal{O}(1,R,-1)} \chi_A(\gamma, r, \varepsilon)}{\sum_{(\gamma,r,\varepsilon)\in\mathcal{O}(1,R,-1)} \chi_B(\gamma, r, \varepsilon)} = \frac{\nu(A)}{\nu(B)}.$$

Proof. We know from Theorem 5.8 that

$$\mathcal{O}(1, R, -1) = \left\{ \left(\frac{a}{b}, \frac{R}{b^2}, \pm 1\right) \,:\, (a, b) = 1\,,\, ab < R \right\} \cup \left\{ \left(\frac{a}{b}, \frac{R}{b^2}, 0\right) \,:\, (a, b) = 1\,,\, ab = R \right\}.$$

We study the distribution of the points of the orbits $\mathcal{O}(1, R, -1)$ as $R \to \infty$ for a set of the form

$$A = \{(\gamma, r, \varepsilon) \in \mathbb{W} \,:\, s < \gamma < t\,,\, u < r < v\,,\, \varepsilon \neq -1\} \tag{5.4}$$

with $s, t, u, v \in \mathbb{R}^+$. The choice $\varepsilon \neq -1$ in A does not reduce the generality of the result as it is evident from the structure of points in $\mathcal{O}(1, R, -1)$ and the fact that the set $\mathbb{W} \cap \{\varepsilon = 0\}$ has vanishing ν-measure. Moreover if the result holds for all open rectangles in (γ, r) it holds for all measurable sets. Finally, it is enough to prove the result for $0 < s < t < 1$, because if a point with $\gamma = a/b$ is in $\mathcal{O}(1, R, -1)$ then also a point with $\gamma = b/a$ is in $\mathcal{O}(1, R, -1)$, and in the r-components of the two points the roles of a and b are exchanged. Hence any set in $\mathbb{W}$ can be split in two subsets, one with γ-component in $(0, 1)$ and the other with γ-component in $(1, \infty)$, and it is enough to prove the result for one of these.

Let then $s, t, u, v \in \mathbb{R}^+$ with $0 < s < t \leq 1$ and A as in (5.4). The set

$$\Pi_A(R) := \left\{ \frac{a}{b} \,:\, (a, b) = 1\,,\, ab \leq R\,,\, s < \frac{a}{b} < t\,,\, u < \frac{R}{b^2} < v \right\}$$

characterizes the γ-component of the points in $A \cap \mathcal{O}(1, R, -1)$, so that

$$\sum_{(\gamma,r,\varepsilon)\in\mathcal{O}(1,R,-1)} \chi_A(\gamma, r, \varepsilon) = \#\,\Pi_A(R).$$

First of all, $\Pi_A(R)$ can be written as

$$\Pi_A(R) = \left\{ \frac{a}{b} \,:\, (a, b) = 1\,,\, ab \leq R\,,\, sb < a < bt\,,\, \sqrt{\frac{R}{v}} < b < \sqrt{\frac{R}{u}} \right\}$$

so that the order of s, t, u, v in $\mathbb{R}^+$ changes the description of the set. One realizes that it is enough to consider the cases

$$s < t = u < v \quad \text{and} \quad s \leq u < v = t.$$

19

as the sets satisfying any of the other possibilities can be reduced to the union of sets falling in one of these two cases. Anyway the computations are similar and the result follows by applying the same ideas[2].

We are now ready to study the asymptotic behaviour of $\#(A \cap \mathcal{O}(1, R, -1))$ as $R \to \infty$ for A defined as in (5.4) with $0 < s < t = u < v$ and $t < 1$. Note that A is a rectangle in $\mathbb{W}$ as the condition $r \geq \gamma$ is satisfied for all points in A. Hence

$$\nu(A) = \int_s^t \left(\int_u^v r^{-2}\, dr \right) d\gamma = (t - s) \left(\frac{1}{u} - \frac{1}{v} \right). \tag{5.5}$$

We have

$$\# \Pi_A(R) = \sum_{\sqrt{R/v} \leq b \leq \sqrt{R/u}} \sum_{\substack{(a,b)=1 \\ bs \leq a \leq bt}} 1\,.$$

The inner sum can be expressed in terms of the *Legendre totient function* $\phi(x, n)$, a generalisation of the Euler totient function which counts the number of positive integers less or equal than x which are prime to n. We get

$$\# \Pi_A(R) = \sum_{\sqrt{R/v} \leq b \leq \sqrt{R/u}} (\phi(bt, b) - \phi(bs, b))\,. \tag{5.6}$$

The Legendre totient function has been studied less extensively than the Euler function. We refer the reader to [30] where estimates of $\phi(x, n)$ are given in terms of $\phi(n)$, such as

$$\phi(x, n) = \frac{x}{n} \varphi(n) + O\left(\sum_{d|n} \mu^2(d) \right)\,. \tag{5.7}$$

Moreover, as shown in the proof of Proposition 5.10, we have

$$\sum_{n=1}^N \left(\sum_{d|n} \mu^2(d) \right) = \frac{6}{\pi^2} N \log N + O(N), \quad N \to \infty.$$

Using this estimate in (5.7) and then inserting the result in (5.6) we get

$$\# \Pi_A(R) = (t - s) \sum_{\sqrt{R/v} \leq b \leq \sqrt{R/u}} \varphi(b) + O\left(\sqrt{R} \log R \right)\,.$$

On the other hand, a classical result in number theory (see [3]) states that:

$$\sum_{n=1}^N \varphi(n) = \frac{3}{\pi^2} N^2 + O(N \log N) \tag{5.8}$$

and therefore

$$\# \Pi_A(R) = \frac{3}{\pi^2} R\,(t - s) \left(\frac{1}{u} - \frac{1}{v} \right) + O\left(\sqrt{R} \log R \right)\,.$$

Comparing with (5.5) have thus proved that in this first case

$$\sum_{(\gamma, r, \varepsilon) \in \mathcal{O}(1, R, -1)} \chi_A(\gamma, r, \varepsilon) = \frac{3}{\pi^2} R \nu(A) + O\left(\sqrt{R} \log R \right)\,. \tag{5.9}$$

Let us now study the asymptotic behaviour of $\#(A \cap \mathcal{O}(1, R, -1))$ as $R \to \infty$ for the second case, namely for a set A defined as in (5.4) with $0 < s \leq u < v = t$ and $t < 1$, which is a trapezoid in $\mathbb{W}$, and

$$\nu(A) = \int_u^v \left(\int_s^r r^{-2}\, d\gamma \right) dr = \log \frac{v}{u} - s \left(\frac{1}{u} - \frac{1}{v} \right)\,. \tag{5.10}$$

---

[2]The result follows also noticing that $\Pi_A(R)$ represents the set of coprime lattice points in a trapezoid $T_A(R)$ and estimating the set of coprime lattice points in $T_A(R)$ with the area of $T_A(R)$ times $6/\pi^2$ plus an error term due to the perimeter of $T_A(R)$. This is a standard technique applied for example in [7, Lemma 2].

In this case we find

$$\# \Pi_A(R) = \sum_{\sqrt{R/v} \le b \le \sqrt{R/u}} \sum_{\substack{(a,b)=1 \\ bs \le a \le R/b}} 1$$

so that using again the Legendre totien function and (5.7) we get

$$\# \Pi_A(R) = \sum_{\sqrt{R/v} \le b \le \sqrt{R/u}} \left(\phi\left(\frac{R}{b}, b\right) - \phi(bs, b) \right) = \sum_{\sqrt{R/v} \le b \le \sqrt{R/u}} \left(\frac{R}{b^2} \varphi(b) - s\, \varphi(b) \right) + O\left(\sqrt{R} \log R\right).$$

Using (5.8) along with [3],

$$\sum_{n=1}^{N} \frac{\varphi(n)}{n^2} = \frac{6}{\pi^2} \log N + O(1)$$

we obtain

$$\# \Pi_A(R) = \frac{3}{\pi^2} R \left(\log \frac{R}{u} - \log \frac{R}{v} \right) - \frac{3}{\pi^2} R\, s \left(\frac{1}{u} - \frac{1}{v} \right) + O\left(\sqrt{R} \log R\right).$$

Comparing with (5.10) we see that (5.9) holds also in this case.

In conclusion, we have proved that all the sets A as in (5.4) satisfy (5.9). This immediately implies the statement of the theorem for all couples of sets A, B of this form. The extension to all measurable sets with finite ν-measure is straighforward. $\qquad \square$

6. Non-periodic orbits

In this section we study the map $\mathcal{P}_h$ on non-periodic orbits. In particular we introduce a "time leap" for $\mathcal{P}_h$ to rewrite it as a map from $X := ([1,\infty) \times [1,\infty) \times \{0,+1\}) \cap \mathbb{W}$ to itself. We need the following result.

