APPLICATIONS OF DYNAMICAL SYSTEMS TO
FAREY SEQUENCES AND CONTINUED FRACTIONS

BY

BYRON HEERSINK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Alexandru Zaharescu, Chair
Professor Florin Boca, Director of Research
Adjunct Associate Professor Jayadev Athreya
Professor Emeritus Joseph Rosenblatt

Abstract

This thesis explores three main topics in the application of ergodic theory and dynamical systems to equidistribution and spacing statistics in number theory. The first is concerned with utilizing the ergodic properties of the horocycle flow in $\mathrm{SL}(2,\mathbb{R})$ to study the spacing statistics of Farey fractions. For a given finite index subgroup $H \subseteq \mathrm{SL}(2,\mathbb{Z})$, we use a process developed by Fisher and Schmidt to lift a cross section of the horocycle flow on $\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$ found by Athreya and Cheung to the finite cover $\mathrm{SL}(2,\mathbb{R})/H$ of $\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$. We then use the properties of this section to prove the existence of the limiting gap distribution of various subsets of Farey fractions. Additionally, to each of these subsets of fractions, we extend solutions by Xiong and Zaharescu, and independently Boca, to a Diophantine approximation problem of Erdős, Szüsz, and Turán.

The latter two topics of this thesis establish properties of the Farey map F by analyzing the transfer operators of F and the Gauss map G, well known maps of the unit interval relating to continued fractions. We first prove an equidistribution result for the periodic points of the Farey map using a connection between continued fractions and the geodesic flow in $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$ illuminated by Series. Specifically, we expand a cross section of the geodesic flow given by Series to produce another section whose first return map under the geodesic flow is a double cover of the natural extension of the Farey map. We then use this cross section to extend the correspondence between the closed geodesics on the modular surface and the periodic points of G to include the periodic points of F. Then, analogous to the work of Pollicott, we find the limiting distribution of the periodic points of F when they are ordered according to the length of their corresponding closed geodesics through the analysis of the transfer operator of G.

Lastly, we provide effective asymptotic results for the equidistribution of sets of the form $F^{-n}([\alpha,\beta])$, where $[\alpha,\beta] \subseteq (0,1]$, and, as a corollary, certain weighted subsets of the Stern-Brocot sequence. To do so, we employ mostly basic properties of the transfer operator of the Farey map

and an application of Freud's effective version of Karamata's Tauberian theorem. This strengthens previous work of Kesseböhmer and Stratmann, who first established the equidistribution results utilizing infinite ergodic theory.

To my parents.

Acknowledgments

First of all, I would like to thank my advisor Florin Boca for his guidance throughout the course of my studies. I would also like to thank the other members of my preliminary exam and dissertation defense committees, Jayadev Athreya, Alexandru Zaharescu, Francesco Cellarosi, and Joseph Rosenblatt, for their advice and suggestions. I thank my coauthors Paul Spiegelhalter and Joseph Vandehey whose collaborations have helped me progress in my research; and I thank Harold Diamond and Claire Merriman for insightful discussions which inspired some of the work in this thesis.

I am grateful for the financial support awarded by the University of Illinois Department of Mathematics. In particular, my work has been supported by National Science Foundation Grant DMS 08-38434 “EMSW21-MCTP: Research Experience for Graduate Students”, Department of Education Grant P200A090062, “University of Illinois GAANN Mathematics Fellowship Project”, and the R. H. Schark Fellowship.

Finally, I thank my family for their love and support, and for bringing me joy and encouragement through the challenging times.

Table of Contents

Chapter 1

Introduction

The central theme of this thesis is the study of the distribution of certain sequences and sets of number theoretical interest. In this introduction, we begin by defining the central equidistribution and spacing statistics notions we use. We then give an overview of the thesis, outlining the number theoretical concepts on which we focus and our main results.

1.1 Equidistribution and spacing statistics

Let X be a metric space, $(A_n)_n$ be an increasing sequence of finite subsets of X, and p be a probability measure on X. We say that $(A_n)_n$ equidistributes with respect to the measure p if

$$\lim_{n\to\infty} \frac{\#(A_n \cap B)}{\#A_n} = p(B) \quad \text{for all } B \subseteq X \text{ Borel such that } p(\partial B) = 0.$$

A useful equivalent condition given by the Portmanteau theorem [11, Theorem 2.1] is that

$$\lim_{n\to\infty} \frac{1}{\#A_n} \sum_{x\in A_n} f(x) = \int_X f\, dp \quad \text{for all bounded continuous functions } f : X \to \mathbb{R}.$$

Intuitively, this means that the density of A_n in X increasingly resembles the measure p as n grows. In cases where the subsets A_n are weighted so that each element $x \in A_n$ has a weight $w_x \in \mathbb{R}$, we then say that the weighted sequence $(A_n)_n$ equidistributes with respect to p if

$$\lim_{n\to\infty} \frac{\sum_{x\in A_n} w_x f(x)}{\sum_{x\in A_n} w_x} = \int_X f\, dp.$$

In every circumstance, the set X we deal with is or can easily be identified with a subset of Euclidean space.

We also wish to consider a notion of equidistribution for infinite subsets of the space X. We consider this notion in Chapter 5 in the specific case where $X = [0,1]$ and the measure p is the Lebesgue measure, which we denote throughout this thesis as λ. We say that a sequence of Borel subsets $(B_n)_n$ of $[0,1]$ equidistributes in $[0,1]$ if

$$\lim_{n\to\infty} \frac{1}{\lambda(B_n)} \int_{B_n} f\, d\lambda = \int_{[0,1]} f\, d\lambda \quad \text{for all } f \in C([0,1]).$$

We also consider a situation of an equidistributing sequence of one dimensional subsets in a three dimensional space, which we detail in Chapter 2.

In Chapter 2, we additionally study the finer spacing statistics of our sequences of finite subsets of $\mathbb{R}$, namely the limiting gap distribution. For a given finite subset $A = \{x_0 \leq x_1 \leq \cdots \leq x_N\}$ of $\mathbb{R}$, we define the gap distribution measure of A to be the finitely supported probability measure ν_A on $[0,\infty)$ such that

$$\nu_A([0,\xi]) := \frac{1}{N}\#\{j \in [1,N] : N(x_j - x_{j-1}) \leq \xi(x_N - x_0)\}, \quad \xi \geq 0.$$

Then for a sequence $(A_n)_n$ of finite subsets of $\mathbb{R}$, we call the weak* limit of $(\nu_{A_n})_n$, if it exists, the limiting gap measure of $(A_n)_n$. We refer to the distribution of this measure as the limiting gap distribution of $(A_n)_n$. Roughly speaking, the limiting gap distribution estimates the distribution of how big the gaps between consecutive elements of A_n are when they are normalized by the average gap length.

Another even finer statistic we look at is the limiting h-spacing distribution. For $h \in \mathbb{N}$ and the set A as above, let $\mathbf{v}_{A,i,h} = (x_{i+j} - x_{i+j-1})_{j=1}^{h} \in \mathbb{R}^h$ for $i \in \{1, \ldots, N-h\}$. We define the h-spacing distribution measure of A to be the finitely supported probability measure $\nu_{A,h}$ on $[0,\infty)^h$ such that

$$\nu_{A,h}\left(\prod_{j=1}^{h}[0,\xi_j]\right) := \frac{1}{N}\#\left\{x_i \in A : N\mathbf{v}_{A,i,h} \in \prod_{j=1}^{h}[0,\xi_j(x_N - x_0)]\right\}.$$

We then call the weak* limit of $(\nu_{A_n,h})_n$, if it exists, the limiting h-spacing measure of $(A_n)_n$. This measure captures the distribution of h consecutive normalized gaps in A_n as $n \to \infty$.

1.2 Homogeneous dynamics and Farey fractions

The Farey sequence of order $Q \in \mathbb{N}$ is the set $\mathcal{F}(Q)$ of fractions $\frac{a}{q} \in [0,1]$ such that $\gcd(a,q) = 1$ and $q \leq Q$. A well known application of Möbius inversion proves the effective equidistribution of sequence $(\mathcal{F}(Q))_Q$ in $[0,1]$ with respect to the Lebesgue measure as $Q \to \infty$. Also, various properties regarding the fine scale spacing statistics of $(\mathcal{F}(Q))_Q$ have been studied. For instance, Hall [30], Augustin et al. [9], and Boca and Zaharescu [17], established the limiting gap distribution, the limiting h-spacing distribution, and the correlations of $(\mathcal{F}(Q))_Q$, respectively.

Additionally, certain subsets of Farey fractions have been considered. For instance, let $\mathcal{F}_{Q,d} \subseteq \mathcal{F}(Q)$ be the set of fractions $\frac{a}{q}$ with $\gcd(q,d) = 1$ and $\widetilde{\mathcal{F}}_{Q,\ell} \subseteq \mathcal{F}(Q)$ be the set of $\frac{a}{q}$ with $\ell \nmid a$. Then the number of pairs $\left(\frac{a}{q}, \frac{a'}{q'}\right)$ of consecutive fractions in $\mathcal{F}_{Q,d}$ with fixed $a'q - aq' = k$ has been estimated by Badziahin and Haynes [10]; the pair correlation function of the sequence $(\mathcal{F}_{Q,d_Q})_Q$ was shown to exist by Xiong and Zaharescu [69], where d_Q varies with Q subject to the constraints $d_{Q_1} \mid d_{Q_2}$ as $Q_1 < Q_2$ and $d_Q \ll Q^{\log\log Q/4}$; and the limiting gap distribution for the sequences $(\mathcal{F}_{Q,d})_Q$ and $(\widetilde{\mathcal{F}}_{Q,\ell})_Q$ were shown to exist for fixed d and ℓ by Boca, Spiegelhalter, and the author [16].

Recently, Athreya and Cheung [7] realized the horocycle flow, that is, the left multiplication of the group

$$\left\{ \begin{pmatrix} 1 & 0 \\ -s & 1 \end{pmatrix} : s \in \mathbb{R} \right\},$$

on $\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$, as a suspension flow over the BCZ map introduced by Boca, Cobeli, and Zaharescu [14] in their study of Farey fractions. Athreya and Cheung used this connection and the equidistribution of horocycles in $\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$ to rederive the limiting gap distribution and other properties of Farey fractions. The process used in [7] to obtain these results was later generalized in [5] to explain the gap distributions of various different sequences. Also, recent work of Fisher and Schmidt [26] was concerned with lifting cross sections of the geodesic flow on the unit tangent bundle of the modular surface to a finite cover, with the primary aim of obtaining statistical properties of continued fractions.

In Chapter 2, we use the process in [26] to explicitly lift, for every finite index subgroup $H \subseteq \mathrm{SL}(2,\mathbb{Z})$, the section of the horocycle flow discovered in [7] to the cover $\mathrm{SL}(2,\mathbb{R})/H$ of

$\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$. As one application, we prove the existence of the limiting gap distribution of certain subsets of Farey fractions following the ideas in [7] and the more general framework of [5, Theorem 2.5]. We additionally obtain results on the h-spacings and numerators of differences of these subsets. A given subset we consider is determined by a finite index subgroup $H \subseteq \mathrm{SL}(2,\mathbb{Z})$, and corresponds to a cross section of $\mathrm{SL}(2,\mathbb{Z})/H$ obtained by intersecting the lift of the section in [7] with certain sheets of the cover $\mathrm{SL}(2,\mathbb{R})/H \to \mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$. As a second application, we establish, for each of the aforementioned subsets of Farey fractions, the existence of the limiting Lebesgue measure of the real numbers in $[0,1]$ that are in some sense well approximated by elements in the subset and having large denominators. This solves an analogue of a Diophantine approximation problem posed by Erdős, Szüsz, and Turán [20]. These results have been published in [35].

This stream of research was inspired by the significant works of Elkies and McMullen [19] and of Marklof and Strömbergsson [55]. In the first work, the equidistribution of horocycles in the space $\mathrm{ASL}(2,\mathbb{R})/\mathrm{ASL}(2,\mathbb{Z})$ of unimodular lattice translates in $\mathbb{R}^2$ was utilized to determine the gap distribution of the sequence $(\{\sqrt{n}\})_n$, where $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part; and in the second, the equidistribution of unipotent flows in the higher dimensional spaces $\mathrm{SL}(d,\mathbb{Z})\backslash\mathrm{SL}(d,\mathbb{R})$ and $\mathrm{ASL}(d,\mathbb{Z})\backslash\mathrm{ASL}(d,\mathbb{R})$ ($d \geq 2$) was used to prove the existence of the limiting distribution of the free path length in the Boltzmann-Grad limit of the periodic Lorentz gas.

1.3 Dynamics and continued fractions

Chapters 3–5 of this thesis are concerned with the regular continued fractions of the form

$$[a_1, a_2, a_3, \ldots] := \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}} \qquad (a_j \in \mathbb{N})$$

In particular, our central object of study is the Farey map F, and we also make significant use of the Gauss map G, which is a speed up of F. Both F and G are dynamical systems on the unit interval to itself which have been studied in connection to continued fractions due to the equalities

$$F([a_1, a_2, \ldots]) = \begin{cases} [a_1 - 1, a_2, \ldots] & \text{if } a_1 \geq 2 \\ [a_2, a_3, \ldots] & \text{if } a_1 = 1 \end{cases} \quad \text{and} \quad G([a_1, a_2, \ldots]) = [a_2, a_3, \ldots].$$

The primary angle through which we study F is via the transfer operators of F and G. The transfer operator can be defined for a dynamical system $T : X \to X$ on a measure space (X, p), where p is a σ-finite measure (which is not necessarily finite) satisfying $p \circ T^{-1} \ll p$, as the operator $\hat{T} : L^1(p) \to L^1(p)$ such that

$$\int_A \hat{T} f \, dp = \int_{T^{-1}(A)} f \, dp \quad \text{for any } A \subseteq X \text{ measurable and } f \in L^1(p).$$

Equivalently,

$$\int (\hat{T} f) \cdot g \, dp = \int f \cdot (g \circ T) \, dp \quad \text{for any } f \in L^1(p) \text{ and } g \in L^\infty(p);$$

so the transfer operator of T is the dual of the Koopman operator $f \mapsto f \circ T$.

We also touch on the natural extensions of the Farey and Gauss maps. The natural extension of a dynamical system, defined originally by Rohlin [61] for probability spaces and extended by Silva [64] to σ-finite spaces, is a minimal invertible dynamical system which has the original system as a factor. Due to the nature of F and G as shift maps on continued fractions, the construction of their natural extensions is intuitive, and can be thought of as two-sided shift operators on continued fraction digits.

We cover more details on some of the basic properties of continued fractions, the Gauss and Farey maps, and their natural extensions and transfer operators in Chapter 3.

In Chapter 4, we study the distribution of the periodic points of F. To do so, we rely on a connection, first noticed by Artin [4] and later lucidly explained by Series [63], between continued fractions and geodesics in the modular surface. Notably, Series found a cross section of the geodesic flow in the tangent bundle of the modular surface, or equivalently, the right action of the group

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

on $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$; and the section's first return map is a double cover of the natural extension of G. This connection gives rise to a correspondence between the periodic orbits of G, consisting of the periodic continued fractions, and the primitive closed geodesics in $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$.

Utilizing this relationship and the work of Mayer [56] on the transfer operator of G, Pollicott [59] determined the limiting distribution of the periodic points of G when ordered according to the length of their corresponding closed geodesics. He then utilized this result to prove the equidistribution of the geodesics themselves. Using Pollicott's technique, Kelmer [45] proved analogous equidistribution results for closed geodesics in the modular surface with a specified linking number, as well as their corresponding periodic continued fractions.

The goal of Chapter 4 is to extend the work of Series and Pollicott to encompass the Farey map. We first enlarge Series' cross section of the geodesic flow to yield another section forming a double cover of the natural extension of F. This allows us to extend the correspondence between closed geodesics and periodic continued fractions to include the periodic points of F. We then utilize the work of Pollicott to establish the equidistribution of the periodic points of F, and of its natural extension, according to their respective invariant measures.

The focus of Chapter 5 is the distribution of the preimages of the form $F^{-n}([\alpha,\beta])$ through the analysis of the transfer operator of F. This study was inspired by the work of Kesseböhmer and Stratmann [49] who, confirming a conjecture of Fiala and Kleban [25], proved that the Lebesgue measure $\lambda(\mathscr{C}_n)$ of the sum-level set for continued fractions

$$\mathscr{C}_n := F^{-(n-1)}\left(\left[\frac{1}{2}, 1\right]\right) = \left\{[a_1, a_2, \ldots] \in [0,1] : \sum_{i=1}^{k} a_i = n \text{ for some } k \in \mathbb{N}\right\}$$

approaches 0 as $n \to \infty$, and furthermore obtained an asymptotic formula for the decay rate. Then in [48], they more generally established, for any subinterval $[\alpha,\beta] \subseteq (0,1]$, the asymptotic decay rate of the Lebesgue measures of the preimages $(F^{-n}([\alpha,\beta]))_n$, in addition to the equidistribution of the preimages with respect to the Lebesgue measure. As a corollary, they also provide an equidistribution result for certain weighted subsets of the Stern-Brocot sequence which can be obtained as preimages of the form $F^{-n}\left(\frac{v}{w}\right)$ with $\frac{v}{w} \in \mathbb{Q}$. To prove these results, Kesseböhmer and Stratmann applied powerful results in infinite ergodic theory to the transfer operator of F with respect to the infinite invariant measure $\frac{dx}{x}$.

In Chapter 5, we obtain effective versions of the results of Kesseböhmer and Stratmann on the equidistribution of both the preimages $(F^{-n}([\alpha, \beta]))_n$ and the subsets of the Stern-Brocot sequence. For our proof, we establish estimates involving the sums of the iterates of the transfer operator of F, and examine more carefully some of the results underlying the machinery used in [49, 48]. Notably, we incorporate Freud's work [27] that provides an effective version of Karamata's Tauberian theorem, an important result utilized in infinite ergodic theory. This work builds upon the author's publication [34].

Throughout this thesis, we make frequent use of asymptotic notation, which we explain here. We write $f(x) = O(g(x))$, or equivalently $f(x) \ll g(x)$, as $x \to \infty$ if there exist constants $M, N > 0$ such that $|f(x)| \leq M|g(x)|$ for all $x \geq N$. The expression $f(x) = O_{c_1,\ldots,c_m}(g(x))$, or equivalently $f(x) \ll_{c_1,\ldots,c_m} g(x)$, means the same thing, though in this case the constants M and N depend on $c_1, \ldots, c_m$. Also, we write $f(x) \sim g(x)$ as $x \to \infty$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Chapter 2

The horocycle flow in $\mathrm{SL}(2,\mathbb{R})$ and Farey fraction statistics

2.1 Introduction

In this chapter, we prove our main results on Farey fraction statistics utilizing dynamics in homogeneous spaces $\mathrm{SL}(2,\mathbb{R})/H$. These results appear in [35]. We begin by giving some basic properties of Farey fractions we use throughout the chapter.

An alternative way of defining the Farey sequence $(\mathcal{F}(Q))_Q$ is by induction through the following process: First, let $\mathcal{F}(1)=\left\{\frac{0}{1},\frac{1}{1}\right\}$. Then, assuming we have defined $\mathcal{F}(Q)$ for a particular $Q\in\mathbb{N}$, we let $\mathcal{F}(Q+1)$ be the set of all fractions in $\mathcal{F}(Q)$, together with all the mediants of consecutive fractions in $\mathcal{F}(Q)$ having denominator $Q+1$, i.e, fractions of the form $\frac{a+a'}{q+q'}$, where $\frac{a}{q}$ and $\frac{a'}{q'}$ are consecutive elements in $\mathcal{F}(Q)$ and $q+q'=Q+1$. A few clear consequences of this is that if $\frac{a}{q}<\frac{a'}{q'}$ are any consecutive elements of $\mathcal{F}(Q)$, then $q+q'>Q$; and for any $Q'\geq Q$, the element succeeding $\frac{a}{q}$ in $\mathcal{F}(Q')$ is of the form $\frac{ma+a'}{mq+q'}$ where $m\in\mathbb{N}\cup\{0\}$, and similarly the element in $\mathcal{F}(Q')$ preceding $\frac{a'}{q'}$ is of the form $\frac{a+a'n}{q+q'n}$ where $n\in\mathbb{N}\cup\{0\}$.

