Athreya, J.S., and Y.Cheung. (2009) "A Poincaré section for the horocycle flow on the space of lattices,"
International Mathematics Research Notices, Vol. 2009, Article ID rnn999, 31 pages.
doi:10.1093/imrn/rnn999

A Poincaré section for the horocycle flow on the space of lattices

Jayadev S. Athreya[1] and Yitwah Cheung[2]

[1]Department of Mathematics, University of Illinois Urbana-Champaign, 1409 W. Green Street, Urbana, IL 61801
[2]Department of Mathematics, San Francisco State University, Thornton Hall 937, 1600 Holloway Ave, San Francisco, CA 94132

Correspondence to be sent to: jathreya@illinois.edu

We construct a Poincaré section for the horocycle flow on the modular surface $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$, and study the associated first return map, which coincides with a transformation (the *BCZ map*) defined by Boca-Cobeli-Zaharescu [8]. We classify ergodic invariant measures for this map and prove equidistribution of periodic orbits. As corollaries, we obtain results on the average depth of cusp excursions and on the distribution of gaps for Farey sequences and slopes of lattice vectors.

1 Introduction

Let $X_2 = SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ be the space of unimodular lattices in $\mathbb{R}^2$. X_2 is a non-compact finite-volume (with respect to Haar measure on $SL(2,\mathbb{R})$) homogeneous space. X_2 can also be viewed as the unit-tangent bundle to the hyperbolic orbifold $\mathbb{H}^2/SL(2,\mathbb{Z})$. The action of various one-parameter subgroups of $SL(2,\mathbb{R})$ on X_2 via left multiplication give several important examples of dynamics on homogeneous spaces, and have close links to geometry and number theory.

For example, the action of the subgroup

$$A := \left\{ g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} : t \in \mathbb{R} \right\}$$

yields the *geodesic flow* on X_2, whose orbits are hyperbolic geodesics when projected to $\mathbb{H}^2/SL(2,\mathbb{Z})$. This flow can be realized as an *suspension flow* over (the natural extension of) the Gauss map $G(x) = \left\{\frac{1}{x}\right\}$, and this connection can be exploited to give many connections between the theory of continued fractions and hyperbolic geometry (see the beautiful articles by Series [33] or Arnoux [1] for very elegant expositions).

In this paper, we study the *horocycle flow*, that is, the action of the subgroup

$$N = \left\{ h_s = \begin{pmatrix} 1 & 0 \\ -s & 1 \end{pmatrix} : s \in \mathbb{R} \right\}.$$

The main result (Theorem 1.1) of this paper displays the horocycle flow as a suspension flow over the *BCZ map*, which was constructed by Boca-Cobeli-Zaharescu [8] in their study of Farey fractions. Equivalently, we construct a transversal to the horocycle flow so that the first return map is the BCZ map. This enables us to use well-known ergodic and equidistribution properties of the horocycle flow to derive equivalent properties for the BCZ map (in particular that it is ergodic, zero entropy, and that periodic orbits equidistribute), and give a unified explanation of several number-theoretic results on the statistical properties of Farey gaps. We also give a 'piecewise-linear' description of the cusp excursions of the horocycle flow, and derive several new results on the geometry of numbers relating to gaps of slopes of lattice vectors.

1.1 Plan of paper

We describe the organization of the paper, and also give a guide for readers.

Received 1 January 2009; Revised 11 January 2009; Accepted 21 January 2009
Communicated by A. Editor

1.1.1 *Organization*

This paper is organized as follows: in the remainder of the introduction (§1), we state our results: in §1.3, we describe the transversal to the horocycle flow, and state Theorem 1.1. In §1.4, we discuss the ergodic properties of the BCZ map (§1.4.2) and the structure and equidistribution of periodic orbits (§1.4.3). A piecewise-linear description of horocycle cusp excursions is given in §1.5; a unified approach to Farey statistics is described in §1.6; and a similar approach to statistics of gaps in slopes of lattice vectors is the subject of §1.7. In §2, we prove Theorem 1.1, and show how to describe it using Euclidean and hyperbolic geometry. The structure of periodic orbits is the subject of §4; and in §5, these structure results are used to obtain equidistribution properties and the corollaries on Farey statistics in §5.2. §3 contains the proof of ergodicity and the calculation of entropy of the BCZ map. We prove our results on cusp excursions in §6, and in §7 we prove our results on geometry of numbers. Finally, in §8, we outline some questions and directions for future research.

1.1.2 *Readers guide*

Since this paper touches on several different topics, it can be read in several different ways. We suggest different approaches for the ergodic-theoretic and number-theoretic minded readers. We recommend that the ergodic-theoretic reader start with sections §1.3 and §1.4 (and perhaps §1.5), and follow it with §2, §3 and §6 before exploring the more number theoretic parts of the paper. The number-theoretic reader should also start with §1.3 and §1.4, but then may be more intrigued by the results of §1.4.3, §1.6, and §1.7 and their proofs in §4,§5.2, and §7 respectively.

1.2 **Acknowledgments and Funding**

The original motivation for this project was the study of cusp excursions for horocycle flow on X_2 (as described in §1.5). We developed the first return map T as a technical tool in order to study these excursions. Later, the first-named author was attending a talk by F. Boca in which the map T was written down in the context of studying gaps in angles between hyperbolic lattice points. After many invaluable and illuminating conversations with F. Boca and A. Zaharescu, we understood the deep connections they had developed between orbits of this map and the study of Farey sequences. It is a pleasure to thank them for their insights and assistance. We would also like to thank Jens Marklof for numerous helpful discussions and insightful comments on an early version of this draft.

We thank the organizers of the Summer School and Conference on Teichmüller dynamics in Marseilles in June 2009, when this project was originally developed. In addition, J.S.A. would like to thank San Francisco State University for its hospitality. J.S.A. supported by National Science Foundation grant number DMS-1069153. Y.C. supported by National Science Foundation CAREER grant number DMS-0956209.

1.3 **Description of transversal**

Recall that X_2 can be explicitly identified with the space of unimodular lattices via the identification

$$gSL(2,\mathbb{Z}) \leftrightarrow g\mathbb{Z}^2.$$

Let

$$P = \left\{ p_{a,b} = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} : a \in \mathbb{R}^*, b \in \mathbb{R} \right\} \tag{1.1}$$

denote the group of upper-triangular matrices in $SL(2,\mathbb{R})$. Let

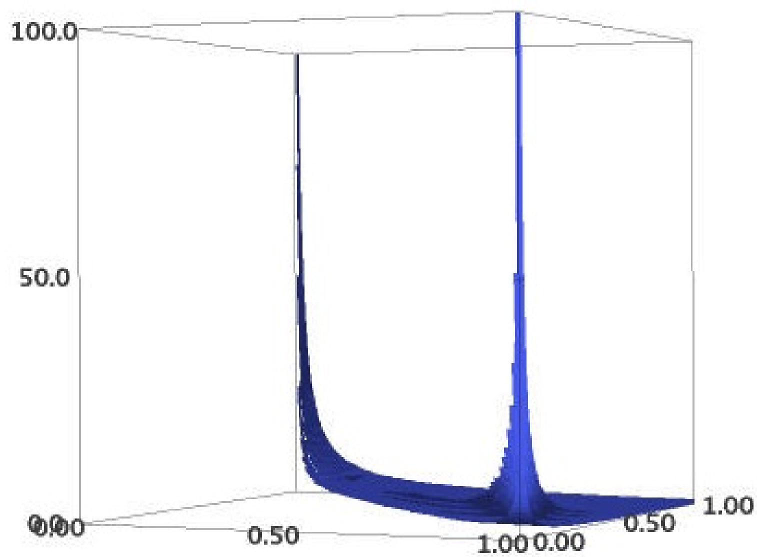
$$\Omega' := \{p_{a,b}SL(2,\mathbb{Z}) : a, b \in (0,1], a+b > 1\} \subset X_2. \tag{1.2}$$

We will use Ω to denote the set

$$\{(a,b) \in \mathbb{R}^2 : a, b \in (0,1], a+b > 1\} \subset \mathbb{R}^2.$$

In §2, we will show that Ω' can be viewed as the set of lattices with a horizontal vector of length at most 1, can also be identified with the subset $\left\{z = x+iy \in \mathbb{H}^2 : -\frac{1}{2} < x \leq \frac{1}{2}, y > 1, \right\} \subset \mathbb{H}^2$ of the upper-half plane. Our main theorem is that the Ω' is a Poincaré section for the horocycle flow, and that the first return map is the BCZ map. We see the space Ω together with the roof in Figure 1.

Fig. 1. A picture of the suspension space over Ω. Trajectories of the flow are vertical lines. Starting from $(a,b) \in \Omega$, h_s trajectories move vertically until they hit the 'roof' (at time $R(a,b)$) upon which they return to the floor at position $T(a,b)$.



Theorem 1.1. $\Omega' \subset X_2$ is a Poincaré section for the action of N on X_2. That is, every N-orbit $\{h_s\Lambda\}_{s\in\mathbb{R}}$ (with the exception of the codimension 1 set of lattices Λ with a length ≤ 1 vertical vector), $\Lambda \in X_2$, intersects Ω' in a non-empty, countable, discrete set of times $\{s_n\}_{n\in\mathbb{Z}}$. Given $\Lambda_{a,b} = p_{a,b}SL(2,\mathbb{Z})$, the first return time $R(a,b) = \min\{s > 0 : h_s\Lambda_{a,b} \in \Omega'\}$ is given by

$$R(a,b) = \frac{1}{ab}. \tag{1.3}$$

The first return map $T : \Omega \to \Omega$ defined implicitly by

$$\Lambda_{T(a,b)} = h_{R(a,b)}x_{a,b}$$

is given explicitly by the BCZ map

$$T(a,b) = \left(b, -a + \left\lfloor \frac{1+a}{b} \right\rfloor b\right) \tag{1.4}$$

$\square$

Remarks:

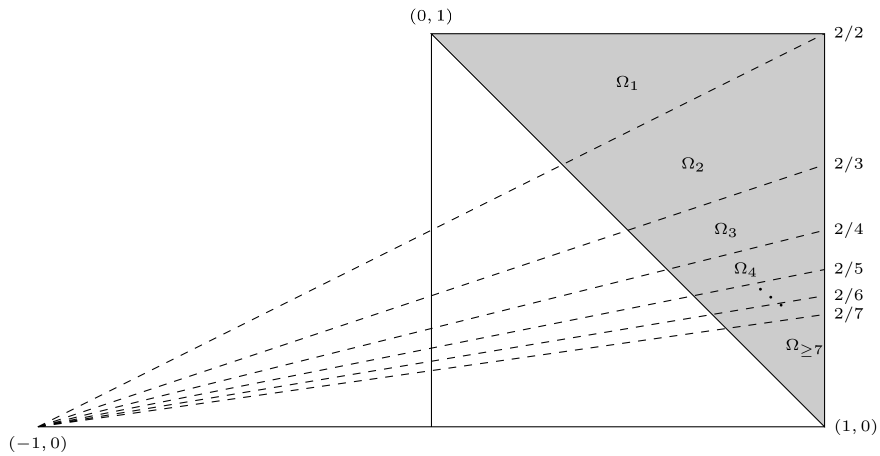
Short periodic orbits We will see below that lattices with a length ≤ 1 vertical vector consist of those whose horocycle orbits are periodic of period at most 1, and are embedded closed horocycles in $\mathbb{H}^2/SL(2,\mathbb{Z})$, foliating the cusp. Thus, from a dynamical point of view, there is no loss in missing them.

Integrability of R A direct calculation shows that $\int_\Omega R2dadb = \frac{\pi^2}{3}$, which is the volume of X_2 when viewed as the unit tangent bundle of $\mathbb{H}^2/SL(2,\mathbb{Z})$ with respect to the measure $\frac{dxdyd\theta}{y^2}$. It is also immediate that $R \in L^p(dm)$ for $p < 2$, where $dm = 2dadb$.

1.4 Return map and applications

The BCZ map has proved to be a powerful technical tool in studying various statistical properties of Farey fractions [5, 8, 10], and the distribution of angles of families of hyperbolic geodesics [11].

Fig. 2. The Farey triangle $\Omega = \bigcup_{k\geq 1} \Omega_k$



1.4.1 *Tiles and images*

We briefly describe the basic structure of the piecewise linear decomposition of T. (1.4) tells us that the map T acts via

$$T(a,b) = (a,b)A_k^{\mathrm{T}},$$

where the superscript T denotes transpose, and

$$A_k = \begin{pmatrix} 0 & 1 \\ -1 & k \end{pmatrix}$$

on the region $\Omega_k := \{(a,b) \in \Omega : \kappa(a,b) = k\}$, where $\kappa(a,b) = \left\lfloor \frac{1+a}{b} \right\rfloor$ (see Figure 2 below). Ω_1 is a triangle with vertices at $(0,1)$, $(1,1)$, and $\left(\frac{1}{3}, \frac{2}{3}\right)$; and for $k \geq 2$, Ω_k is a quadrilateral with vertices at $\left(1, \frac{2}{k}\right)$, $\left(1, \frac{2}{k+1}\right)$, $\left(\frac{k-1}{k+1}, \frac{2}{k+1}\right)$, and $\left(\frac{k}{k+2}, \frac{2}{k+2}\right)$. Note that area$(\Omega_1) = \frac{1}{6}$, andfor $k \geq 2$, area$(\Omega_k) = \frac{4}{k(k+1)(k+2)} = O(k^{-3})$ so that $\kappa \in L^p(m)$ for $1 \leq p < 2$.

The image $T\Omega_k$ is the reflection of Ω_k about the line $a = b$ (see Figure 3). However, T is *orientation-preserving*, and thus does not act by the reflection. For $k = 1$, T acts by the elliptic matrix A_1, for $k = 2$ the parabolic matrix A_2, and for $k \geq 3$, by the hyperbolic matrices, since trace$(A_k) = k$.