Proposition 6.1. *For the map $\mathcal{P}_h : \mathbb{W} \to \mathbb{W}$ we have:*

(i) *for each $(\gamma, r, \varepsilon) \in \mathbb{W}$ with $\gamma \in (0,1)$ and $\varepsilon \ne -1$, there exists $n_1(\gamma, r)$ such that $(\gamma', r', \varepsilon') = \mathcal{P}_h^{n_1(\gamma,r)}(\gamma, r, \varepsilon)$ satisfies $\gamma' \in [1,\infty)$;*

(ii) *for each $(\gamma, r, \varepsilon) \in \mathbb{W}$ with $\varepsilon = -1$, there exists $n_2(\gamma, r)$ such that $\mathcal{P}_h^{n_2(\gamma,r)}(\gamma, r, -1) = (\gamma, r, +1)$, and*

$$n_2(\gamma, r) = 2 \cdot \# \left\{ (a,b) \in \mathbb{N} \times \mathbb{N} : (a,b) = 1, (q\gamma + p)(q'\gamma + p') < r \right\} +$$
$$+ \# \left\{ (a,b) \in \mathbb{N} \times \mathbb{N} : (a,b) = 1, (q\gamma + p)(q'\gamma + p') = r \right\} - 1$$

where

$$M\left(\frac{a}{b}\right) = \begin{pmatrix} q & p \\ q' & p' \end{pmatrix};$$

(iii) *for each $(\gamma, r, \varepsilon) \in \mathbb{W}$ with $\gamma \notin \mathbb{N}$ and $\varepsilon \ne -1$, we have $\mathcal{P}_h^{\lfloor \gamma \rfloor}(\gamma, r, \varepsilon) = (\gamma - \lfloor \gamma \rfloor, r, \varepsilon)$;*

(iv) *for each $(\gamma, r, \varepsilon) \in \mathbb{W}$ with $\gamma \in \mathbb{N}$, $\gamma \ge 2$ and $\varepsilon \ne -1$, we have $\mathcal{P}_h^{(\gamma-1)}(\gamma, r, \varepsilon) = (1, r, +1)$.*

Proof. (i). Let $(\gamma_0, r_0, \varepsilon_0) := (\gamma, r, \varepsilon) \in \mathbb{W}$ with $\gamma \in (0,1)$ and $\varepsilon \ne -1$ and use the notation $(\gamma_n, r_n, \varepsilon_n) = \mathcal{P}_h^n(\gamma, r, \varepsilon)$ for all $n \ge 1$. To obtain $(\gamma_1, r_1, \varepsilon_1)$ we use Table 1 or Table 3. Let $\varepsilon_0 = +1$. If $r_0 \ge 1 - \gamma_0$ we obtain $\gamma_1 = 1/(1 - \gamma_0) > 1$ and we are done with $n_1(\gamma, r) = 1$. If $r_0 < 1 - \gamma_0$, we obtain $(\gamma_1, r_1, \varepsilon_1) = (\gamma_0/(1 - \gamma_0), r_0/(1 - \gamma_0)^2, +1)$, hence either $\gamma_1 \ge 1$ and we are done with $n_1(\gamma, r) = 1$, or $\gamma_1 \in (\gamma_0, 1)$ and we consider $(\gamma_2, r_2, \varepsilon_2)$. In the case $\varepsilon_0 = 0$, when $\gamma_0 = r_0 \ge 1/2$ then as before $\gamma_1 = 1/(1 - \gamma_0) > 1$ and we are done with $n_1(\gamma, r) = 1$. If $\gamma_0 = r_0 < 1/2$ we obtain again $(\gamma_1, r_1, \varepsilon_1) = (\gamma_0/(1 - \gamma_0), r_0/(1 - \gamma_0)^2, +1)$, now $\gamma_1 \in (\gamma_0, 1)$ and we consider $(\gamma_2, r_2, \varepsilon_2)$. By repeating the argument, if we are not done at the n-th step, we have $\gamma_n = \gamma_0/(1 - n\gamma_0) \in (\gamma_{n-1}, 1)$, $r_n = r_0/(1 - n\gamma_0)^2$ and $\varepsilon_n = +1$. At this point, if $r_n \ge 1 - \gamma_n$, which is equivalent to $r_0 \ge (1 - (n+1)\gamma_0)(1 - n\gamma_0)$, then $\gamma_{n+1} = 1/(1 - \gamma_n) > 1$. Instead, if $r_n < 1 - \gamma_n$, we have $\gamma_{n+1} = \gamma_n/(1 - \gamma_n) = \gamma_0/(1 - (n+1)\gamma_0)$ and again there are two cases: either $\gamma_{n+1} \ge 1$ and we are done, or $\gamma_{n+1} < 1$ and we proceed with the $(n+1)$-th step.

We can conclude that either there exists $\bar{n}$ such that $r_0 \geq (1 - (\bar{n} + 1)\gamma_0)(1 - \bar{n}\gamma_0)$, and then $\gamma_{\bar{n}+1} > 1$, or the sequence $\{\gamma_n\}$ is given by $\gamma_n = \gamma_0/(1 - n\gamma_0)$ for all $n \geq 1$, and since for all $\gamma_0 \in (0,1)$ there exists $\tilde{n}$ such that $\gamma_0 > 1/\tilde{n}$, we have $\gamma_{\tilde{n}+1} > 1$. Therefore

$$n_1(\gamma, r) = 1 + \min\left\{ \min\left\{ k \geq 1 \,:\, r \geq (1 - (k+1)\gamma)(1 - k\gamma) \right\} \,,\, \min\left\{ j \geq 1 \,:\, \frac{\gamma}{1 - j\gamma} \geq 1 \right\} \right\}.$$

(ii). If $\gamma = 1$, we are looking at the periodic point $(1, r, -1)$, and we have argued that $\mathcal{P}_h^{\mathrm{per}(r)-1}(1, r, -1) = (1, r, +1)$ where $\mathrm{per}(r)$ is defined in Theorem 5.8. Hence $n_2(1, r) = \mathrm{per}(r) - 1$. For a general $\gamma > 0$ we need to count the intersections of the horocycle tangent to $\mathbb{R}$ at γ and of radius r, with the sets $\mathcal{I}_n$ with $n < 0$ and with the geodesics $\mathfrak{g}(p/q, p'/q')$ defined in (5.1). The number of these intersections give $n_2(\gamma, r) - 1$.

The intersections with the lines $\mathcal{I}_n$ exist for all n such that $\gamma - r \leq n < 0$. In the formula for $n_2(\gamma, r)$ they correspond to the cases $s = 1$ and $m = |n|$, for which

$$M\left(\frac{|n|}{1}\right) = \begin{pmatrix} 1 & |n| \\ 0 & 1 \end{pmatrix}$$

and the condition is $\gamma + |n| \leq r$. By Lemma 5.6, the intersections with the geodesic $\mathfrak{g}(p/q, p'/q')$ occur if $r \geq (q\gamma + p)(q'\gamma + p')$, and by Lemmas 5.5 and 4.1, to such an intersection it corresponds a point $(\gamma', r', \varepsilon') \in \mathbb{W}$ with

$$\gamma' = \hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'}\right)(\gamma), \quad r' = \mathrm{def}\left(\hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'}\right)(\gamma)\right) r.$$

Since in Section 2 we have shown that for all $a, b \in \mathbb{N}$ with $(a, b) = 1$ we have

$$M\left(\frac{a}{b}\right) = M\left(\frac{p}{q} \oplus \frac{p'}{q'}\right) = \begin{pmatrix} p' & p \\ q' & q \end{pmatrix} \quad \text{and} \quad \hat{M}\left(\frac{p}{q} \oplus \frac{p'}{q'}\right) = \begin{pmatrix} q & p \\ q' & p' \end{pmatrix},$$

we obtain the formula for $n_2(\gamma, r)$.