Another elementary property we wish to note is that for any nonnegative integers a,a',q,q' such that $0\leq\frac{a}{q}<\frac{a'}{q'}\leq 1$, then $\frac{a}{q}<\frac{a'}{q'}$ are consecutive in the Farey sequence of order $\max\{q,q'\}$ if and only if $a'q-aq'=1$. It is then easy to see that if $q,q'\in\{1,\ldots,Q\}$ such that $\gcd(q,q')=1$ and $q+q'>Q$, then there exist nonnegative integers a,a' such that $\frac{a}{q}<\frac{a'}{q'}$ are consecutive elements in $\mathcal{F}(Q)$, which establishes a one-to-one correspondence $\frac{a}{q}\leftrightarrow(q,q')$ between $\mathcal{F}(Q)\backslash\{1\}$ and the set

$$\{(q,q')\in\{1,\ldots,Q\}^2 : q+q'>Q, \gcd(q,q')=1\}.$$

A property of great importance is that if $\frac{a}{q}<\frac{a'}{q'}<\frac{a''}{q''}$ are three consecutive elements in $\mathcal{F}(Q)$, then $a+a''=Ka'$ and $q+q''=Kq'$ for some positive integer K which is commonly refered to as the index of $\frac{a'}{q'}$. More precisely, we have

$$K = \left\lfloor \frac{Q+q}{q'} \right\rfloor, \quad a'' = \left\lfloor \frac{Q+q}{q'} \right\rfloor a' - a, \quad q'' = \left\lfloor \frac{Q+q}{q'} \right\rfloor q' - q. \tag{2.1}$$

(See, e.g., [32, Chapter 3] and [31].) This led to the study of the BCZ map defined by Boca, Cobeli, and Zaharescu in [14]. The BCZ map is a function $T : \Omega \to \Omega$ on the Farey triangle $\Omega = \{(a,b) \in (0,1]^2 : a+b > 1\}$ defined by the equality

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

As a result of (2.1), we have

$$T\left(\frac{q}{Q}, \frac{q'}{Q}\right) = \left(\frac{q'}{Q}, \frac{q''}{Q}\right).$$

Let

$$\mathcal{F}(Q) = \left\{\gamma_0 = \frac{a_0}{q_0} = \frac{0}{1} < \gamma_1 = \frac{a_1}{q_1} < \cdots < \gamma_{N(Q)} = \frac{a_{N(Q)}}{q_{N(Q)}} = \frac{1}{1}\right\}$$

with $\gcd(a_i, q_i) = 1$. The properties given above motivate the correspondence $\frac{a_i}{q_i} \leftrightarrow \left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right)$ between $\mathcal{F}(Q)\backslash\{1\}$ and the subset

$$\left\{\left(\frac{q_i}{Q}, \frac{q_{i+1}}{Q}\right) : \gamma_i \in \mathcal{F}(Q)\right\} = \left\{\left(\frac{q}{Q}, \frac{q'}{Q}\right) : q, q' \in \{1, \ldots, Q\}, \gcd(q,q') = 1, q + q' > Q\right\}$$

of Ω that we refer to as the set of Farey points of order Q, and which is a periodic orbit of T. In the next section, we see how the distribution of these points in Ω as $Q \to \infty$ impacts the gap distribution of the Farey sequence.

Let $G = \mathrm{SL}(2,\mathbb{R})$ and $\Gamma = \mathrm{SL}(2,\mathbb{Z})$. (We caution the reader that in all other chapters, G denotes the Gauss map.) The central result of this chapter is the following:

Theorem 1. *Let H be a finite index subgroup of Γ and $M \subseteq \Gamma/H$ be a nonempty subset, closed under left multiplication by $\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$. Also, for $Q \in \mathbb{N}$, let $\mathcal{F}_M(Q) \subseteq \mathcal{F}(Q)$ be the set of fractions $\frac{a}{q}$ such that*

$$\begin{pmatrix} q' & a' \\ -q & -a \end{pmatrix} H \in M,$$

where $\frac{a'}{q'}$ is the successor of $\frac{a}{q}$ in $\mathcal{F}(Q)$. Then the sequence $(\mathcal{F}_M(Q))_Q$ becomes equidistributed in $[0,1]$ with respect to the Lebesgue measure as $Q \to \infty$. Furthermore, if $I \subseteq [0,1]$ is a given subinterval and $\mathcal{F}_{I,M}(Q) = \mathcal{F}_M(Q) \cap I$, then the limiting gap measure $\nu_{I,M}$ of $(\mathcal{F}_{I,M}(Q))_Q$ exists and has a continuous and piecewise real analytic density.

We include the hypothesis that M is closed under left multiplication by $\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$ in order to ensure that $(\mathcal{F}_M(Q))_Q$ is an increasing sequence of sets. Indeed, let $\frac{a}{q} \in \mathcal{F}_M(Q)$ so that

$$\begin{pmatrix} q' & a' \\ -q & -a \end{pmatrix} H \in M,$$

where $\frac{a'}{q'}$ is the successor of $\frac{a}{q}$ in $\mathcal{F}(Q)$. If $Q' \geq Q$, then by the mediant property of Farey fractions, the successor of $\frac{a}{q}$ in $\mathcal{F}(Q')$ is equal to $\frac{ma+a'}{mq+q'}$ for some $m \geq 0$. We then have

$$\begin{pmatrix} mq+q' & ma+a' \\ -q & -a \end{pmatrix} H = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^m \begin{pmatrix} q' & a' \\ -q & -a \end{pmatrix} H \in M,$$

implying that $\frac{a}{q} \in \mathcal{F}_M(Q')$, and hence $(\mathcal{F}_M(Q))_Q$ is increasing.

Applying Theorem 1 with $H = \Gamma(m)$, where m is a positive integer and $\Gamma(m)$ is the congruence subgroup

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod m : a,b,c,d \in \mathbb{Z}, ad-bc=1 \right\}$$

of Γ, and with

$$M = \left\{ \begin{pmatrix} n_4 & n_3 \\ -n_2 & -n_1 \end{pmatrix} H \in \Gamma/H : (n_1,n_2) \bmod m \in A \right\}$$

where $A \subseteq (\mathbb{Z}/m\mathbb{Z})^2$ is such that M is nonempty, i.e., there is some $(n_1,n_2) \in A$ such that $\gcd(n_1,n_2,m) = 1$, we have the following result:

Corollary 1. *Let $A \subseteq (\mathbb{Z}/m\mathbb{Z})^2$ contain some (n_1,n_2) such that $\gcd(n_1,n_2,m) = 1$, and let $I \subseteq [0,1]$ be a subinterval. Then for $Q \in \mathbb{N}$, let $\mathcal{F}_{m,A}(Q)$ be the set of fractions $\frac{a}{q} \in \mathcal{F}(Q)$ such that $(a,q) \equiv (n_1,n_2) \bmod m$ for some $(n_1,n_2) \in A$, and let $\mathcal{F}_{I,m,A}(Q) = \mathcal{F}_{m,A}(Q) \cap I$. Then $(\mathcal{F}_{m,A}(Q))_Q$ becomes equidistributed in $[0,1]$ with respect to the Lebesgue measure as $Q \to \infty$, and* *the limiting gap measure $\nu_{I,m,A}$ of $(\mathcal{F}_{I,m,A}(Q))_Q$ exists and has a continuous and piecewise real analytic density.*

Corollary 1 includes the existence of the limiting gap measures of $(\mathcal{F}_{Q,d})_Q$ and $(\widetilde{\mathcal{F}}_{Q,\ell})_Q$ proven in [16] as special cases since $\mathcal{F}_{Q,d} = \mathcal{F}_{d,A}(Q)$, where $A \subseteq (\mathbb{Z}/d\mathbb{Z})^2$ is the subset consisting of all pairs (n_1, n_2) such that $\gcd(n_2, d) = 1$, and $\widetilde{\mathcal{F}}_{Q,\ell} = \mathcal{F}_{\ell,A'}$, where $A' \subseteq (\mathbb{Z}/\ell\mathbb{Z})^2$ is the subset having all pairs (n_1, n_2) with $n_1 \not\equiv 0 \bmod \ell$. Congruence subgroups also appear in the study [55, Corollary 2.7] of the related problem of proving the existence of the limiting gap measure for the angles of visible points in $\mathbb{Z}^2$ with respect to an observer at a rational point. See [6] and [68] for similar applications of the ergodic properties of the horocycle flow on homogeneous spaces G/H to the computation of the limiting gap measure of slopes on the golden L and the octagon, respectively.

In [20], Erdős, Szüsz, and Turán introduced the Diophantine problem concerning the sets

$$S(n, \alpha, c) := \left\{ \xi \in [0,1] : \text{there exists } a, q \in \mathbb{Z} \text{ such that } (a,q) = 1, n \leq q \leq nc, |q\xi - a| \leq \frac{\alpha}{q} \right\}$$

where $n \in \mathbb{N}$, $\alpha > 0$, and $c \geq 1$. They computed the limit of the Lebesgue measures

$$\lim_{n \to \infty} \lambda(S(n, \alpha, c))$$

when $\alpha \leq \frac{c}{1+c^2}$, and the posed the problem of deciding the existence of the limit, and computing it if it does exist, for all α and c. The limit was computed in the expanded range $\alpha c \leq 1$ by Kesten [50] and shown to exist in all cases by Kesten and Sós [51] using probabilistic methods. Later, Xiong and Zaharescu [70], and independently Boca [12], gave a direct proof for the existence of the limit and expressed it in all cases explicitly in terms of iterates of the BCZ map. This type of problem was recently investigated in higher dimensions as well as in the setting of translation surfaces by Athreya and Ghosh [8]. The second main result of this chapter is the following theorem, which establishes the limiting measure for sets defined in the same way as $S(n, \alpha, c)$, with the restriction that the pairs (a, q) are such that $\frac{a}{q} \in \mathcal{F}_M(Q)$ for some $Q \in \mathbb{N}$.

Theorem 2. *Let $\mathcal{F}_M(Q)$ be the set of Farey fractions defined as in Theorem 1. Then for $n \in \mathbb{N}$,*

$\alpha > 0$, and $c \geq 1$, let

$$S_M(n,\alpha,c) := \left\{\xi \in [0,1] : \textit{there exists } \frac{a}{q} \in \mathcal{F}_M(\lfloor nc \rfloor) \textit{ in lowest terms} \right.$$
$$\left. \textit{such that } q \geq n, |q\xi - a| \leq \frac{\alpha}{q} \right\},$$

and for a given subinterval $I \subseteq [0,1]$, $S_{I,M}(n,\alpha,c) := I \cap S_M(n,\alpha,c)$. Then the limits

$$\lim_{n\to\infty} \lambda(S_M(n,\alpha,c)) \quad \textit{and} \quad \lim_{n\to\infty} \lambda(S_{I,M}(n,\alpha,c))$$

exist, and if we denote $\varrho_M(\alpha,c) := \lim_{n\to\infty} \lambda(S_M(n,\alpha,c))$, then

$$\lim_{n\to\infty} \lambda(S_{I,M}(n,\alpha,c)) = |I|\varrho_M(\alpha,c).$$

Again, letting $H = \Gamma(m)$ and $M \subseteq \Gamma/H$ be defined as before Corollary 1, we obtain the following:

Corollary 2. *Let $A \subseteq (\mathbb{Z}/m\mathbb{Z})^2$ contain some (n_1,n_2) such that $\gcd(n_1,n_2,m) = 1$. Then for $n \in \mathbb{N}$, $\alpha > 0$, and $c \geq 1$, let*

$$S(n,\alpha,c,A) := \left\{\xi \in [0,1] : \textit{there exists } a,q \in \mathbb{Z}_{\geq 0} \textit{ such that } (a,q) = 1, \right.$$
$$\left. (a,q) \bmod m \in A, n \leq q \leq nc, |q\xi - a| \leq \frac{\alpha}{q} \right\}$$

and for a given subinterval $I \subseteq [0,1]$, $S_I(n,\alpha,c,A) := I \cap S(n,\alpha,c,A)$. Then the limits

$$\lim_{n\to\infty} \lambda(S(n,\alpha,c,A)) \quad \textit{and} \quad \lim_{n\to\infty} \lambda(S_I(n,\alpha,c,A))$$

exist, and

$$\lim_{n\to\infty} \lambda(S_I(n,\alpha,c,A)) = |I| \lim_{n\to\infty} \lambda(S(n,\alpha,c,A)).$$

In Section 2.2, we review the work of Athreya and Cheung [7] in showing the horocycle flow as a suspension flow by naturally identifying the Farey triangle Ω with a cross section Ω' of the horocycle flow whose first return map corresponds to the BCZ map. We also outline how Athreya and Cheung used the ergodic properties of the horocycle flow to prove the equidistribution of the Farey points in Ω', which in turn yields results about Farey fraction gaps. Then in Sections 2.3–2.6, we prove Theorem 1 using the same process. We start in Section 2.3 by proving that $\bigcup_{Q\in\mathbb{N}}\mathcal{F}_M(Q)$ is dense in $[0,1]$, and this involves proving an elementary lemma regarding representatives for cosets in Γ/H. In Section 2.4, we use results in [26] to construct a cross section Ω_M of the horocycle flow on G/H analogous to Ω' that relates to the gaps in $(\mathcal{F}_M(Q))_Q$. In Section 2.5, we prove some important properties of the first return time function of Ω_M which have an effect on the existence and properties of $\nu_{I,M}$ mentioned in Theorem 1. Then in Section 2.6, we prove the equidistribution of certain points in Ω_M, analogous to the Farey points in Ω'. From this equidistribution we can conclude the existence of the limiting gap measure of $(\mathcal{F}_{I,M}(Q))_Q$. In Section 2.7, we examine a particular property that we call the repulsion gap of $(\mathcal{F}_{I,M}(Q))_Q$, which is the infimum of the support of $\nu_{I,M}$. In the particular case where $(\mathcal{F}_{I,M}(Q))_Q$ is the sequence of Farey fractions $\frac{a}{q}$ with $q\equiv 1 \bmod m$ for some fixed $m\in\mathbb{N}$, the repulsion gap is explicitly computed as

$$\frac{3}{\pi^2 m}\prod_{\substack{p\mid m\\ p\text{ prime}}}(1-p^{-2})^{-1}. \tag{2.2}$$

Lastly, in Section 2.8, we prove Theorem 2 also as a corollary of the work in Sections 2.3–2.6.

2.2 The cross section of Athreya and Cheung

In [7], Athreya and Cheung viewed the Farey triangle Ω as a subset of G/Γ by letting

$$P := \left\{ p_{a,b} = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} : (a,b)\in\Omega \right\}$$

and considering the set

$$\Omega' := P\Gamma/\Gamma = \{\Lambda_{a,b} = p_{a,b}\Gamma : (a,b)\in\Omega\},$$

which was found to be a cross section for the horocycle flow, the action by left multiplication on G/Γ of

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

This means that for almost every $\Lambda \in G/\Gamma$ with respect to the Haar measure, the set $\{s \in \mathbb{R} : h_s\Lambda \in \Omega'\}$ of times the orbit of Λ under the horocycle flow meets Ω' is nonempty, countable, and discrete. The first return time function $R : \Omega' \to \mathbb{R}$ defined by $R(\Lambda_{a,b}) = \min\{s > 0 : h_s\Lambda_{a,b} \in \Omega'\}$ is $R(\Lambda_{a,b}) = (ab)^{-1}$, and the first return map $r : \Omega' \to \Omega'$ defined by $r(\Lambda_{a,b}) = h_{r(\Lambda_{a,b})}\Lambda_{a,b}$ is $r(\Lambda_{a,b}) = \Lambda_{T(a,b)}$, where T is the BCZ map. The last equality can be seen by the calculation

$$\begin{aligned} h_{R(\Lambda_{a,b})}\Lambda_{a,b} &= \begin{pmatrix} 1 & 0 \\ -(ab)^{-1} & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \Gamma = \begin{pmatrix} a & b \\ -b^{-1} & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & \lfloor \frac{1+a}{b} \rfloor \end{pmatrix} \Gamma \\ &= \begin{pmatrix} b & \lfloor \frac{1+a}{b} \rfloor b - a \\ 0 & b^{-1} \end{pmatrix} \Gamma = \Lambda_{T(a,b)}. \end{aligned}$$

Also, if we identify G/Γ with the set $\{(a, b, s) : (a, b) \in \Omega, 0 \le s < (ab)^{-1}\}$ via the correspondence $h_s\Lambda_{a,b} \leftrightarrow (a, b, s)$, then $d\mu_{G/\Gamma} = 2\, da\, db\, ds$, where $\mu_{G/\Gamma}$ is the Haar measure on G/Γ such that $\mu_{G/\Gamma}(G/\Gamma) = \frac{\pi^2}{3}$.

Letting $I \subseteq [0, 1]$ be a subinterval, $\mathcal{F}_I(Q) = \mathcal{F}(Q) \cap I$, and $N_I(Q) = \#\mathcal{F}_I(Q)$, we define on Ω' the measure

$$\rho_{Q,I} := \frac{1}{N_I(Q)} \sum_{i:\gamma_i \in I} \delta_{r^i(\Lambda_{1,1/Q})} = \frac{1}{N_I(Q)} \sum_{i:\gamma_i \in I} \delta_{\Lambda_{q_i/Q, q_{i+1}/Q}},$$

where δ_x denotes the Dirac measure of measure one concentrated at $x \in \Omega'$. Now notice that $R(\Lambda_{q_i/Q, q_{i+1}/Q}) = \frac{Q^2}{q_i q_{i+1}} = Q^2(\gamma_{i+1} - \gamma_i)$, and as a result,

$$\frac{\#\{\gamma_i \in \mathcal{F}_I(Q) : \frac{3}{\pi^2}Q^2(\gamma_{i+1} - \gamma_i) \in [0, \xi]\}}{N_I(Q)} = \rho_{Q,I}\left(R^{-1}\left(\left[0, \frac{\pi^2}{3}\xi\right]\right)\right) \tag{2.3}$$

for all $\xi \ge 0$. By the equidistribution of $(\mathcal{F}(Q))_Q$, $N_I(Q) \sim \frac{3}{\pi^2}|I|Q^2$, and if $\gamma_{I,l}$ and $\gamma_{I,g}$ are the least and greatest elements in $\mathcal{F}_I(Q)$, respectively (we suppress the dependence on Q), then $\gamma_{I,g} - \gamma_{I,l} \to |I|$ as $Q \to \infty$. Thus the limit of the left side of (2.3) as $Q \to \infty$ is the measure of $[0, \xi]$ under the limiting gap measure of $(\mathcal{F}_I(Q))_Q$. So to show that the limiting gap measure of $(\mathcal{F}_I(Q))_Q$ exists, it suffices to prove that the limit of the right side of (2.3) exists. To do so, Athreya and Cheung proved that the sequence $(\rho_{Q,I})_Q$ of measures converges in the weak* topology to the measure m on Ω' given by $dm = 2\,da\,db$. They first noticed that if $\rho^R_{Q,I}$ is the measure on G/Γ defined by $d\rho^R_{Q,I} = d\rho_{Q,I}\,ds$, where we are viewing G/Γ as the set $\{(\Lambda_{a,b}, s) \in \Omega' \times \mathbb{R} : 0 \leq s < (ab)^{-1}\}$ by the correspondence $h_s\Lambda_{a,b} \leftrightarrow (\Lambda_{a,b}, s)$, then $\rho^R_{Q,I} \to \mu_{G/\Gamma}$ in the weak* topology. This convergence is a consequence of the equidistribution of closed horocycles in G/Γ (see, e.g., [71, 62, 21, 36, 65]).

It then follows that if $\pi_{\Omega'} : G/\Gamma \to \Omega'$ is the projection $(\Lambda_{a,b}, s) \mapsto \Lambda_{a,b}$ (we are again viewing G/Γ as $\{(\Lambda_{a,b}, s) \in \Omega' \times \mathbb{R} : 0 \leq s < (ab)^{-1}\}$), then

$$\frac{1}{R}\pi_{\Omega'*}\rho^R_{Q,I} \to \frac{1}{R}\pi_{\Omega'*}\mu_{G/\Gamma} \quad (Q \to \infty)$$

in the weak* topology. It is easy to see that $\rho_{Q,I} = \frac{1}{R}\pi_{\Omega'*}\rho^R_{Q,I}$ and $m = \frac{1}{R}\pi_{\Omega'*}\mu_{G/\Gamma}$, and so $\rho_{Q,I} \to m$.

Remark 1. The convergence $\rho_{Q,I} \to m$ was proven in [43] in the case $I = [0, 1]$. This can actually be proven in an elementary way using Möbius summation. For shrinking subintervals $I = I(Q) \subseteq [0, 1]$ with $|I(Q)| \gg Q^{-1/2+\epsilon}$, this convergence can be deduced using a corollary of the Weil bound for Kloosterman sums (see, e.g., [15, Section 2]). For fixed I, this convergence also follows from [53, Theorem 6], in which the equidistribution of Farey points of arbitrary dimension was proven. See [54] for results regarding the spacing statistics of higher dimensional Farey fractions. Using $\rho_{Q,I} \to m$, Athreya and Cheung [7] not only explained the gap distribution of $(\mathcal{F}_I(Q))_Q$, but through finding appropriate functions $f : \Omega' \to \mathbb{R}$ such that

$$\lim_{Q\to\infty} \int_{\Omega'} f\, d\rho_{Q,I} = \int_{\Omega'} f\, dm,$$

they were able to recast in this unified setting many previously known results about Farey fractions, including results on their h-spacings and indices.