1.4.2 *Ergodic properties of the BCZ map*

In [7, §3], Boca-Zaharescu posed a series of questions on the ergodic properties of T:

Question. Is T ergodic (with respect to Lebesgue measure)? Weak mixing? What is the entropy of T? □

A corollary of Theorem 1.1 and the ergodicity and entropy properties of the horocycle flow is:

Theorem 1.2. T is an ergodic, zero-entropy map with respect to the Lebesgue probability measure $dm = 2dadb$. Moreover, m is the unique absolutely continuous invariant probability measure, and in fact is the unique ergodic invariant measure not supported on a periodic orbit. □

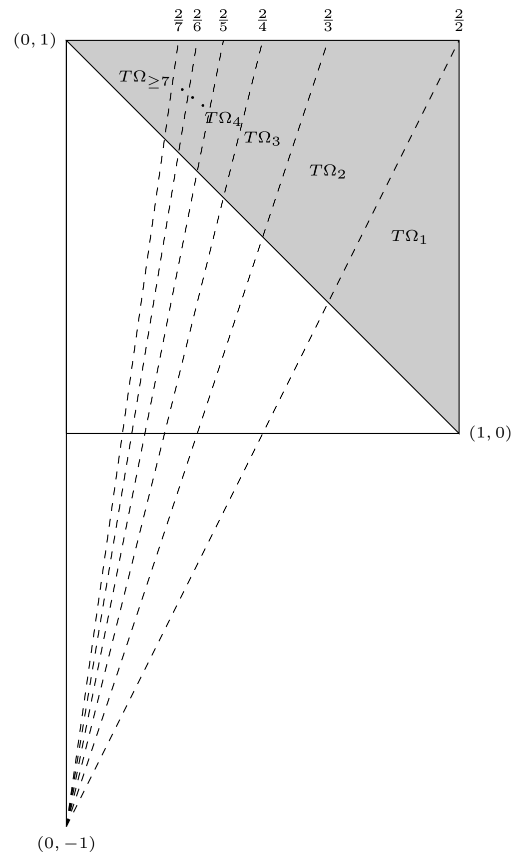
1.4.3 *Periodic orbits*

The map T has a rich and intricate structure of periodic orbits, closely related to the Farey sequences. A direct calculation shows that for $Q \in \mathbb{N}$ the point $\Lambda_{\frac{1}{Q},1} = p_{\frac{1}{Q},1}\mathbb{Z}^2$ is h_s-periodic with (minimal) period Q^2.

Given $Q \in \mathbb{N}$, the Farey sequence $\mathcal{F}(Q)$ is the collection (in increasing order) of fractions $0 < \frac{p}{q} < 1$ with $q \leq Q$. Let $N = N(Q)$ denote the cardinality of $\mathcal{F}(Q)$. We write $\mathcal{F}(Q) = \{\frac{0}{1} = \gamma_1 < \gamma_2 = \frac{1}{Q} < \ldots < \gamma_N\}$. For notational convenience, we write $\gamma_{N+1} = 1 = \frac{1}{1}$. Writing $\gamma_i = \frac{p_i}{q_i}$, with $q_i \leq Q$, we have $q_i + q_{i+1} > Q$, and

$$p_{i+1}q_i - p_i q_{i+1} = 1.$$

Fig. 3. The images $T\Omega_k$



The following fundamental observation is due to Boca-Cobeli-Zaharescu [8]:

$$T\left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right) = \left(\frac{q_{i+1}}{Q}, \frac{q_{i+2}}{Q}\right). \tag{1.5}$$

The indices are viewed cyclically in $\mathbb{Z}/N\mathbb{Z}$. We give a geometric explanation for (1.5) in §4.1 below. Thus, $(\frac{1}{Q}, 1)$ is a periodic point of order N for T. To relate this to the periodic orbit of $\Lambda_{1,\frac{1}{Q}}$ for h_s, we first record two simple observations. First,

$$\left(\frac{q_1}{Q}, \frac{q_2}{Q}\right) = \left(\frac{1}{Q}, 1\right),$$

and second,

$$\sum_{i=1}^{N} \frac{1}{q_i q_{i+1}} = \sum_{i=1}^{N} (\gamma_{i+1} - \gamma_i) = 1$$

Multiplying both sides by Q^2, and applying (1.5), we obtain

$$\sum_{i=1}^{N} \frac{Q^2}{q_i q_{i+1}} = \sum_{i=1}^{N} \left(R \circ T^{i-1}\right) \left(\frac{1}{Q}, 1\right) = Q^2.$$

Thus, as we sum the roof function R over the periodic orbit for T, we obtain the length of the associated periodic orbit for the flow.

We recall Sarnak [32] showed that long periodic orbits for h_s (i.e., long closed horocycles) become equidstributed with respect to μ_2, the Haar measure on X_2. Our main result on the distribution of periodic orbits is the following discrete corollary of Sarnak's result: let $\rho_{Q,I}$ denote the probability measure supported on (a long piece of) the orbit of $(\frac{1}{Q}, 1)$: given $I = [\alpha, \beta] \subset [0, 1]$, let $N_I(Q) := |\mathcal{F}(Q) \cap I|$, and define

$$\rho_{Q,I} = \frac{1}{N_I(Q)} \sum_{i:\gamma_i \in I} \delta_{T^i(\frac{1}{Q},1)}.$$

If $I = [0,1]$, we write $\rho_{Q,I} = \rho_Q$.

Theorem 1.3. For any (non-empty) interval $I = [\alpha, \beta] \subset [0,1]$, the measures $\rho_{Q,I}$ become equidistributed with respect to m as $Q \to \infty$. That is, $\rho_{Q,I} \to m$, where convergence is in the weak-* topology. □

Remark. The case $I = [0,1]$ of Theorem 1.3 was proven in [27] using different methods. Also, Theorem 1.3 can be deduced from Theorem 6 in [28], which is proven in the more general context of horospherical flows using a section that reduces to the same one in Theorem 1.1.

Theorem 1.4. A point $(a,b) \in \Omega$ is T-periodic if and only if it has rational slope. In particular, the set of periodic points is dense. □

While periodic points are abundant, there are strong restrictions on the *lengths* of periodic orbits, and the associated matrices, governed by the relationship between the (discrete) period $P(a,b)$ under T of a point (a,b) and the (continuous) period $s(a,b)$ of the associated lattice $p_{a,b}\mathbb{Z}^2$ under h_s. We fix notation: for $n \geq 1$, we write

$$A_n(a,b) = A\left(T^{n-1}(a,b)\right) A\left(T^{n-2}(a,b)\right) \dots A(a,b),$$

so

$$T^n(a,b) = (a,b)A_n(a,b)^{\mathrm{T}}.$$

As a starting point for our observations, note that the diagonal $(a,a) \in \Omega$ consists of *fixed* (i.e, period $P(a,a) = 1$) points for T, the associated horocycle period is the value roof function $R(a,a) = \frac{1}{a^2}$, and the associated matrix

$$A_{P(a,a)}(a,a) = A(a,a) = A_2 = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$$

is parabolic. Our main result on periodic points is that appropriate versions of these observations hold for all (segments of) periodic points.

Theorem 1.5. For any periodic point $(a,b) \in \Omega$, we have

$$P(a,b) = N\left(\left\lfloor \sqrt{s(a,b)} \right\rfloor\right).$$

Thus the set of possible periods is given by the cardinalities $\{N(Q) : Q \in \mathbb{N}\}$ of Farey sequences. Moreover, the matrix

$$A_{P(a,b)}(a,b) = A\left(T^{P(a,b)-1}(a,b)\right) A\left(T^{P(a,b)-2}(a,b)\right) \dots A(a,b)$$

associated to the periodic orbit is always parabolic , and in fact constant along the segment $\left\{(ta,tb) : t \in \left(\frac{1}{a+b}, 1\right]\right\}$. Precisely, for $k,l \in \mathbb{N}$, $k \leq l$ relatively prime, and $a \in \left(\frac{l}{l+r}, \frac{l}{l+r-1}\right]$, $1 \leq r \leq k$, we have

$$P\left(a, a\frac{k}{l}\right) = N(l+r-1),$$

and for $a \in (\frac{l}{l+k}, 1]$,

$$A_{P\left(a,a\frac{k}{l}\right)}\left(a, a\frac{k}{l}\right) = \begin{pmatrix} 1-kl & l^2 \\ -k^2 & 1+kl \end{pmatrix}.$$

□

1.5 Piecewise linear description of horocycles

In this section, we describe the results that originally motivated this project, on cusp excursions and returns to compact sets for the horocycle flow. Let $\|.\|$ denote the supremum norm on $\mathbb{R}^2$. For $\Lambda = g\mathbb{Z}^2 \in X_2$, define

$$\ell(\Lambda) := \inf_{0 \neq \mathbf{v} \in g\mathbb{Z}^2} \|\mathbf{v}\|,$$

and let $\alpha : X_2 \to \mathbb{R}^+$ be given by

$$\alpha(\Lambda) = \frac{1}{\ell(\Lambda)}.$$

By Mahler's compactness criterion, a subset A of X_2 is precompact if and only if there is an $\epsilon > 0$ so that for all $\Lambda \in A$,

$$\ell(\Lambda) > \epsilon,$$

or, equivalently if $\alpha|_A$ is bounded.

Dani [15] showed that for any lattice $\Lambda \in X_2$, the orbit $\{h_s\Lambda\}_{s\geq 0}$ is either closed or uniformly distributed with respect to the Haar probability measure μ on X_2. Thus for an Λ so that $\{h_s\Lambda\}_{s\geq 0}$ is not closed, we have

$$\limsup_{s\to\infty} \alpha_1(h_s\Lambda) = \infty$$

and

$$\limsup_{s\to\infty} \ell(h_s\Lambda) = 0.$$

In fact, this does not require equidistribution but simply density (due to Hedlund [24]). $\{h_s\Lambda\}$ is closed if and only if Λ has vertical vectors.

A natural question is to understand the *rate* at which these excursions to the non-compact part ('cusp') of X_2 occur. The first-named author, in joint work with G. Margulis [4], showed that for any $\Lambda \in X_2$ without vertical vectors,

$$\limsup_{s\to\infty} \frac{\log \alpha_1(h_s\Lambda)}{\log s} \geq 1,$$

and related the precise limit to Diophantine properties of the lattice Λ (see also [2]).

Our results concern the average behavior of *all* visits to the cusp, as defined by local minima of the function $\ell_\Lambda(s) = \ell(h_s\Lambda)$ (or, equivalently, local maxima of $\alpha_\Lambda(s) = \alpha(h_s\Lambda)$). The function $\ell_\Lambda(s)$ is a piecewise-linear function of s, and helps give a picture of the 'height' of the horocycle at time s. In this way it is similar to work of the second author [14], where a piecewise linear description of diagonal flows on $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$ was studied.

We also consider returns to the compact part of X_2, given by local minima of $\ell_\Lambda(s)$. Given $\Lambda \in X_2$ so that $\{h_s\Lambda\}_{s\geq 0}$ is not closed, let $\{s_n\}$ and $\{S_n\}$ denote the sequences of minima and maxima of $\ell_\Lambda(s)$ respectively. To be completely precise, the local minima for the supremum norm occur in intervals, and we take s_n to be the *midpoint* of the n^{th} interval. Define the averages

$$a_N(\Lambda) := \frac{1}{N}\sum_{n=1}^{N} \alpha_\Lambda(s_n) \ ; \ A_N(\Lambda) := \frac{1}{N}\sum_{n=1}^{N} \alpha_\Lambda(S_n),$$

$$l_N(\Lambda) := \frac{1}{N}\sum_{n=1}^{N} \ell_\Lambda(s_n) \ ; \ L_N(\Lambda) := \frac{1}{N}\sum_{n=1}^{N} \ell_\Lambda(S_n).$$

Theorem 1.6. For any Λ without vertical vectors, we have

$$\lim_{N\to\infty} a_N(\Lambda) = 2 \tag{1.6}$$

$$\lim_{N\to\infty} l_N(\Lambda) = \frac{2}{3} \tag{1.7}$$

□

Let $M : \Omega \to [0,1]$ be given by

$$M(a,b) = \max\left\{a, b, \frac{1}{a+b}\right\} \tag{1.8}$$

Theorem 1.7. For any Λ without vertical vectors, we have

$$\lim_{N\to\infty} A_N(\Lambda) = \int_\Omega \frac{1}{M} dm = \frac{2}{3}\left(13 - 8\sqrt{2}\right) \approx 1.2 \tag{1.9}$$

$$\lim_{N\to\infty} L_N(\Lambda) = \int_\Omega M dm = \frac{2}{3}\left(7 - 4\sqrt{2}\right) \approx .73 \tag{1.10}$$

□

Theorem 1.6 and Theorem 1.7 follow from applying the ergodic theorem to the BCZ transformation $T : \Omega \to \Omega$. The assumption that Λ does not have vertical vectors is to ensure it does not have a periodic orbit under T (equivalently, h_s). There is a version of this result for periodic orbits (Corollary 1.12) given in §1.6.5 The function M gives the maximum of the function ℓ on a sojourn from the transversal Ω. The limits (1.6) and (1.7) in Theorem 1.6 are the integrals of the functions $g(a,b) = \Lambda$ and $h(a,b) = \frac{1}{\Lambda}$ over Ω respectively. The times s_n correspond to the return times of $\{h_s\Lambda\}$ to Ω.

1.6 Farey Statistics

Theorem 1.3 has several number theoretic corollaries. We record results on spacing and indices of Farey fractions, originally proved using analytic methods in [5, 8, 20]. Our results give a unified explanation for these equidistribution phenomena. For similar applications in the context of higher dimensional generalizations of Farey sequences we refer the reader to [30] and [29].