(iii) and (iv). It is enough to repeatedly apply $\mathcal{P}_h$ as defined in Tables 1 and 3. $\qquad \square$

Theorem 6.2. *Let* $X := ([1, \infty) \times [1, \infty) \times \{0, +1\}) \cap \mathbb{W}$. *For all* $(\gamma, r, \varepsilon) \in \mathbb{W}$ *there exists* $\bar{n}(\gamma, r) \in \mathbb{N}$ *such that* $\mathcal{P}_h^{\bar{n}(\gamma, r)}(\gamma, r, \varepsilon) \in X$. *In particular it is well defined the map* $\mathcal{T}_h : X \to X$ *given by*

$$\mathcal{T}_h(\gamma, r, \varepsilon) = \mathcal{P}_h^{\tau(\gamma, r)}(\gamma, r, \varepsilon) \quad \text{with} \quad \tau(\gamma, r) = \begin{cases} \lfloor \gamma \rfloor + 1 + n_2\left(\frac{1}{1 + \lfloor \gamma \rfloor - \gamma}, \frac{r}{(1 + \lfloor \gamma \rfloor - \gamma)^2}\right), & \text{if } \gamma \notin \mathbb{N}; \\ \gamma - 1, & \text{if } \gamma \in \mathbb{N}; \end{cases}$$

and $n_2(\gamma, r)$ *defined as in Proposition 6.1. More specifically*

$$\mathcal{T}_h(\gamma, r, \varepsilon) = \left(\frac{1}{1 + \lfloor \gamma \rfloor - \gamma}, \frac{r}{(1 + \lfloor \gamma \rfloor - \gamma)^2}, \tilde{\varepsilon}\right),$$

with $\tilde{\varepsilon} = +1$ *if* $\gamma \neq 1$ *and* $\tilde{\varepsilon} = \varepsilon$ *if* $\gamma = 1$.

Proof. From Proposition 6.1 we know that for all $(\gamma, r, -1)$ we end up at $(\gamma, r, +1)$ with $n_2(\gamma, r)$ iterations of $\mathcal{P}_h$, whereas if $\gamma \in (0, 1)$ and $\varepsilon \neq -1$ a point with $\gamma' \geq 1$ is reached with $n_1(\gamma, r)$ iterations. Together, these two steps imply the first part of the statement.

It is then possible to define a map $\mathcal{T}_h$ from X to itself. Let us check that the times $\tau(\gamma, r)$ may be chosen as in the statement. Considering first the case $\gamma \notin \mathbb{N}$, Proposition 6.1-(iii) entails that the points $(\gamma, r, \varepsilon) \in X$ hit $(\gamma - \lfloor \gamma \rfloor, r, +1)$ after $\lfloor \gamma \rfloor$ iterations of $\mathcal{P}_h$. Since $\gamma - \lfloor \gamma \rfloor \in (0, 1)$ and $r \geq \gamma > 1$, we can apply Proposition 6.1-(i) with $n_1(\gamma - \lfloor \gamma \rfloor, r) = 1$, to get

$$\mathcal{P}_h^{\lfloor \gamma \rfloor + 1}(\gamma, r, \varepsilon) = \left(\frac{1}{1 + \lfloor \gamma \rfloor - \gamma}, \frac{r}{(1 + \lfloor \gamma \rfloor - \gamma)^2}, -1\right).$$

We can now apply Proposition 6.1-(ii) and conclude the computation of $\tau(\gamma, r)$ for $\gamma \notin \mathbb{N}$. If $\gamma \in \mathbb{N}$ the result immediately follows from Proposition 6.1-(iv), with the standard convention that $\mathcal{P}_h^0$ is the identity map. $\qquad\square$

Remark 6.3. We point out that the map $\mathcal{T}_h$ is not the first return map of $\mathcal{P}_h$ onto X. For (γ, r, ε) with $\gamma \in \mathbb{N}$ and $\gamma \geq 2$, the map $\mathcal{T}_h$ simply corresponds to $(\gamma - 1)$ applications of the translation R^{-1} on the horocycle tangent at γ. One is then reduced to the case $(1, r, \varepsilon)$ which is fixed by $\mathcal{T}_h$ and corresponds to going once round the periodic orbit. If $\gamma \notin \mathbb{N}$, the action of $\mathcal{T}_h$ has three steps. First one applies the translation R^{-1} on the horocycle $\lfloor \gamma \rfloor$ times, obtaining $(\gamma', r', \varepsilon') = (\gamma - \lfloor \gamma \rfloor, r, +1)$ which is not in X since $\gamma - \lfloor \gamma \rfloor \in (0, 1)$, and then applies SR^{-1} according to Table 1 to get $(\gamma'', r'', \varepsilon'') = (1/(1-\gamma'), r'/(1-\gamma')^2, -1)$ which again is not in X. The horocycle tangent at γ'' is close to horocycles tangent at rational approximation of γ'' which describe periodic orbits for the horocycle flow on $S\mathcal{M}$, and the last step brings $(\gamma'', r'', \varepsilon'')$ to $\mathcal{T}_h(\gamma, r, \varepsilon) = (\gamma'', r'', +1) \in X$ by shadowing these periodic orbits.

Looking at the first component of the map $\mathcal{T}_h$, we are led to study the map

$$T : [1, \infty) \to [1, \infty), \quad T(t) = \frac{1}{1 + \lfloor t \rfloor - t}$$

for which an easy computation yields the following result.

Proposition 6.4. *The dynamical system $([1, \infty), T)$ is topologically conjugate to the backward continued fraction system $([0, 1), F)$ given by*

$$F : [0, 1) \to [0, 1), \quad F(s) = \frac{1}{1 - s} \pmod 1$$

with conjugacy $\phi : [1, \infty) \to [0, 1)$ given by $\phi(t) = 1 - \frac{1}{t}$. As a consequence the map T preserves the infinite measure $d\mu(t) = dt/(t(t-1))$. In particular, if $t \in \mathbb{Q}$ there exists $k_t \geq 0$ such that $T^{k_t}(t) = 1$, and therefore $T^k(t) = 1$ for all $k \geq k_t$.

Let us now consider the action of $\mathcal{T}_h$ on the second component. One notes that if (γ, r, ε) is not periodic, the second component is strictly increasing since $(1 + \lfloor \gamma \rfloor - \gamma) < 1$ for all $\gamma \notin \mathbb{N}$. Hence if we wait enough time, the map $\mathcal{P}_h$ makes the radius of the horocycle to increase, thus converging to the cusp of $\mathcal{M}$.

Remark 6.5. The map F defined in Proposition 6.4 gets its name because can be obtained by flipping the standard continued fraction transformation about the vertical line $x = 1/2$. Adler and Flatto showed in [2] that F is the map obtained by inducing on $(0, 1)$ the first component of the Poincaré map on $\mathcal{C}$ of the geodesic flow $\tilde{g}_t$ (see the map $\mathcal{P}_g$ constructed in Appendix A, Equation (A.2)). On the other hand, we showed above that F is (conjugate to) the map on $[1, \infty)$ obtained by a time-reparameterization of the first component of the Poincaré map on $\mathcal{C}$ of the horocycle flow $\tilde{h}_s^+$. We believe that this result helps to highlight how the parabolic horocycle flow also contains an "expanding component", which can be detected by an appropriate reparameterization.

7. RELATIONS WITH [6]

In [6] the authors have introduced a Poincaré section for the horocycle flow on $\mathcal{M}$ using the identification of $S\mathcal{M}$ with the space of unimodular lattices. By adapting their argument to our notations, let

$$\tilde{\mathcal{C}} := \{(z, \zeta) \in S\mathcal{M} \,:\, y \geq 1,\, 0 < x \leq 1,\, \theta(\zeta) = \pi\} \,,$$

that is all the vectors pointing downwards with base point in the subset of the fundamental domain[3] of $\mathcal{M}$ with imaginary part greater than or equal to 1. As in Section 4 we set

$$\tilde{C} := \left\{(z, \zeta) \in S\mathbb{H} \,:\, \pi_*(z, \zeta) \in \tilde{\mathcal{C}}\right\} .$$

---

[3]Here we are thinking of the fundamental domain in the strip $x \in (0, 1]$.