2.3 The density of $\bigcup_{Q\in\mathbb{N}} \mathcal{F}_M(Q)$ in $[0,1]$

Throughout this section and Sections 2.4–2.6, let $H \subseteq \Gamma$ be a subgroup of finite index, $M = \{m_1H, \dots, m_kH\} \subseteq \Gamma/H$ be a nonempty subset closed under left multiplication by $\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$, and $I = [t_1, t_2] \subseteq [0,1]$ be a subinterval. We now set out to prove that the limiting gap measure $\nu_{I,M}$ of $(\mathcal{F}_{I,M}(Q))_Q$ exists and start in this section by proving that $\bigcup_{Q\in\mathbb{N}} \mathcal{F}_M(Q)$ is dense in $[0,1]$. We first prove the following elementary lemma:

Lemma 1. *Let gH be any coset in Γ/H. Then there exist positive integers a, b, c, d such that*

$$\begin{pmatrix} a & b \\ -c & -d \end{pmatrix} \in gH.$$

Proof. First note that since $[\Gamma : H] < \infty$, there exists an integer $N \geq 2$ such that

$$U_N = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}, \quad L_N = \begin{pmatrix} 1 & 0 \\ N & 1 \end{pmatrix} \in H.$$

Let $A = \left(\begin{smallmatrix} a_0 & b_0 \\ c_0 & d_0 \end{smallmatrix}\right) \in gH$. If $a_0 > 0$ and $b_0 = 0$, then $A = \left(\begin{smallmatrix} 1 & 0 \\ c_0 & 1 \end{smallmatrix}\right)$. Replacing A by AL_N^{-j} for a large enough j replaces c_0 by a number less than -1, and so we can assume that $c_0 < -1$. We then have $AU_N = \left(\begin{smallmatrix} 1 & N \\ c_0 & c_0N+1 \end{smallmatrix}\right)$, which is a matrix of the desired form. So the proof is complete in this case.

So assume that $a_0 \leq 0$ or $b_0 \neq 0$. If $b_0 = 0$, we must have $a_0 < 0$, and if $a_0 = 0$ so that $b_0 \neq 0$, multiplying A on the right by L_N or L_N^{-1} replaces a_0 by a negative number. So we can assume that $a_0 < 0$. Then multiplying A on the right by U_N^{-j} for a large enough j replaces b_0 by a positive number, and so assume that $b_0 > 0$. Now since $a_0d_0 - b_0c_0 = 1$, we clearly have $c_0d_0 \leq 0$. Suppose that $d_0 = 0$, implying that $A = \left(\begin{smallmatrix} a_0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Multiplying A on the right by $L_N^jU_N$ yields $\left(\begin{smallmatrix} a_0+jN & N(a_0+jN)+1 \\ -1 & -N \end{smallmatrix}\right)$. Choosing j so that $a_0 + jN > 0$ yields a matrix of the desired form, and so the proof is complete in this case. If $c_0 = 0$, then $A = \left(\begin{smallmatrix} -1 & b_0 \\ 0 & -1 \end{smallmatrix}\right)$, and multiplying A on the right by L_N^{-1} yields $\left(\begin{smallmatrix} -1-b_0N & b_0 \\ N & -1 \end{smallmatrix}\right)$.

Thus we have reduced the case where $c_0 = 0$ to the situation where $A = \left(\begin{smallmatrix} -a & b \\ c & -d \end{smallmatrix}\right)$ with $a, b, c, d > 0$, which we now consider. Let $\frac{\gamma}{\alpha} < \frac{\delta}{\beta}$ be fractions such that $\alpha\delta - \beta\gamma = 1$ and $\frac{a}{b} < \frac{\gamma}{\alpha}$. Matrix multiplication reveals that any power of $B = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ is of the form $\left(\begin{smallmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{smallmatrix}\right)$ where $\frac{\gamma}{\alpha} \leq \frac{\gamma'}{\alpha'} < \frac{\delta'}{\beta'} \leq \frac{\delta}{\beta}$.

So noting that $[\Gamma : H] < \infty$, we can replace B by some power of B that is in H. We then have $AB \in gH$ and

$$AB = \begin{pmatrix} -a & b \\ c & -d \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} -a\alpha + b\gamma & -a\beta + b\delta \\ c\alpha - d\gamma & c\beta - d\delta \end{pmatrix},$$

which is a matrix of the desired form since $\frac{c}{d} < \frac{a}{b} < \frac{\gamma}{\alpha} < \frac{\delta}{\beta}$.

The last case we need to consider is when $A = \left(\begin{smallmatrix} -a & b \\ -c & d \end{smallmatrix}\right)$ with $a, b, c, d > 0$. Noting that $\frac{a}{b} < \frac{c}{d}$ since $-ad + bc = 1$, we can find fractions $\frac{\gamma}{\alpha} < \frac{\delta}{\beta}$ such that $\alpha\delta - \beta\gamma = 1$ and $\frac{a}{b} < \frac{\gamma}{\alpha} < \frac{\delta}{\beta} < \frac{c}{d}$. As in the previous case, we may assume that $B = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in H$. Then $AB \in gH$ and

$$AB = \begin{pmatrix} -a & b \\ -c & d \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} -a\alpha + b\gamma & -a\beta + b\delta \\ -c\alpha + d\gamma & -c\beta + d\delta \end{pmatrix}$$

is a matrix of the desired form. This completes the proof. $\square$

So by Lemma 1, there exist $a, b, c, d \in \mathbb{N}$ such that $\left(\begin{smallmatrix} a & b \\ -c & -d \end{smallmatrix}\right) H \in M$. Multiplying $\left(\begin{smallmatrix} a & b \\ -c & -d \end{smallmatrix}\right)$ on the right by L_N yields a matrix of the form $\left(\begin{smallmatrix} q' & a' \\ -q & -a \end{smallmatrix}\right)$, where $\frac{a}{q} < \frac{a'}{q'}$ are consecutive Farey fractions of some order. We have $\left(\begin{smallmatrix} q' & a' \\ -q & -a \end{smallmatrix}\right) H \in M$, proving that $\bigcup_{Q \in \mathbb{N}} \mathcal{F}_M(Q)$ is nonempty.

Lemma 2. *There exists constants $Y > 0$ and $Q_0 > 0$, depending only on the subgroup $H \subseteq G$, such that for any $Q \geq Q_0$ and $x \in [0, 1]$,*

$$\min_{\beta \in \mathcal{F}_M(Q)} |x - \beta| \leq \frac{Y}{Q}.$$

Proof. By Lemma 1, there exists a matrix of the form $\left(\begin{smallmatrix} a & b \\ -c & -d \end{smallmatrix}\right)$, with $a, b, c, d > 0$, in each coset gH in Γ/H. By multiplying each on the right by a sufficiently large power of $\left(\begin{smallmatrix} 1 & -N \\ 0 & 1 \end{smallmatrix}\right)$, we can say that each coset in Γ/H also contains a matrix of the form $\left(\begin{smallmatrix} a & -b \\ -c & d \end{smallmatrix}\right)$, with $a, b, c, d > 0$. Let $k_H \in \mathbb{N}$ be an upper bound on the entries of coset representatives in Γ/H of the forms $\left(\begin{smallmatrix} a & b \\ -c & -d \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} a & -b \\ -c & d \end{smallmatrix}\right)$.

Now let $x \in [0, 1]$ be given. Then for $Q \in \mathbb{N}$, let $\frac{a}{q} < \frac{a'}{q'}$ be consecutive in $\mathcal{F}(Q)$ such that $\frac{a}{q} \leq x \leq \frac{a'}{q'}$, and note that the difference between x and both $\frac{a}{q}$ and $\frac{a'}{q'}$ is at most Q^{-1}. Thus any fraction between $\frac{a}{q}$ and $\frac{a'}{q'}$ is at most Q^{-1} away from x.

By our comments above, the coset $\left(\begin{smallmatrix} q & a \\ q' & a' \end{smallmatrix}\right) H$ contains an element of the form $\left(\begin{smallmatrix} a_0 & -b_0 \\ -c_0 & d_0 \end{smallmatrix}\right)$, with $0 < a_0, b_0, c_0, d_0 \leq k_H$. Thus H contains

$$\begin{pmatrix} a_0 & -b_0 \\ -c_0 & d_0 \end{pmatrix}^{-1} \begin{pmatrix} q & a \\ q' & a' \end{pmatrix} = \begin{pmatrix} d_0 & b_0 \\ c_0 & a_0 \end{pmatrix} \begin{pmatrix} q & a \\ q' & a' \end{pmatrix} = \begin{pmatrix} v & u \\ v' & u' \end{pmatrix},$$

where $\frac{u}{v} < \frac{u'}{v'}$ are consecutive Farey fractions of some order between $\frac{a}{q}$ and $\frac{a'}{q'}$ such that $v, v' \leq 2k_H Q$. If $\left(\begin{smallmatrix} a_1 & b_1 \\ -c_1 & -d_1 \end{smallmatrix}\right)$ is a representative of any coset $m_i H$ in M with $0 < a_1, b_1, c_1, d_1 \leq k_H$, then $m_i H$ contains

$$\begin{pmatrix} a_1 & b_1 \\ -c_1 & -d_1 \end{pmatrix} \begin{pmatrix} v & u \\ v' & u' \end{pmatrix} = \begin{pmatrix} a_1 v + b_1 v' & a_1 u + b_1 u' \\ -c_1 v - d_1 v' & -c_1 u - d_1 u' \end{pmatrix}.$$

Thus $\mathcal{F}_M(Q')$ contains $\frac{c_1 u + d_1 u'}{c_1 v + d_1 v'}$ for all $Q' \geq \max\{a_1 v + b_1 v', c_1 v + d_1 v'\}$. Since $\max\{a_1 v + b_1 v', c_1 v + d_1 v'\} \leq 4k_H^2 Q$, $\mathcal{F}_M(Q')$ contains $\frac{c_1 u + d_1 u'}{c_1 v + d_1 v'}$ for $Q' \geq 4k_H^2 Q$. Thus the minimum distance between x and an element in $\mathcal{F}_M(4k_H^2 Q)$ is at most Q^{-1}. Replacing Q by $\left\lfloor \frac{Q}{4k_H^2} \right\rfloor$ in this argument, and letting $Q \geq 8k_H^2$, we see that the minimum distance between x and an element in $\mathcal{F}_M(Q)$ is at most $\frac{1}{(Q/4k_H^2)-1} \leq \frac{8k_H^2}{Q}$. This completes the proof with $Y = Q_0 = 8k_H^2$. □

This immediately implies the density of $(\mathcal{F}_M(Q))_Q$.

Corollary 3. *The set* $\bigcup_{Q \in \mathbb{N}} \mathcal{F}_M(Q)$ *is dense in* $[0, 1]$.

With Lemma 2 and the horocycle equidistribution results of Hejhal [36], and later Strömbergsson [65], it is conceivable that one can obtain results while allowing the subinterval I to shrink with Q and n in Theorems 1 and 2, respectively.

2.4 A cross section for G/H

In the same way the properties of Ω' as a cross section of the horocycle flow on G/Γ were used in [7] to deduce many consequences for the gaps in $(\mathcal{F}(Q))_Q$, we now find a new space Ω_M that is a cross section of the horocycle flow on G/H which can be used to analyze the gaps in $\mathcal{F}_M(Q)$. One step toward this goal is to lift the cross section Ω' to G/H via the natural projection $\pi_H : G/H \to G/\Gamma$.

In order to do this, we use the work of Fisher and Schmidt [26] on the behavior of a lifted cross section for the geodesic flow from $F\backslash\mathrm{PSL}(2,\mathbb{R})$, $F \subseteq \mathrm{PSL}(2,\mathbb{R})$ being a Fuchsian group of finite covolume, to a finite cover $F'\backslash\mathrm{PSL}(2,\mathbb{R})$ of $F\backslash\mathrm{PSL}(2,\mathbb{R})$. In particular, we apply [26, Lemma 2, Theorem 3] to the finite cover $\pi_H : G/H \to G/\Gamma$, lifting the cross section Ω' to $\Omega_H := \pi_H^{-1}(\Omega')$. In applying [26, Theorem 3], we make the slight modifications of working in the left coset space instead of the right coset space, replacing the geodesic flow with the horocycle flow, and allowing the possibility that $-\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right) \notin H$, so that G/H is not necessarily of the form $\mathrm{PSL}(2,\mathbb{R})/F'$. We summarize the results we need from this application in the following theorem:

Theorem 3. *The set*

$$\Omega_H := \pi_H^{-1}(\Omega') = P\Gamma H/H = \{p_{a,b}gH : p_{a,b} \in P, g \in \Gamma\}$$

is a cross section for the action of $\mathcal{N}$ on G/H with first return time function $R \circ \pi_H : \Omega_H \to \mathbb{R}$ and first return map $r_H : \Omega_H \to \Omega_H$ such that, for all $p_{a,b} \in P$ and $g \in \Gamma$,

$$\begin{aligned} r_H(p_{a,b}\gamma H) &= h_{R(\pi_H(p_{a,b}\gamma H))}p_{a,b}\gamma H \\ &= \begin{pmatrix} 1 & 0 \\ -(ab)^{-1} & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & \lfloor\frac{1+a}{b}\rfloor \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & \lfloor\frac{1+a}{b}\rfloor \end{pmatrix}^{-1} \gamma H \\ &= \begin{pmatrix} b & \lfloor\frac{1+a}{b}\rfloor b - a \\ 0 & b^{-1} \end{pmatrix} \begin{pmatrix} \lfloor\frac{1+a}{b}\rfloor & 1 \\ -1 & 0 \end{pmatrix} \gamma H. \end{aligned} \tag{2.4}$$

Identify Ω_H with $P \times \Gamma/H$ via the correspondence $p_{a,b}\gamma H \leftrightarrow (p_{a,b}, \gamma H)$, and let μ_{Ω_H} be the measure on Ω_H corresponding in this way to the product measure on $P \times \Gamma/H$ of $2\,da\,db$ with the counting measure on Γ/H. Then identify G/H with $\{(x,s) \in \Omega_H \times \mathbb{R} : 0 \leq s < (R\circ\pi_H)(x)\}$ via $(x,s) \leftrightarrow h_s x$. Then the Haar measure $\mu_{G/H}$ on G/H, normalized so that $\mu_{G/H}(G/H) = \frac{\pi^2}{3}[\Gamma : H]$, is given by $d\mu_{G/H} = d\mu_{\Omega_H}\,ds$.

Our next step is to find a correspondence of $\mathcal{F}(Q)$ with points in Ω_H analogous to that of $\mathcal{F}(Q)$ with the Farey points in Ω. So suppose that $\frac{a}{q} < \frac{a'}{q'} < \frac{a''}{q''}$ are consecutive fractions in $\mathcal{F}(Q)$. Then letting $K = \left\lfloor \frac{Q+q}{q'} \right\rfloor$ and noting that $a'' = Ka' - a$ and $q'' = Kq' - q$, by (2.4) we have

$$\begin{aligned} r_H\left(\begin{pmatrix} \frac{q}{Q} & \frac{q'}{Q} \\ 0 & \frac{Q}{q} \end{pmatrix}\begin{pmatrix} q' & a' \\ -q & -a \end{pmatrix} H\right) &= \begin{pmatrix} \frac{q'}{Q} & \frac{Kq'-q}{Q} \\ 0 & \frac{Q}{q'} \end{pmatrix}\begin{pmatrix} Kq'-q & Ka'-a \\ -q' & -a' \end{pmatrix} H \\ &= \begin{pmatrix} \frac{q'}{Q} & \frac{q''}{Q} \\ 0 & \frac{Q}{q'} \end{pmatrix}\begin{pmatrix} q'' & a'' \\ -q' & -a' \end{pmatrix} H. \end{aligned}$$

Therefore, if we associate each fraction $\frac{a}{q} \in \mathcal{F}(Q)$ to the element

$$W_{H,Q}\left(\frac{a}{q}\right) := \begin{pmatrix} \frac{q}{Q} & \frac{q'}{Q} \\ 0 & \frac{Q}{q} \end{pmatrix}\begin{pmatrix} q' & a' \\ -q & -a \end{pmatrix} H$$

of Ω_H, where $\frac{a'}{q'}$ is the element succeeding $\frac{a}{q}$ in $\mathcal{F}(Q)$, then $r_H\left(W_{H,Q}\left(\frac{a}{q}\right)\right) = W_{H,Q}\left(\frac{a'}{q'}\right)$. The map $W_{H,Q}$ gives the correspondence of $\mathcal{F}(Q)$ with Ω_H that we are seeking, and $W_{H,Q}$ sends a fraction in $\mathcal{F}(Q)$ to

$$\Omega_M := PM = \{pm_iH : p \in P, m_iH \in M\}$$

if and only if the fraction is in $\mathcal{F}_M(Q)$. The set Ω_M is the cross section in G/H we set out to find at the beginning of this section.

To see that Ω_M is in fact a cross section for the horocycle flow on G/H, notice that the set

$$h_{[-1,0]}\Omega_M = \{h_s pm_iH : s \in [-1,0], pm_iH \in \Omega_M\} \subseteq G/H$$

has positive $\mu_{G/H}$-measure. Now by the ergodicity of the horocycle flow [33], $\mu_{G/H}$-a.e. $x \in G/H$ is sent to $h_{[-1,0]}\Omega_M$ by $\{h_s : s > 0\}$. Clearly all of $h_{[-1,0]}\Omega_M$ is sent to Ω_M by $\{h_s : s \geq 0\}$, and so a.e. $x \in G/H$ is sent to Ω_M by $\{h_s : s > 0\}$. The discreteness of $\{s \in \mathbb{R} : h_s x \in \Omega_M\}$ for a.e. $x \in G/H$ follows from the fact that $\Omega_M \subseteq \Omega_H$. This proves that Ω_M is a cross section for the action of $\mathcal{N}$ on G/H.

Let $R_M : \Omega_M \to (0,\infty]$ be the first return time function $R_M(x) := \min\{s > 0 : h_s x \in \Omega_M\}$ and r_M be the first return map on Ω_M defined by $r_M(x) := h_{R_M(x)}x$. Here we note that if $\mu_{\Omega_M} := \frac{1}{\#M}\mu_{\Omega_H}|_{\Omega_M}$, then $d\mu_{G/H} = (\#M)\, d\mu_{\Omega_M}\, ds$, where we identify G/H with $\{(x,s) \in \Omega_M \times \mathbb{R} :$

$0 \le s < R_M(x)\}$ by $h_s x \leftrightarrow (x,s)$.

Now that we have identified a cross section in G/H and a correspondence $W_{H,Q}$ of $\mathcal{F}_M(Q)$ to a subset of Ω_M analogous to the set of Farey points in Ω, we wish to see if information about the gaps in $\mathcal{F}_M(Q)$ can be deduced from Ω_M. So let $\gamma_i \in \mathcal{F}_M(Q)$ such that there are $i' > i$ with $\gamma_{i'} \in \mathcal{F}_M(Q)$. Since $\Omega_M \subseteq \Omega_H$, for each $x \in \Omega_M$ in which $r_M(x)$ is defined, $r_M(x) = r_H^j(x)$ for some $j \in \mathbb{N}$. So

$$r_M(W_{H,Q}(\gamma_i)) = r_H^j(W_{H,Q}(\gamma_i)) = W_{H,Q}(\gamma_{i+j}),$$

where $j \in \mathbb{N}$ is the least element such that $W_{H,Q}(\gamma_{i+j}) \in \Omega_M$, i.e., $\gamma_{i+j} \in \mathcal{F}_M(Q)$. We then have

$$\begin{aligned} R_M(W_{H,Q}(\gamma_i)) &= \sum_{i'=0}^{j-1} (R \circ \pi_H)(r_H^{i'}(W_{H,Q}(\gamma_i))) = \sum_{i'=0}^{j-1} \frac{Q^2}{q_{i+i'} q_{i+i'+1}} \\ &= Q^2(\gamma_{i+j} - \gamma_i). \end{aligned} \tag{2.5}$$

So just as the return time function R on Ω' contained information about the gaps in $\mathcal{F}(Q)$, R_M contains information about the gaps in $\mathcal{F}_M(Q)$.