1.6.1 *Spacings and h-spacings*

We now fix an interval $I \subset [0,1]$. It is well known that the Farey sequences $\mathcal{F}_I(Q)$ become equidistributed in $[0,1)$ as $Q \to \infty$. A natural statistical question is to understand the distribution of the *spacings* $(\gamma_{i+1} - \gamma_i)$. Since there are $N_I(Q)$ points, and $N_I(Q) \sim |I|\frac{3}{\pi^2}Q^2$, the natural normalization yields the following question: given $0 \le c \le d$, what is the limiting $(Q \to \infty)$ behavior of

$$\frac{|\{\gamma_i \in \mathcal{F}_I(Q) : \frac{3}{\pi^2}|I|Q^2(\gamma_{i+1} - \gamma_i) \in [c,d]\}|}{N_I(Q)}?$$

Since $Q^2(\gamma_{i+1} - \gamma_i) = \left(R \circ T^{i-1}\right)\left(\frac{1}{Q}, 1\right)$, the above quantity reduces to

$$\rho_{Q,I}\left(R^{-1}\left(\left[\frac{\pi^2}{3|I|}c, \frac{\pi^2}{3|I|}d\right]\right)\right). \tag{1.11}$$

Since $R^{-1}\left(\left[\frac{\pi^2}{3|I|}c, \frac{\pi^2}{3I|}d\right]\right)$ is *compact*, we can directly apply Theorem 1.3 to (1.11) obtain a result of R. R. Hall [20]:

Corollary 1.8. As $Q \to \infty$,

$$\frac{|\{\gamma_i \in \mathcal{F}_I(Q) : \frac{3}{\pi^2}|I|Q^2(\gamma_{i+1} - \gamma_i) \in (c,d)\}|}{N_I(Q)} \to m\left(R^{-1}\left(\frac{\pi^2}{3|I|}c, \frac{\pi^2}{3|I|}d\right)\right).$$

□

We call this distribution *Hall's distribution*. A generalization of this result to higher-order spacings was considered by Augustin-Boca-Cobeli-Zaharescu [5]. They considered *h-spacings*: the vector of h-tuples $(h \ge 1)$ of spacings $\mathbf{v}_{i,h} = \mathbf{v}_{i,h}(Q) = (\gamma_{i+j} - \gamma_{i+j-1})_{j=1}^h \in \mathbb{R}^h$. Given $\mathcal{B} = \prod_{i=1}^k [c_i, d_i]$, we have, as above,

$$\frac{\left|\left\{\gamma_i \in \mathcal{F}_I(Q) : \frac{3}{\pi^2}|I|Q^2\mathbf{v}_{i,h} \in \mathcal{B}\right\}\right|}{N_I(Q)} = \rho_{Q,I}\left(R_h^{-1}\left(\tilde{B}\right)\right), \tag{1.12}$$

where $R_h : \Omega \to \mathbb{R}^h$ is given by

$$R^h(a,b) = \left(R\left(T^{j-1}(a,b)\right)\right)_{j=1}^h,$$

and

$$\tilde{B} = \frac{\pi^2}{3|I|}B = \prod_{i=1}^k \left[\frac{\pi^2}{3|I|}c_i, \frac{\pi^2}{3|I|}d_i\right].$$

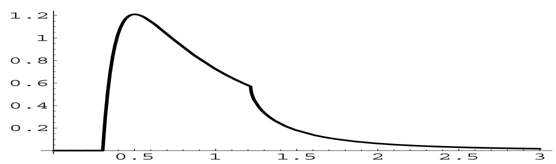
Applying our equidistribution result to (1.12), we recover Theorem 1.1 of [5]:

Corollary 1.9. As $Q \to \infty$,

$$\frac{|\{\gamma_i \in \mathcal{F}_I(Q) : \frac{3}{\pi^2}Q^2\mathbf{v}_{i,h} \in \mathcal{B}\}|}{N_I(Q)} \to m\left(R_h^{-1}\left(\tilde{B}\right)\right).$$

□

Fig. 4. The distribution function for the gaps [7, Figure 2]



1.6.2 *Density of the limiting distribution*

Corollary 1.8 allows one to explicitly calculate the density of the limiting distribution of consecutive gaps, that is, of Hall's distribution. For $d > 0$, we define the distribution function $G_c(d)$ to be the asymptotic proportion of (normalized) gaps of size at most b, that is,

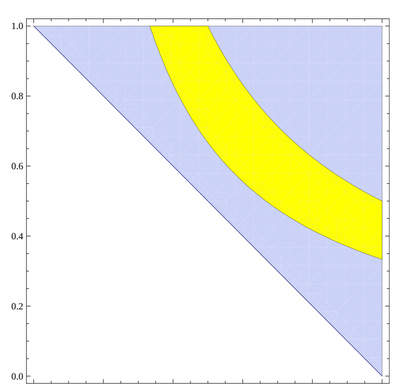
$$G_c(d) = \lim_{Q\to\infty} \frac{|\{\gamma_i \in \mathcal{F}_I(Q) : \frac{3}{\pi^2}|I|Q^2(\gamma_{i+1} - \gamma_i) \in (0, d)\}|}{N_I(Q)}.$$

By Corollary 1.8, we can write

$$G_c(d) = m\left(R^{-1}\left(0, \frac{\pi^2}{3|I|}d\right)\right).$$

Since for $d < \frac{3|I|}{\pi^2}$, the curve $R(a, b) = \frac{\pi^2}{3|I|}d$ does not intersect Ω, we have $G_c(d) = 0$. In particular, the distribution does not have any support at 0. For $d \geq \frac{3|I|}{\pi^2}$, $G_c(d) > 0$. The distribution has another point of non-differentiability at $d = 4\frac{3}{\pi^2}$, when the curve $R(a, b) = \frac{3|I|}{\pi^2}d$ intersects the bottom line $a + b = 1$. The picture of the distribution for $|I| = [0, 1)$ is given in Figure 4, reproduced from [7, Figure 2]. In Figure 5, we display a picture of the region $R^{-1}([c, d])$.

Fig. 5. The region $R^{-1}([c, d]) \subset \Omega$.



1.6.3 *Indices of Farey fractions*

Following [21], we define the *index* of the Farey fraction $\gamma_i = \frac{a_i}{q_i} \in \mathcal{F}(Q)$ by

$$\nu(\gamma_i) := \frac{q_{i-1} + q_{i+1}}{q_i} = \left\lfloor \frac{Q + q_{i-1}}{q_i} \right\rfloor.$$

In [8], this was reinterpreted in terms of the map T and the function $\kappa : \Omega \to \mathbb{N}$. Recall that $\kappa(x,y) = \left\lfloor \frac{1+x}{y} \right\rfloor$. Precisely, we have

$$\nu(\gamma_i) = \kappa\left(\frac{q_{i-1}}{Q}, \frac{q_i}{Q}\right) = \left(\kappa \circ T^i\right)\left(\frac{1}{Q}, 1\right).$$

Fix a (non-trivial) interval $I \subset [0,1)$, and an exponent $\alpha \in (0,2)$. We consider the average

$$\rho_{Q,I}(\kappa^\alpha) = \frac{1}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \nu(\gamma_i)^\alpha.$$

Applying (a strengthening of)Theorem 1.3 to κ^α (for which we need to estimate the integrals of κ^α, see §5.2.2), we obtain a result originally due to Boca-Gologan-Zaharescu [10]:

Corollary 1.10. As $Q \to \infty$,

$$\frac{1}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \nu(\gamma_i)^\alpha \to \int_\Omega \kappa^\alpha dm = 2 \sum_{k=1}^\infty k^\alpha \operatorname{area}(\Omega_k).$$

Here $\Omega_k = \kappa^{-1}(k)$. □

1.6.4 *Powers of denominators*

We can also obtain general results on sums associated to Farey fractions by applying Theorem 1.3 to various functions in $L^1(\Omega, m)$. For example, if we define $f_{s,t} : \Omega \to \mathbb{C}$ by

$$f_{s,t}(a,b) = a^s b^t,$$

for $s,t \in \mathbb{C}$ with $\Re s, \Re t > -1$, and define

$$B_{s,t} := \|f_{s,t}\|_1 = \int_\Omega f_{s,t}(x,y) dm = 2\left(\frac{1}{(s+1)(t+1)} - \frac{\Gamma(s+1)\Gamma(t+1)}{\Gamma(s+t+3)}\right),$$

we obtain results originally due to Hall-Tanenbaum [22].

Theorem 1.11. Fix a non-trivial interval $I \subset [0,1)$. Let $s,t \in \mathbb{C}$. Then, if $\Re s, \Re t > -1$,

$$\lim_{Q\to\infty} \frac{1}{N_I(Q)Q^{s+t}} \sum_{\gamma_i \in \mathcal{F}_I(Q)} q_i^s q_{i+1}^t = \int_\Omega x^s y^t dm = B_{s,t}. \tag{1.13}$$

In particular, for $s = 1$ and $t = 0$, we have

$$\lim_{Q\to\infty} \frac{1}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \frac{q_i}{Q} = \int_\Omega x dm = \frac{2}{3}. \tag{1.14}$$

In addition, for $s = -1$ and $t = 0$,

$$\lim_{Q\to\infty} \frac{1}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \frac{Q}{q_i} = \int_\Omega \frac{1}{x} dm = 2, \tag{1.15}$$

and for $s = t = -1$, we have

$$\lim_{Q\to\infty} \frac{Q^2}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \frac{1}{q_i q_{i+1}} = \int_\Omega \frac{1}{xy} dm = \frac{\pi^2}{3}, \tag{1.16}$$

which yields the classical result

$$N_I(Q) \sim |I| \frac{3}{\pi^2} Q^2.$$

□

1.6.5 *Excursions*

Finally, we remark that applying the equisitribution result Theorem 1.3 to the functions M (and $\frac{1}{M}$) defined in (1.8) above, we obtain an amusing statistical result on Farey fractions:

Corollary 1.12. Let $I \subset [0,1)$ be a non-trivial interval. Then

$$\lim_{Q\to\infty} \frac{1}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \min\left(\frac{Q}{q_i}, \frac{Q}{q_{i+1}}, \frac{q_i+q_{i+1}}{Q}\right) = \frac{2}{3}\left(13 - 8\sqrt{2}\right)$$

$$\lim_{Q\to\infty} \frac{1}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \max\left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}, \frac{Q}{q_i+q_{i+1}}\right) = \frac{2}{3}\left(7 - 4\sqrt{2}\right)$$

□

1.7 **Geometry of numbers**

Let $\Lambda \in X_2$ be a unimodular lattice, and fix $t > 0$. Let

$$S_t(\Lambda) = \{s_1 < s_2 < \ldots < s_n < \ldots\}$$

denote the slopes of the lattice vectors in the vertical strip $V_t \subset \mathbb{R}^2$ given by

$$V_t := \{(x, y) : x \in (0, t], y > 0\},$$

written in increasing order. Let

$$G_{N,t}(\Lambda) = \{s_{n+1} - s_n : 0 \le n \le N\}$$

denote the sequence of gaps in this sequence, viewed as a *set*, so $|G_{N,t}(\Lambda)| \le N$. Our main geometry of numbers result states that for lattices Λ without vertical vectors, this sequences has the same limiting distribution as the gaps for Farey fractions, namely, Hall's distribution. That is:

Theorem 1.13. Suppose Λ does not have vertical vectors, and let $0 \le c \le d \le \infty$. Then

$$\lim_{N\to\infty} \frac{1}{N} |G_{N,t}(\Lambda) \cap (c,d)| = 2m(R^{-1}(c,d)).$$

□

This result will follow from the application of the Birkhoff ergodic theorem to the orbit of Λ under the BCZ map, with observable given by the indicator function of the set $R^{-1}(c,d)$. If Λ does have a vertical vector, and thus a periodic orbit under the BCZ map, we have that for any $t > 0$, there is an $N_0 = N_0(t) > 0$ so that $G_{N,t}(\Lambda) = G_{N_0,t}(\Lambda)$ (as sets of numbers) for all $N > N_0$. We can use Theorem 1.3 to obtain the following:

Corollary 1.14. Let $0 \le c \le d \le \infty$. Then

$$\lim_{t\to\infty} \frac{1}{N_0(t)} |G_{N_0(t),t}(\Lambda) \cap (c,d)| = 2m(R^{-1}(c,d)).$$

□

2 **Construction of Transversal**

In this section, we prove Theorem 1.1. We first give an interpretation of the transversal Ω in terms of geometry of numbers in §2.1; and prove Theorem 1.1 in §2.1.1. We also record some observations on slopes in §2.1.2. We show how Ω can be interpreted in terms of hyperbolic geometry in §2.2. In §2.3, we show a certain self-similarity property of the BCZ map T.

2.1 Short horizontal vectors

Fix $t > 0$. We say a lattice $\Lambda \in X_2$ is *t-horizontally short* if it contains a non-zero horizontal vector $\mathbf{v} = (a, 0)^{\mathrm{T}}$ so that $|a| \le t$. Note that since Λ is a group, we can assume $a > 0$. We will call 1-horizontally short lattices simply *horizontally short*. We also have the analgous notions of t-vertically and vertically short.To prove Theorem 1.1, we break it up into several lemmas. Our first lemma is:

Lemma 2.1. $\Omega = \{\Lambda_{a,b} : a, b \in (0, 1], a + b > 1\}$ is the set of horizontally short lattices. □

Proof. Since the horizontal vector $(a, 0)^{\mathrm{T}}$ is in $\Lambda_{a,b}$, and $a \le 1$, clearly every lattice in Ω is short. To show the reverse containment, suppose Λ is short. Then we can write $\Lambda = \Lambda_{a,b'} = p_{a,b'}\mathbb{Z}^2$, with $0 < a \le 1$, and $b' \ne 0$. Let $m \in \mathbb{Z}$ be such that

$$1 - a < ma + b' \le 1,$$

i.e., $m = \left\lfloor \frac{1-a}{b'} \right\rfloor$. Set $b = ma + b'$. Then, since $p_{1,m} \in SL(2, \mathbb{Z})$,

$$\Lambda = \Lambda_{a,b'} = p_{a,b'}\mathbb{Z}^2 = p_{a,b'}p_{1,m}\mathbb{Z}^2 = \Lambda_{a,b} = \Lambda_{a,b} \in \Omega'.$$

■

Next, we show that for any $(a, b) \in \Omega$, that there is a $s \in (0, \infty)$ so that $h_s\Lambda_{a,b} \in \Omega'$, and give a formula for the minimum s.