Setting
$$\Omega := \left\{ (\alpha, \beta) \in \mathbb{R}^2 \ : \ \alpha, \beta \in (0, 1], \ \alpha + \beta > 1 \right\},$$
the set $\tilde{\mathcal{C}}$ may be identified with Ω by

$$z = \begin{cases} \frac{\beta}{\alpha} - \left\lfloor \frac{\beta}{\alpha} \right\rfloor + i \frac{1}{\alpha^2} & \text{if } \frac{\beta}{\alpha} \notin \mathbb{Z} \\ 1 + i \frac{1}{\alpha^2} & \text{if } \frac{\beta}{\alpha} \in \mathbb{N} \end{cases}. \tag{7.1}$$

Theorem 7.1 ([6]). *The Poincaré map of the horocycle flow $\tilde{h}_s^+$ with negative s on the section $\tilde{\mathcal{C}}$ in the variables $(\alpha, \beta) \in \Omega$ is the Boca-Cobeli-Zaharescu map studied in [7] and defined to be*

$$V : \Omega \to \Omega, \quad V(\alpha, \beta) = \left(\beta, -\alpha + \left\lfloor \frac{1+\alpha}{\beta} \right\rfloor \beta \right).$$

One of the main features of the map V is its relation with the Farey sequences. We recall that for each fixed $Q \in \mathbb{N}$ the *Farey sequence* $\mathcal{F}(Q)$ is the collection of fractions in $[0, 1]$ which written in reduced form have denominator less than or equal to Q. If

$$\mathcal{F}(Q) = \left\{ \frac{0}{1} = \frac{p_0}{q_0} < \frac{1}{Q} = \frac{p_1}{q_1} < \frac{p_2}{q_2} < \cdots < \frac{p_{N-1}}{q_{N-1}} < \frac{1}{1} = \frac{p_N}{q_N} \right\},$$

the relation between V and $\mathcal{F}(Q)$ introduced in [7] is that

$$V \left(\frac{q_{i-1}}{Q}, \frac{q_i}{Q} \right) = \left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q} \right)$$

for all $i = 1, \ldots, N$, with the convention $q_{N+1} = q_0$. Hence the denominators of the fractions in the Farey sequence $\mathcal{F}(Q)$ describe a periodic orbit of V of period $N = N(Q)$, one less than the cardinality of $\mathcal{F}(Q)$. Moreover, it is shown in [6] that this periodic orbit corresponds to a closed horocycle of length Q^2.

Let us find the corresponding periodic orbit for the Poincaré map $\mathcal{P}_h$. As explained in Remark 5.2, the closed horocycle of length Q^2 corresponds to the $\mathcal{P}_h$-periodic point $(1, Q^2/2, -1) \in \mathbb{W}$ for $Q \geq 2$, with orbit

$$\mathcal{O} \left(1, \frac{Q^2}{2}, -1 \right) = \left\{ \left(\frac{a}{b}, \frac{Q^2}{2b^2}, \pm 1 \right) \ : \ (a, b) = 1, \ ab < \frac{Q^2}{2} \right\} \cup \left\{ \left(\frac{a}{b}, \frac{Q^2}{2b^2}, 0 \right) \ : \ (a, b) = 1, \ ab = \frac{Q^2}{2} \right\}.$$

Therefore if one follows the closed horocycle $W^+_{Q^2/2}$, tangent to $\mathbb{R}$ at 1 and of radius $Q^2/2$, from its lowest point on $\mathcal{I}_0$ until it closes on $S\mathcal{M}$, then one finds $N(Q)$ passages through the section $\tilde{\mathcal{C}}$. Each of these passages occurs between two consecutive returns to the section $\mathcal{C}$ we have used to define the map $\mathcal{P}_h$. In particular, we use the suspension flow ϕ_s defined in (4.4) in terms of the map $\mathcal{P}_h$ to identify the passages through $\tilde{\mathcal{C}}$.

Proposition 7.2. *For all $Q \geq 2$, for a point $(\gamma, r, \varepsilon) \in \mathcal{O}(1, Q^2/2, -1)$ there exists $s_* \in (s_h(\gamma, r, \varepsilon), 0)$ such that $\phi_{s_*}(\gamma, r, \varepsilon, 0) \in \tilde{\mathcal{C}}$ if and only if $\gamma = a/b$ and $r = Q^2/(2b^2)$, with $(a, b) \in \mathbb{N} \times \mathbb{N}$ in the set*

$$\tilde{\mathcal{F}}(Q) := \left\{ (a, b) = 1, \ a \leq b \leq Q, \ ab \leq \frac{Q^2}{2} \right\} \cup \left\{ (a, b) = 1, \ a > b, \ a + b \leq Q, \ a(a+b) > \frac{Q^2}{2} \right\}.$$

Hence $\# \tilde{\mathcal{F}}(Q) = N(Q)$ for all $Q \geq 2$. In addition, the point $\phi_{s_*}(a/b, Q^2/(2b^2), \varepsilon, 0)$ corresponds to the point $(\alpha(a, b), \beta(a, b)) \in \Omega$ given by

$$(\alpha(a, b), \beta(a, b)) = \begin{cases} \left(\frac{b}{Q}, \frac{a + \left\lfloor \frac{Q-a}{b} \right\rfloor b}{Q} \right) & \text{if } \frac{a}{b} \leq 1 \\ \left(\frac{a+b}{Q}, \frac{a}{Q} \right) & \text{if } \frac{a}{b} > 1 \end{cases}.$$

Proof. Let $Q \geq 2$ and consider all points $(\gamma, r, \varepsilon) \in \mathcal{O}(1, Q^2/2, -1)$. Looking at the part of the horocycle flow from the point in $\mathcal{C}$ corresponding to (γ, r, ε) to its first return to $\mathcal{C}$, we need to determine whether there is a passage through $\tilde{\mathcal{C}}$.

Let us first consider the case $\varepsilon = +1$. If $\gamma = a/b > 1$ then, by Table 1, the first return to $\mathcal{C}$ occurs on $\mathcal{I}_1$ with $\mathcal{P}_h(a/b, Q^2/(2b^2), +1) = ((a-b)/b, Q^2/(2b^2), +1)$. Therefore along this portion of the horocycle flow there is no passage through $\tilde{\mathcal{C}}$, since the vector never points downward. If $\gamma = a/b = 1$, then $a = b = 1$ and the first return to $\mathcal{C}$ occurs on $\mathcal{I}_2$. In this case, before the return to $\mathcal{I}_2$, the horocycle $W^+_{Q^2/2}$ intersects $\mathcal{I}_1$ and on this intersection the vector points downward. The intersection point is $z = 1 + iQ^2$, so that from (7.1)

$$\alpha(1,1) = \frac{1}{Q}$$

and $\beta/\alpha \in \mathbb{N}$, which implies

$$\beta(1,1) = 1 = \frac{Q}{Q}.$$

If $\gamma = a/b < 1$ and $b \leq Q$, then the return to $\mathcal{C}$ occurs after a passage through the point

$$z = \frac{a}{b} + i\frac{Q^2}{b^2}$$

with vector pointing downward. This yields a passage through $\tilde{\mathcal{C}}$ with

$$\alpha(a,b) = \frac{b}{Q}, \quad \frac{\beta(a,b)}{\alpha(a,b)} - \left\lfloor \frac{\beta(a,b)}{\alpha(a,b)} \right\rfloor = \frac{a}{b}$$

by (7.1), from which

$$\beta(a,b) = \frac{a + \lfloor \frac{Q-a}{b} \rfloor b}{Q}.$$

If instead $b > Q$, then there is no passage through $\tilde{\mathcal{C}}$ before the first return to $\mathcal{C}$. Otherwise, there would be a point $(\bar{z}, \bar{\zeta}) \in SM$ with $\bar{z} \in \mathcal{I}^+$, that is with imaginary part greater or equal than 1, and $\theta(\bar{\zeta}) \in (-\pi, 0)$, equivalent in SM to $(z, \zeta) = \mathcal{W}^{-1}(a/b, Q^2/(2b^2), +1)$. But this is impossible, since $S_*(z, \zeta)$ has base point in $\mathcal{I}^+$ and angle in $(0, \pi)$.

For the case $\varepsilon = 0$ we can repeat exactly the same argument. We have thus proved half of the result. Indeed, we have proved that for $\varepsilon = 0, +1$, for a point $(\gamma, r, \varepsilon) \in \mathcal{O}(1, Q^2/2, -1)$ there exists $s_* \in (s_h(\gamma, r, \varepsilon), 0)$ such that $\phi_{s_*}(\gamma, r, \varepsilon, 0) \in \tilde{\mathcal{C}}$ if and only if $\gamma = a/b$ and $r = Q^2/(2b^2)$, with $(a, b) \in \mathbb{N} \times \mathbb{N}$ in the set

$$\left\{ (a,b) = 1, \ a \leq b \leq Q, \ ab \leq \frac{Q^2}{2} \right\}.$$

In addition the point $\phi_{s_*}(a/b, Q^2/(2b^2), \varepsilon, 0)$ corresponds to the point

$$(\alpha(a,b), \beta(a,b)) = \left(\frac{b}{Q}, \frac{a + \lfloor \frac{Q-a}{b} \rfloor b}{Q} \right) \in \Omega.$$

We now consider the case $\varepsilon = -1$. Looking at Table 2, if the radius is big enough the first return to $\mathcal{C}$ occurs on $\mathcal{J}_{-1}$. Hence we are dealing with the portion of the flow on the horocycle W^+ tangent to $\mathbb{R}$ at a point γ and of radius r between $\mathcal{I}_0$ and $\mathcal{J}_{-1}$. Let us consider the action of the matrix $S \in SL(2, \mathbb{Z})$ on this horocycle. Since $S(\mathcal{I}_0) = \mathcal{I}_0$ and $S(\mathcal{J}_{-1}) = \mathcal{I}_1$, we get the flow on $S(W^+)$ between $\mathcal{I}_0$ and $\mathcal{I}_1$. In addition, if $(z, \zeta) = \mathcal{W}^{-1}(\gamma, r, -1)$ (see (4.3)), then $(\bar{z}, \bar{\zeta}) = S_*(z, \zeta)$ satisfies $\theta(\bar{\zeta}) \in (\pi/2, \pi)$ since $\theta(\zeta) \in (-\pi/2, 0)$. Therefore along the portion of the flow on $S(W^+)$ we are looking at, there is no passage through $\tilde{\mathcal{C}}$ because the vector never points downward. Since W^+ and $S(W^+)$ are equivalent in SM, this shows that in this first case there is no passage through $\tilde{\mathcal{C}}$ before the first return to $\mathcal{C}$.