Let $N_{I,M}(Q) := \#\mathcal{F}_{I,M}(Q) - 1$ and

$$\mathcal{F}_{I,M}(Q) = \{\beta_0 < \beta_1 < \cdots < \beta_{N_{I,M}(Q)}\}.$$

For notational convenience, we suppress the dependence of the β_i on Q. Then define the measure $\rho_{Q,I,M}$ on Ω_M by

$$\rho_{Q,I,M} := \frac{1}{N_{I,M}(Q)} \sum_{i=0}^{N_{I,M}(Q)-1} \delta_{W_{H,Q}(\beta_i)}.$$

By (2.5), we have

$$\frac{\#\{0 \le i \le N_{I,M}(Q) - 1 : Q^2(\beta_{i+1} - \beta_i) \in [0,\xi]\}}{N_{I,M}(Q)} = \rho_{Q,I,M}\left(R_M^{-1}([0,\xi])\right) \tag{2.6}$$

for all $\xi \ge 0$. To show that the limit of the left side, and hence the right side, exists, we prove in Section 2.5 that the boundary $\partial\left(R_M^{-1}([0,\xi])\right)$ of the set $R_M^{-1}([0,\xi])$ has μ_{Ω_M}-measure 0, and then prove in Section 2.6 that $(\rho_{Q,I,M})_Q$ converges weakly to μ_{Ω_M} as $Q \to \infty$. These results imply that

$$\lim_{Q\to\infty} \rho_{Q,I,M}\left(R_M^{-1}([0,\xi])\right) = \mu_{\Omega_M}\left(R_M^{-1}([0,\xi])\right)$$

by the Portmanteau theorem. Note that the right side of (2.6) is not the relevant limit to prove the existence of the limiting gap measure for $(\mathcal{F}_{I,M}(Q))_Q$ because of the gap normalization factor Q^2 and the missing factor $\beta_{N_{I,M}(Q)} - \beta_0$ to be multiplied by ξ. However, we have $\beta_{N_{I,M}(Q)} - \beta_0 \to |I|$ as $Q \to \infty$ by the density of $\bigcup_{Q\in\mathbb{N}} \mathcal{F}_M(Q)$ in $[0,1]$, and we show that $N_{I,M}(Q) \sim \frac{|I|(\#M)Q^2}{\mu_{G/H}(G/H)} = \frac{3|I|(\#M)Q^2}{\pi^2[\Gamma:H]}$ as $Q \to \infty$ in the course of our work in Section 2.6. It then follows that the limiting gap measure $\nu_{I,M}$ exists and satisfies

$$\nu_{I,M}([0,\xi]) = \mu_{\Omega_M}\left(R_M^{-1}\left(\left[0, \frac{\pi^2[\Gamma:H]}{3(\#M)}\xi\right]\right)\right) \tag{2.7}$$

for all $\xi \geq 0$. Another corollary is that $\lim_{Q\to\infty} \frac{N_{I,M}(Q)}{N_{[0,1],M}(Q)} = |I|$ for every subinterval $I \subseteq [0,1]$, implying that $(\mathcal{F}_M(Q))_Q$ becomes equidistributed in $[0,1]$ as $Q \to \infty$.

2.5 The return time function R_M

In this section, we prove important properties of the first return time function R_M. Specifically, we show that $\mu_{\Omega_M}\left(\partial\left(R_M^{-1}([0,\xi])\right)\right) = 0$ for every $\xi \geq 0$, and the function $\mathcal{A}_M : [0,\infty) \to [0,1]$ defined by

$$\mathcal{A}_M(\xi) = \mu_{\Omega_M}\left(R_M^{-1}([0,\xi])\right)$$

has a continuous, piecewise real analytic derivative. These results, together with our work in Section 2.6, prove that the limiting gap measure $\nu_{I,M}$ exists and has a continuous and piecewise real analytic density. To do this, we show that R_M is a piecewise rational function, viewing each component Pm_iH/H of Ω_M as a copy of the Farey triangle Ω by the correspondence $p_{a,b}m_iH \leftrightarrow (a,b)$, and that the region in a given component Pm_iH/H over which R_M is equal to a particular rational function is a polygon. This allows us to say that $R_M^{-1}([0,\xi])$ is the union of regions, each being obtained by intersecting a polygon with a region bounded below by a hyperbola which depends in a smooth way on ξ. In particular, we show that $R_M^{-1}([0,\xi])$ is a finite union of these regions, which grants us the properties of $\mathcal{A}_M$ we seek.

2.5.1 R_M is piecewise rational

First let $p_{a,b} \in P$ and $m_iH \in M$ so that $p_{a,b}m_iH \in \Omega_M$, and suppose $s > 0$. We have $h_s p_{a,b}m_iH \in \Omega_M$ if and only if there exist $p_{c,d} \in P$ and $m_jH \in M$ such that $h_s p_{a,b}m_iH = p_{c,d}m_jH$. This means there exists $h \in H$ such that $h_s p_{a,b}m_ihm_j^{-1} = p_{c,d}$. Letting $m_ihm_j^{-1} = \left(\begin{smallmatrix} c_1 & c_2 \\ c_3 & c_4 \end{smallmatrix}\right)$, this equality is

$$\begin{pmatrix} ac_1 + bc_3 & ac_2 + bc_4 \\ a^{-1}c_3 - s(ac_1 + bc_3) & a^{-1}c_4 - s(ac_2 + bc_4) \end{pmatrix} = \begin{pmatrix} c & d \\ 0 & c^{-1} \end{pmatrix}.$$

Thus $h_s p_{a,b}m_iH \in \Omega_M$ if and only if there exists $\left(\begin{smallmatrix} c_1 & c_2 \\ c_3 & c_4 \end{smallmatrix}\right) \in \bigcup_{j=1}^k m_iHm_j^{-1}$ such that $(ac_1+bc_3, ac_2+bc_4) \in \Omega$ and $s = \frac{c_3}{a(ac_1+bc_3)} = \frac{1}{a(b+c_1a/c_3)}$. The latter conditions and $s > 0$ imply that $c_3 > 0$, and hence $c_1 \leq 0$ since $ac_1 + bc_3 \leq 1$ and $a + b > 1$. In particular, if $R_M(p_{a,b}m_iH) < \infty$, then $R_M(p_{a,b}m_iH) = \frac{1}{a(b+c_1a/c_3)}$, where $\frac{c_1}{c_3}$ is the greatest fraction such that $c_1 \leq 0$, $c_3 > 0$, and there exists $\left(\begin{smallmatrix} c_1 & c_2 \\ c_3 & c_4 \end{smallmatrix}\right) \in \bigcup_{j=1}^k m_iHm_j^{-1}$ with $(ac_1 + bc_3, ac_2 + bc_4) \in \Omega$. Note that such a greatest fraction exists since if $\frac{c_1}{c_3}$ and $\frac{c_1'}{c_3'}$ are distinct fractions satisfying the conditions of the previous sentence, then $\left|\frac{1}{a(b+c_1a/c_3)} - \frac{1}{a(b+c_1'a/c_3')}\right| = \frac{|c_1'c_3 - c_3'c_1|}{(c_3b+c_1a)(c_3'b+c_1'a)} > 1$. So R_M can be written not only as an infimum, but as a minimum, of the expressions $\frac{1}{a(b+c_1a/c_3)}$. We have thus proven the following result:

Proposition 1. *The function R_M is a piecewise rational function. Specifically,*

$$R_M = \min_{C \in \bigcup_{i,j=1}^k m_iHm_j^{-1}} f_C,$$

where for each $C = \left(\begin{smallmatrix} c_1 & c_2 \\ c_3 & c_4 \end{smallmatrix}\right) \in \bigcup_{i,j=1}^k m_iHm_j^{-1}$, $f_C : \Omega_M \to [0, \infty]$ is defined by

$$f_C(p_{a,b}m_iH) := \begin{cases} \dfrac{1}{a(b + c_1a/c_3)} & \text{if } C \in \bigcup_{j=1}^k m_iHm_j^{-1}, c_3 > 0, (\, a \;\, b\,)\left(\begin{smallmatrix} c_1 & c_2 \\ c_3 & c_4 \end{smallmatrix}\right) \in \Omega \\ \infty & \text{otherwise.} \end{cases}$$

Next, for a given fraction $\frac{c_1}{c_3}$ with $c_1 \leq 0$ and $c_3 > 0$, we wish to better understand the region

$$\mathcal{R}_{-c_1/c_3} := \left\{ p_{a,b}m_iH \in \Omega_M : R_M(p_{a,b}m_iH) = \frac{1}{a(b + c_1a/c_3)} \right\}.$$

In particular, we want to prove the following technical result, which is a great aid in showing that $\mu_{\Omega_M}\left(\partial\left(R_M^{-1}([0,\xi])\right)\right)=0$ for $\xi\geq 0$ and that $\mathcal{A}'_M$ is continuous and piecewise real analytic.

Proposition 2. *For each $i\in\{1,\ldots,k\}$, $\mathcal{R}_{-c_1/c_3}\cap Pm_iH/H$ is either empty or a polygon when we view Pm_iH/H as the Farey triangle Ω.*

Proof. We first examine the region

$$\mathcal{R}_C:=\left\{p_{a,b}m_iH\in\Omega_M:(ac_1+bc_3,ac_2+bc_4)\in\Omega, C\in\bigcup_{j=1}^k m_iHm_j^{-1}\right\}$$

for a given $C=\left(\begin{smallmatrix}c_1&c_2\\c_3&c_4\end{smallmatrix}\right)\in\bigcup_{i,j=1}^k m_iHm_j^{-1}$. Note that $\mathcal{R}_{-c_1/c_3}$ is a subset of the union of all $R_{C'}$ such that $C'\in\bigcup_{i,j=1}^k m_iHm_j^{-1}$ is a matrix having $\left(\begin{smallmatrix}c_1\\c_3\end{smallmatrix}\right)$ as its first column. If $c_4=0$, then $C=\left(\begin{smallmatrix}c_1&-1\\1&0\end{smallmatrix}\right)$ and for a given index i, $\mathcal{R}_C\cap Pm_iH/H$ is a subset of $\{p_{a,b}m_iH:(ac_1+b,-a)\in\Omega\}=\emptyset$. Thus $\mathcal{R}_C$ is empty in this case. Next, suppose $c_4\leq -1$. Then $\mathcal{R}_C\cap Pm_iH/H$ consists of the elements $p_{a,b}m_iH$ which must satisfy

$$-\frac{c_1}{c_3}a<b\leq\frac{1-c_1a}{c_3}\quad\text{and}\quad\frac{1-c_2a}{c_4}\leq b<-\frac{c_2}{c_4}a.$$

However, we have $c_1c_4-c_2c_3=1$, which implies that $\frac{c_1}{c_3}-\frac{c_2}{c_4}=(c_3c_4)^{-1}<0$, and hence $-\frac{c_2}{c_4}<-\frac{c_1}{c_3}$. Thus the above conditions on $p_{a,b}m_iH$ cannot be satisfied, and therefore $\mathcal{R}_C=\emptyset$. So $\mathcal{R}_C$ is nonempty only when $c_1\leq 0$ and $c_3,c_4\geq 1$, which then implies that $c_2=\frac{c_1c_4-1}{c_3}<0$.

As a result of our work above, we reset notation so that $C=\left(\begin{smallmatrix}-c_1&-c_2\\c_3&c_4\end{smallmatrix}\right)$ and we assume that $c_1\geq 0$ and $c_2,c_3,c_4\geq 1$. Now a point $(a,b)\in\Omega$ satisfies $(-ac_1+bc_3,-ac_2+bc_4)\in\Omega$ if and only if

$$\frac{c_1}{c_3}a<b\leq\frac{1+c_1a}{c_3},\quad\frac{c_2}{c_4}a<b\leq\frac{1+c_2a}{c_4},\quad\text{and}\quad b>\frac{1+(c_1+c_2)a}{c_3+c_4}.$$

Since $\frac{c_1}{c_3}a<\frac{c_2}{c_4}a\leq\frac{1+(c_1+c_2)a}{c_3+c_4}$ for $a\in(0,1]$, the above conditions reduce to $\frac{1+(c_1+c_2)a}{c_3+c_4}<b\leq\min\left\{\frac{1+c_1a}{c_3},\frac{1+c_2a}{c_4}\right\}$. We therefore have

$$\mathcal{R}_C=\left\{p_{a,b}m_iH\in\Omega_M:\begin{array}{c}\frac{1+(c_1+c_2)a}{c_3+c_4}<b\leq\min\left\{\frac{1+c_1a}{c_3},\frac{1+c_2a}{c_4}\right\}\\ C\in\bigcup_{j=1}^k m_iHm_j^{-1}\end{array}\right\}.$$

Since M is closed under left multiplication by $\left(\begin{smallmatrix}1 & -1\\ 0 & 1\end{smallmatrix}\right)$, $\bigcup_{j=1}^{k} m_i H m_j^{-1}$ is closed under right multiplication by $\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)$ for each $i \in \{1, \ldots, k\}$. So if $C \in \bigcup_{j=1}^{k} m_i H m_j^{-1}$, then for every $n \in \mathbb{N}$,

$$C \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -c_1 & -nc_1 - c_2 \\ c_3 & nc_3 + c_4 \end{pmatrix} \in \bigcup_{j=1}^{k} m_i H m_j^{-1}.$$

We have

$$\mathcal{R}_{C\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right)} = \left\{ p_{a,b} m_i H \in \Omega_M : \begin{array}{c} \frac{1+((n+1)c_1+c_2)a}{(n+1)c_3+c_4} < b \leq \frac{1+(nc_1+c_2)a}{nc_3+c_4} \\ C \in \bigcup_{j=1}^{k} m_i H m_j^{-1} \end{array} \right\},$$

noting that $\frac{1+(nc_1+c_2)a}{nc_3+c_4} < \frac{1+c_3 a}{c_4}$. Thus for each i such that $C \in \bigcup_{j=1}^{k} m_i H m_j^{-1}$, the regions $\left\{\mathcal{R}_{C\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right)} \cap P m_i H/H : n \in \mathbb{N}\right\}$ paste together to form

$$\left\{ p_{a,b} m_i H \in P m_i H/H : \frac{c_1}{c_3} a < b \leq \min\left\{ \frac{1+c_1 a}{c_3}, \frac{1+c_2 a}{c_4} \right\} \right\}$$

since the sequence $\left(\frac{1+(nc_1+c_2)a}{nc_3+c_4}\right)_n$ decreases to $\frac{c_1}{c_3} a$ for each $a \in (0,1]$. Therefore, if we assume that $\frac{c_{2,i}}{c_{4,i}}$ is the largest fraction such that $\left(\begin{smallmatrix}-c_1 & -c_{2,i}\\ c_3 & c_{4,i}\end{smallmatrix}\right) \in \bigcup_{j=1}^{k} m_i H m_j^{-1}$, then

$$\mathcal{R}_{c_1/c_3} \cap P m_i H/H \subseteq \left\{ p_{a,b} m_i H : \frac{c_1}{c_3} a < b \leq \min\left\{ \frac{1+c_1 a}{c_3}, \frac{1+c_{2,i} a}{c_{4,i}} \right\} \right\}.$$

(Every candidate for $\frac{c_{2,i}}{c_{4,i}}$ is a fraction immediately succeeding $\frac{c_1}{c_3}$ in $\mathcal{F}(Q)$ for some $Q \in \mathbb{N}$, and thus a largest such fraction exists.) We use $\mathcal{R}^{(i)}_{c_1/c_3}$ to denote the set on the right for each $c_1 \geq 0$, $c_3 \geq 1$, and $i \in \{1, \ldots, k\}$ such that $\left(\begin{smallmatrix}-c_1\\ c_3\end{smallmatrix}\right)$ is the first column of a matrix in $\bigcup_{j=1}^{k} m_i H m_j^{-1}$, and $\frac{c_{2,i}}{c_{4,i}}$ ($c_{2,i}, c_{4,i} \geq 1$) is the largest fraction with $\left(\begin{smallmatrix}-c_1 & -c_{2,i}\\ c_3 & c_{4,i}\end{smallmatrix}\right) \in \bigcup_{j=1}^{k} m_i H m_j^{-1}$. If c_1, c_3, and i do not satisfy these conditions, we let $\mathcal{R}^{(i)}_{c_1/c_3} = \emptyset$. Then for all c_1 and c_3, we have

$$\mathcal{R}_{c_1/c_3} = \bigcup_{i=1}^{k} \left(\mathcal{R}^{(i)}_{c_1/c_3} \setminus \bigcup_{\substack{s \in \mathbb{Q} \\ 0 \leq s < c_1/c_3}} \mathcal{R}^{(i)}_s \right).$$

Assume that c_1, c_3, and i are such that $\mathcal{R}^{(i)}_{c_1/c_3} \neq \emptyset$. In order to show that $\mathcal{R}_{c_1/c_3} \cap P m_i H/H$ is either empty or a polygon, it is sufficient to prove that $\mathcal{R}_{c_1/c_3} \cap P m_i H/H$ can be written in the form $\mathcal{R}^{(i)}_{c_1/c_3} \setminus \left(\bigcup_{\ell=1}^{n} \mathcal{R}^{(i)}_{s_\ell} \right)$ for some $s_\ell \in \mathbb{Q}$. Note that if $\frac{c_1}{c_3} = 0$, then $\mathcal{R}_{c_1/c_3} \cap Pm_iH/H = \mathcal{R}^{(i)}_0$, which is a triangle if nonempty. So assume that $\frac{c_1}{c_3} > 0$. We now consider two cases.

Case 1. There exists $s \in \mathbb{Q}$ with $s < \frac{c_1}{c_3}$ such that $\mathcal{R}^{(i)}_s$ contains the lower-right border $\left\{ p_{a,b}m_iH : b = \frac{c_1}{c_3}a \right\}$ of $\mathcal{R}^{(i)}_{c_1/c_3}$ in its interior. Then clearly there exists $N \in \mathbb{N}$ such that

$$\mathcal{R}^{(i)}_{c_1/c_3} \setminus \mathcal{R}^{(i)}_s \subseteq \left\{ p_{a,b}m_iH : \frac{1}{N} + \frac{c_1}{c_3}a < b \leq \frac{1 + c_1 a}{c_3} \right\}.$$

If there is another $\frac{c_1'}{c_3'} \in \mathbb{Q}$ with $\frac{c_1'}{c_3'} < \frac{c_1}{c_3}$ such that $\mathcal{R}^{(i)}_{c_1'/c_3'}$ intersects $\mathcal{R}^{(i)}_{c_1/c_3} \setminus \mathcal{R}^{(i)}_s$, then there exists $(a,b) \in \Omega$ such that $\frac{1}{N} + \frac{c_1}{c_3}a < b \leq \frac{1+c_1'a}{c_3'}$. This implies that $\left(\frac{c_1}{c_3} - \frac{c_1'}{c_3'} \right) a < \frac{1}{c_3'} - \frac{1}{N}$, which can only hold if $c_3' < N$. Thus there are finitely many $s' \in \mathbb{Q}$ such that $\mathcal{R}^{(i)}_{s'}$ intersects $\mathcal{R}^{(i)}_{c_1/c_3} \setminus \mathcal{R}^{(i)}_s$. So

$$\mathcal{R}_{c_1/c_3} \cap Pm_iH/H = \mathcal{R}^{(i)}_{c_1/c_3} \setminus \left(\mathcal{R}^{(i)}_s \cup \bigcup_{\ell=1}^{n} \mathcal{R}^{(i)}_{s_\ell} \right)$$

for some $s_\ell \in \mathbb{Q}$, completing the proof that $\mathcal{R}_{c_1/c_3} \cap Pm_iH/H$ is a polygon in this case.

Case 2. There exists no $s \in \mathbb{Q}$ with $s < \frac{c_1}{c_3}$ such that $\mathcal{R}^{(i)}_s$ contains the lower-right border of $\mathcal{R}^{(i)}_{c_1/c_3}$ in its interior. Let $n_1 \geq 0$ and $n_2 \geq 1$ be integers such that $\left\{ p_{a,b}m_iH : b = \frac{1+n_1a}{n_2} \right\}$ is an upper border for $\mathcal{R}^{(i)}_s$ for some $s \in \mathbb{Q}$ with $s < \frac{c_1}{c_3}$. If $\frac{n_1}{n_2} = \frac{c_1}{c_3}$, then the line $\left\{ (a,b) \in \mathbb{R}^2 : b = \frac{1+n_1a}{n_2} \right\}$ is above and parallel to $\left\{ (a,b) \in \mathbb{R}^2 : b = \frac{c_1}{c_3}a \right\}$, in which case the lower-right border of $\mathcal{R}^{(i)}_{c_1/c_3}$ is contained in the interior of $\mathcal{R}^{(i)}_s$, a contradiction. So $\frac{n_1}{n_2} \neq \frac{c_1}{c_3}$, and $\left\{ (a,b) \in \mathbb{R}^2 : b = \frac{1+n_1a}{n_2} \right\}$ intersects $\left\{ (a,b) \in \mathbb{R}^2 : b = \frac{c_1}{c_3}a \right\}$ at the point $\left(\frac{c_3}{n_2c_1 - n_1c_3}, \frac{c_1}{n_2c_1 - n_1c_3} \right)$. If $\frac{n_1}{n_2} > \frac{c_1}{c_3}$, then since $\frac{c_3}{n_2c_1 - n_1c_3} < 0$ and the slope of $b = \frac{1+n_1a}{n_2}$ is greater than that of $b = \frac{c_1}{c_3}a$, the lower-right border of $\mathcal{R}^{(i)}_{c_1/c_3}$ is again contained in the interior of $\mathcal{R}^{(i)}_s$, another contradiction.