Lemma 2.2. Let $(a, b) \in \Omega$, so $\Lambda_{a,b} \in \Omega'$. Let $s_0 = R(a, b) = \frac{1}{ab}$. Then $h_{s_0}\Lambda \in \Omega'$, and for every $0 < s < s_0$, $h_s\Lambda_{a,b} \notin \Omega'$. Furthermore,

$$h_{s_0}\Lambda_{a,b} = \Lambda_{T(a,b)}.$$

□

Proof. Since Ω' consists of horizontally short lattices, we first note that the first time $h_s\Lambda_{a,b}$ will have a horizontal vector is given by the equation

$$-sb + \frac{1}{a} = 0.$$

Thus we set $s_0 = \frac{1}{ab}$, and a direct calculation shows

$$h_{s_0}p_{a,b} = \begin{pmatrix} a & b \\ -\frac{1}{b} & 0 \end{pmatrix}.$$

Let $\kappa(a, b) = \lfloor \frac{1+a}{b} \rfloor$. Applying the matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & \kappa(a,b) \end{pmatrix} = \left(A(a,b)^{-1}\right)^{\mathrm{T}}$$

we obtain

$$h_{s_0}p_{a,b}\left(A(a,b)^{-1}\right)^{\mathrm{T}} = \begin{pmatrix} b & -a + \kappa(a,b)b \\ 0 & b^{-1} \end{pmatrix} = p_{T(a,b)}. \tag{2.1}$$

Thus, $h_{s_0}\Lambda_{a,b} = \Lambda_{T(a,b)}$, as desired. ■

Finally, we show that for any lattice $\Lambda \in X_2$ which is not vertically short, the orbit under $\{h_s\}_{s\ge 0}$ will intersect Ω'.

Lemma 2.3. Let $\Lambda \in X_2$ so that Λ is not vertically short. Then there is a $s_1 \in \mathbb{R}$ so that $h_{s_1}\Lambda \in \Omega'$. □

Proof. Let $S = [-1, 1]^2 \subset \mathbb{R}^2$. S is a square centered at 0, so is convex, centrally symmetric, and has area 4. Thus by the Minkowski convex body theorem, any lattice Λ must have a nonzero vector $\mathbf{v} \in S$. Since Λ is not vertically short, we can assume that $\mathbf{v} = (a, b)^{\mathrm{T}}$ is not vertical, that is, $a \ne 0$. Further, we can assume that $a \ge 0$, otherwise we multiply by -1. Let $s_1 = \frac{b}{a}$. Then $(a, 0)^{\mathrm{T}} \in h_{s_1}\Lambda$, so $h_{s_1}\Lambda \in \Omega'$.

■

2.1.1 *Proof of Theorem 1.1*

To prove Theorem 1.1, we combine the above lemmas. Lemma 2.3 guarantees that all non-vertically short lattice horocycle orbits $\{h_s\Lambda\}_{s\in\mathbb{R}}$ intersect Ω', and Lemma 2.2 shows that if an h_s-orbit intersects Ω' once, it must intersect Ω' infinitely often (both forward and backward in time). To see that the set of intersection times is discrete, observe that the roof function $R(a,b) = \frac{1}{ab}$ is bounded below by 1 on Ω, so visits must be spread out at least time 1 apart. The calculation of the roof function R and the return map T are also given by Lemma 2.2.

Finally, we address the remark following Theorem 1.1. A direct calculation shows that if a lattice Λ has a vertical vector, then it is periodic under h_s with period t^2, where t is the length of the shortest vertical vector. If Λ is vertically short, this means that it is periodic under h_s of period ≤ 1, so the h_s-orbit of Λ cannot intersect Ω'. We will see in §2.2 below that these orbits correspond to embedded closed horocycles, the family of which foliates the cusp of $\mathbb{H}^2/SL(2,\mathbb{Z})$. ∎

2.1.2 *Observations on slopes*

Given a unimodular lattice Λ, let

$$\{0 \leq s_1 < s_2 < \ldots < s_N \ldots\}$$

denote the sequence of slopes of vectors in $\Lambda \cap V_1$, where

$$V_1 = \{0 < x \leq 1, y > 0\} \subset \mathbb{R}^2$$

is the vertical strip as in §1.7. Then we see that s_1 is in fact the first hitting time of Ω' of the positive orbit $h_s\Lambda$, and the s_n are the subsequent hitting times, since Ω' is the set of horizontally short lattices. The slopes of vectors decrease by s under the horocycle flow h_s (in particular, while h_s does not preserve slopes, it preserves *differences of slopes*), and the BCZ map records them when they become horizontal, while also keeping track of their horizontal component (the coordinate a) and the *next* vector in the strip to become horizontal (the vector $(b, a^{-1})^{\mathrm{T}}$). The differences in slopes are the return times R, that is, for $n \geq 1$,

$$s_{n+1} - s_n = R(T^n(h_{s_1}\Lambda)). \tag{2.2}$$

This observation will be crucial for the proofs our results on the geometry of numbers, see §7.

2.2 **Hyperbolic geometry**

We recall that the there is a natural identification of the the upper-half plane

$$\mathbb{H}^2 = \{z = x + iy : y > 0\}.$$

with the space of unimodular lattices with a horizontal vector, via

$$z \mapsto \frac{1}{\sqrt{y}} \begin{pmatrix} 1 & x \\ 0 & y \end{pmatrix} \mathbb{Z}^2 = \Lambda_{\frac{1}{\sqrt{y}}, \frac{x}{\sqrt{y}}}.$$

That is, we identify z with the unimodular lattice homothetic to the lattice generated by 1 and z. This extends to an identification of all unimodular lattices with the unit-tangent bundle $T^1\mathbb{H}^2$, where the point (z, i) (where i denotes the upward pointing tangent vector) is identified to the lattice with a horizontal vector, and the point $(z, e^{i\theta}i)$ is identified to the lattice rotated by angle θ. This identification is well-defined up to the action of $SL(2,\mathbb{Z})$ by isometries on $\mathbb{H}^2$. In the standard fundamental domain for $SL(2,\mathbb{Z})$,

$$\left\{z = x + iy \in \mathbb{H}^2 : |z| > 1, |x| < \frac{1}{2}\right\},$$

the set of horizontally short lattices can be identified with the strip

$$C_1 := \left\{z = x + iy \in \mathbb{H}^2 : y > 1, |x| < \frac{1}{2}\right\},$$

together with their upward pointing tangent vectors, see Figure 6.

The explicit map between this region and Ω' is given by $(x,y) \mapsto \Lambda_{I(x,y)}$, where

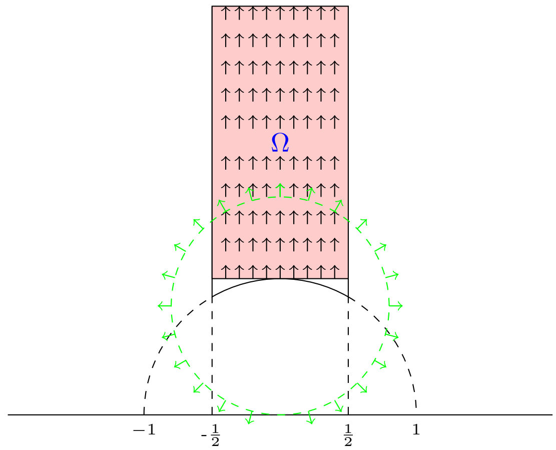
$$I(x,y) = (y^{-1/2}, xy^{-1/2} + my^{-1/2}),$$

where

$$m = \left\lfloor \frac{1 - y^{-1/2}}{xy^{-1/2}} \right\rfloor = \left\lfloor xy^{1/2} - \frac{1}{x} \right\rfloor.$$

The choice of m is as in the proof of Lemma 2.3, which guarantees that the image $I(C_1) = \Omega'$, since (x,y) is already naturally identified with $\Lambda_{y^{-1/2}, xy^{-1/2}}$.

Fig. 6. A hyperbolic picture of $\Omega' \equiv C_1$. The green circle is an h_s-orbit. Vertical lines are g_t-orbits and horizontal lines are u_s-orbits.



2.2.1 *Geodesics and horocycles*

The orbits of the one-parameter subgroups

$$A := \left\{ g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} : t \in \mathbb{R} \right\}$$

and

$$U := \left\{ u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : t \in \mathbb{R} \right\},$$

the *geodesic* and *opposite horocycle* flows respectively, also have nice interpretations in terms of hyperbolic geometry. Under our choices, the action of $\{g_t : t \leq 0\}$ moves the upward pointing tangent vectors vertically upward in $\mathbb{H}^2$, and the flow $\{u_s\}$ moves them horizontally. In particular, Ω' is preserved by the action of $\{g_t : t \leq 0\}$ and $\{u_s\}$. In dynamical language, orbits of h_s are leaves of the *stable foliation* for $\{g_t : t \leq 0\}$, and $\{u_s\}$ are leaves of the *strong unstable foliation*. We are constructing a cross section of h_s by taking a piece of the *unstable foliation* AU. That is, $\Omega' \subset AU$.

In the Euclidean picture, direct calculations show that

$$g_t \Lambda_{a,b} = \Lambda_{e^{-t/2}a, e^{-t/2}b} \text{ and } u_s \Lambda_{a,b} = \Lambda_{a, b+sa^{-1}},$$

so geodesic orbits are radial lines and opposite horocyclic orbits are vertical lines, see Figure 7. We will use these observations in crucial ways in §2.3 below.

2.2.2 *Identifications*

Note that in the fundamental domain

$$\{z = x + iy : |z| \geq 1, |x| \leq \frac{1}{2}\}$$

for $SL(2, \mathbb{Z})$ acting on $\mathbb{H}^2$, there is a natural identification of the vertical sides $x = \frac{1}{2}$ and $x = -\frac{1}{2}$ via the transformation

$$z \mapsto z + 1,$$

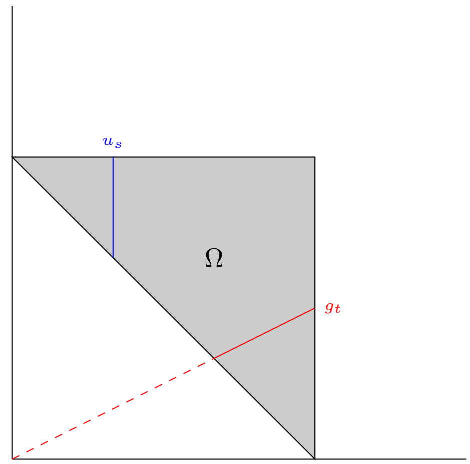
which is a fractional linear transformation associated to the unipotent matrix

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

In the Euclidean picture, we can also identify two boundaries of the region Ω via the above unipotent matrix, namely the lines $\{x + y = 1 : 0 < x \leq 1\}$ and $\{y = 1 : 0 < x \leq 1\}$. The resulting loop formed by the vertical segment

$$\{x = 1 : 0 < y \leq 1\}$$

Fig. 7. Geodesic and opposite horocyclic orbits in the Euclidean picture of Ω.



corresponds to the closed horocycle

$$\{z = 1 + iy : 0 \leq y \leq 1\}$$

in the hyperbolic picture. Thus, the topology of Ω is that of a sphere with one puncture (corresponding to the point at ∞ in the hyperbolic picture and the point $(0, 1)$ in the Euclidean picture) and one boundary component (corresponding to the loop described above).

2.3 Self-similarity

The BCZ map has an extraordinary self-similarity property which can be seen naturally in both the hyperbolic and Euclidean pictures. Let $t > 0$, and let Ω'_t denote the set of t-horizontally short lattices. Arguing as in Lemma 2.1, we have the identification with the set

$$\Omega_t := \{(a, b) \in (0, t] : a + b > t\}. \tag{2.3}$$

Appropriately modifying Lemma 2.3, we can define a return map (the *t-BCZ map*)

$$T_t : \Omega_t \to \Omega_t,$$

which captures the h_s-orbits of all lattices except those of $\frac{1}{t}$-vertically short lattices. Modifying the argument in Lemma 2.2, we have

$$T_t(a, b) = \left(b, -a + \left\lfloor \frac{t + a}{b} \right\rfloor b\right). \tag{2.4}$$

A direct calculation shows that the t-BCZ map is conjugate to the original (1-) BCZ map via the linear transformation $M_t : \Omega \to \Omega_t$ given by

$$M_t(a, b) = (ta, tb),$$

that is

$$T_t \circ M_t = M_t \circ T. \tag{2.5}$$

Now assume $t < 1$ (if $t > 1$, the roles of t and 1 below should be reversed). Then the set of t-horizontally short lattices can also be identified with the subset $\Omega^{(t)} \subset \Omega$ given by

$$\Omega^{(t)} := \{(a, b) \in \Omega : a < t\}.$$

Let $T^{(t)} : \Omega^{(t)} \to \Omega^{(t)}$ be the first return map of the BCZ map T to $\Omega^{(t)}$. Thus, $T^{(t)}$ also represents the first return map of h_s to the set of t-horizontally short lattices, and so $T^{(t)}$ and T_t are conjugate. This conjugacy can be made explicit via the map $L_t : \Omega^{(t)} \to \Omega_t$ given by

$$L_t(a, b) = \left(a, b + \left\lfloor \frac{t - b}{a} \right\rfloor b\right),$$

Fig. 8. Ω_t and $\Omega^{(t)}$. Ω can be identified with Ω_t via the scaling M_t, and Ω_t and $\Omega^{(t)}$ via the map L_t.