By Table 2 we are thus left with two cases to consider. The first corresponds to $\gamma = a/b > 1$, $r = Q^2/(2b^2) < \gamma(1 + \gamma) = a(a + b)/b^2$, and first return to $\mathcal{C}$ on $\mathcal{I}_{-1}$. We argue as before by looking at the action of the matrix $L \in SL(2, \mathbb{Z})$. If W^+ is the horocycle tangent to $\mathbb{R}$ at γ and of radius r, then $L(W^+)$ is the horocycle tangent to $\mathbb{R}$ at $L(\gamma) = \gamma/(1 + \gamma) \in (0, 1)$ and of radius $r/(1 + \gamma)^2 < L(\gamma)$. This new horocycle does not

intersect $\mathcal{I}_0$, and since $L(\mathcal{I}_0) = \mathcal{J}_0$ and $L(\mathcal{I}_{-1}) = \mathcal{I}_1$, we look at the portion of the flow along this horocycle from $\mathcal{J}_0$ to $\mathcal{I}_1$. Since $L(\gamma) \in (0,1)$, there exists a point

$$z = \frac{\gamma}{1+\gamma} + i\frac{2r}{(1+\gamma)^2} = \frac{a}{a+b} + i\frac{Q^2}{(a+b)^2}$$

at which the vector points downward. Therefore, if $Q^2/(a+b)^2 \geq 1$, we have found a passage through $\tilde{\mathcal{C}}$. If instead $Q^2/(a+b)^2 < 1$ there is no passage through $\tilde{\mathcal{C}}$ as can be shown by arguing as in the case $\varepsilon = +1$, $\gamma > 1$ and $b > Q$ before. Therefore, in this first case we obtain a passage through $\tilde{\mathcal{C}}$ if and only if (a,b) is in the set

$$\left\{ (a,b) = 1, \, a > b, \, a+b \leq Q, \, ab \leq \frac{Q^2}{2} < a(a+b) \right\}$$

and one realises that the condition $(a+b) \leq Q$ implies $ab \leq Q^2/2$. In addition, using (7.1), we obtain

$$\alpha(a,b) = \frac{a+b}{Q}, \quad \frac{\beta(a,b)}{\alpha(a,b)} - \left\lfloor \frac{\beta(a,b)}{\alpha(a,b)} \right\rfloor = \frac{a}{Q}.$$

It follows that $a \geq Q/2$, hence $2a + b > Q$, so that $\beta(a,b) = a/Q$.

The second case we need to consider corresponds to $r < 1+\gamma$ and the first return to $\mathcal{C}$ occuring on $\mathcal{I}_0$. In this case, the horocycle W^+ along which we are flowing is equivalent to $L(W^+)$, which needs to intersect $\mathcal{J}_0$ in two points, without intersecting other copies of $\mathcal{I}_0$ in the meantime. Again we are using that $L(\mathcal{I}_0) = \mathcal{J}_0$. It means that the radius $r/(1+\gamma)^2$ of $L(W^+)$ is smaller than $1/2$. Hence no point along the horocycle $L(W^+)$ is in $\tilde{\mathcal{C}}$.

We have thus examined all the points in the orbit of $(1, Q^2/2, -1)$, finding all the passages through $\tilde{\mathcal{C}}$. The proof is finished. $\qquad\square$

The ideas in the proof of Proposition 7.2 may be used also for non-periodic orbits. One can show that following the orbit of a point $(\gamma_0, r_0, \varepsilon_0) \in \mathbb{W}$, there is a passage through the section $\tilde{\mathcal{C}}$ between two consecutive returns to $\mathcal{C}$, $(\gamma_n, r_n, \varepsilon_n) = \mathcal{P}_h^n(\gamma, r, \varepsilon)$ and $(\gamma_{n+1}, r_{n+1}, \varepsilon_{n+1}) = \mathcal{P}_h^{n+1}(\gamma, r, \varepsilon)$ for $n \geq 0$, if and only if

$$(\gamma_n, r_n, \varepsilon_n) \in \left\{ \gamma \leq 1, \, r \geq \frac{1}{2}, \, \varepsilon \neq -1 \right\} \cup \left\{ \gamma > 1, \, \frac{(1+\gamma)^2}{2} \leq r < \gamma(1+\gamma), \, \varepsilon = -1 \right\}.$$

APPENDIX A. THE POINCARÉ MAP FOR THE GEODESIC FLOW AND ITS SUSPENSION

In this appendix we show for completeness the construction of the Poincaré map for the geodesic flow with respect to the Poincaré section $\mathcal{C}$ defined in (4.1). This is similar in spirit to the results in [1, 27] and uses essentially the same Poincaré section as in [21].

Let $(z, \zeta) \in \mathcal{C}$ and $\gamma_{z,\zeta}(t)$ be the geodesic tangent to ζ at z. Since $\theta(\zeta) \in (-\pi, 0)$ we have $\gamma_{z,\zeta}^+ \in \mathbb{R}^+$ and $\gamma_{z,\zeta}^- \in \mathbb{R}^-$. We let

$$\mathbb{G} := \left\{ (\gamma, \eta) \, : \, \gamma, \eta \in \mathbb{R}^+ \right\}$$

and consider the map

$$\mathcal{G} : \mathcal{C} \to \mathbb{G}, \qquad (z, \zeta) \mapsto \mathcal{G}(z, \zeta) = \left(\gamma_{z,\zeta}^+, -\gamma_{z,\zeta}^- \right).$$

Using the identification of $S\mathbb{H}$ with $PSL(2, \mathbb{R})$, the map $\mathcal{G}$ can be defined also as a map from $SL(2, \mathbb{R})$ to $\mathbb{G}$. Following the notation of Section 3, we have

$$(z, \zeta) = \mathcal{G}^{-1}(\gamma, \eta) \quad \Rightarrow \quad \Gamma_{(z,\zeta)} = \begin{pmatrix} \left(\frac{\eta}{\gamma}\right)^{\frac{1}{4}} \frac{\gamma}{\sqrt{\gamma+\eta}} & -\left(\frac{\gamma}{\eta}\right)^{\frac{1}{4}} \frac{\eta}{\sqrt{\gamma+\eta}} \\ \left(\frac{\eta}{\gamma}\right)^{\frac{1}{4}} \frac{1}{\sqrt{\gamma+\eta}} & \left(\frac{\gamma}{\eta}\right)^{\frac{1}{4}} \frac{1}{\sqrt{\gamma+\eta}} \end{pmatrix}. \tag{A.1}$$

We now consider the Poincaré map on $\mathcal{C}$ of the geodesic flow $\tilde{g}_t$ starting from a point $(z, \zeta) \in \mathcal{C}$ and along $\gamma_{z,\zeta}(t)$ for positive times t. Let $(\gamma, \eta) = \mathcal{G}(z, \zeta)$. Given (z', ζ') the point in $\mathcal{C}$ of the first return, we can describe the Poincaré map as a map

$$\mathcal{P}_g : \mathbb{G} \to \mathbb{G}, \quad \mathcal{P}_g(\gamma, \eta) = (\gamma', \eta') := \mathcal{G}(z', \zeta').$$

Our aim is now to describe the map $\mathcal{P}_g$ on the points of $\mathbb{G}$, also in relation with the $\{L, R\}$ coding of $\gamma \in \mathbb{R}^+$ introduced in Section 2. The map $\mathcal{P}_g$ depends on the set of C defined in (4.2) on which the first return occurs. The following analogue of Lemma 4.1 holds and the proof is similar.