So $\frac{n_1}{n_2} < \frac{c_1}{c_3}$, and furthermore, the intersection point $\left(\frac{c_3}{n_2c_1 - n_1c_3}, \frac{c_1}{n_2c_1 - n_1c_3} \right)$ cannot be above or to the right of Ω. If the intersection point is in the interior of Ω or is on or below its border $b = 1 - a$, then $\mathcal{R}^{(i)}_{c_1/c_3} \setminus \mathcal{R}^{(i)}_s$ contains a set of the form

$$\left\{ p_{a,b}m_iH : b - \frac{c_1}{c_3}a \in (0, \epsilon), a \in (w_1, w_2) \right\}$$

where $\epsilon, w_1, w_2 \in (0,1)$ and $w_1 < w_2$. Since $\frac{1}{a(b - c_1a/c_3)}$ is not μ_{Ω_M}-integrable over the above set for any $\epsilon, w_1, w_2 \in (0,1)$ with $w_1 < w_2$ and R_M is μ_{Ω_M}-integrable, the lower-right border of $\mathcal{R}^{(i)}_{c_1/c_3}$ is

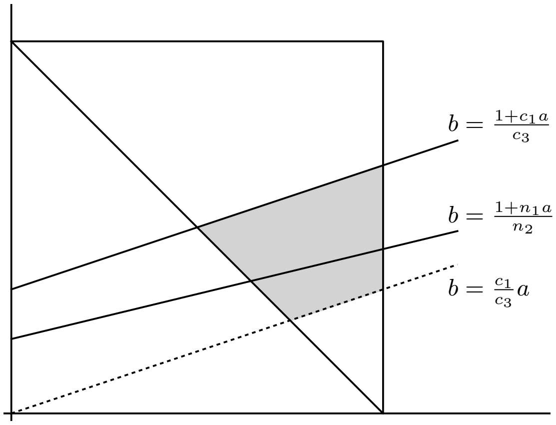

Figure 2.1: $\mathcal{R}_{c_1/c_3}$ in Case 1

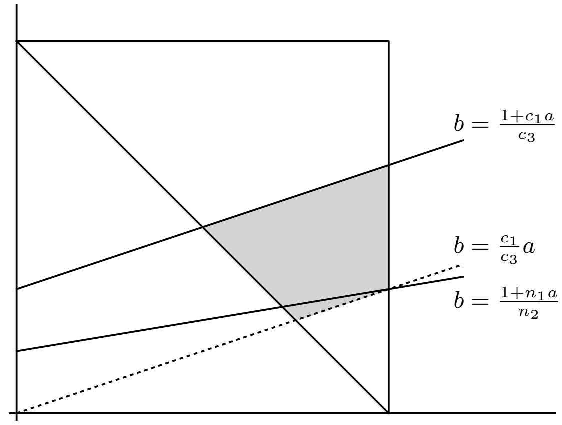

Figure 2.2: $\mathcal{R}_{c_1/c_3}$ in Case 2

contained in $\bigcup_{s\in\mathbb{Q},0\le s<c_1/c_3} \mathcal{R}_s^{(i)}$. Since the set $\left\{\left(\frac{c_3}{n}, \frac{c_1}{n}\right) : n \in \mathbb{N}\right\} \cap \Omega$ of possible intersection points in Ω of $b = \frac{c_1}{c_3}a$ with a line $b = \frac{1+n_1 a}{n_2}$ corresponding to the upper-left border of a set $\mathcal{R}_s^{(i)}$ is finite, there must be some $s \in \mathbb{Q}$ with $s < \frac{c_1}{c_3}$ such that $\mathcal{R}_s^{(i)}$ contains the lower-right border of $\mathcal{R}_{c_1/c_3}^{(i)}$, and in this case, the line determining the upper-left border of $\mathcal{R}_s^{(i)}$, say $b = \frac{1+n_1 a}{n_2}$, intersects $b = \frac{c_1}{c_3}a$ at the border of Ω at the line $a = 1$ or $b = 1$.

Now let $s' \in \mathbb{Q}$ with $s' < \frac{c_1}{c_3}$ such that the upper-left border of $\mathcal{R}_{s'}^{(i)}$ is determined by $b = \frac{1+n_1' a}{n_2'}$. Since $\mathcal{R}_{s'}^{(i)}$ does not contain the lower-right border of $\mathcal{R}_{c_1/c_3}^{(i)}$, $\frac{n_1'}{n_2'} < \frac{c_1}{c_3}$ and $b = \frac{1+n_1' a}{n_2'}$ intersects $b = \frac{c_1}{c_3}a$ at or to the left of the intersection point of $b = \frac{1+n_1 a}{n_2}$ with $b = \frac{c_1}{c_3}a$. If $\frac{n_1'}{n_2'} > \frac{n_1}{n_2}$, then it is clear that the line $b = \frac{1+n_1' a}{n_2'}$ passes under the set $\left\{(a,b) \in \Omega : b > \max\left\{\frac{1+n_1 a}{n_2}, \frac{c_1}{c_3}a\right\}\right\}$, and thus $\mathcal{R}_{s'}^{(i)}$ does not intersect $\mathcal{R}_{c_1/c_3}^{(i)} \backslash \mathcal{R}_s^{(i)}$. If $\frac{n_1'}{n_2'} < \frac{n_1}{n_2}$ and $\mathcal{R}_{s'}^{(i)}$ does intersect $\mathcal{R}_{c_1/c_3}^{(i)} \backslash \mathcal{R}_s^{(i)}$, then there exists $(a,b) \in \Omega$ such that $\frac{1+n_1 a}{n_2} < b \le \frac{1+n_1' a}{n_2'}$, implying that $\left(\frac{n_1}{n_2} - \frac{n_1'}{n_2'}\right) a < \frac{1}{n_2'} - \frac{1}{n_2}$. This inequality holds only if $n_2' < n_2$. This shows that there are finitely many $s' \in \mathbb{Q}$ such that $\mathcal{R}_{s'}^{(i)}$ intersects $\mathcal{R}_{c_1/c_3}^{(i)} \backslash \mathcal{R}_s^{(i)}$, and thus completes the proof that $\mathcal{R}_{c_1/c_3} \cap Pm_iH/H$ is a polygon.

$\square$

So by our work above, we have proven that R_M is a piecewise rational function on Ω_M, and for a given $\frac{c_1}{c_3} \in \mathbb{Q}$, the region $\mathcal{R}_{c_1/c_3}$ over which $R_M(p_{a,b} m_i H) = \frac{1}{a(b - c_1 a/c_3)}$ is either empty or a union of polygons, one polygon being in each component Pm_iH/H such that $\mathcal{R}_{c_1/c_3} \cap Pm_iH/H \neq \emptyset$. Note also that if $C = \begin{pmatrix} -c_1 & -c_2 \\ c_3 & c_4 \end{pmatrix} \in m_i H m_j^{-1}$, then the restriction of the return map r_M to the polygon $\mathcal{R}_{c_1/c_3} \cap \mathcal{R}_C \cap Pm_iH/H$ is given by

$$r_M(p_{a,b}m_iH) = p_{-c_1a+c_3b,-c_2a+c_4b}m_jH.$$

Therefore, each component Pm_iH/H of Ω_M can be divided into a countable number of polygons $\mathcal{P}$ such that r_M maps each $\mathcal{P}$ linearly from Pm_iH/H to Pm_jH/H, for some j depending on $\mathcal{P}$.

2.5.2 The boundary of $R_M^{-1}([0,\xi])$ has measure 0

Next, we prove that for a given $\xi > 0$, the boundary of $R_M^{-1}([0,\xi])$ has μ_{Ω_M}-measure 0. First notice that R_M is continuous μ_{Ω_M}-a.e. Because the set $\overline{R_M^{-1}([0,\xi])} \backslash R_M^{-1}([0,\xi])$ contains only points of discontinuity of R_M, it has μ_{Ω_M}-measure 0. Now for a given $s \in \mathbb{Q}$, define $f_s : \Omega_M \to [0,\infty]$ so that

$$f_s(p_{a,b}m_iH) = \begin{cases} \frac{1}{a(b-sa)} & \text{for all } p_{a,b}m_iH \in R_s \\ \infty & \text{otherwise.} \end{cases}$$

We then have

$$\begin{aligned} R_M^{-1}([0,\xi]) \backslash \left(R_M^{-1}([0,\xi])\right)^o &= \bigcup_{s\in\mathbb{Q}} f_s^{-1}([0,\xi]) \backslash \left(\bigcup_{s\in\mathbb{Q}} f_s^{-1}([0,\xi])\right)^o \\ &\subseteq \bigcup_{s\in\mathbb{Q}} f_s^{-1}([0,\xi]) \backslash \left(\bigcup_{s\in\mathbb{Q}} (f_s^{-1}([0,\xi]))^o\right) \\ &\subseteq \bigcup_{s\in\mathbb{Q}} \left(\left(f_s^{-1}([0,\xi])\right) \backslash \left(f_s^{-1}([0,\xi])\right)^o\right). \end{aligned}$$

Each set $f_s^{-1}([0,\xi])$ is either empty or a finite union of sets of the form

$$\left\{p_{a,b}m_iH \in Pm_iH : \frac{c_1}{c_3}a < b \leq \min\left\{\frac{1+c_1a}{c_3}, \frac{1+c_{2,i}a}{c_{4,i}}\right\}, b \geq \frac{c_1}{c_3}a + \frac{1}{\xi a}\right\}.$$

So $f_s^{-1}([0,\xi])\backslash\left(f_s^{-1}([0,\xi])\right)^o$ is essentially a set of finitely many line and curve segments for each $s\in\mathbb{Q}$. Hence it is clear that $R_M^{-1}([0,\xi])\backslash\left(R_M^{-1}([0,\xi])\right)^o$ is of μ_{Ω_M}-measure 0. As a consequence,

$$\begin{aligned}\partial\left(R_M^{-1}([0,\xi])\right) &= \overline{R_M^{-1}([0,\xi])}\backslash\left(R_M^{-1}([0,\xi])\right)^o \\ &= \left(\overline{R_M^{-1}([0,\xi])}\backslash R_M^{-1}([0,\xi])\right)\cup\left(R_M^{-1}([0,\xi])\backslash\left(R_M^{-1}([0,\xi])\right)^o\right)\end{aligned}$$

has μ_{Ω_M}-measure 0. This, along with Section 2.6, proves the existence of the limiting gap measure $\nu_{I,M}$.

2.5.3 $\mathcal{A}'_M$ is continuous and piecewise real analytic

Next, we want to prove that the function $\mathcal{A}_M$ has a continuous, piecewise real analytic derivative. Note first that since $R_M^{-1}([0,\xi])$ is the disjoint union of the sets $f_s^{-1}([0,\xi])$, we have

$$\mathcal{A}_M(\xi)=\sum_{s\in\mathbb{Q}}\mu_{\Omega_M}\left(f_s^{-1}([0,\xi])\right).$$

So it suffices to show that $\xi\mapsto\mu_{\Omega_M}\left(f_s^{-1}([0,\xi])\right)$ has a continuous, piecewise real analytic derivative for every $s\in\mathbb{Q}$, and that for a given $\xi>0$, there are at most finitely many $s\in\mathbb{Q}$ for which $f_s^{-1}([0,\xi])$ is nonempty.

The first claim is easy to see. Indeed, by triangulating the polygons that make up the region $\mathcal{R}_s$, we can write $\xi\mapsto\mu_{\Omega_M}\left(f_s^{-1}([0,\xi])\right)$ as a finite sum of functions $\mathcal{A}_{\mathcal{T}}:[0,\infty)\to\mathbb{R}$ defined by

$$\mathcal{A}_{\mathcal{T}}(\xi)=\frac{1}{\#M}m\left(\left\{(a,b)\in\mathcal{T}:b\geq sa+\frac{1}{\xi a}\right\}\right),$$

where $\mathcal{T}\subseteq\Omega$ is a triangle and as in Section 2.2, $dm=2\,da\,db$. It is then straightforward to show that each function $\mathcal{A}_{\mathcal{T}}$ has a continuous, piecewise real analytic derivative, implying that $\xi\mapsto\mu_{\Omega_M}\left(f_s^{-1}([0,\xi])\right)$ has the same property.

To prove the latter assertion, let $\xi>0$ be given and suppose that there exists $\frac{c_1}{c_3}\in\mathbb{Q}$ such that $f_{c_1/c_3}^{-1}([0,\xi])$ is nonempty (assume $c_1\geq 0$ and $c_3\geq 1$). Then there is some index i such that

$f^{-1}_{c_1/c_3}([0,\xi]) \cap Pm_iH/H \neq \emptyset$. We have

$$f^{-1}_{c_1/c_3}([0,\xi]) \cap Pm_iH/H \subseteq \left\{p_{a,b}m_iH : \frac{c_1}{c_3}a < b \leq \frac{1+c_1a}{c_3}, b \geq \frac{c_1}{c_3}a + \frac{1}{\xi a}\right\},$$

and hence the latter set is nonempty. So there exists $p_{a,b}m_iH \in Pm_iH/H$ such that $\frac{c_1}{c_3}a + \frac{1}{\xi a} \leq b \leq \frac{1+c_1a}{c_3}$, which implies that $\xi \geq \frac{c_3}{a}$. We have

$$\sup\left\{a \in \mathbb{R} : (a,b) \in \Omega, \frac{c_1}{c_3}a < b \leq \frac{1+c_1a}{c_3}\right\} = \begin{cases} 1 & \text{if } \frac{c_1}{c_3} \leq 1 \\ \frac{c_3}{c_1} & \text{if } \frac{c_1}{c_3} > 1 \end{cases},$$

and therefore $\xi \geq c_3$ if $\frac{c_1}{c_3} \leq 1$ and $\xi \geq c_1$ if $\frac{c_1}{c_3} > 1$; i.e., $\xi \geq \max\{c_1, c_3\}$. There are clearly finitely many positive fractions $\frac{c_1}{c_3}$ satisfying this condition, and thus satisfying $f^{-1}_{c_1/c_3}([0,\xi]) \neq \emptyset$. This completes the proof that $\mathcal{A}'_M$ is continuous and piecewise real analytic.

One more property of R_M we wish to mention is that for any $\epsilon > 0$, there exist bounded continuous functions $\psi_{1,\epsilon}, \psi_{2,\epsilon}, \psi_{3,\epsilon} : \Omega_M \to [0,\infty)$ such that $\psi_{1,\epsilon} \leq R_M$, $\psi_{2,\epsilon} \leq \frac{1}{R_M} \leq \psi_{3,\epsilon}$, and

$$\int_{\Omega_M} (R_M - \psi_{1,\epsilon})\, d\mu_{\Omega_H}, \int_{\Omega_M} \left(\frac{1}{R_M} - \psi_{2,\epsilon}\right) d\mu_{\Omega_H}, \int_{\Omega_M} \left(\psi_{3,\epsilon} - \frac{1}{R_M}\right) d\mu_{\Omega_H} < \epsilon.$$

This follows easily from the properties of R_M proven in Section 2.5.1, and the fact that $\frac{1}{R_M} \leq \frac{1}{R} \leq 1$.

2.5.4 The h-spacings and numerators of differences of $(\mathcal{F}_{I,M}(Q))_Q$

In this section, we briefly see how our work through Section 2.6 yields results on the limiting distribution of the h-spacings and the numerators of differences of $(\mathcal{F}_{I,M}(Q))_Q$. First, from the equality $(R_M \circ r_M^{j-1})(W_{H,Q}(\beta_i)) = Q^2(\beta_{i+j} - \beta_{i+j-1})$, $i \in \{0,\ldots,N_{I,M}(Q)-h\}$, we have

$$\frac{\#\{\beta_i \in \mathcal{F}_{I,M}(Q) : Q^2\mathbf{v}_{\mathcal{F}_{I,M}(Q),i,h} \in \prod_{j=1}^h [0,\xi_j]\} + O(h)}{N_{I,M}(Q)} = \rho_{Q,I,M}\left(\bigcap_{j=1}^h (R_M \circ r_M^{j-1})^{-1}([0,\xi_j])\right).$$

(Recall the notation $\mathbf{v}_{\mathcal{F}_{I,M}(Q),i,h}$ from Section 1.1.) For a given $j \in \mathbb{N}$, the function $R_M \circ r_M^j$, like R_M itself, is piecewise rational and the domain on which $R_M \circ r_M^j$ is defined by a given rational function is a union of polygons. Indeed, we have seen in Section 2.5.1 that Ω_M can be divided into a countable number of polygons on each of which r_M is linear, which implies that the same property holds for r_M^j for any $j \in \mathbb{N}$. (This is in fact true for all $j \in \mathbb{Z}$ since r_M is invertible.) This, together with the piecewise rationality of R_M on polygons implies the same for $R_M \circ r_M^j$. We also note that the level sets of the rational functions defining $R_M \circ r_M^j$ are hyperbolas since the same is true for R_M and r_M^j is linear. As a consequence, the sets $\left(R_M \circ r_M^j\right)^{-1}([0, \xi_j])$ have boundaries of measure 0. Hence, by our work in Section 2.6, the limiting h-spacing measure $\nu_{I,M,h}$ of $(\mathcal{F}_{I,M}(Q))_Q$ exists and satisfies

$$\nu_{I,M,h}\left(\prod_{j=1}^{h}[0,\xi_j]\right) = \mu_{\Omega_M}\left(\bigcap_{j=1}^{h}\left(R_M \circ r_M^{j-1}\right)^{-1}\left(\left[0, \frac{\pi^2[\Gamma : H]}{3(\#M)}\xi_j\right]\right)\right).$$

Lastly, note that if $\frac{a}{q} < \frac{b}{p}$ are consecutive elements in $\mathcal{F}_{I,M}(Q)$, $\frac{a'}{q'}$ succeeds $\frac{a}{q}$ in $\mathcal{F}(Q)$, and $W_{H,Q}\left(\frac{a}{q}\right) \in \mathcal{R}_{c_1/c_3}$, then $p = -c_1 q + c_3 q'$ and

$$bq - ap = qp\left(\frac{b}{p} - \frac{a}{q}\right) = \frac{q(-c_1 q + c_3 q')}{Q^2} R_M\left(W_{H,Q}\left(\frac{a}{q}\right)\right) = \frac{q(-c_1 q + c_3 q')}{Q^2}\frac{Q^2}{q(q' - \frac{c_1}{c_3}q)} = c_3.$$

Using this fact and Section 2.6, one can show that for every $c_3 \in \mathbb{N}$,

$$\lim_{Q\to\infty}\frac{\#\{\frac{a}{q} < \frac{b}{p} \text{ consecutive in } \mathcal{F}_{I,M}(Q) : bq - ap = c_3\}}{N_{I,M}(Q)} = \mu_{\Omega_M}\left(\bigcup_{\substack{c_1 \in \mathbb{N}\cup\{0\} \\ \gcd(c_1,c_3)=1}} \mathcal{R}_{c_1/c_3}\right),$$

recovering a result Badziahin and Haynes proved for the sequence $(\mathcal{F}_{Q,d})_Q$ in [10].

2.6 The convergence $\rho_{Q,I,M} \to \mu_{\Omega_M}$

In this section, we prove the weak convergence $\rho_{Q,I,M} \to \mu_{\Omega_M}$, and hence complete the proof of Theorem 1. We first consider the measures $\left(\rho^R_{Q,I,M}\right)_Q$ on G/H defined by

$$d\rho^R_{Q,I,M} = \frac{N_{I,M}(Q)}{Q^2}\, d\rho_{Q,I,M}\, ds.$$

In other words, $\rho^R_{Q,I,M}$ is a measure concentrated on segments of the horocycle flow connecting $W_{H,Q}(\beta_i)$ to $W_{H,Q}(\beta_{i+1})$ for $0 \le i \le N_{I,M}(Q) - 1$. These segments connect to give one segment from $W_{H,Q}(\beta_0)$ to $W_{H,Q}(\beta_{N_{I,M}(Q)})$. So for a bounded, measurable function $f : G/H \to \mathbb{R}$, we have

$$\int f \, d\rho^R_{Q,I,M} = \frac{1}{Q^2} \int_{Q^2\beta_0}^{Q^2\beta_{N_{I,M}(Q)}} f\left(\begin{pmatrix} 1 & 0 \\ -s & 1 \end{pmatrix} W_{H,Q}(0)\right) ds, \tag{2.8}$$

noting that $h_{Q^2\beta_i} W_{H,Q}(0) = W_{H,Q}(\beta_i)$. We wish to show that the sequence $(\rho^R_{Q,I,M})_Q$ converges weakly to $\frac{|I|\mu_{G/H}}{\mu_{G/H}(G/H)}$. Notice that

$$\begin{pmatrix} 1 & 0 \\ -s & 1 \end{pmatrix} W_{H,Q}(0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & Q^{-1} \end{pmatrix} \begin{pmatrix} 1 & \frac{s}{Q^2} \\ 0 & 1 \end{pmatrix} H.$$

So (2.8) can be written as

$$\int f \, d\rho^R_{Q,I,M} = \int_{\beta_0}^{\beta_{N_{I,M}(Q)}} (\tilde{f} \circ \tilde{g}_Q)\left(\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} H\right) dt,$$

where $\tilde{f} : G/H \to \mathbb{R}$ is the composition of left multiplication on G/H by $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ followed by f, and $\tilde{g}_Q : G/H \to G/H$ is left multiplication by $\left(\begin{smallmatrix} Q & 0 \\ 0 & Q^{-1} \end{smallmatrix}\right)$. Since $\bigcup_{Q\in\mathbb{N}} \mathcal{F}_M(Q)$ is dense in $[0,1]$, $\beta_0 \to t_1$ and $\beta_{N_{I,M}(Q)} \to t_2$ as $Q \to \infty$. So if we define the measure $\rho^{R'}_{Q,I,M}$ on G/H such that

$$\int f \, d\rho^{R'}_{Q,I,M} = \int_{t_1}^{t_2} (\tilde{f} \circ \tilde{g}_Q)\left(\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} H\right) dt,$$

for all bounded, measurable functions $f : G/H \to \mathbb{R}$, then it is clear that $\rho^R_{Q,I,M} - \rho^{R'}_{Q,I,M} \to 0$ weakly. Thus to show that $\rho^R_{Q,I,M} \to \frac{|I|\mu_{G/H}}{\mu_{G/H}(G/H)}$ weakly, it suffices to prove that $\rho^{R'}_{Q,I,M} \to \frac{|I|\mu_{G/H}}{\mu_{G/H}(G/H)}$ weakly.