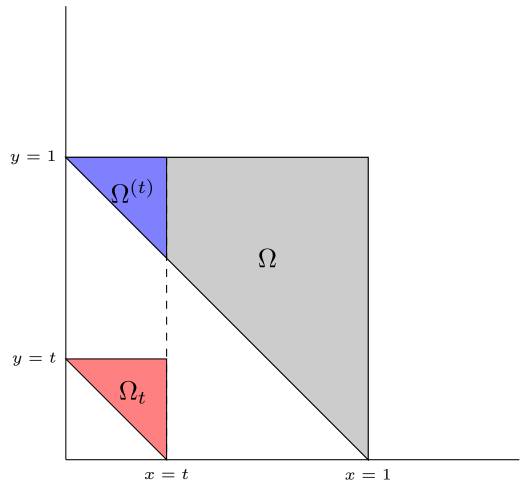
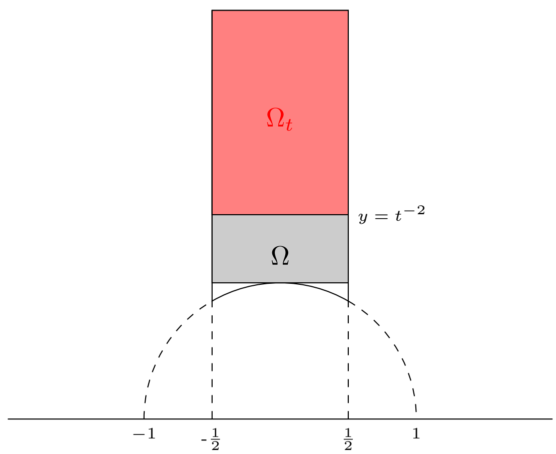


Fig. 9. A hyperbolic picture of $\Omega_t \subset \Omega$, that is, $C_t \subset C_1$.



that is,

$$T_t \circ L_t = L_t \circ T^{(t)}.$$

In the hyperbolic picture, the set of t-horizontally short lattices can be identified with the subset

$$C_t := \left\{ z = x + iy \in \mathbb{H}^2 : y > t^{-2}, |x| < \frac{1}{2} \right\}.$$

Figures 8 and 9 show the Euclidean and hyperbolic pictures respectively. We can summarize this discussion with the following:

Observation 1. T is self-similar in the following sense: For any $0 < t < 1$, the BCZ map T is conjugate to its own first-return map $T^{(t)}$. □

Essentially, this self-similarity is a consequence of the following conjugation relation for g_t and h_s

$$g_t h_s g_{-t} = h_{se^{-t}}, \tag{2.6}$$

which in particular implies that the time-1 map h_1 and time-s map h_s are conjugate for any $s > 0$ (via the map $g_{\log s}$).

3 Ergodic properties of the BCZ map

In this section, we prove Theorem 1.2, using Theorem 1.1 and well-known properties of suspension flows. We first recall these properties, which can be found, e.g., in [31].

3.1 Suspension Flows

Let (X, μ) be a measure space, and $T : X \to X$ a measure-preserving bijection. Let $R : X \to \mathbb{R}^+$ be in $L^1(X, \mu)$. The *suspension flow* over T with roof function R is defined on the space

$$X^R := \{(x, t) : x \in X, t \in [0, R(x))\}/\sim \tag{3.1}$$

where $(x, R(x)) \sim (T(x), 0)$. The *suspension flow* $\phi^{R,T}$ is given by

$$\phi_s^{R,T}(x, t) = (x, s + t).$$

The measure $d\nu_{R,T} = \frac{1}{\|R\|_1} d\mu dt$ is a $\phi^{R,T}$-invariant probability measure on X^R. It is ergodic if and only if μ is T-ergodic.The natural dual construction to suspension flow is the construction of a *first return map for a flow.* Given a measure-preserving flow $\phi : (Y, \nu) \to (Y, \nu)$, and a subset $X \subset Y$ so that for almost all $y \in Y$, $\{t \in \mathbb{R} : \phi^t y \in X\}$ is discrete, we define the *first return map* $T : X \to X$ by $T(x) = \phi^{R(x)}(x)$, where

$$R(x) = \inf\{t > 0 : \phi^t(x) \in X\}$$

is the *first return time*. There is a natural (possibly infinite) T-invariant measure μ on X so that $R \in L^1(X, \mu)$. and the suspension flow $\phi^{R,T} : (X^R, \nu_{R,T}) \to (X^R, \nu_{R,T})$ is naturally isomorphic to the original flow ϕ. There is also a map between the sets of invariant measures for the flow $\phi^{R,T}$ and the map T, whose properties are summarized in the following:

Lemma 3.1. Let (X, μ) be a measure space, and let $T : (X, \mu) \to (X, \mu)$ be a μ-preserving bijection. Let R be a positive function in $L^1(X, \mu)$. The map $\eta \longmapsto \eta_{R,T}$, with $d\eta_{R,T} = \frac{1}{\|R\|_{1,\eta}} d\eta dt$ is a bijection between the set of T-invariant measures η on X so that $R \in L^1(X, \eta)$, and the set of $\phi^{R,T}$ invariant probability measures on X^R (and thus also between the ergodic invariant measures). Moreover, we have *Abramov's formula* relating the entropy of the flow $\phi^{R,T}$ (that is, of its time 1-map) to the entropy of the map T:

$$h_{\eta^{R,T}}(\phi_1^{R,T}) = \frac{h_\eta(T)}{\|R\|_{1,\eta}}. \tag{3.2}$$

□

3.2 Ergodicity

In this section, we prove the ergodicity and measure-classification parts of Theorem 1.2. The ergodicity of the BCZ map with respect to $dm = 2dadb$ follows from

Ergodicity of the horocycle flow with respect to Haar measure μ_2 on X_2.

Suspension The measure $2dadbds$ on X_2 (viewing it as the suspension space over Ω) is an absolutely continuous invariant measure for $\{h_s\}$.

Measure classification By Dani's measure classification [15], this must be (a constant multiple of) μ_2.

To show uniqueness, suppose ν was another ergodic T invariant measure on Ω. Then it must be supported on a periodic orbit for T, since the measure $d\nu ds$ on X_2 is a non-Haar ergodic invariant measure for $\{h_s\}$ and thus, by Dani's measure classification, must be supported on a periodic orbit for $\{h_s\}$. ■

3.3 Entropy

We now prove the fact that the entropy $h_m(T)$ of T with respect to the Lebesgue probability measure m is 0, completing the proof of Theorem 1.2. We use Lemma 3.1), which states that the entropy $h_m(T)$ of the BCZ map with respect to m is proportional to the entropy of the horocycle flow $h_{\mu_2}(h_1)$ with respect to Haar measure. To show that $h_{\mu_2}(h_1) = 0$, we record a fact which seems to be well-known but does not have a complete proof in the literature as far as we are aware:

Theorem 3.1. Let $\Gamma \subset SL(2,\mathbb{R})$ be a lattice. Let μ_2 denote the finite measure on $SL(2,\mathbb{R})/\Gamma$ induced by Haar measure. Then

$$h_{\mu_2}(h_1) = 0.$$

In fact, the topological entropy

$$h_{top}(h_1) = 0.$$

□

Proof. When Γ is a *uniform* lattice, that is, the quotient $\mathbb{H}^2/\Gamma$ is compact, the second assertion is a theorem of Gurĕvic [19], and together with the standard variational principle (see, e.g., Walters [35]), which states that the topological entropy is the supremum of the measure-theoretic entropies. When Γ is non-uniform, we can use a version of the variational principle for locally compact spaces due to Handel-Kitchens [23], to again reduce the measure-theoretic statement to a topological statement. Since the horocycle flow is C^∞, we can apply a theorem of Bowen [12] to conclude that

$$h_{top}(h_1) < \infty.$$

On the other hand, since h_s is conjugate to h_1 for any $s > 0$, we have

$$h_{top}(h_1) = h_{top}(h_s) = s h_{top}(h_1),$$

which implies that $h_{top}(h_1) = 0$. ■

4 Structure of Periodic Orbits

In this section we prove Theorems 1.4 and 1.5 and describe in detail the relationship between periodic orbits for T and h_s. We first record (in §4.1) our geometric proof of the key observation (1.5) that certain periodic orbits of the BCZ map parameterize Farey fractions.

4.1 Farey fractions and lattices

In this section, we give a geometric proof the following

Lemma 4.1. Let Q be a positive integer, and let

$$\mathcal{F}(Q) = \{\frac{0}{1} = \gamma_1 < \gamma_2 = \frac{1}{Q} < \ldots < \gamma_i = \frac{p_i}{q_i} < \ldots < \gamma_N = \frac{1}{1}\},$$

$N = N(Q) = \sum_{q \le Q} \varphi(q)$ denote the Farey sequence. Then (1.5) holds, that is,

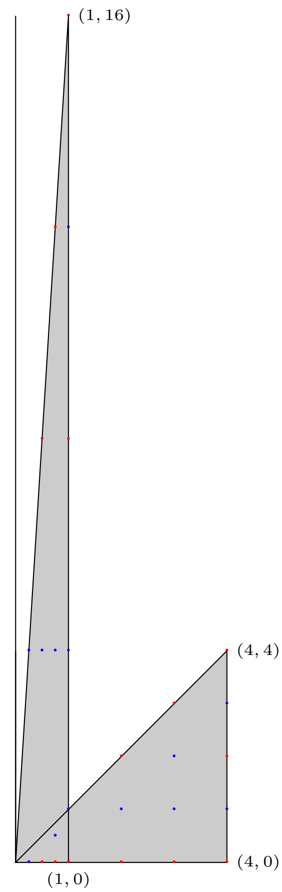
$$T^i\left(\frac{1}{Q}, 1\right) = \left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right).$$

□

Remark: Here, i is interepreted cyclically (i.e., in $\mathbb{Z}/N\mathbb{Z}$). In particular, this is a *periodic orbit* for T (and for h_s). Note also that the return time gives (normalized) gaps between the Farey fractions, that is,

$$R \circ T^i\left(\frac{1}{Q}, 1\right) = \frac{Q^2}{q_i q_{i+1}} = Q^2(\gamma_{i+1} - \gamma_i).$$

Fig. 10. The triangles T_4 and $g_{t_4}T_4$, with lattice points. Primitive lattice points in blue, others are in red .



Proof. Geometrically, the sequence $\mathcal{F}(Q)$ correspond to the slopes of primitive integer vectors $\begin{pmatrix} q_i \\ p_i \end{pmatrix}$ in the (closed) triangle T_Q with vertices at $(0,0)$, $(Q,0)$, and (Q,Q). Now note that since $\mathbb{Z}^2$ contains a vertical vector, it is *periodic* under h_s, with period 1. Using the conjugation relation (2.6), we have that the lattice $g_t\mathbb{Z}^2$ has period e^{-t} under h_s, since

$$h_{e^{-t}}g_t\mathbb{Z}^2 = g_t h_1\mathbb{Z}^2 = g_t\mathbb{Z}^2.$$

Note that g_t scales slopes (and their differences) by e^{-t}. For $Q \in \mathbb{N}$, setting $t_Q = -2\ln Q$, consider the lattice

$$g_{t_Q}\mathbb{Z}^2 = \begin{pmatrix} Q^{-1} & 0 \\ 0 & Q \end{pmatrix}\mathbb{Z}^2$$

Note that we can also write this as

$$\begin{pmatrix} Q^{-1} & 0 \\ 0 & Q \end{pmatrix}\mathbb{Z}^2 = \begin{pmatrix} Q^{-1} & 1 \\ 0 & Q \end{pmatrix}\begin{pmatrix} 1 & -Q \\ 0 & 1 \end{pmatrix}\mathbb{Z}^2 = \begin{pmatrix} Q^{-1} & 1 \\ 0 & Q \end{pmatrix}\mathbb{Z}^2$$

This lattice corresponds to the point $(Q^{-1}, 1)$ in the Farey triangle Ω. By our earlier observation, this has period Q^2 under h_s. We are interested in the behavior of the orbit of this point under T. Recall that the BCZ map only 'sees' primitive vectors with horizontal component less than 1 that is, in the strip $V = V_1$, since h_s does not change the horizontal component of vectors. Originally we were interested in Farey fractions of level Q, that is, the slopes of (primitive) integer vectors in the region T_Q. Applying the matrix g_{t_Q}, the region T_Q is transformed into a subset V. Thus, the vectors $\begin{pmatrix} q_i \\ p_i \end{pmatrix} \in T_Q$ correspond to vectors $\begin{pmatrix} Q^{-1}q_i \\ Qp_i \end{pmatrix} \in g_{t_Q}T_Q \subset V$. Therefore, they are seen by the BCZ map. Precisely, $g_{t_Q}T_Q$ is the triangle with vertices at $(0,0)$, $(1,0)$, and $(1,Q^2)$. See Figure 10.