Lemma A.1. *For $(z, \zeta) \in \mathcal{C}$, let $(\gamma, \eta) = \mathcal{G}(z, \zeta)$. If the first return to C for $\tilde{g}_t$ with positive t along $\gamma_{z,\zeta}(t)$ occurs on $\Gamma^{-1}(\mathcal{I}_0)$ for some $\Gamma \in PSL(2, \mathbb{Z})$, then $(\gamma', \eta') = \mathcal{P}_g(\gamma, \eta)$ satisfies*

$$\gamma' = \Gamma(\gamma) \quad and \quad \eta' = -\Gamma(-\eta).$$

Different cases have to be considered (see Figure 6). Given $(z, \zeta) = \mathcal{G}^{-1}(\gamma, \eta)$, if $\gamma > 1$ the first return to C for g_t with t positive along $\gamma_{z,\zeta}(t)$ occurs on $\mathcal{I}_1 = R(\mathcal{I}_0)$. On the other hand, if $\gamma < 1$ it occurs on $\mathcal{J}_0 = L(\mathcal{I}_0)$. Finally, when $\gamma = 1$ there are no returns to C along $\gamma_{z,\zeta}(t)$ for positive t. However we can extend the map also to this case, and using Lemma A.1 we obtain

$$(\gamma', \eta') = \mathcal{P}_g(\gamma, \eta) = \begin{cases} (\gamma - 1, \eta + 1), & \text{if } \gamma > 1; \\ \left(\frac{\gamma}{1-\gamma}, \frac{\eta}{1+\eta}\right), & \text{if } \gamma < 1; \\ (0, \infty), & \text{if } \gamma = 1; \end{cases} \tag{A.2}$$

so that we recognise the map U defined in (2.1) as the factor map of $\mathcal{P}_g$ on the first coordinate.
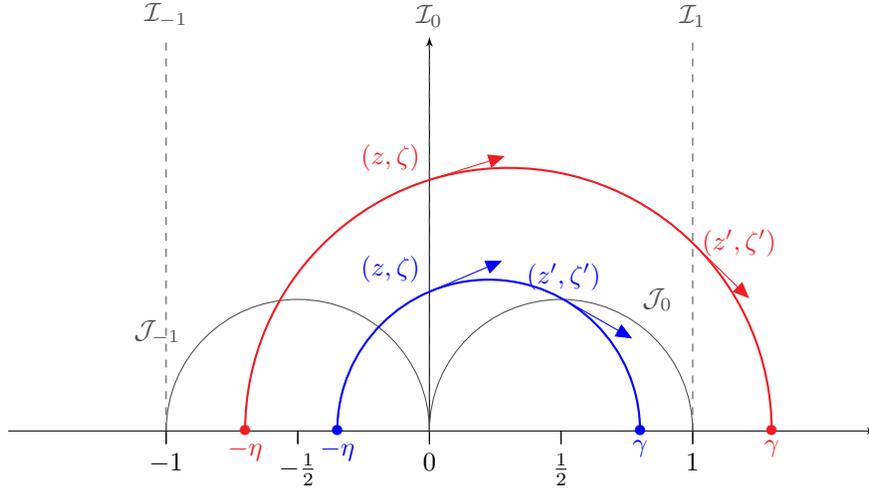

Figure 6. Action of the first return map $\mathcal{P}_g$. The red geodesic has $\gamma > 1$ and thus the first return to C of the geodesic flow occurs on $\mathcal{I}_1$; the blue geodesic has $\gamma < 1$, so that the first return to C of the geodesic flow occurs on $\mathcal{J}_0$.

Let us now consider the $\{L, R\}$ coding of $\gamma \in \mathbb{R}^+$. Using that $M(\gamma) = R \cdots$ when $\gamma > 1$, $M(\gamma) = L \cdots$ when $\gamma < 1$, and $M(\gamma) = I$ when $\gamma = 1$, we obtain

$$M(\gamma') = \begin{cases} \sigma(M(\gamma)), & \text{if } \gamma \neq 1; \\ \lambda, & \text{if } \gamma = 1; \end{cases}$$

where λ denotes the empty word.

We can now construct the suspension flow associated to $(\mathbb{G}, \mathcal{P}_g)$. Let $(z, \zeta) \in \mathcal{C}$ and apply the geodesic flow along $\gamma_{z,\zeta}(t)$ for t positive. Letting $(\gamma, \eta) := \mathcal{G}(z, \zeta)$, the first return of the geodesic flow $g_t(z, \zeta)$ on C occurs at a time $t_g(z, \zeta) > 0$, which can then be written as a function of (γ, η) as $\rho_g(\gamma, \eta) := t_g(\mathcal{G}^{-1}(\gamma, \eta))$. Let us now compute $\rho_g(\gamma, \eta)$. When $\gamma > 1$, we have seen that the first return on C along $\gamma_{z,\zeta}(t)$ for t positive occurs on $\mathcal{I}_1$. Hence, using the identification between $S\mathbb{H}$ and $PSL(2, \mathbb{R})$ and (3.3), we write

$$\Gamma_{g_t(z,\zeta)} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix} \quad \text{where } \Gamma_{(z,\zeta)} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

from which it follows that $t_g(z, \zeta)$ is the positive solution of

$$\operatorname{Re}\left(\frac{a\, e^{\frac{t}{2}} i + b\, e^{-\frac{t}{2}}}{c\, e^{\frac{t}{2}} i + d\, e^{-\frac{t}{2}}} \right) = 1\,.$$

A straightforward computation and (A.1) give

$$t_g(z, \zeta) = \frac{1}{2} \log\left(\frac{ac + d^2}{ac - c^2} \right) \quad \text{and} \quad \rho_g(\gamma, \eta) = \frac{1}{2} \log\left(\frac{1 + \frac{1}{\eta}}{1 - \frac{1}{\gamma}} \right) \quad \text{if } \gamma > 1.$$

When $\gamma < 1$, we have seen that the first return on C along $\gamma_{z,\zeta}(t)$ for t positive occurs on $\mathcal{J}_0$. Hence, as before it follows that $t_g(z, \zeta)$ is the positive solution of

$$\left| \left(\frac{a\, e^{\frac{t}{2}} i + b\, e^{-\frac{t}{2}}}{c\, e^{\frac{t}{2}} i + d\, e^{-\frac{t}{2}}} \right) - \frac{1}{2} \right| = \frac{1}{2}\,.$$

A straightforward computation and (A.1) give

$$t_g(z, \zeta) = \frac{1}{2} \log\left(\frac{b^2 - bd}{ac - a^2} \right) \quad \text{and} \quad \rho_g(\gamma, \eta) = \frac{1}{2} \log\left(\frac{1 + \eta}{1 - \gamma} \right) \quad \text{if } \gamma < 1.$$

Let now

$$\Sigma_\rho := \left\{ (\gamma, \eta, \tau) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \; : \; \gamma \neq 1\,, \, 0 \leq \tau \leq \rho_g(\gamma, \eta) \right\}$$

and consider the flow $\phi_t : \Sigma_\rho \to \Sigma_\rho$ defined by

$$\phi_t(\gamma, \eta, \tau) = \begin{cases} (\gamma, \eta, \tau + t)\,, & \text{if } 0 \leq \tau + t < \rho_g(\gamma, \eta); \\ (\mathcal{P}_g(\gamma, \eta), 0)\,, & \text{if } \tau + t = \rho_g(\gamma, \eta); \end{cases}$$

and analogously for other values of t. Then (Σ_ρ, ϕ_t) is the suspension flow over the map $\mathcal{P}_g$, for which the following proposition holds.

Proposition A.2. *The following properties hold:*

(i) *The map $\mathcal{P}_g : \mathbb{G} \to \mathbb{G}$ preserves the infinite measure μ which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{G}$ with density $h(\gamma, \eta) = (\gamma + \eta)^{-2}$.*

(ii) *The flow $\phi_t : \Sigma_\rho \to \Sigma_\rho$ preserves the measure $\tilde{\mu}$ which is absolutely continuous with respect to the Lebesgue measure on Σ_ρ with density $\tilde{h}(\gamma, \eta, \tau) = (\gamma + \eta)^{-2}$. Moreover $\tilde{\mu}(\Sigma_\rho) = \pi^2/3$.*

(iii) *The dynamical system $(\Sigma_\rho, \tilde{\mu}, \phi_t)$ is isomorphic to the system $(S\mathcal{M}, \tilde{m}, \tilde{g}_t)$, where $\tilde{m}$ is the projection on $S\mathcal{M}$ of the Liouville measure $dm(x, y, \theta) = y^{-2}\, dx\, dy\, d\theta$ on $S\mathbb{H}$, where $z = x + iy$ and $\theta = \theta(\zeta)$.*