Like the convergence $\rho^R_{Q,I} \to \mu_{G/\Gamma}$, this is a consequence of the equidistribution of closed horocycles in G/H. In particular, the argument for [21, Theorem 7] can be used to prove

$$\lim_{Q\to\infty} \int_{t_1}^{t_2} (f \circ \tilde{g}_Q)\left(\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} H\right) dt = \frac{t_2 - t_1}{\mu_{G/H}(G/H)} \int_{G/H} f \, d\mu_{G/H} \tag{2.9}$$

for all functions $f : G/H \to \mathbb{R}$ that are bounded and uniformly continuous.

We give this argument in detail. Let $f : G/H \to \mathbb{R}$ be bounded and uniformly continuous. For $\theta, t, y \in \mathbb{R}$ with $y > 0$, we make the following definitions:

$$k_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad a_y = \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}, \quad u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

By the uniform continuity of f, for a given $\epsilon > 0$, there is a $\delta \in (0,1)$ such that if $|\theta|, |y-1| \leq \delta$, then

$$|(f \circ \tilde{g}_Q)(k_\theta a_y u_t H) - (f \circ \tilde{g}_Q)(u_t H)| < \epsilon$$

for all $Q \in \mathbb{N}$ and $t \in [0,1]$. So if we let

$$B = \{k_\theta a_y u_t H : \theta \in [0,\delta], t \in [t_1, t_2], y \in [1-\delta, 1]\}$$

and define $\overline{f \circ \tilde{g}_Q} : B \to \mathbb{R}$ by $\overline{f \circ \tilde{g}_Q}(k_\theta a_y u_t H) = f \circ \tilde{g}_Q(u_t H)$, then

$$\left| \int_B \overline{f \circ \tilde{g}_Q} \, d\mu_{G/H} - \int_B f \circ \tilde{g}_Q \, d\mu_{G/H} \right| \leq \epsilon \cdot \mu_{G/H}(B)$$

for all $Q \in \mathbb{N}$. Now notice that

$$\begin{aligned} \int_B \overline{f \circ \tilde{g}_Q} \, d\mu_{G/H} &= \int_0^\delta \int_{1-\delta}^1 \int_{t_1}^{t_2} \overline{f \circ \tilde{g}_Q} \, (k_\theta a_y u_t H) \, dt \, dy \, d\theta \\ &= \frac{\mu_{G/H}(B)}{t_2 - t_1} \int_{t_1}^{t_2} (f \circ \tilde{g}_Q) \, (u_t H) \, dt, \end{aligned}$$

and so

$$\left| \int_{t_1}^{t_2} (f \circ \tilde{g}_Q) \, (u_t H) \, dt - \frac{t_2 - t_1}{\mu_{G/H}(B)} \int_B f \circ \tilde{g}_Q \, d\mu_{G/H} \right| \leq \epsilon (t_2 - t_1). \tag{2.10}$$

By the Howe-Moore theorem [37], the geodesic flow $\{\tilde{g}_s : s > 0\}$ is mixing on G/H, and so

$$\lim_{Q \to \infty} \int_B f \circ \tilde{g}_Q \, d\mu_{G/H} = \frac{\mu_{G/H}(B)}{\mu_{G/H}(G/H)} \int_{G/H} f \, d\mu_{G/H}.$$

Hence by (2.10), we have

$$\lim_{Q\to\infty}\int_{t_1}^{t_2}(f\circ\tilde{g}_Q)\,(u_tH)\,dt=\frac{t_2-t_1}{\mu_{G/H}(G/H)}\int_{G/H}f\,d\mu_{G/H},$$

noting that $\epsilon>0$ was chosen arbitrarily. This completes the proof of (2.9).

Next, since $\mu_{G/H}$ is left G-invariant, we have

$$\lim_{Q\to\infty}\int f\,d\rho^{R'}_{Q,I,M}=\frac{|I|}{\mu_{G/H}(G/H)}\int_{G/H}\tilde{f}\,d\mu_{G/H}=\frac{|I|}{\mu_{G/H}(G/H)}\int_{G/H}f\,d\mu_{G/H},$$

for every bounded, uniformly continuous function $f:G/H\to\mathbb{R}$. By the Portmanteau theorem, this is equivalent to saying that $\rho^{R'}_{Q,I,M}\to\frac{|I|\mu_{G/H}}{\mu_{G/H}(G/H)}$ weakly, which then implies that $\rho^{R}_{Q,I,M}\to\frac{|I|\mu_{G/H}}{\mu_{G/H}(G/H)}$ weakly.

Our next step is to prove that if $\pi_M:G/H\to\Omega_M$ is the projection $(x,s)\mapsto x$, where we are viewing G/H as $\{(x,s)\in\Omega_M\times\mathbb{R}:0\le s<R_M(x)\}$, then $\pi_{M*}\rho^R_{Q,I,M}\to\frac{|I|\pi_{M*}\mu_{G/H}}{\mu_{G/H}(G/H)}$ weakly. So let $f\in C(\Omega_M)$ be nonnegative and bounded. For a given $\epsilon>0$, let $\psi_\epsilon:\Omega_M\to\mathbb{R}$ be a bounded continuous function such that $\psi_\epsilon\le R_M$ and $\int_{\Omega_M}(R_M-\psi_\epsilon)\,d\mu_{\Omega_H}<\frac{\epsilon}{2}$. Then

$$O_\epsilon=\{h_sp_{a,b}m_iH:(a,b)\in\Omega^o,m_iH\in M,0<s<\psi_\epsilon(p_{a,b}m_iH)\}$$

is an open subset of G/H in which $\mu_{G/H}((G/H)\backslash O_\epsilon)<\frac{\epsilon}{2}$. So by the inner regularity of $\mu_{G/H}$ and Urysohn's lemma, there is a continuous function $\chi_\epsilon:G/H\to[0,1]$ such that $\operatorname{Supp}\chi_\epsilon\subseteq O_\epsilon$ and $\chi_\epsilon^{-1}(\{1\})$ is a compact subset of O_ϵ with $\mu_{G/H}((G/H)\backslash\chi_\epsilon^{-1}(\{1\}))<\epsilon$.

Now notice that π_M is continuous on O_ϵ, and therefore $f_{\epsilon,1}=\chi_\epsilon\cdot(f\circ\pi_M)$, $f_{\epsilon,2}=N-\chi_\epsilon\cdot(N-f\circ\pi_M)\in C(G/H)$, where $N>0$ is a constant such that $f\le N$. Thus

$$\lim_{Q\to\infty}\int_{G/H}f_{\epsilon,j}\,d\rho^R_{Q,I,M}=\frac{|I|}{\mu_{G/H}(G/H)}\int_{G/H}f_{\epsilon,j}\,d\mu_{G/H},\quad j=1,2.$$

Since $f_{\epsilon,1} \leq f \circ \pi_M \leq f_{\epsilon,2}$, we also have

$$\liminf_{Q\to\infty} \int_{G/H} f \circ \pi_M \, d\rho^R_{Q,I,M} \geq \frac{|I|}{\mu_{G/H}(G/H)} \int_{G/H} f_{\epsilon,1} \, d\mu_{G/H} \text{ and}$$
$$\limsup_{Q\to\infty} \int_{G/H} f \circ \pi_M \, d\rho^R_{Q,I,M} \leq \frac{|I|}{\mu_{G/H}(G/H)} \int_{G/H} f_{\epsilon,2} \, d\mu_{G/H}.$$

By the properties of χ_ϵ,

$$\int_{G/H} f \circ \pi_M \, d\mu_{G/H} \leq \int_{G/H} f_{\epsilon,1} \, d\mu_{G/H} + N\epsilon \text{ and}$$
$$\int_{G/H} f \circ \pi_M \, d\mu_{G/H} \geq \int_{G/H} f_{\epsilon,2} \, d\mu_{G/H} - N\epsilon,$$

and therefore the following two inequalities hold:

$$\liminf_{Q\to\infty} \int_{G/H} f \circ \pi_M \, d\rho^R_{Q,I,M} \geq \frac{|I|}{\mu_{G/H}(G/H)} \left(\int_{G/H} f \circ \pi_M \, d\mu_{G/H} - N\epsilon \right),$$
$$\limsup_{Q\to\infty} \int_{G/H} f \circ \pi_M \, d\rho^R_{Q,I,M} \leq \frac{|I|}{\mu_{G/H}(G/H)} \left(\int_{G/H} f \circ \pi_M \, d\mu_{G/H} + N\epsilon \right).$$

Letting $\epsilon \to 0$ then yields

$$\lim_{Q\to\infty} \int_{G/H} f \circ \pi_M \, d\rho^R_{Q,I,M} = \frac{|I|}{\mu_{G/H}(G/H)} \int_{G/H} f \circ \pi_M \, d\mu_{G/H},$$

proving that $\pi_{M*}\rho^R_{Q,I,M} \to \frac{|I|\pi_{M*}\mu_{G/H}}{\mu_{G/H}(G/H)}$ weakly.

As noted in Section 2.5, $\frac{1}{R_M}$ can be well approximated in $L^1(\Omega_M, \mu_{\Omega_M})$ from above and below by continuous functions $\psi_{2,\epsilon}$ and $\psi_{3,\epsilon}$, and so one can easily show that $\frac{1}{R_M}\pi_{M*}\rho^R_{Q,I,M} \to \frac{|I|\pi_{M*}\mu_{G/H}}{R_M\mu_{G/H}(G/H)}$ weakly using the fact that

$$\psi_{2,\epsilon}\pi_{M*}\rho^R_{Q,I,M} \to \frac{\psi_{2,\epsilon}|I|\pi_{M*}\mu_{G/H}}{\mu_{G/H}(G/H)} \quad \text{and} \quad \psi_{3,\epsilon}\pi_{M*}\rho^R_{Q,I,M} \to \frac{\psi_{3,\epsilon}|I|\pi_{M*}\mu_{G/H}}{\mu_{G/H}(G/H)}$$

weakly. Notice that

$$\frac{1}{R_M}\pi_{M*}\rho^R_{Q,I,M} = \frac{N_{I,M}(Q)}{Q^2}\rho_{Q,I,M} \quad \text{and} \quad \frac{|I|\pi_{M*}\mu_{G/H}}{R_M\mu_{G/H}(G/H)} = \frac{|I|(\#M)\mu_{\Omega_M}}{\mu_{G/H}(G/H)},$$

and hence $\frac{N_{I,M}(Q)}{Q^2}\rho_{Q,I,M} \to \frac{|I|(\#M)\mu_{\Omega_M}}{\mu_{G/H}(G/H)}$ weakly. Since $\rho_{Q,I,M}$ is a probability measure for all $Q \in \mathbb{N}$, we have

$$\lim_{Q\to\infty} \frac{N_{I,M}(Q)}{Q^2} = \lim_{Q\to\infty} \frac{N_{I,M}(Q)}{Q^2}\rho_{Q,I,M}(\Omega_M) = \frac{|I|(\#M)\mu_{\Omega_M}(\Omega_M)}{\mu_{G/H}(G/H)} = \frac{|I|(\#M)}{\mu_{G/H}(G/H)},$$

implying that $N_{I,M}(Q) \sim \frac{|I|(\#M)Q^2}{\mu_{G/H}(G/H)} = \frac{3|I|(\#M)Q^2}{\pi^2[\Gamma:H]}$ as $Q \to \infty$. This proves the equidistribution of $(\mathcal{F}_M(Q))_Q$ in $[0,1]$ and the weak convergence

$$\rho_{Q,I,M} \to \mu_{\Omega_M},$$

completing the proof of Theorem 1.

2.7 The repulsion gap for Farey fractions $\frac{a}{q}$ with $q \equiv 1 \bmod m$

In this section, we determine the repulsion gap for Farey fractions with denominators congruent to 1 modulo m. For a given increasing sequence $\mathscr{A} := (A_n)$ of subsets of $[0,1]$ with limiting gap measure $\nu_{\mathscr{A}}$, we define the repulsion gap of $\mathscr{A}$ to be

$$K_{\mathscr{A}} := \sup\{\xi \geq 0 : \nu_{\mathscr{A}}([0,\xi]) = 0\}.$$

This means that if $\Delta_{av}(A_n)$ is the average gap between consecutive elements in A_n, then for a given $\epsilon \in (0, K_{\mathscr{A}})$,

$$\lim_{n\to\infty} \frac{\#\{x, x' \text{ consecutive in } A_n : x' - x \leq \epsilon\Delta_{av}(A_n)\}}{\#A_n - 1} = 0.$$

In other words, the proportion of the number of gaps of elements in A_n that are smaller than $\epsilon\Delta_{av}(A_n)$ approaches 0 as $n \to \infty$. So $K_{\mathscr{A}}$ provides a measure for how big a large proportion of the gaps in A_n must be for large n.

Let $I \subseteq [0,1]$ be a subinterval, $m \in \mathbb{N}$, and

$$A = \{(a,1) \bmod m : a \in \{0,\ldots,m-1\}\} \subseteq (\mathbb{Z}/m\mathbb{Z})^2$$

so that $\mathcal{F}_{I,m,A}(Q)$ is the set of fractions $\frac{a}{q} \in \mathcal{F}(Q) \cap I$ with $q \equiv 1 \bmod m$. We now compute the repulsion gap for the sequence $(\mathcal{F}_{I,m,A}(Q))_Q$. First note that $\mathcal{F}_{I,m,A}(Q) = \mathcal{F}_{I,M}(Q)$, where M is the set of cosets of the form $\left(\begin{smallmatrix} a & b \\ -1 & d \end{smallmatrix}\right)\Gamma(m)$ in $\Gamma/\Gamma(m)$, where a, b, and d are any integers such that $ad+b=1$. It is well known that $[\Gamma:\Gamma(m)] = m^3 \prod_{p|m,\, p\,\text{prime}}(1-p^{-2})$. Also, since the congruence $ad+b \equiv 1 \bmod m$ has m^2 solutions and each coset of $\Gamma/\Gamma(m)$ is completely determined by the congruence classes modulo m of the entries of one of its elements, there are m^2 cosets in M. So by (2.7), the repulsion gap of $(\mathcal{F}_{I,m,A}(Q))_Q$ is

$$\frac{3\xi'}{\pi^2 m} \prod_{\substack{p|m \\ p\text{ prime}}} (1-p^{-2})^{-1}$$

where $\xi' = \sup\left\{\xi \geq 0 : \mu_{\Omega_H}\left(R_M^{-1}([0,\xi])\right) = 0\right\}$. We have previously found that for a given non-negative fraction $\frac{c_1}{c_3}$, $f_{c_1/c_3}^{-1}([0,\xi])$ is nonempty only if $\xi \geq \max\{c_1, c_3\}$. Also, $f_0^{-1}([0,\xi])$ is a subset of

$$\left\{p_{a,b} m_i H : b \geq \frac{1}{\xi a}, i \in \{1,\ldots,k\}\right\}$$

which clearly has positive μ_{Ω_H}-measure if and only if $\xi > 1$. Thus, we have $\xi' \geq 1$. On the other hand, notice that if $m_1 = \left(\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right)$ and $m_2 := \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, then $m_1\Gamma(m), m_2\Gamma(m) \in M$ and

$$m_1 m_2^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \in m_1\Gamma(m)m_2^{-1}.$$

This implies that $f_0^{-1}([0,\xi])$ contains the set $\left\{p_{a,b}m_1 H : b \geq \frac{1}{\xi a}\right\}$, which has positive μ_{Ω_H}-measure when $\xi > 1$. This proves that $\xi' = 1$, and hence the repulsion gap, $K_{m,A}$, of $(\mathcal{F}_{I,m,A}(Q))_Q$ is given by (2.2).

Figure 2.3 depicts numerical approximations of densities of the revised measures $\nu'_{I,m,A}$ given by $\nu'_{I,m,A}([0,\xi]) = \nu_{I,m,A}\left(\left[0, \frac{3m^2}{\pi^2[\Gamma:\Gamma(m)]}\xi\right]\right)$ for $m = 3, 6, 11$. The multiplication by $\frac{3m^2}{\pi^2[\Gamma:\Gamma(m)]}$ makes $\nu'_{I,m,A}$ the limiting measure corresponding to (2.6) in which the normalization of the gaps is Q^2, which allows for an even comparison of the gaps in the sequences. The initial interval $[0,1]$ on which the densities are zero in Figure 2.3 reflect the fact that the constant ξ' above equals 1 for all three sequences.

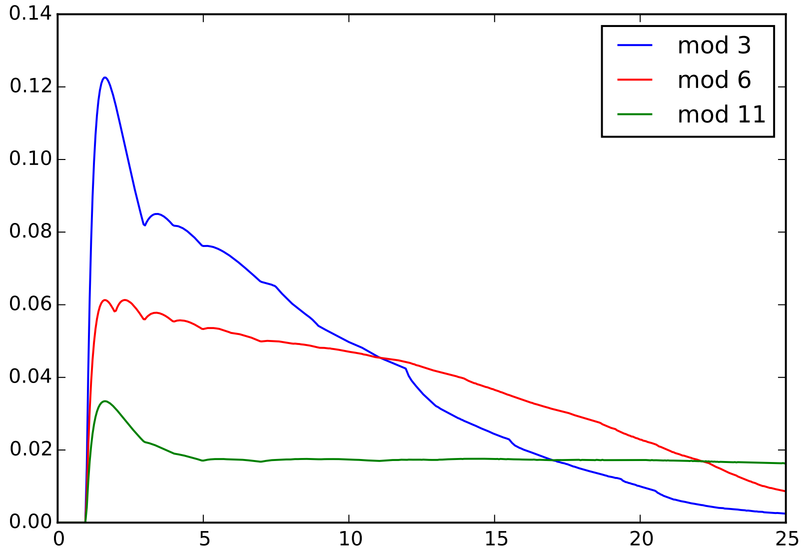

Figure 2.3: Revised gap distribution densities for fractions with denominators congruent to 1 modulo 3, 6, and 11

2.8 Proof of Theorem 2

As in Sections 2.3–2.6, we let $H \subseteq \Gamma$ be a finite index subgroup, $M = \{m_1H, \ldots, m_kH\} \subseteq \Gamma/H$ be nonempty and closed under left multiplication by $\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$, and $I \subseteq [0,1]$ be a subinterval. Recall that $S_{I,M}(n,\alpha,c)$ is the set of $\xi \in I$ for which there exists $\frac{a}{q} \in \mathcal{F}_M(\lfloor nc \rfloor)$ such that $q \geq n$ and $|q\xi - a| \leq \frac{\alpha}{q}$; and that we aim to show that the limits of the sequences $(\lambda(S_M(n,\alpha,c)))_n$ and $(\lambda(S_{I,M}(n,\alpha,c)))_n$ exist, and

$$\lim_{n\to\infty} S_{I,M}(n,\alpha,c) = |I|\varrho_M(\alpha,c), \quad \text{where} \quad \varrho_M(\alpha,c) = \lim_{n\to\infty} \lambda(S_M(n,\alpha,c)).$$

The idea of the proof is to show that the measures $\lambda(S_{I,M}(n,\alpha,c))$ can, up to a small change, be written as sums of expressions of the form

$$\int_{\mathscr{H}(n/\lfloor nc \rfloor)} f \, d\rho_{\lfloor nc \rfloor, I, M}, \tag{2.11}$$

where $\mathscr{H}(\frac{n}{\lfloor nc \rfloor}) \subseteq \Omega_M$ is a subset having a boundary of μ_{Ω_M}-measure 0, and f is a piecewise real analytic function which is bounded on $\mathscr{H}(C)$ for any $C > 0$. We then use the convergence $\rho_{Q,I,M} \to \mu_{\Omega_M}$, in addition to the fact that the region $\mathscr{H}\big(\frac{n}{\lfloor nc \rfloor}\big)$ becomes $\mathscr{H}(c^{-1})$ as $n \to \infty$, to show that the above integral approaches

$$\frac{3|I|(\#M)}{\pi^2[\Gamma : H]}\int_{\mathscr{H}(c^{-1})} f\, d\mu_{\Omega_M}.$$

We begin by following the process in [70] of using the inclusion-exclusion principle to rewrite $\lambda(S_{I,M}(n,\alpha,c))$ as approximately a linear combination of measures of intervals. For $\alpha > 0$, $c \geq 1$, and $n \in \mathbb{N}$, let $Q = \lfloor nc \rfloor$ and

$$\mathcal{F}_{I,M}(Q) = \left\{ \beta_0 = \frac{b_0}{p_0} < \beta_1 = \frac{b_1}{p_1} < \cdots < \beta_{N_{I,M}(Q)} = \frac{b_{N_{I,M}(Q)}}{p_{N_{I,M}(Q)}} \right\},$$

with $\gcd(b_i, p_i) = 1$. Then for every $\beta_i \in \mathcal{F}_{I,M}(Q)$, let

$$J(\beta_i) = \left[\frac{b_i}{p_i} - \frac{\alpha}{p_i^2}, \frac{b_i}{p_i} + \frac{\alpha}{p_i^2} \right].$$

and then define

$$S'_{I,M}(n,\alpha,c) = \bigcup_{\substack{\beta_i \in \mathcal{F}_{I,M}(Q) \\ p_i \geq n}} J(\beta_i).$$

Now $S_{I,M}(n,\alpha,c) \backslash S'_{I,M}(n,\alpha,c)$ is contained in the union of the intervals of the form $\left[\frac{b}{p} - \frac{\alpha}{p^2}, \frac{b}{p} + \frac{\alpha}{p^2}\right] \cap I$, where $\frac{b}{p}$ is a fraction in $\mathcal{F}_M(Q) \backslash I$ with $p \geq n$. On the other hand, $S'_{I,M}(n,\alpha,c) \backslash S_{I,M}(n,\alpha,c)$ is contained in the union of the intervals of the form $\left[\frac{b}{p} - \frac{\alpha}{p^2}, \frac{b}{p} + \frac{\alpha}{p^2}\right] \backslash I$, where $\frac{b}{p}$ is a fraction in $\mathcal{F}_M(Q) \cap I$ with $p \geq n$. It is clear that the measure of each of these unions cannot exceed $\frac{2\alpha}{n^2}$, and so the sequences $(\lambda(S_{I,M}(n,\alpha,c)))_n$ and $(\lambda(S'_{I,M}(n,\alpha,c)))_n$ converge and have the same limit if one converges. Thus from now on, we examine the sets $S'_{I,M}(n,\alpha,c)$.