As noted above, the (primitive) vectors $\begin{pmatrix} Q^{-1}q_i \\ Qp_i \end{pmatrix} \in g_{t_Q}\mathbb{Z}^2$ will be seen by the BCZ map in the order of their slopes (that is, at time $Q^2\gamma_i$), and the time in between the i^{th} and $(i+1)^{st}$ vectors will precisely be the difference in their slopes

$$Q^2(\gamma_{i+1} - \gamma_i).$$

At time 0 we are seeing the horizontal vector $\begin{pmatrix} Q^{-1} \\ 0 \end{pmatrix}$ (corresponding to γ_0), and the next vector to get short is $\begin{pmatrix} 1 \\ Q \end{pmatrix}$ (corresponding to γ_1), which will become horizontal in time

$$R(Q^{-1}, 1) = Q = Q^2\left(\frac{1}{Q} - 0\right) = Q^2(\gamma_1 - \gamma_0).$$

More generally, we have that $T^i(Q^{-1}, 1)$ corresponds to the lattice $h_{Q^2\gamma_i}g_{t_Q}\mathbb{Z}^2$, and noting

$$h_{Q^2\gamma_i}g_{t_Q}\begin{pmatrix} q_i & q_{i+1} \\ p_i & p_{i+1} \end{pmatrix} = \begin{pmatrix} Q^{-1}q_i & Q^{-1}q_{i+1} \\ 0 & Qq_i^{-1} \end{pmatrix}.$$

we obtain, as desired

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

and

$$R(T^i(Q^{-1}, 1)) = Q^2(\gamma_{i+1} - \gamma_i).$$

In particular, the point $(Q^{-1}, 1)$ is T-periodic with period $N(Q)$. Also note that

$$\sum_{i=0}^{N(Q)-1} R(T^i(Q^{-1}, 1)) = Q^2,$$

and thus, the sum of the return times over the discrete periodic orbit correspond, as they must, to the continuous period. ∎

4.2 **Proof of Theorem 1.4**

We first note that any periodic point $(a, b) \in \Omega$ for T corresponds to a periodic point $p_{a,b}\mathbb{Z}^2 \in \tilde{Y}$ for h_s on X_2 and vice-versa. Precisely, suppose $(a, b) \in \Omega$ is periodic under T, and $n = P(a, b) > 0$ is the minimal period, so

$$T^n(a, b) = (a, b)$$

and $T^j(a, b) \neq (a, b)$ for $j < n$. Then

$$h_{s(a,b)}p_{a,b}\mathbb{Z}^2 = p_{a,b}\mathbb{Z}^2,$$

where

$$s(a, b) = \sum_{i=0}^{n-1} R(T^i(a, b)).$$

Note that the points $(a, a) \in \Omega$ have period $P(a, a) = 1$, and that a direct calculation shows that $s(a, a) = R(a, a) = \frac{1}{a^2}$.

Periodic orbits for h_s occur in natural families, due the conjugation relation (2.6). In particular, if Λ is periodic with period s_0,

$$h_{s_0 e^{-t}}g_t\Lambda = g_t h_{s_0}\Lambda = g_t\Lambda, \tag{4.1}$$

so the flow period of $g_t g\mathbb{Z}^2$ is $s_0 e^{-t}$. We have

$$g_t\Lambda_{a,b} = \Lambda_{e^{t/2}a, e^{t/2}b},$$

so we see that if (a, b) is periodic for T, so is the entire line segment

$$\left\{(ta, tb), 1 \geq t > \frac{1}{a+b}\right\}.$$

Thus, to analyze which points have periodic orbits, it suffices to consider points of the form $(1,b)$ or $(a,1)$. Since $T(a,1) = (1, 1-a)$, we consider points of the first kind, or, equivalently, lattices of the form $\Lambda_{1,b}$. This is a periodic point under h_s if and only if there exists an s_0 so that $h_{s_0}\Lambda_{1,b} = \Lambda_{1,b}$, or, equivalently,

$$p_{1,b}^{-1} h_{s_0} p_{1,b} \in SL(2,\mathbb{Z}).$$

A direct calculation shows that this is impossible if y is irrational, and if $y = \frac{k}{l}$ the minimal such $s_0 = l^2$, and we have

$$p_{1,b}^{-1} h_{s_0} p_{1,b} = \begin{pmatrix} 1+kl & k^2 \\ -l^2 & 1-kl \end{pmatrix}. \tag{4.2}$$

This proves Theorem 1.4, since we have shown that any point with rational slope is periodic. ∎

4.3 Proof of Theorem 1.5

As above, we consider the point $(1,b)$. If $b = \frac{k}{l}$, $k \le l \in \mathbb{N}$ relatively prime, then the period under h_s is l^2. The line segment associated to this point is

$$\left\{ \left(t, t\frac{k}{l}\right) : t \in \left(\frac{l}{l+k}, 1\right] \right\}.$$

Applying (2.6) to $p_{t,t\frac{k}{l}}\mathbb{Z}^2$, we can calculate the period

$$s\left(t, t\frac{k}{l}\right) = \frac{l^2}{t^2}.$$

To calculate the discrete period, we analyze the scalings of the Farey periodic orbits

$$\left\{ T^i\left(1, \frac{1}{Q}\right) \right\}_{i=0}^{N(Q)-1}$$

using the following proposition, a special case of Theorem 1.5 which we will use to prove the general result:

Proposition 4.1. For $t \in \left(\frac{Q}{Q+1}, 1\right]$,

$$P\left(t, \frac{t}{Q}\right) = N(Q).$$

□

This will follow from showing that T in fact is *linear* along the orbit of the segment $\left\{ \left(t, \frac{t}{Q}\right) : t \in \left(\frac{Q}{Q+1}, 1\right] \right\}$, that is, the segment does not 'break up' into pieces. Precisely, we have:

Lemma 4.2. For $t \in \left(\frac{Q}{Q+1}, 1\right]$, $1 \le i \le N(Q)$,

$$\kappa\left(T^i\left(t, \frac{t}{Q}\right)\right) = \kappa\left(T^i\left(1, \frac{1}{Q}\right)\right).$$

□

Proof. Since $T^i\left(1, \frac{1}{Q}\right) = \left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right)$, we need to show, for $t \in \left(\frac{Q}{Q+1}, 1\right]$

$$\kappa\left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right) = \kappa\left(\frac{tq_i}{Q}, \frac{tq_{i+1}}{Q}\right).$$

We have

$$\begin{aligned}
\kappa\left(\frac{tq_i}{Q}, \frac{tq_{i+1}}{Q}\right) &= \left\lfloor \frac{\frac{Q}{t}+q_i}{q_{i+1}} \right\rfloor \\
&\leq \frac{\frac{Q}{t}+q_i}{q_{i+1}} \\
&< \frac{Q+1+q_i}{q_{i+1}} \\
&\leq \frac{q_{i+1}+q_{i+2}+q_i}{q_{i+1}} \\
&= 1+\left\lfloor \frac{Q+q_i}{q_{i+1}} \right\rfloor
\end{aligned} \tag{4.3}$$

where in the last line we are using the identity

$$\frac{q_i+q_{i+2}}{q_{i+1}} = \left\lfloor \frac{Q+q_i}{q_{i+1}} \right\rfloor .$$

On the other hand,

$$\frac{\frac{Q}{a}+q_i}{q_{i+1}} \geq \frac{Q+q_i}{q_{i+1}} \geq \left\lfloor \frac{Q+q_i}{q_{i+1}} \right\rfloor .$$

Thus,

$$\left\lfloor \frac{Q+q_i}{q_{i+1}} \right\rfloor \leq \kappa\left(\frac{tq_i}{Q}, \frac{tq_{i+1}}{Q}\right) < 1+\left\lfloor \frac{Q+q_i}{q_{i+1}} \right\rfloor ,$$

so

$$\kappa\left(\frac{tq_i}{Q}, \frac{tq_{i+1}}{Q}\right) = \left\lfloor \frac{Q+q_i}{q_{i+1}} \right\rfloor = \kappa\left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right)$$

as desired. ■

Now Proposition 4.1 follows from $P\left(1, \frac{1}{Q}\right) = N(Q)$. ■

To prove Theorem 1.5, we observe that the point $\left(1, \frac{k}{l}\right)$ is contained in the periodic orbit corresponding to the point $\left(\frac{1}{l}, 1\right)$, since we can find consecutive Farey fractions $\gamma_i = \frac{a_i}{l}$ and $\gamma_{i+1} = \frac{a_{i+1}}{k}$ in $\mathcal{F}(l)$ (we are assuming that $\gcd(k,l)=1$). By Proposition 4.1, for $a \in (\frac{l}{l+1}, 1]$,

$$P\left(a, a\frac{k}{l}\right) = N(l).$$

If we put $a = \frac{l}{l+1}$, we have the point $\left(\frac{l}{l+1}, \frac{k}{l+1}\right)$, which is contained in the periodic orbit of the point $\left(\frac{1}{l+1}, 1\right)$, by similar reasoning as above. Applying Proposition 4.1 to this point, we have that for $b \in \left(\frac{1}{l+1}, 1\right]$,

$$P\left(\frac{bl}{l+1}, \frac{bk}{l+1}\right) = N(l+1).$$

Rewriting, we have that for $a \in \left(\frac{l}{l+2}, \frac{l}{l+1}\right]$,

$$P\left(t, t\frac{k}{l}\right) = N(l+1).$$

Continuing to reason in this fashion, we have that for $1 \leq r \leq k$, $t \in \left(\frac{l}{l+r}, \frac{l}{l+r-1}\right]$,

$$P\left(t, t\frac{k}{l}\right) = N(l+r-1).$$

Finally, we need to calculate $A_n\left(t, t\frac{k}{l}\right)$, where $n = P\left(t, t\frac{k}{l}\right)$. Letting $s_0 = s\left(a, a\frac{k}{l}\right) = \frac{l^2}{a^2}$, we have, by a similar calculation to (4.2) that

$$p_{t,t\frac{k}{l}}^{-1} h_{s_0} p_{t,t\frac{k}{l}} = \begin{pmatrix} 1+kl & k^2 \\ -l^2 & 1-kl \end{pmatrix} .$$

On the other hand, (2.1) can be re-written as the conjugation

$$p_{T(a,b)}^{-1} h_{R(a,b)} p_{a,b} = A(a,b)^{\mathrm{T}}.$$

Iterating, we obtain, for any $m > 0, (a,b) \in \Omega$,

$$p_{T^m(a,b)}^{-1} h_{R(a,b)} p_{a,b} = A_m(a,b)^{\mathrm{T}}.$$

Applying this to $(a,b) = \left(t, t\frac{k}{l}\right)$ and $m = n$, we have

$$\left(A_n(a,b)^{-1}\right)^{\mathrm{T}} = \begin{pmatrix} 1+kl & k^2 \\ -l^2 & 1-kl \end{pmatrix},$$

since $T^n\left(t, t\frac{k}{l}\right) = \left(t, t\frac{k}{l}\right)$. Inverting and taking transposes, we obtain, as desired

$$A_{P\left(t,t\frac{k}{l}\right)}\left(t, t\frac{k}{l}\right) = \begin{pmatrix} 1-kl & l^2 \\ -k^2 & 1+kl \end{pmatrix},$$

completing the proof. ■

4.4 Segments and the shearing matrix

Note that the matrix

$$A_{P\left(a,a\frac{k}{l}\right)}\left(a, a\frac{k}{l}\right) = \begin{pmatrix} 1-kl & l^2 \\ -k^2 & 1+kl \end{pmatrix}$$

is parabolic, and fixes the segment $\left\{\left(a, a\frac{k}{l}\right) : a \in \left(\frac{l}{l+k}, 1\right]\right\}$. In fact, it *shears* along this segment in the following fashion. Note that

$$\frac{1}{k^2+l^2}\begin{pmatrix} l & k \\ -k & l \end{pmatrix}\begin{pmatrix} 1-kl & l^2 \\ -k^2 & 1+kl \end{pmatrix}\begin{pmatrix} l & -k \\ k & l \end{pmatrix} = \begin{pmatrix} 1 & l^2+k^2 \\ 0 & 1 \end{pmatrix}.$$

Thus for points (a,b) sufficiently near the segment, they will be sheared along the segment under the appropriate power of T (which may vary depending on the period of the subsegment (a,b) is close to). As it moves along the segment, the period will change. This is illustrated in Figure 11, which shows the periodic segment $\left\{\left(a, a\frac{2}{3}\right) : a \in \left(\frac{3}{5}, 1\right]\right\}$, which breaks up into two pieces, one of period 5 and one of period 6. The shearing matrix for this line is given by

$$\begin{pmatrix} -5 & 9 \\ -4 & 7 \end{pmatrix},$$

which is conjugate to

$$\begin{pmatrix} 1 & 13 \\ 0 & 1 \end{pmatrix}.$$

4.5 Hierarchies of periodic orbits

Theorem 1.5 has a nice geometric explanation, which we describe informally: If we start with the points $(a,a) \in \Omega$ of period 1 under T, and move down the line, we hit $\left(\frac{1}{2}, \frac{1}{2}\right)$, which is not in Ω. However, we can identify $\Lambda_{\frac{1}{2},\frac{1}{2}}$ with $\Lambda_{\frac{1}{2},1}$, since, in general, we have

$$\Lambda_{a,1-a} = \Lambda_{a,1}.$$

The orbit of $\left(\frac{1}{2}, 1\right)$ has period $N(2) = 2$, and continuing to move down the line, Proposition 4.1 shows that for $a \in (2/3, 1]$,

$$P\left(\frac{a}{2}, a\right) = N(2).$$

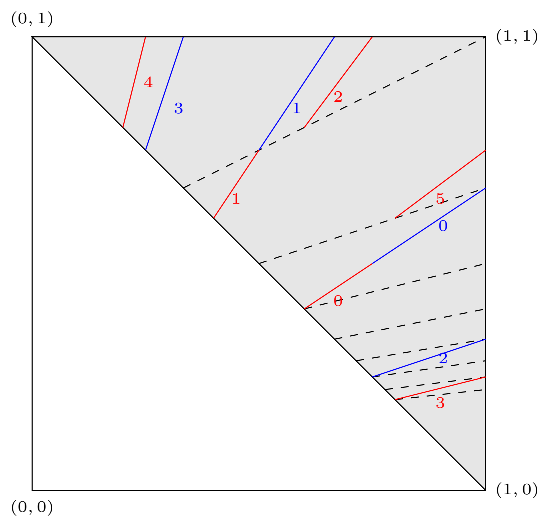
At $a = \frac{2}{3}$, we identify $\left(\frac{1}{3}, \frac{2}{3}\right)$ with the point $\left(\frac{1}{3}, 1\right)$, which has period $N(3) = 5$. Continuing this process, we obtain all the Farey periodic orbits $\left(\frac{1}{Q}, 1\right)$ and the scaling segments

$$\left\{\left(\frac{t}{Q}, t\right) : t \in \left(\frac{Q}{Q+1}, 1\right]\right\}.$$

The set of T-periodic points in Ω is the union

$$\bigcup_{Q\in\mathbb{N}} \bigcup_{1\le i\le N(Q)} \left\{T^i\left(\frac{t}{Q}, t\right) : t \in \left(\frac{Q}{Q+1}, 1\right]\right\}.$$

Fig. 11. The periodic segment $\left\{\left(a, a\frac{2}{3}\right) : a \in \left(\frac{3}{5}, 1\right]\right\}$. The segment $a \in \left(\frac{3}{4}, 1\right)$ is in blue and the segment $a \in \left(\frac{3}{5}, \frac{3}{4}\right]$ is in red. The labels next to the segments correspond to the power of T. The red segment has period $N(4) = 6$, and the blue segment has period $N(3) = 4$.