Proof. Given $(z, \zeta) \in S\mathbb{H}$ with $\theta(\zeta) \notin \{0, \pi\}$, consider its coordinates (γ, η, τ) defined by $(\gamma, \eta) = \mathcal{G}(z, \zeta)$, and $\tau \in \mathbb{R}$ such that $(z, \zeta) = g_\tau(z', \zeta')$ where z' is the point of the geodesic $\gamma_{z,\zeta}(t)$ with maximum value of the y component. We obtain the following expressions for the coordinates (x, y, θ) of (z, ζ), where $z = x + iy$ and $\theta = \theta(\zeta)$:

$$x(\gamma, \eta, \tau) = \frac{\gamma\, e^\tau - \eta\, e^{-\tau}}{e^\tau + e^{-\tau}}\,, \quad y(\gamma, \eta, \tau) = \frac{\gamma + \eta}{e^\tau + e^{-\tau}}\,, \quad \theta(\gamma, \eta, \tau) = -2 \arctan(e^\tau)\,. \tag{A.3}$$

From (A.3) it follows that the Liouville measure $dm(x, y, \theta) = y^{-2}\, dx\, dy\, d\theta$ becomes

$$2\, d\tilde{\mu} = \frac{2}{(\gamma + \eta)^2}\, d\gamma\, d\eta\, d\tau$$

in the coordinates (γ, η, τ). Moreover by the definition of τ, the geodesic flow acts in the coordinates (γ, η, τ) simply by translation on τ, that is

$$g_t(\gamma, \eta, \tau) = (\gamma, \eta, \tau + t).$$

These facts, together with $\tilde{m}(S\mathcal{M}) = 2\pi^2/3$, prove (ii) and (iii). Finally (i) follows from (ii). $\qquad\square$

Appendix B. The dynamics of the periodic points of $\mathcal{P}_h$ on the permuted Stern-Brocot tree

Here we describe the algorithm to find the points of the periodic orbit of $(1, r, -1)$ for $\mathcal{P}_h$ in their dynamical order up to $(1, r, +1)$ (refer to Figure 7 where the case $r \in (10, 11)$ is shown):

(a) the initial condition is the point $(1, r, -1) \in \mathbb{W}$ and for completeness we consider the case $r \in \mathbb{N}$, with $r \geq 2$ (the case $r = 1$ has already been studied and for $r \in [1, 2)$ no fraction other than $1/1$ in $\hat{\mathcal{T}}$ has energy small enough to be in the periodic orbit). If $r \notin \mathbb{N}$ then the points with $\varepsilon = 0$ should be excluded in the next steps. Then we start with the fraction $1/1$;

(b) go down on the left along $\hat{\mathcal{T}}$ finding the fractions $1/n$ for $n \leq r$ (here we are applying L to the γ-components). These fractions correspond to the points $(1/n, r/n^2, -1) \in \mathbb{W}$ if $2 \leq n < r$, and to the point $(1/n, 1/n, 0) \in \mathbb{W}$ if $n = r$. Note that for the energy $\xi(1/r) = r$, and it is an elementary remark that there is only another fraction at the same level in $\hat{\mathcal{T}}$ with energy greater than or equal to r, namely $r/1$;

(c) if $r = 2$, it is clear that the only other fraction in $\hat{\mathcal{T}}$ with energy less than or equal to $r = 2$ is $2/1$, hence we obtain the point $(2/1, 2, 0)$, from which we get $(1, r, +1)$;

(d) if $r \geq 3$ go up back to the fraction $1/(r - 1)$ (applying L^{-1}), to find the corresponding point $(1/(r - 1), r/(r-1)^2, +1) \in \mathbb{W}$. The right sister of $1/(r-1)$ is the fraction $(r-1)/(r-2)$, and $\xi((r-1)/(r-2)) > r$ for all $r \geq 4$. If $r = 3$, the right sister of $1/2$ is $2/1$, we get the point $(2/1, 3, -1) \in \mathbb{W}$, and we continue as in step (i);

(e) if $r \geq 4$, keep going up back to the fractions $1/n$ with $2 \leq n < r - 1$ (applying L^{-1}), to find the corresponding point $(1/n, r/n^2, +1) \in \mathbb{W}$, and stop going up at n_* if the right sister $n_*/(n_* - 1)$ satisfies $\xi(n_*/(n_* - 1)) = n_*(n_* - 1) \leq r$. In this case we go to the right (applying $RL^{-1} = SR^{-1}$) to the fraction $n_*/(n_* - 1)$, corresponding to the point $(n_*/(n_* - 1), r/(n_* - 1)^2, -1) \in \mathbb{W}$ if $n_*(n_* - 1) < r$, and to the point $(n_*/(n_* - 1), n_*/(n_* - 1), 0) \in \mathbb{W}$ if $n_*(n_* - 1) = r$;

(f) repeat the same algorithm: first go left until possible with points with $\varepsilon = -1$ or $\varepsilon = 0$; then go back up with points with $\varepsilon = +1$ until the energy of the right sister of the fraction is small enough to go right with $\varepsilon = -1$; eventually we reach a fraction which is a right descendant (hence it has no right sisters) and in correspondence with a point with $\varepsilon = +1$, from which we cannot go down; then we go up back visiting all the already seen fractions, this time with $\varepsilon = +1$, until we get to the fraction $n_*/(n_* - 1)$ of step (e), in correspondence with the point $(n_*/(n_* - 1), r/(n_* - 1)^2, +1) \in \mathbb{W}$;

(g) move up to the fraction $1/(n_* - 1)$, the direct parent of $1/n_*$ and $n_*/(n_* - 1)$, (applying R^{-1}) obtaining the point $(1/(n_* - 1), r/(n_* - 1)^2, +1)$. The same pattern described in (e) and (f) repeats, that is we move to the right to the fraction $(n_* - 1)/(n_* - 2)$ with $\varepsilon = -1$, and then apply step (f), until we reach again $(n_* - 1)/(n_* - 2)$, this time with $\varepsilon = +1$, and then go back to $1/(n_* - 2)$;

(h) repeat step (g) moving up until reaching $1/2$ in correspondence with the point $(1/2, r/4, +1) \in \mathbb{W}$, then go right to $2/1$ with $\varepsilon = -1$;

(i)   now apply the steps in (b)-(g) by interchanging left with right. This follows from the symmetry of the Stern-Brocot trees, for which a reflection of the tree along a vertical line passing through $1/1$ sends the fraction p/q to the fraction q/p. In particular the energy of these specular fractions is the same. Then we visit all the fractions with energy less than or equal to r which lie on this right side of $\hat{\mathcal{T}}$, until we reach back $2/1$ in correspondence with the point $(2/1, r, +1) \in \mathbb{W}$. The last step is to go back to $1/1$ and to $(1, r, +1)$, the last point in the periodic orbit of $(1, r, -1)$.

References

[1] R. L. Adler, L. Flatto, *Cross section map for the geodesic flow on the modular surface*. In "Conference in modern analysis and probability" (New Haven, Conn., 1982), 9–24, Contemp. Math., 26, Amer. Math. Soc., Providence, RI, 1984.

[2] R. L. Adler, L. Flatto, *The backward continued fraction map and the geodesic flow*. Ergodic Theory Dynam. Systems **4** (1984), no. 4, 487–492.

[3] T. M. Apostol, "Introduction to analytic number theory". Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1976.

[4] J. S. Athreya, *Gap distributions and homogeneous dynamics*. In "Geometry, topology, and dynamics in negative curvature", 1–31, London Math. Soc. Lecture Note Ser., 425, Cambridge Univ. Press, Cambridge, 2016.

[5] J. S. Athreya, J. Chaika, S. Lelièvre, *The gap distribution of slopes on the golden L*. In "Recent trends in ergodic theory and dynamical systems", 47–62, Contemp. Math., 631, Amer. Math. Soc., Providence, RI, 2015.

[6] J. S. Athreya, Y. Cheung, *A Poincaré section for the horocycle flow on the space of lattices*. Int. Math. Res. Not. IMRN **2014** (2014), no. 10, 2643–2690.

[7] F. Boca, C. Cobeli, A. Zaharescu, *A conjecture of R. R. Hall on Farey points*. J. Reine Angew. Math. **535** (2001), 207–236.

[8] C. Bonanno, S. Isola, *Orderings of the rationals and dynamical systems*. Colloq. Math. **116** (2009), no. 2, 165–189.

[9] N. Calkin, H. S. Wilf, *Recounting the rationals*. Amer. Math. Monthly **107** (2000), no. 4, 360–363.

[10] S. G. Dani, *Divergent trajectories of flows on homogeneous spaces and Diophantine approximation*. J. Reine Angew. Math. **359** (1985), 55–89.

[11] M. Einsiedler, T. Ward, "Ergodic theory with a view towards number theory". Graduate Texts in Mathematics, 259. Springer-Verlag London, Ltd., London, 2011.