By the inclusion-exclusion principle, we have

$$\begin{aligned} \lambda(S'_{I,M}(n,\alpha,c)) &= \lambda\left(\bigcup_{\substack{\beta_i \in \mathcal{F}_{I,M}(Q) \\ q_i \geq n}} J(\beta_i) \right) \\ &= \sum_{r=0}^{N_{I,M}(Q)-1} (-1)^r \sum_{0 = j_0 < \cdots < j_r \leq N_{I,M}(Q)} \sum_{\substack{i=0 \\ p_{i+j_s} \geq n, 0 \leq s \leq r}}^{N_{I,M}(Q)-j_r} \lambda\left(\bigcap_{s=0}^{r} J(\beta_{i+j_s}) \right). \end{aligned} \tag{2.12}$$

By [70, Lemma 3], there exists an integer K, depending only on α and c, such that if $\frac{a}{q}, \frac{a'}{q'} \in \mathcal{F}(Q)$ such that $q, q' \geq n$ and $J(\frac{a}{q}) \cap J(\frac{a'}{q'}) \neq \emptyset$, then there are at most $K-1$ elements in $\mathcal{F}(Q)$ between $\frac{a}{q}$ and $\frac{a'}{q'}$. It follows that if $\beta_i, \beta_j \in \mathcal{F}_M(Q)$ with $p_i, p_j \geq n$ and $J(\beta_i) \cap J(\beta_j) \neq \emptyset$, then $|i-j| \leq K$. We can thus rewrite (2.12) as

$$\sum_{r=0}^{K} (-1)^r \sum_{0=j_0<\cdots<j_r \leq K} \sum_{\substack{i=1 \\ p_{i+j_s} \geq n, 0 \leq s \leq r}}^{N_{I,M}(Q)-j_r} \lambda\left(\bigcap_{s=0}^{r} J(\beta_{i+j_s})\right). \tag{2.13}$$

Next, again analogous to [70], we construct a region $\mathscr{H}_{j_1,\ldots,j_r}(\frac{n}{Q})$ in Ω_M having the property that $W_{H,Q}(\beta_i) \in \mathscr{H}_{j_1,\ldots,j_r}(\frac{n}{Q})$ if and only if $p_{i+j_s} \geq n$ for $0 \leq s \leq r$, and then write

$$\lambda\left(\bigcap_{s=0}^{r} J(\beta_{i+j_s})\right)$$

as a piecewise real analytic function of $W_{H,Q}(\beta_i)$. In this way, we will rewrite (2.13), up to a small change, as a linear combination of expressions in the form (2.11) as mentioned above.

The fraction $\beta_i \in \mathcal{F}_{I,M}(Q)$ satisfies $p_{i+j_s} \geq n$ if and only if

$$W_{H,Q}(\beta_{i+j_s}) = r_M^{j_s}(W_{H,Q}(\beta_i)) \in \left\{p_{a,b} m_i H \in \Omega_M : a \geq \frac{n}{Q}\right\}.$$

So the set of β_i such that $p_{i+j_s} \geq n$ for $0 \leq s \leq r$ are such that $W_{H,Q}(\beta_i) \in \mathscr{H}_{j_1,\ldots,j_r}(\frac{n}{Q})$, where for $t \in (0,1]$,

$$\mathscr{H}_{j_1,\ldots,j_r}(t) = \bigcap_{s=0}^{r} r_M^{-j_s}\{p_{a,b} m_i H \in \Omega_M : a \geq t\}.$$

Note that $\mathscr{H}_{j_1,\ldots,j_r}(t)$ is a countable union of polygons. This follows easily from our observation in Section 2.5.1 that for a given $j \in \mathbb{Z}$, Ω_M can be divided into a countable number of polygons $\mathcal{P}$ such that r_M^j is linear on $\mathcal{P}$. Hence any set of the form $r_M^{-j}\{p_{a,b} m_i H \in \Omega_M : a \geq t\}$ is a countable union of polygons, and thus $\mathscr{H}_{j_1,\ldots,j_r}(t)$ is as well.

Next, for $j \in \mathbb{N}$, let $R_M^{(j)} : \Omega_M \to \mathbb{R}$ be the jth return time function defined by

$$R_M^{(j)} = \sum_{i=0}^{j-1} R_M \circ r_M^i.$$

(Let $R_M^{(0)} \equiv 0$.) We then have $R_M^{(j)}(W_{H,Q}(\beta_i)) = Q^2(\beta_{i+j} - \beta_i)$. Also, define the function $\pi_1 : \Omega_M \to \mathbb{R}$ by $\pi_1(p_{a,b} m_i H) = a$. We then have

$$\begin{aligned}
\lambda\left(\bigcap_{s=0}^{r} J(\beta_{i+j_s})\right) &= \max\left\{0, \min_{0\le s\le s'\le r}\left\{\left(\frac{b_{i+j_s}}{p_{i+j_s}} + \frac{\alpha}{p_{i+j_s}^2}\right) - \left(\frac{b_{i+j_{s'}}}{p_{i+j_{s'}}} - \frac{\alpha}{p_{i+j_{s'}}^2}\right)\right\}\right\} \\
&= \frac{1}{Q^2}\max\left\{0, \min_{0\le s\le s'\le r}\left\{\alpha\left(\frac{Q^2}{p_{i+j_s}^2} + \frac{Q^2}{p_{i+j_{s'}}^2}\right) - Q^2\left(\frac{b_{i+j_{s'}}}{p_{i+j_{s'}}} - \frac{b_{i+j_s}}{p_{i+j_s}}\right)\right\}\right\} \\
&= \frac{1}{Q^2}\max\Big\{0, \min_{0\le s\le s'\le r}\{\alpha\big((\pi_1\circ r_M^{j_s})(W_{H,Q}(\beta_i))^{-2} + (\pi_1\circ r_M^{j_{s'}})(W_{H,Q}(\beta_i))^{-2}\big) \\
&\qquad\qquad - (R_M^{(j_{s'}-j_s)}\circ r_M^{j_s})(W_{H,Q}(\beta_i))\}\Big\} \\
&= \frac{1}{Q^2} f_{j_1,\dots,j_r}^{(\alpha)}(W_{H,Q}(\beta_i)),
\end{aligned}$$

where $f_{j_1,\dots,j_r}^{(\alpha)} : \Omega_M \to \mathbb{R}$ is given by

$$f_{j_1,\dots,j_r}^{(\alpha)} = \max\Big\{0, \min_{0\le s\le s'\le r}\{\alpha\big((\pi_1\circ r_M^{j_s})^{-2} + (\pi_1\circ r_M^{j_{s'}})^{-2}\big) - (R_M^{(j_{s'}-j_s)}\circ r_M^{j_s})\}\Big\}.$$

We can now rewrite (2.13) as

$$\begin{aligned}
&\sum_{r=0}^{K}(-1)^r \sum_{0=j_0<\cdots<j_r\le K} \frac{1}{Q^2} \sum_{\substack{\beta_i\in\mathcal{F}_{I,M}(Q),\\ i\le N_{I,M}(Q)-j_r}} f_{j_1,\dots,j_r}^{(\alpha)}(W_{H,Q}(\beta_i)) \\
&\qquad = \sum_{r=0}^{K}(-1)^r \sum_{0=j_0<\cdots<j_r\le K}\Bigg(\frac{N_{I,M}(Q)}{Q^2}\int_{\mathscr{H}_{j_1,\dots,j_r}(n/Q)} f_{j_1,\dots,j_r}^{(\alpha)}\, d\rho_{Q,I,M} \\
&\qquad\qquad + O\bigg(\frac{K\|f_{j_1,\dots,j_r}^{(\alpha)}|_{\mathscr{H}_{j_1,\dots,j_r}(n/Q)}\|_\infty}{Q^2}\bigg)\Bigg).
\end{aligned}$$

For $C > 0$, we have $\|f_{j_1,\dots,j_r}^{(\alpha)}|_{\mathscr{H}_{j_1,\dots,j_r}(C)}\|_\infty \le \frac{2\alpha}{C^2}$, and if $C = \frac{n}{Q}$, $\frac{2\alpha}{C^2} \le 2\alpha c^2$. Thus the error term above is negligible, and to complete the proof, it remains to show the existence of

$$\lim_{n\to\infty}\int_{\mathscr{H}_{j_1,\dots,j_r}(n/Q)} f_{j_1,\dots,j_r}^{(\alpha)}\, d\rho_{Q,I,M}$$

for all $j_1,\dots,j_r$.

Now by the properties of R_M, r_M, and π_1, it is clear that $f_{j_1,\dots,j_r}^{(\alpha)}$ is a piecewise real analytic function. We have proven above that for a fixed $C>0$, $f^{(\alpha)}_{j_1,\ldots,j_r}$ is bounded on $\mathscr{H}_{j_1,\ldots,j_r}(C)$. Also, it is clear that $\mathscr{H}_{j_1,\ldots,j_r}(C)$ has a boundary of μ_{Ω_M}-measure 0, implying that

$$\lim_{n\to\infty}\int_{\mathscr{H}_{j_1,\ldots,j_r}(C)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\rho_{Q,I,M} = \int_{\mathscr{H}_{j_1,\ldots,j_r}(C)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\mu_{\Omega_M}.$$

Since $\frac{n}{Q}\geq c^{-1}$, and thus $\mathscr{H}_{j_1,\ldots,j_r}(\frac{n}{Q})\subseteq \mathscr{H}_{j_1,\ldots,j_r}(c^{-1})$, we have

$$\begin{aligned}\limsup_{n\to\infty}\int_{\mathscr{H}_{j_1,\ldots,j_r}(n/Q)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\rho_{Q,I,M} &\leq \lim_{n\to\infty}\int_{\mathscr{H}_{j_1,\ldots,j_r}(c^{-1})} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\rho_{Q,I,M}\\ &= \int_{\mathscr{H}_{j_1,\ldots,j_r}(c^{-1})} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\mu_{\Omega_M}.\end{aligned}$$

On the other hand, for a given $\epsilon>0$, we have $\frac{n}{Q}\leq c^{-1}+\epsilon$ for large n. Therefore,

$$\begin{aligned}\liminf_{n\to\infty}\int_{\mathscr{H}_{j_1,\ldots,j_r}(n/Q)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\rho_{Q,I,M} &\geq \lim_{n\to\infty}\int_{\mathscr{H}_{j_1,\ldots,j_r}(c^{-1}+\epsilon)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\rho_{Q,I,M}\\ &= \int_{\mathscr{H}_{j_1,\ldots,j_r}(c^{-1}+\epsilon)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\mu_{\Omega_M}.\end{aligned}$$

By the continuity of measure from below, letting $\epsilon\to 0$ yields

$$\liminf_{n\to\infty}\int_{\mathscr{H}_{j_1,\ldots,j_r}(n/Q)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\rho_{Q,I,M} \geq \int_{\mathscr{H}'_{j_1,\ldots,j_r}(c^{-1})} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\mu_{\Omega_M},$$

where

$$\mathscr{H}'_{j_1,\ldots,j_r}(c^{-1}) = \bigcap_{s=0}^{r} r_M^{-j_s}\{p_{a,b}m_iH\in\Omega_M : a>c^{-1}\}.$$

We clearly have

$$\mu_{\Omega_M}\left(\bigcap_{s=0}^{r} r_M^{-j_s}\{p_{a,b}m_iH\in\Omega_M : a=c^{-1}\}\right)=0,$$

and thus

$$\lim_{n\to\infty}\int_{\mathscr{H}_{j_1,\ldots,j_r}(n/Q)} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\rho_{Q,I,M} = \int_{\mathscr{H}_{j_1,\ldots,j_r}(c^{-1})} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\mu_{\Omega_M}.$$

Noting again that $N_{I,M}(Q)\sim\frac{3|I|(\#M)Q^2}{\pi^2[\Gamma:H]}$ as $Q\to\infty$, we have completed the proof of Theorem 2, with

$$\varrho_M(\alpha,c)=\frac{3(\#M)}{\pi^2[\Gamma:H]}\sum_{r=0}^{K}(-1)^r\sum_{0=j_0<\cdots<j_r\leq K}\int_{\mathscr{H}_{j_1,\ldots,j_r}(c^{-1})} f^{(\alpha)}_{j_1,\ldots,j_r}\, d\mu_{\Omega_M}.$$

Chapter 3

Continued fractions and the Gauss and Farey maps

We now turn our attention to the dynamics of continued fractions, and in particular, the Farey map. We begin in this chapter by reviewing some elementary properties of continued fractions which we need in the following chapters. We then define the Gauss and Farey maps, in addition to their natural extensions and transfer operators, and give some of their properties.

3.1 Continued fractions

Recall that we denote a regular continued fraction by

$$[a_1, a_2, \ldots] := \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}} \qquad (a_j \in \mathbb{N})$$

We shall also make use of the notation

$$[a_0; a_1, a_2, \ldots] := a_0 + [a_1, a_2, \ldots]. \qquad (a_0 \in \mathbb{Z}, a_j \in \mathbb{N})$$

For a given sequence $a = (a_j)_{j=1}^{\infty}$ of positive integers, define the nonnegative, coprime integers $p_n = p_n(a) = p_n(a_1, \ldots, a_n)$, $q_n = q_n(a) = q_n(a_1, \ldots, a_n)$ by

$$\frac{p_n}{q_n} := [a_1, a_2, \ldots, a_n].$$

Denoting also $p_0 = p_0(a) := 0$ and $q_0 = q_0(a) := 1$, one can find that for $n \geq 2$, $p_n = a_n p_{n-1} + p_{n-2}$ and $q_n = a_n q_{n-1} + q_{n-2}$, which imply

$$\begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} q_n & p_n \\ q_{n-1} & p_{n-1} \end{pmatrix}. \tag{3.1}$$

This in turn gives $p_{n-1}q_n - p_n q_{n-1} = (-1)^n$, and hence $\frac{p_{n-1}}{q_{n-1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_{n-1}q_n}$.

For $j, n \in \mathbb{N}$ with $j \leq n$, we define $p_{j,n} = p_{j,n}(a)$ and $q_{j,n} = q_{j,n}(a)$ by

$$\frac{p_{j,n}}{q_{j,n}} = [a_j, a_{j+1}, \ldots, a_n].$$

Taking transposes in (3.1) reveals that $q_n(a_1, \ldots, a_n) = q_n(a_n, \ldots, a_1)$, which then implies that $q_n(a) = a_1 q_{2,n}(a) + q_{3,n}(a)$, and more generally, $q_{j,n}(a) = a_j q_{j+1,n}(a) + q_{j+2,n}(a)$ for $j \leq n-2$. This equality extends to $j = n-1, n$ once we denote $q_{n+1,n} = q_{n+1,n}(a) := 1$ and $q_{n+2,n} = q_{n+2,n}(a) := 0$. Another subtle property of the values $q_{j,n}$ we wish to note is that

$$\prod_{j=1}^{n} \frac{p_{j,n}}{q_{j,n}} = \frac{1}{q_n}. \tag{3.2}$$

(See [22, Lemma 2.1, Theorem 3.6] for proof.)

Lastly, we define for a finite tuple $b = (b_1, \ldots, b_n) \in \mathbb{N}^n$ the set

$$I_b = [\![b_1, \ldots, b_n]\!] := \{[b_1, \ldots, b_n + t] : t \in [0,1]\} = \{[a_1, a_2, \ldots] \in [0,1] : a_j = b_j, j = 1, \ldots n\},$$

which is the closed interval between $\frac{p_{n+1}(b,1)}{q_{n+1}(b,1)}$ and $\frac{p_n(b)}{q_n(b)}$. We thus have

$$\lambda(I_b) = \left| \frac{p_{n+1}(b,1)}{q_{n+1}(b,1)} - \frac{p_n(b,1)}{q_n(b,1)} \right| = \frac{1}{q_{n+1}(b,1)q_n(b,1)} = \frac{1}{q_n(b)(q_n(b) + q_{n-1}(b))}. \tag{3.3}$$

3.2 The Gauss map

The Gauss map $G : [0,1] \to [0,1]$ is defined by

$$G(x) := \begin{cases} \{x^{-1}\} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

This map is invariant with respect to the Gauss measure ν given by

$$d\nu := \frac{dx}{(1+x)\log 2}.$$

We define the natural extension $\tilde{G} : [0,1]^2 \to [0,1]^2$ of the Gauss map by

$$\tilde{G}(x,y) := \begin{cases} \left(G(x), \frac{1}{\lfloor x^{-1} \rfloor + y}\right) & \text{if } x \neq 0 \\ (0,0) & \text{if } x = 0 \end{cases}$$

which is invariant with respect to the measure $\tilde{\nu}$ given by

$$d\tilde{\nu} := \frac{dx\,dy}{(1+xy)^2 \log 2}.$$

The action of G and $\tilde{G}$ on continued fractions is as follows:

$$G([a_1, a_2, \ldots]) = [a_2, a_3, \ldots]; \tag{3.4}$$

$$\tilde{G}([a_1, a_2, \ldots], [b_1, b_2, \ldots]) = ([a_2, a_3, \ldots], [a_1, b_1, b_2, \ldots]). \tag{3.5}$$

In other words, G and $\tilde{G}$ act respectively as the one and two-sided shifts on the continued fraction expansions of their arguments.

Next, we let $\hat{G}_\lambda$ and $\hat{G}_\nu$ be the transfer operators of the Gauss map with respect to the measures λ and ν, respectively, so that

$$\int_{[0,1]} (\hat{G}_\lambda f) \cdot g \, d\lambda = \int_{[0,1]} f \cdot (g \circ G)\, d\lambda \quad \text{and} \quad \int_{[0,1]} (\hat{G}_\nu f) \cdot g \, d\nu = \int_{[0,1]} f \cdot (g \circ G)\, d\nu$$

for all $f \in L^1(\lambda) = L^1(\nu)$ and $g \in L^\infty(\lambda) = L^\infty(\nu)$. We can calculate $\hat{G}_\lambda$ as follows: Fix $f \in L^1(\lambda)$ and $x \in [0,1]$. Then we have

$$\int_0^x (\hat{G}_\lambda f)(t)\, dt = \int_{G^{-1}([0,x])} f\, d\lambda = \sum_{n=1}^{\infty} \int_{1/(x+n)}^{1/n} f(t)\, dt.$$

Taking the derivative with respect to x yields

$$(\hat{G}_\lambda f)(x) = \sum_{n=1}^{\infty} f\left(\frac{1}{x+n}\right) \frac{1}{(x+n)^2}.$$

One can similarly find an explicit formula for $\hat{G}_\nu$. Also, it is easy to see by definition that if $\xi : [0,1] \to \mathbb{R}$ is the Radon-Nikodym derivative $\frac{d\nu}{d\lambda}$ so that $\xi(x) = \frac{1}{(1+x)\log 2}$ and $M_\xi : L^1(\nu) \to L^1(\lambda)$ is the (invertible) multiplicative operator $M_\xi f = \xi \cdot f$, then $M_\xi \circ \hat{G}_\nu = \hat{G}_\lambda \circ M_\xi$. Hence $\hat{G}_\nu$ and $\hat{G}_\lambda$ are similar.