4.6 Hyperbolic Geometry of Periodic Orbits

What we have done in X_2 is to push the cuspidal horocycle correpsonding to $\Lambda_{a,a}$, $a > 1$ down using g_t. For $a > 1$, the lattice $\Lambda_{a,a}$ is vertically short, and the h_s trajectory is an embedded loop in X_2 which does not intersect Ω. For $1 > a > \frac{1}{2}$, the h_s-orbit it intersects our transversal once, on the point $\Lambda_{a,a}$, which is a fixed point for the BCZ map T. As we push it down, the h_s trajectory intersects the transversal more and more times. However, it does not simply pick up one intersection at a time, but rather $N(Q+1) - N(Q) = \varphi(Q+1)$ intersections when we transition from $\left(\frac{1}{Q+1}, \frac{Q}{Q+1}\right)$ to $\left(\frac{1}{Q+1}, 1\right)$.

5 Equidistribution of periodic orbits

5.1 Proof of Theorem 1.3

In this section, we prove Theorem 1.3, which is the key result for our number theoretic applications. This is essentially a direct consequence of the well-known equidistribution principle for closed horocycles on X_2, originally due to Sarnak [32] and reproved by Eskin-McMullen [16] using ergodic theoretic techniques. In our notation, and in terms of the map T, this result can be reformulated as follows. Recall that given a (non-empty) interval $I = [\alpha, \beta] \subset [0, 1]$, and $Q > 1$, we defined $\rho_{Q,I}$ as the probability measure supported on (a long piece of) the orbit of $(\frac{1}{Q}, 1)$. That is, setting $N_I(Q) := |\mathcal{F}(Q) \cap I|$, we had

$$\rho_{Q,I} = \frac{1}{N_I(Q)} \sum_{i:\gamma_i \in I} \delta_{T^i(\frac{1}{Q},1)}.$$

Define the measure $\rho^R_{Q,I}$ on X_2 by setting

$$d\rho^R_{Q,I} = d\rho_{Q,I} dt,$$

where we are identifying X_2 with the suspension space over Ω.

Theorem 5.1. [[32],[16],[25]] As $Q \to \infty$,

$$\rho^R_{Q,I} \to \mu_2,$$

where the convergence is taken in the weak-* topology. $\square$

Remark. In [32] and [16] the result is only established for the case $\alpha = 0, \beta = 1$, although it is clear that the same arguments in [16] work for the case of fixed $0 < \alpha < \beta < 1$. A proof of this was outlined in [25]. Stronger results where the difference $\beta - \alpha$ is permitted to tend to zero with Q have been obtained by Hejhal [26] and Strömbergsson [34].

Let π be the projection map from the suspension onto Ω. Note that π is continuous except for a set of measure zero with respect to the measures $\rho^R_{Q,I}$ and μ_2. Thus, we have as $Q \to \infty$,

$$\rho_{Q,I} = \frac{1}{R}\pi_*\rho^R_{Q,I} \to \frac{1}{R}\pi_*\mu_2 = m,$$

which completes the proof of Theorem 1.3.

5.2 **Farey Statistics**

Our results on the statistics of Farey fractions are all immediate corollaries of Theorem 1.3, being statements of the form

$$\int_\Omega G d\rho_{Q,I} \to \int_\Omega G dm \tag{5.1}$$

for appropriate choices of functions G on Ω. In the applications to spacings and h-spacings, the function G is in fact an indicator function, and so the convergence is immediate. In the setting of indices and moments, the functions G are $L^1(m)$ functions, we need the following lemma.

Lemma 5.1. For $Q \in \mathbb{N}$, set $m_Q = \rho_{Q,I}$ and let $m_\infty = m$. Let W be the set of measurable functions $G : \Omega \to \mathbb{R}$ satisfying

$$\|G\|_W := \sup_{Q \in \mathbb{N}\cup\{\infty\}} \int_\Omega |G| dm_Q < \infty \tag{5.2}$$

where two functions in W are identified if their difference has zero norm. Let $C_0(\Omega)$ be the space of continuous functions vanishing at infinity equipped with the uniform norm and let W_0 be the closure of its image in W. Then (5.1) holds for any $G \in W_0$. □

Proof. Let $\eta : C_0(\Omega) \to W$ be the natural inclusion map, which we note is a continuous injection that fails to be an embedding. It induces a map $\eta^* : W^* \to M(\Omega)$ between the dual W^* with the weak-* topology and the space $M(\Omega)$ of Radon measures on Ω. Each m_Q determines a linear map $\eta(C_0(\Omega)) \to \mathbb{R}$ that is readily seen to be continuous even with the subspace topology on $\eta(C_0(\Omega))$ coming from W. This linear map extends uniquely to a linear map $W_0 \to \mathbb{R}$ of norm at most one. Given $\varepsilon > 0$ and $G \in W_0$, there is sequence $G_k \in C_0(\Omega)$ converging to G. Choose k so that $\|G_k - G\| < \varepsilon/3$, then choose Q_0 such that for all $Q > Q_0$ we have $|m_Q(G_k) - m(G_k)| < \varepsilon/3$. The triangle inequality now gives

$$|m_Q(G) - m(G)| \leq 2\|G - G_k\| + |m_Q(G_k) - m(G_k)| < \varepsilon.$$

This shows that (5.1) holds. ■

Below, we indicate which functions go with which corollaries, and show how to verify the condition (5.2) in these settings.

5.2.1 *Spacings and h-spacings*

Corollary 1.8 follows from applying (5.1) to the indicator function of the set $R^{-1}\left(\left[\frac{3}{\pi^2|I|}c, \frac{3}{\pi^2|I|}d\right]\right)$. Similarly, Corollary 1.9 follows from applying (5.1) to the indicator function of the set $R_h^{-1}(\tilde{B})$. SInce indicator functions are bounded, (5.2) holds immediately. ■

5.2.2 *Indices*

To verify (5.2) for κ^α, we need:

Lemma 5.2 ([9], Lemma 3.4)**.** Let $r \in \mathbb{N}$, $n \geq 4r + 2$. Suppose $\kappa(a, b) = n$. Then

1. For $i = \pm 1$, $\kappa(T^i(a, b)) = 1$
2. For $1 < |i| \leq r$, $\kappa(T^i(a, b)) = 2$

□

That is, any large value n of κ must be followed by roughly $n/4$ small values. Furthermore, by the definition of κ, we can see that the maximum of κ along the support of m_Q is at most $2Q$, which is on the order of the square root of the length of the orbit $N_I(Q)$, since

$$\kappa\left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right) = \left\lfloor \frac{1+\frac{q_i}{Q}}{\frac{q_{i+1}}{Q}} \right\rfloor = \left\lfloor \frac{Q+q_i}{q_{i+1}} \right\rfloor \leq \frac{2Q}{1} = 2Q.$$

Combining these two facts, we have that there is a constant $c = c_\beta$, depending on the interval I and the power β so that for any $0 < \beta < 2$, and any $Q \in \mathbb{N}$,

$$m_Q(\kappa > N) \leq c_\beta N^{-\beta}.$$

Let $0 < \alpha < \beta < 2$. Then

$$m_Q(\kappa^\alpha) = \sum_{N=1}^{\infty} m_Q\left(\kappa > N^{\frac{1}{\alpha}}\right) < c_\beta \sum_{N=1}^{\infty} N^{-\frac{\beta}{\alpha}} < c_\beta \zeta\left(\beta/\alpha\right).$$

Since $m(\kappa = k) = \frac{8}{k(k+1)(k+2)}$ for $k \geq 2$, we have

$$m(\kappa^\alpha) < 8\zeta(3-\alpha),$$

so we have verified (5.2) for κ^α, proving Corollary 1.10. ■

5.2.3 *Moments*

Finally, to obtain Theorem 1.11, we apply (5.1) to the function

$$G_{s,t}(a,b) = a^s b^t,$$

$t, s \in \mathbb{C}$, $\Re s, \Re t \geq -1$. To verify (5.2), we use that for $(a,b) \in \Omega$,

$$|G_{s,t}(a,b)| < |G_{-1,-1}(a,b)|,$$

so it suffices to verify the condition in the case $(s,t) = (-1,-1)$. We need the following:

Lemma 5.3. For any interval $I \subset [0,1)$ there is a constant c_1 such that for any Q with $N_I(Q) > 0$, we have

$$N_I(Q) \geq c_1 Q^2.$$

□

Proof. $N_I(Q)$ counts the number of primitive lattice points in the triangle:

$$\left\{(x,y) \in \mathbb{R}^2_+ : \frac{y}{x} \in I, x \leq Q\right\},$$

which has area $\frac{1}{2}Q^2|I|$. There are two cases. If $N_{Q/2}(I) \geq 1$, then Theorem 3 of [13] (with the sup norm) implies

$$N_Q(I) \geq \left(\frac{4}{27\pi}\right) \cdot \frac{3}{8} Q^2|I| = \frac{|I|}{18\pi} Q^2.$$

On the other hand, if $N_{Q/2}(I) = 0$, the length of I is bounded above by the largest gap in $F_{Q/2}$, which is at most $2/Q$, so that

$$N_Q(I) \geq 1 \geq \frac{Q^2|I|^2}{4}.$$

Thus, the lemma follows with the choice

$$c_1 = \min\left(\frac{|I|}{18\pi}, \frac{|I|^2}{4}\right).$$

■

We have

$$m_Q(G_{-1,-1}) = \frac{1}{N_I(Q)} \sum_{\gamma_i \in \mathcal{F}_I(Q)} \frac{Q^2}{q_i q_{i+1}} = \frac{|I| Q^2}{N_I(Q)}.$$

Applying Lemma 5.3, we have that $m_Q(G_{-1,1})$ is unfiormly bounded in Q, and we have that $m(G_{-1,1}) = \frac{\pi^2}{3}$, completing the proof. ■

5.2.4 *Excursions*

To prove Corollary 1.12, we apply (5.1) to the functions

$$G_{\max}(a,b) = \max\left\{a, b, \frac{1}{a+b}\right\},$$

and

$$G_{\min}(a,b) = \min\left\{\frac{1}{a}, \frac{1}{b}, a+b\right\}$$

respectively. Equation (5.2) is satisfied since $G_{\max}$ is bounded by 1 and $G_{\min}$ is bounded by 2. ■

6 Piecewise linear description of horocycles

Before proving Theorems 1.6 and 1.7, we record an elementary lemma on the behavior of the length of vectors under h_s, whose proof we leave to the reader.

Lemma 6.1. Let $\mathbf{v} = (x,y)^{\mathrm{T}} \in \mathbb{R}^2$ be a vector with $x > 0$. Let $\sigma^{\pm} = \sigma \pm |x|^{-1}$ where $\sigma = \frac{y}{x}$ is the slope of $\mathbf{v}$. Then $f(s) = \|h_s\mathbf{v}\|_{\sup}$ is the continuous piecewise linear function given by

(i) $f'(s) = -|x|$ for $s < \sigma^-$,
(ii) $f(s) = |x|$ for $\sigma^- \le s \le \sigma^+$, and
(ii) $f'(s) = +|x|$ for $s > \sigma^+$.