[12] A. Eskin, C. McMullen, *Mixing, counting, and equidistribution in Lie groups*. Duke Math. J. **71** (1993), no. 1, 181–209.

[13] L. Flaminio, G. Forni, *Invariant distributions and time averages for horocycle flows*. Duke Math. J. **119** (2003), no. 3, 465–526.

[14] G. H. Hardy, E. M. Wright, "An introduction to the theory of numbers". Fourth edition. The Clarendon Press, Oxford University Press, London, 1975.

[15] B. Heersink, *Poincaré sections for the horocycle flow in covers of $SL(2, \mathbb{R})/SL(2, \mathbb{Z})$ and applications to Farey fraction statistics*. Monatsh. Math. **179** (2016), no. 3, 389–420.

[16] B. Heersink, *Distribution of the periodic points of the Farey map*. Comm. Math. Phys. **365** (2019), no. 3, 971–1003.

[17] D. A. Hejhal, *On the uniform equidistribution of long closed horocycles*. Asian J. Math. **4** (2000), no. 4, 839–853.

[18] J. Marklof, *The asymptotic distribution of Frobenius numbers*. Invent. Math. **181** (2010), no. 1, 179–207.

[19] J. Marklof, M. Welsh, *Fine-scale distribution of roots of quadratic congruences*. Duke Math. J. **172** (2023), no. 12, 2303–2364.

[20] D. H. Mayer, *The thermodynamic formalism approach to Selberg's zeta function for $PSL(2, \mathbb{Z})$*. Bull. Amer. Math. Soc. (N.S.) **25** (1991), no. 1, 55–60.

[21] D. H. Mayer, *Transfer operators, the Selberg zeta function and the Lewis-Zagier theory of periodic functions*. In "Hyperbolic geometry and applications in quantum chaos and cosmology", 146–174, London Math. Soc. Lecture Note Ser., 397, Cambridge Univ. Press, Cambridge, 2012.

[22] M. Ratner, *Rigidity of horocycle flows*. Ann. of Math. (2) **115** (1982), no. 3, 597–614.

[23] M. Ratner, *The rate of mixing for geodesic and horocycle flows*. Ergodic Theory Dynam. Systems **7** (1987), no. 2, 267–288.

[24] D. Ravotti, *Asymptotics and limit theorems for horocycle ergodic integrals à la Ratner*. J. Éc. polytech. Math. **10** (2023), 305–334.

[25] O. Sarig, M. Schmoll, *Adic Flows, transversal Flows, and horocycle Flows*. In "Ergodic Theory and Dynamical Systems", 241–259, proceedings of the Ergodic Theory Workshops at University of North Carolina, Chapel Hill, NC, March 17–21, 2011–March 22–25, 2012. Edited by Idris Assani. De Gruyter Proc. Math. De Gruyter, Berlin, 2014.

[26] P. Sarnak, *Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series*. Comm. Pure Appl. Math. **34** (1981), no. 6, 719–739.

[27] C. Series, *The modular surface and continued fractions*. J. London Math. Soc. (2) **31** (1985), no. 1, 69–80.

[28] A. Strömbergsson, *On the uniform equidistribution of long closed horocycles.* Duke Math. J. **123** (2004), no. 3, 507–547.

[29] A. Strömbergsson, *On the deviation of ergodic averages for horocycle flows.* J. Mod. Dyn. 7 (2013), no. 2, 291–328.

[30] D. Suryanarayana, *On* $\triangle(x, n) = \phi(x, n) - x\phi(n)/n$. Proc. Amer. Math. Soc. **44** (1974), 17–21.

[31] D. Taha, *The Boca-Cobeli-Zaharescu map analogue for the Hecke triangle groups* G_q. arXiv:1810.10668 [math.DS]

[32] J. Tseng, *Shrinking target equidistribution of horocycles in cusps.* Math. Z. **302** (2022), no. 4, 2005–2035.

[33] C. Uyanik, G. Work, *The distribution of gaps for saddle connections on the octagon.* Int. Math. Res. Not. IMRN **2016**, (2016) no. 18, 5569–5602.

[34] D. Zagier, *Eisenstein series and the Riemann zeta function.* In "Automorphic forms, representation theory and arithmetic" (Bombay, 1979), pp. 275–301, Tata Inst. Fund. Res. Studies in Math., 10, Tata Institute of Fundamental Research, Bombay, 1981.
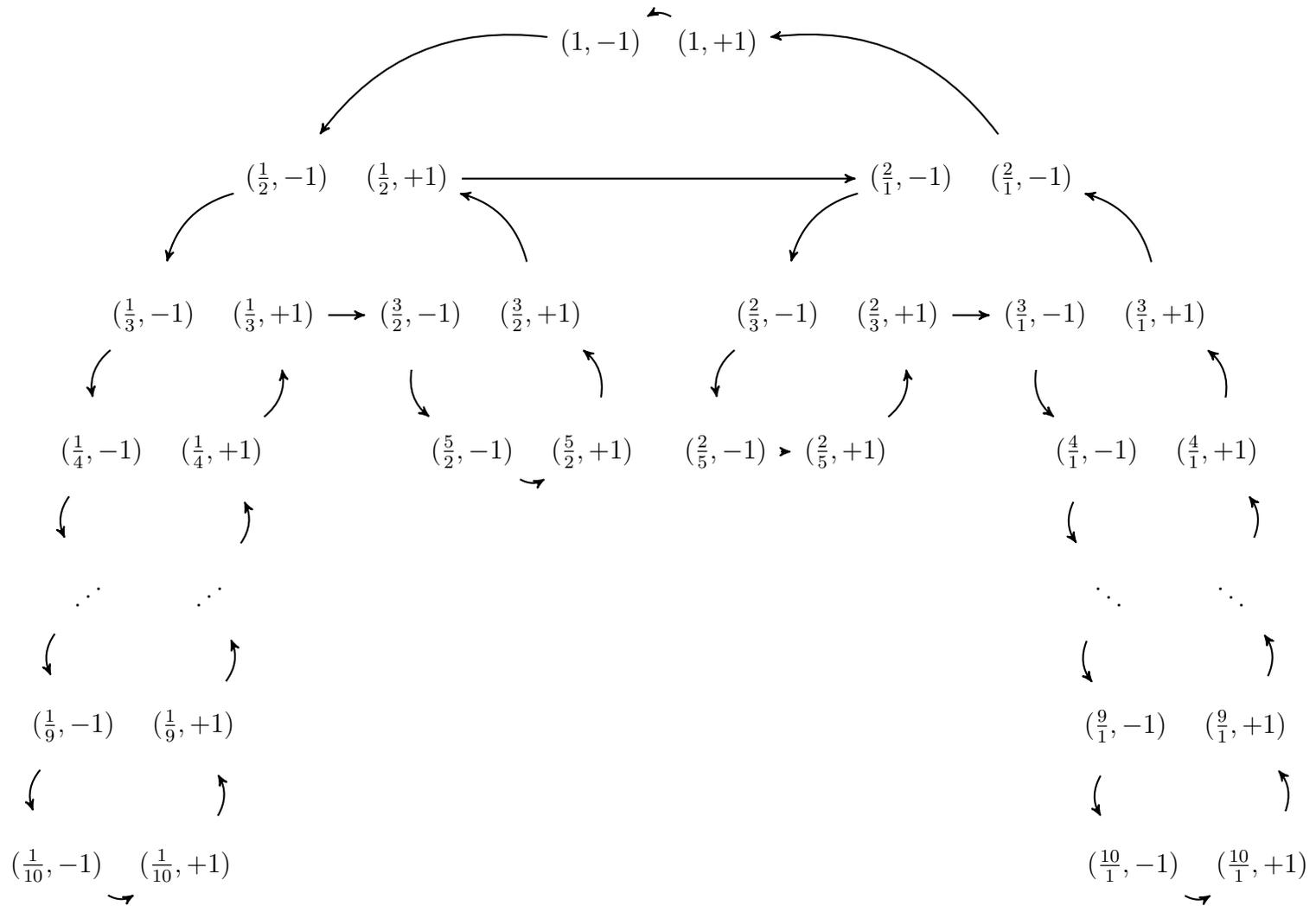
Figure 7. Representation on the permuted Stern-Brocot tree of the periodic orbit of $(1, r, -1)$ for $\mathcal{P}_h$ in the case $r \in (10, 11)$.

Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, 56127 Pisa, Italy
Email address: claudio.bonanno@unipi.it

Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, 56127 Pisa, Italy
Email address: delvigna@mail.dm.unipi.it

Scuola di Scienze e Tecnologie, Università degli Studi di Camerino, via Madonna delle Carceri, 62032 Camerino, Italy
Email address: stefano.isola@unicam.it