□

6.1 Minima and maxima

Recall that Lemma 2.1 states that Ω' consists of lattices with a short (length ≤ 1) horizontal vector. This observation shows that the visits of the trajectory $\{h_s\Lambda\}_{s\ge 0}$ to the transversal Ω' are precisely the times $s_n(\Lambda)$ of local minima of $\ell_\Lambda(s)$, since a lattice cannot contain more than 1 vector of supremum norm at most 1, and under h_s, vectors are shorter when they are horizontal. On Ω', we have

$$\ell\left(\Lambda_{a,b}\right) = a.$$

By abuse of notation we write $\ell : \Omega \to (0,1]$ with

$$\ell(a,b) = a.$$

Similarly we write $\alpha(a,b) = \frac{1}{a}$. Thus when $h_{s_n}\Lambda = \Lambda_{a_n,b_n}$, the length of the shortest vector in $h_{s_n}\Lambda$ is given by the horizontal vector $\begin{pmatrix} a_n \\ 0 \end{pmatrix}$. Setting

$$s_1(\Lambda) = \min\{s \ge 0 : h_s\Lambda \in \Omega'\},$$

and writing $\Lambda_{a_1,b_1} = h_{s_1}\Lambda$, we have that

$$l_N(\Lambda) = \frac{1}{N}\sum_{n=1}^{N} \ell_\Lambda(s_n) = \frac{1}{N}\sum_{n=1}^{N} \ell\left(T^{n-1}\left(a_1,b_1\right)\right),$$

and

$$a_N(\Lambda) := \frac{1}{N}\sum_{n=1}^{N} \alpha_\Lambda(s_n) = \frac{1}{N}\sum_{n=1}^{N} \alpha\left(T^{n-1}\left(a_1,b_1\right)\right).$$

Since both α and ℓ are in $L^1(\Omega, dm)$, and we have assumed that Λ is not periodic under h_s (and so, (a_1,b_1) is not under T), we can apply the Birkhoff Ergodic Theorem (and the fact that all non-periodic orbits are generic for m, by the measure classification result) to conclude

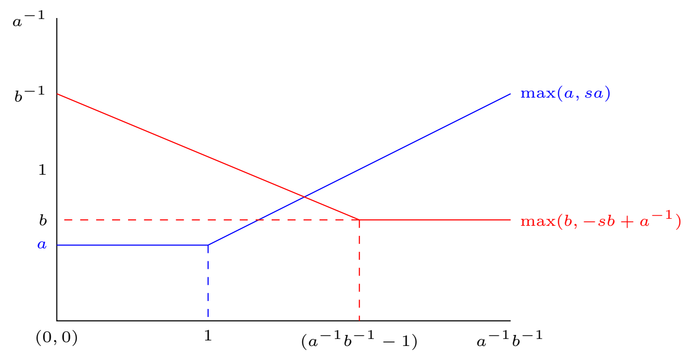
$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \ell_\Lambda(s_n) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \ell\left(T^{n-1}\left(a_1,b_1\right)\right) = \int_\Omega \ell dm = \frac{2}{3}.$$

and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \alpha_\Lambda(s_n) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \alpha\left(T^{n-1}\left(a_1,b_1\right)\right) = \int_\Omega \alpha dm = 2,$$

proving Theorem 1.6. ■

Fig. 12. The hand-off between the vectors $(a, -sb)^{\mathrm{T}}$ and $(b, -sa + b^{-1})^{\mathrm{T}}$.



6.2 Returns

For the corresponding results for local *maxima* $S_n(\Lambda)$, we need to consider the 'hand-off' between the basis vectors $(a, 0)^{\mathrm{T}}$ and $(b, a^{-1})^{\mathrm{T}}$ which happens after hitting the point $\Lambda_{a,b} \in \Omega$. This hand-off happens when the vectors $h_s(a, 0)^{\mathrm{T}} = (a, -sa)^{\mathrm{T}}$ and $h_s(b, a^{-1})^{\mathrm{T}} = (b, -sb + a^{-1})^{\mathrm{T}}$ have the same length. That is, we need to calculate the time $s \in (0, R(a, b))$ so that

$$\max(a, sa) = \max(b, -sb + a^{-1}),$$

where we have taken absolute values of the coordinates (see Figure 12). A case-by-case analysis shows that the length of the vector at the hand-off is given by $f(x, y) = \max\left\{a, b, \frac{1}{a+b}\right\}$. By a similar analysis as above, we apply the ergodic theorem to this function (and to its reciprocal) to obtain Theorem 1.7. ■

7 Geometry of Numbers

In this section we prove Theorem 1.13 and Corollary 1.14.

7.1 Slope gap distribution

We will prove Theorem 1.13 by noting that the observations of §2.1.2 yield the following:

Lemma 7.1. Let $t > 0$, and let T_t denote the t-BCZ map. Let Λ denote a lattice without t-short vertical vectors, and let $\Lambda_1 = h_{s_1}\Lambda$ be the first intersection of the orbit $\{h_s\Lambda\}_{s>0}$ with the transversal Ω_t. Let $0 \leq c \leq d \leq \infty$. Let $\chi_{c,d}$ denote the indicator function of the set

$$R^{-1}(c, d) = \left\{(x, y) \in \Omega_t : \frac{1}{d} < xy < \frac{1}{c}\right\}.$$

Then

$$\frac{1}{N}|G_{N,t}(\Lambda) \cap (c, d)| = \frac{1}{N}\sum_{i=0}^{N} \chi_{c,d}(T_t^n(\Lambda_1)).$$

□

Proof. The lemma follows from the fact that h_s preserves differences in slopes, and the function R captures these differences, for vectors in the strip V_t. In particular,

$$R(T_t^n(\Lambda_1)) = s_{n+1} - s_n,$$

which yields the lemma. ■

The proof of the theorem then follows as in §6.1, by applying the Birkhoff ergodic theorem and Dani's measure classification. ■

7.2 Slope gap statistics

To prove Corollary 1.14, we first observe that if Λ has a vertical vector, its periodicity under h_s implies that the sequence of slope gaps in V_t must repeat, so in particular there is an $N_0 = N_0(t)$ so that $G_{N,t}(\Lambda) = G_{N_0,t}(\Lambda)$ for all $N \geq N_0$. For example, in §4.1 we see that for the lattice $\mathbb{Z}^2$, $N_0(t) = N(\lfloor t \rfloor)$, where $N(Q)$ denotes the cardinality of the Farey sequence. Again using the conjugation relation (2.6), we see that $G_{N,t}(\Lambda) = G_{N,1}(g_{-2\log t}\Lambda)$. Using Lemma 7.1, we see that $\frac{1}{N}|G_{N,t}(\Lambda) \cap (c,d)|$ now corresponds to the integral of $\chi_{c,d}$ along a periodic orbit for T, which, as $t \to \infty$, is getting longer and longer. By Theorem 1.3, this converges to the integral of $\chi_{c,d}$ with respect to Lebesgue measure m, as desired. ■

8 Further Questions

8.1 BCZ maps for other lattices

It was communicated to us by O. Sarig that it is well-known that the horocycle flow on $SL(2,\mathbb{R})/\Gamma$ for any lattice Γ can be realized as a suspension flow over an adic transformation. A natural conjecture, then is:

Conjecture. The BCZ map is an adic transformation. □

To our knowledge, the BCZ map is the first *explicit* description of a section and a return map for the horocycle flow. We have the following immediate

Question. How explicitly can the return maps for h_s on $SL(2,\mathbb{R})/\Gamma$ for $\Gamma \neq SL(2,\mathbb{Z})$ be given? □

The first author, in joint work with J. Chaika and S. Lelievre [3], has calculated a BCZ-type map for $\Gamma = \Delta(2,5,\infty)$, the $(2,5,\infty)$-Hecke triangle group (note that $SL(2,\mathbb{Z})$ is the $(2,3,\infty)$ triangle group).

8.2 Slope gaps for translation surfaces

The motivation for [3] was to generalize the calculation of the gap distribution for Farey sequences, which correspond to closed trajectories for geodesic flow on the torus, to a particular example of a higher-genus translation surface, given by a particular L-shaped polygon known as the *golden L*. A more general question, for which more details can be found in [2] is:

Question. Explicitly describe a BCZ-type map and the return time function for the space of translation surfaces. Use this to explicitly compute gap distributions for saddle connection slopes. □

8.3 Mixing properties

While the horocycle flow h_s is known to be mixing on X_2 with respect to the measure μ_2 (by, e.g., the Howe-Moore theorem [6]), mixing is a property that does *not* pass between flows and return maps. In fact, there are many well-known constructions of mixing suspension flows over non-mixing base transformations. Thus, we have the natural

Question. Is the BCZ map T mixing? □

8.4 Rates of convergence

Flaminio-Forni [17] and Strömbergsson [34] have proved power law upper bounds for the *deviation* of ergodic averages for trajectories of h_s on X_2. These pass to the return map (with different leading constants), and it would be interesting to give a direct proof of this phenomenon. Zagier [36] showed that proving an optimal rate of equidistribution for long periodic trajectories for h_s on $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ (that is, an optimal error term in Sarnak's theorem [32]) is equivalent to the classical Riemann hypothesis. There is also an equivalent formulation of the Riemann hypothesis in terms of the distribution of the Farey sequence due to Franel-Landau [18]. This leads to

Question. Is there an optimum bound on the error term in Theorem 1.3 that is equivalent to the Riemann hypothesis? □

Jens Marklof has informed us that some progress toward the last question has been described in [29].

References

[1] P. Arnoux, *Le codage du flot géodésique sur la surface modulaire*, L'Enseignement Math?matique, Vol.40 (1994)

[2] J. S. Athreya, *Cusp excursions on parameter spaces*, to appear, Journal of London Mathematical Society. arXiv:1104.2797.

[3] J. S. Athreya, J. Chaika, and S. Lelievre, *The distribution of gaps for saddle connections on the golden L*, in preparation.

[4] J. S. Athreya and G. A. Margulis, *Logarithm laws for unipotent flows, I*, Journal of Modern Dynamics **3** (2009), no. 3, 359-378.

[5] V. Augustin, F. P. Boca, C. Cobeli, and A. Zaharescu, *The h-spacing distribution between Farey points.* Math. Proc. Cambridge Philos. Soc. **131** (2001), no. 1, 23 - 38.

[6] M. B. Bekka and M. Mayer, *Ergodic Theory and Topological Dynamics of Group Action on Homogeneous Spaces*, London Mathematical Society Lecture Notes v. 269, Cambridge University Press, 2000.

[7] F. P. Boca and A. Zaharescu, *Farey fractions and two-dimensional tori*, in *Noncommutative Geometry and Number Theory* (C. Consani, M. Marcolli, eds.), Aspects of Mathematics E37, Vieweg Verlag, Wiesbaden, 2006, pp. 57-77.

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

[9] F. Boca, C. Cobeli, and A. Zaharescu, *On the distribution of the Farey sequence with odd denominators.* Michigan Math. J. **51** (2003), no. 3, 557 - 573.

[10] F. Boca, R. Gologan, and A. Zaharescu, *On the index of Farey sequences.* Q. J. Math. **53** (2002), no. 4, 377 - 391.

[11] F. Boca, V. Pasol, A. Popa, and A. Zaharescu, *Pair correlation of angles between reciprocal geodesics on the modular surface*, preprint, arXiv:1102.0328.

[12] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces,* Trans. AMS **153** (1971), 401 - 414.

[13] Y. Cheung, *Slowly divergent geodesics in moduli space,* Conf. Geom. Dyn. **8** (2004), 167–189.

[14] Y. Cheung, *Hausdorff dimension of the set of singular pairs*, Ann. Math. **173**, (2011), 127 - 167.

[15] S. G. Dani, *On uniformly distributed orbits of certain horocycle flows.* Ergodic Theory Dynamical Systems **2** (1982), no. 2, 139 - 158 (1983).

[16] A. Eskin and C. McMullen, *Mixing, counting, and equidistribution in Lie groups,* Duke Math. J. **71** (1993), 181 - 209.

[17] L. Flaminio and G. Forni, *Invariant distributions and time averages for horocycle flows*, Duke Math. J., v. 119, 465-526, 2003.

[18] J. Franel and E. Landau, *Les suites de Farey et le probl?me des nombres premiers,* Göttinger Nachr., 198?206, 1924.

[19] B. M. Gurěvic, *The entropy of a horocycle flow*, Soviet Math. Doklady **2** (1961), 124 - 130.

[20] R. R. Hall, *A note on Farey series.* J. London Math. Soc. (2) **2** 1970 139 - 148.

[21] R. R. Hall and P. Shiu, *The index of a Farey sequence.* Michigan Math. J. **51** (2003), no. 1, 209?223.

[22] R. R. Hall and G. Tenenbaum, *On consecutive Farey arcs.* Acta Arith. **44** (1984), no. 4, 397 - 405.

[23] M. Handel and B. Kitchens, *Metrics and entropy for non-compact spaces. With an appendix by Daniel J. Rudolph.* Israel J. Math. **91** (1995), no. 1-3, 253 - 271.

[24] G. Hedlund, *Fuchsian groups and transitive horocycles*, Duke Math. J. **2** (1936), no. 3, 530 - 542.

[25] D.A. Hejhal, *On value distribution properties of automorphic functions along closed horocycles* in *XVI*th Rolf Nevanlinna Colloquium (Joensuu, Finland, 1995), de Gruyter, Berlin, 1996, 39 - 52.

[26] D.A. Hejhal, *On the uniform equidistribution of long closed horocycles* in Loo-Keng Hua: A Great Mathematician of the Twentieth Century, Asian J. Math. **4**, Int. Press, Somerville, Mass., 2000, 839 ? 853.

[27] P.P. Kargaev and A. A. Zhigljavsky, *Asymptotic distribution of the distance function to the Farey points*, J. Number Theory **65** (1997) 130-149.

[28] J. Marklof, *The asymptotic distribution of Frobenius numbers*, Invent. Math. **181** (2010) 179 - 207.

[29] J. Marklof, *Fine-scale statistics for the multidimensional Farey sequence*, preprint: arXiv:1207.0954, 2012.

[30] J. Marklof and A. Str?mbergsson, *The distribution of free path lengths in the periodic Lorentz gas and related lattice point problems*, Ann. of Math. **172** (2010) 1949 - 2033.

[31] M. G. Nadkarni, *Basic ergodic theory.* Second edition. Birkh?user Advanced Texts, Birkh?user Verlag, Basel, 1998.

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

[33] C. Series, *The Modular Surface and Continued Fractions*, J. London Math. Soc. (2), **31** (1985), 69-80.

[34] A. Strömbergsson, *On the uniform equidistribution of long closed horocycles,* Duke Math. J. **123** (2004), 507 - 547.

[35] P. Walters, *An introduction to ergodic theory*, in Graduate Texts in Mathematics 79 (Springer-Verlag, New York, 1982).

[36] D. Zagier, *Eisenstein series and the Riemann zeta function* in *Automorphic Forms, Representation Theory and Arithmetic* (Bombay, 1979), Tata Inst. Fund. Res. Studies in Math. 10, Tata Inst. Fund. Res., Bombay, 1981, 275 ? 301.