The operators $\hat{G}_\lambda$ and $\hat{G}_\nu$ have garnered a lot of study over the years. A central problem motivating this attention, posed by Gauss in a letter to Laplace in 1812, was to estimate the error

$$\lambda(\{[a_1, a_2, \ldots] \in [0,1] : [a_{n+1}, a_{n+2}, \ldots] \le u\}) - \frac{\log(1+u)}{\log 2}, \tag{3.6}$$

with $u \in (0,1)$ fixed, as $n \to \infty$. This is equivalent to estimating

$$\lambda(G^{-n}([0,u])) - \nu([0,u]), \qquad (n \to \infty)$$

which can then be written as

$$\int_{[0,u]} (\hat{G}_\lambda^n(1) - \xi)\, d\lambda \quad \text{or} \quad \int_{[0,u]} (\hat{G}_\nu^n(\xi^{-1}) - 1)\, d\nu.$$

Since ν is G-invariant, the constant function 1 is an eigenfunction of $\hat{G}_\nu$ with eigenvalue 1, and $\hat{G}_\lambda$ has the corresponding eigenfunction ξ with eigenvalue 1. Wirsing found the optimal decay rate of (3.6) by finding a spectral gap in $\hat{G}_\nu$ as an operator on $C^1([0,1])$ below 1, and an implication of his work is [38, Theorem 2.2.6]

$$\hat{G}_\nu^n f - \int_{[0,1]} f\, d\nu \sim C_f(-\psi)^n \quad \text{for } f \in C^1([0,1]),$$

where $C_f \in \mathbb{R}$ is a constant dependent on f and $\psi = 0.30366\ldots$ is the absolute value of the second largest eigenvalue of $\hat{G}_\nu$. Thus, letting $f = \xi^{-1}$ in the above equality, we see that (3.6) is commensurate with ψ^n as $n \to \infty$. An exact solution to Gauss's problem was first given by Babenko who analyzed an operator similar to $\hat{G}_\nu$ and $\hat{G}_\lambda$ which is compact on a certain Hilbert space of analytic functions. This work was later extended by Mayer and Roepstorff. See [38] for history and details.

The discovery of the spectral gap in the transfer operator of G has further implications on the ergodic properties of G, and in particular its rate of mixing, which we interpret as the rate at which its preimages equidistribute. For $f \in C^1([0,1])$ and a subinterval $I \subseteq [0,1]$, we have

$$\begin{aligned} \int_{G^{-n}(I)} f\, d\lambda = \int_I \hat{G}_\nu^n(\xi^{-1} f)\, d\nu &= \int_I \left(\int_{[0,1]} \xi^{-1} f\, d\nu + O_f(\psi^n) \right) d\nu \\ &= \nu(I) \left(\int_{[0,1]} f\, d\lambda + O_f(\psi^n) \right). \end{aligned} \tag{3.7}$$

This shows that the preimages $(G^{-n}(I))_n$ equidistribute in $[0,1]$ with respect to the Lebesgue measure at an exponential rate.

The transfer operator of G has also been utilized in the study of the periodic points of G. Utilizing Mayer's work [56], Pollicott [59] determined the limiting distribution of the periodic points. He studied a Ruelle-Perron-Frobenius operator which forms an analytic perturbation of $\hat{G}_\lambda$, and in particular, he related the determinants of this operator to the Laplace transform of the sum of functions over the periodic points. Finding the growth rate of the sum was, by the Wiener-Ikehara Tauberian theorem, reduced to studying the behavior of the leading eigenvalue of the operator that determined the residue of the pole in the Laplace transform. We detail this process more in Section 4.3, where we analyze an appropriate alteration to the Ruelle-Perron-Frobenius operator in establishing the distribution of the periodic points of the Farey map.

3.3 The Farey map

The Farey map $F : [0,1] \to [0,1]$ is defined by

$$F(x) := \begin{cases} \dfrac{x}{1-x} & \text{if } x \in \left[0, \frac{1}{2}\right] \\ \dfrac{1-x}{x} & \text{if } x \in \left(\frac{1}{2}, 1\right]. \end{cases}$$

An invariant measure μ for F is given by

$$d\mu := \frac{dx}{x}.$$

The natural extension $\tilde{F} : [0,1]^2 \to [0,1]^2$ of F is defined by

$$\tilde{F}(x,y) := \begin{cases} \left(\dfrac{x}{1-x}, \dfrac{y}{1+y}\right) & \text{if } x \in \left[0, \frac{1}{2}\right] \\ \left(\dfrac{1-x}{x}, \dfrac{1}{1+y}\right) & \text{if } x \in \left(\frac{1}{2}, 1\right], \end{cases}$$

and has invariant measure $\tilde{\mu}$ given by

$$d\tilde{\mu} := \frac{dx\, dy}{(x+y-xy)^2}.$$

The fact that both measures μ and $\tilde{\mu}$ are infinite leads F and $\tilde{F}$ to have significantly different behavior than G and $\tilde{G}$. For future reference, we note that μ is the natural projection of $\tilde{\mu}$ onto the first coordinate, i.e.,

$$\int_{[0,1]^2} f(x)\, d\tilde{\mu}(x,y) = \int_{[0,1]} f\, d\mu. \qquad (f \in C([0,1])) \tag{3.8}$$

The maps F and $\tilde{F}$ act on continued fractions according to

$$F([a_1, a_2, \ldots]) = \begin{cases} [a_1 - 1, a_2, \ldots] & \text{if } a_1 \geq 2 \\ [a_2, a_3, \ldots] & \text{if } a_1 = 1; \end{cases} \tag{3.9}$$

$$\tilde{F}([a_1, a_2, \ldots], [b_1, b_2, \ldots]) = \begin{cases} ([a_1 - 1, a_2, \ldots], [b_1 + 1, b_2, \ldots]) & \text{if } a_1 \geq 2 \\ ([a_2, a_3, \ldots], [1, b_1, b_2, \ldots]) & \text{if } a_1 = 1. \end{cases} \tag{3.10}$$

So, like G and $\tilde{G}$, the Farey map and its extension act as shifts on the continued fraction expansions of its arguments, though in a slower manner, shifting a 1 in a digit instead of a whole digit at a time. In fact, F is a slowdown of G as demonstrated by the equality

$$F^{\lfloor 1/x \rfloor}(x) = G(x). \qquad (x \neq 0)$$

Also, by [18, Theorem 1], G is isomorphic to the induced transformation $F_A : A \to A$ of the Farey map on $A := \left[\frac{1}{2}, 1\right] = \{[1, a_1, a_3, \ldots] : a_j \in \mathbb{N}\}$ defined by

$$F_A(x) := F^{\phi_A(x)}(x), \quad \text{where} \quad \phi_A(x) = \min\{n \in \mathbb{N} : F^n(x) \in A\}.$$

This can be easily seen from the equality

$$F_A([1, a_1, a_2, \ldots]) = [1, a_2, a_3, \ldots].$$

Similarly, $\tilde{G}$ can be seen as isomorphic to the induced transformation $\tilde{F}_{\tilde{A}} : \tilde{A} \to \tilde{A}$ of $\tilde{F}$ on $\tilde{A} = (0, 1] \times \left(\frac{1}{2}, 1\right]$ defined by

$$\tilde{F}_{\tilde{A}}(x) := \tilde{F}^{\phi_{\tilde{A}}(x)}(x), \quad \text{where} \quad \phi_{\tilde{A}}(x) = \min\{n \in \mathbb{N} : \tilde{F}^n(x) \in \tilde{A}\} \tag{3.11}$$

from the equality

$$\tilde{F}_{\tilde{A}}([a_1, a_2, \ldots], [1, b_1, b_2, \ldots]) = ([a_2, a_3, \ldots], [1, a_1, b_1, b_2, \ldots]). \tag{3.12}$$

Early studies of the Farey map include [23, 24] in the context of thermodynamics, and [40], where the natural extension $\tilde{F}$ was also introduced, in examining mediant continued fraction convergents. See also [18] for more details on the above properties of F and $\tilde{F}$ and their relationships to G and $\tilde{G}$, in addition to continued fraction applications.

Next, let $\hat{F} : L^1(\mu) \to L^1(\mu)$ be the transfer operator of F with respect to the invariant measure μ, which is easily seen to satisfy

$$\hat{F}f(x) = \frac{f\left(\frac{x}{1+x}\right) + xf\left(\frac{1}{1+x}\right)}{1+x}.$$

Analogous to the mixing results for the Gauss map given in the previous section, we establish an effective equidistribution result for the preimages of the Farey map utilizing the transfer operator $\hat{F}$. However, we make no appeal to the spectral properties of $\hat{F}$, as such a process appears to be difficult by the examinations of the spectrum by Isola [39] and Prellberg [60].

Chapter 4

Distribution of the periodic points of the Farey map

4.1 Introduction

In this chapter, we study the distribution of the periodic points of the Farey map. In this introductory section, we characterize the periodic points of both F and $\tilde{F}$ in relation to those of G and $\tilde{G}$, and formulate our main equidistribution results. Then in Section 4.2, we review the connection between the modular surface and continued fractions. In particular, we recall Series' cross section of the geodesic flow. We then enlarge this cross section to yield another whose first return map under the geodesic flow is a double cover of the Farey map's natural extension. We then use this new section to extend the correspondence between closed geodesics in the modular surface and the periodic points of the Gauss map to those of the Farey map. Lastly, in Section 4.3, we prove our main equidistribution result utilizing the relationship between the periodic points of the Farey and Gauss maps to essentially reduce the problem to proving the equidistribution of the Gauss periodic points over certain continuous functions on $(0, 1]$ which are allowed to have a vertical asymptote at 0. We thus adapt Pollicott's work [59] on the Ruelle-Perron-Frobenius operator of the Gauss map, being careful to account for a possible asymptote in a function used to define the operator.

First, notice that from the equalities (3.4) and (3.5), it is easy to see that the periodic points of G are exactly the periodic continued fractions of the form

$$[\overline{a_1, \ldots, a_n}] := [a_1, \ldots, a_n, a_1, \ldots, a_n, \ldots],$$

i.e., the reduced quadratic irrationals $x \in [0, 1]$ with conjugate root $\bar{x} < -1$; and the periodic points of $\tilde{G}$ other than $(0,0)$ are of the form

$$([\overline{a_1, a_2, \ldots, a_n}], [\overline{a_n, a_{n-1}, \ldots, a_1}]),$$

where the continued fraction expansion of the second argument is the reverse of that of the first. Alternatively, the nonzero periodic points of $\tilde{G}$ are of the form $(x, -\bar{x}^{-1})$, where x is a reduced quadratic irrational. Notice that $x \leftrightarrow (x, -\bar{x}^{-1})$ gives a natural one-to-one correspondence between the nonzero periodic points of G and $\tilde{G}$.

Let Q_G denote the set of nonzero periodic points of G. To each $x \in Q_G$ with minimal even periodic expansion $x = [\overline{a_1, \ldots, a_{2n}}]$, we associate the value

$$\ell(x) := -2 \sum_{j=1}^{2n} \log(G^j(x)) \tag{4.1}$$

which is the length of a corresponding geodesic in the modular surface (see Section 4.2.2). For future reference, we analogously define, for a given tuple $a = (a_1, \ldots, a_n) \in \mathbb{N}^n$ of any length,

$$\ell(a) := -2 \sum_{j=1}^{n} \log(G^j[\overline{a_1, a_2, \ldots, a_n}]).$$

We then let

$$Q_G(T) := \{x \in Q_G : \ell(x) \leq T\}. \qquad (T > 0)$$

The result of Pollicott [59, Theorem 3] states that for all $f \in C([0,1])$, we have

$$\lim_{T\to\infty} \frac{1}{|Q_G(T)|} \sum_{x \in Q_G(T)} f(x) = \int_{[0,1]} f \, d\nu; \tag{4.2}$$

and it then follows from Kelmer's result [45, Lemma 17] that for all $f \in C([0,1]^2)$,

$$\lim_{T\to\infty} \frac{1}{|Q_G(T)|} \sum_{x \in Q_G(T)} f(x, -\bar{x}^{-1}) = \int_{[0,1]^2} f \, d\tilde{\nu}. \tag{4.3}$$

The main goal of this chapter is to formulate and prove results analogous to (4.2) and (4.3) for the periodic points of the Farey map and its natural extension. We have the following characterizations, the first originally discovered by Claire Merriman, of the periodic points of F and $\tilde{F}$ which provide connections to those of G and $\tilde{G}$.

Proposition 3 (C. Merriman)**.** *A number $x \in (0,1]$ is a periodic point of the Farey map if and only if $x = F^k(y)$, where $y \in Q_G$ and $k \in \mathbb{Z}_{\geq 0}$. Equivalently, x is a periodic point of F if and only if*

$$x = [a_1 - k, \overline{a_2, \ldots, a_n, a_1}], \tag{4.4}$$

for some $a_j \in \mathbb{N}$ and $k \in \{0, \ldots, a_1 - 1\}$.

Proof. It is clear that if x is the continued fraction (4.4), then $x = F^{a_1+\cdots+a_n}(x)$. So we only need to show that every periodic point of F can be written in the form (4.4). So assume that $x \in (0,1]$ with $F^j(x) = x$. Let $x = [a_1, a_2, \ldots]$ be the continued fraction expansion of x. Then the equality $F^j(x) = x$ is equivalent to

$$\left[\sum_{m=1}^{n} a_m - j, a_{n+1}, a_{n+2}, \ldots\right] = [a_1, a_2, a_3, \ldots],$$

where $n \in \mathbb{N}$ is the least index such that $\sum_{m=1}^{n} a_m > j$; hence $\sum_{m=1}^{n-1} a_m \leq j$. Therefore, we have $x = [a_1, \overline{a_2, \ldots, a_n}]$, and

$$a_1 = \sum_{m=1}^{n} a_m - j = a_n + \left(\sum_{m=1}^{n-1} a_m - j\right) \leq a_n,$$

that is, $a_1 = a_n - k$ for some $k \in \{0, \ldots, a_n - 1\}$. This shows that $x = F^k([\overline{a_n, a_2, a_3 \ldots, a_{n-1}}])$, which completes the proof. $\square$

Proposition 4. *A point in $[0,1]^2 \backslash \{(0,0)\}$ is a periodic point of $\tilde{F}$ if and only if it is of the form*

$$\tilde{F}^k([\overline{a_1, a_2, \ldots, a_n}], [1, \overline{a_n, a_{n-1} \ldots, a_1}]) = ([a_1 - k, \overline{a_2, a_3, \ldots, a_n, a_1}], [1 + k, \overline{a_n, a_{n-1} \ldots, a_1}]) \tag{4.5}$$

for some $a_j \in \mathbb{N}$ and $k \in \{0, \ldots, a_1 - 1\}$. In other words, with the exception of $(0,0)$, the periodic points of $\tilde{F}$ are exactly those of the form $\tilde{F}^k(x)$, where x is a periodic point of the induced map $\tilde{F}_{\tilde{A}}$.

Proof. We clearly have

$$\tilde{F}^{a_1+\cdots+a_n}([\overline{a_1,\ldots,a_n}],[1,\overline{a_n,\ldots,a_1}]) = ([\overline{a_1,\ldots,a_n}],[1,\overline{a_n,\ldots,a_1}]);$$

thus it suffices to prove that every periodic point of $\tilde{F}$ in $[0,1]^2\backslash\{(0,0)\}$ is of the form (4.5). So suppose that $(x,y) \in [0,1]^2\backslash\{(0,0)\}$ is a point such that $\tilde{F}^j(x,y) = (x,y)$ for some $j \in \mathbb{N}$. From the definition of $\tilde{F}$, it is clear that $x \neq 0$, for if $x = 0$, and hence $y \neq 0$, then the second coordinate of the sequence $(\tilde{F}^j(0,y))_{j=0}^{\infty}$ would be strictly decreasing. Also, letting $x = [a_1, a_2, \ldots]$ (possibly terminating), we have

$$\tilde{F}(x,0) = \begin{cases} ([a_1-1,a_2,\ldots],[1]) & \text{if } a_1 \geq 2 \\ ([a_2,a_3,\ldots],[1]) & \text{if } a_1 = 1, \end{cases}$$

and subsequent iterations of $(x,0)$ under $\tilde{F}$ are then determined by (3.10). If the continued fraction expansion of x does not terminate, then iterations of $\tilde{F}$ simply add continued fraction digits to the second coordinate, and periodicity never occurs. If the expansion of x does terminate, then for $k \in \mathbb{N}$ sufficiently large, $\tilde{F}^k(x,0)$ has first coordinate 0, and thus, again periodicity of the orbit of $(x,0)$ doesn't occur. Therefore, in order for (x,y) to be periodic, we must have $y \neq 0$.

So we have $(x,y) \in (0,1]^2$, and furthermore, the above shows that the first coordinate of every iterate $\tilde{F}^k(x,y)$ is nonzero, i.e., the continued fraction expansion of x is nonterminating. Now let $y = [b_0, b_1, \ldots]$. By taking the image of (x,y) under an iterate of $\tilde{F}^{-1}$, we may assume without loss of generality that $b_0 = 1$. Then $\tilde{F}^j(x,y) = (x,y)$ implies that

$$(x,y) = \left(\left[\sum_{m=1}^{n} a_m - j, a_{n+1}, a_{n+2}, \ldots\right], \left[1 + j - \sum_{m=1}^{n-1} a_m, a_{n-1}, a_{n-2}, \ldots, a_1, b_1, b_2, \ldots\right]\right),$$

where $n \in \mathbb{N}$ is the least index such that $\sum_{m=1}^{n} a_m > j$. Equating the first continued fraction digits of the second coordinates in the above equality reveals that $\sum_{m=1}^{n-1} a_m = j$; hence the first digit of the first coordinate equals a_n. Then equating the rest of the digits yields

$$x = [\overline{a_1, a_2, \ldots, a_{n-1}}] \quad \text{and} \quad y = [1, \overline{a_{n-1}, a_{n-2}, \ldots, a_1}].$$

$\square$

Notice that, analogous to the periodic points of G and $\tilde{G}$, there is a natural correspondence

$$[a_1 - k, \overline{a_2, \ldots, a_n, a_1}] \leftrightarrow ([a_1 - k, \overline{a_2, \ldots, a_n, a_1}], [1 + k, \overline{a_n, \ldots, a_1}])$$

between the periodic points of F and $\tilde{F}$. We let Q_F be the set of all nonzero periodic points of F, and for a given $x \in Q_F$, we let $\tilde{x} \in [0, 1]$ be such that $(x, \tilde{x})$ is the periodic point of $\tilde{F}$ corresponding to x.

Using Proposition 3, we extend the definition of the length function ℓ on Q_G to Q_F by letting

$$\ell(F^k(x)) := \ell(x)$$

for all $x \in Q_G$ and $k \in \mathbb{N}$. We shall see that this definition follows naturally from the correspondence between the primitive closed geodesics in the modular surface and the periodic points of the Farey map which we develop in Section 4.2. Also, define the set

$$Q_F(T) = \{x \in Q_F : \ell(x) \leq T\}.$$

We can now formulate our main theorem. It is best expressed in terms of proving the equidistribution in $[0, 1]$ of the weighted points in $Q_F(T)$ as $T \to \infty$. For $T > 0$, we define the measure m_T on $[0, 1]$ by the equality

$$\int_{[0,1]} f \, dm_T := \frac{\sum_{x \in Q_F(T)} x f(x)}{\sum_{x \in Q_F(T)} x}. \qquad (f \in C([0, 1]))$$

In other words, m_T is the sum of the Dirac measures over the points $x \in Q_F(T)$ with weight x, normalized to be a probability measure. Our main result of this chapter is the following:

Theorem 4. *For $f \in C([0, 1])$, we have*

$$\lim_{T \to \infty} \int_{[0,1]} f \, dm_T = \int_0^1 f(x) \, dx.$$

In other words, the weighted set of periodic points of F given by the measure m_T equidistributes with respect to the Lebesgue measure on $[0, 1]$.

Alternatively, one may view this theorem as saying that the unweighted periodic points of F equidistribute according to the invariant measure μ. However, one must restrict the functions over which equidistribution can be tested to those of the form $x \mapsto xf(x)$, with $f \in C([0,1])$. Additionally, one must maintain the normalizing factor $\sum_{x \in Q_F(T)} x$; because μ has infinite measure, normalizing by $\#Q_F(T)$ would yield 0 in the limit. Indeed, we see below that the growth rate of $\sum_{x \in Q_F(T)} x$ is commensurate with e^T. However, the growth of $\#Q_F(T)$ is given by

$$\#Q_F(T) = \frac{1}{4\zeta(2)} T e^T + \frac{1}{2\zeta(2)} \left(\gamma - \frac{3}{2} - \frac{\zeta'(2)}{\zeta(2)} \right) e^T + O(T^4 e^{3T/4}), \qquad (T \to \infty) \tag{4.6}$$

where ζ is the Riemann zeta function and γ is Euler's constant. This follows from the fact that

$$\#Q_F(T) = \sum_{x \in Q_G(T)} \left\lfloor \frac{1}{x} \right\rfloor$$

(for each $x = [\overline{a_1, \ldots, a_n}] \in Q_G(T)$ there exist $a_1 = \left\lfloor \frac{1}{x} \right\rfloor$ corresponding elements of $Q_F(T)$); and in analyzing the growth of the number of products of the matrices $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ with bounded trace, Kallies et al. [41] provided an asymptotic expression for the sum on the right, providing the main term in (4.6). The second term was extracted by Boca [13], who then obtained the error term $O(e^{(7/8+\epsilon)T})$. Ustinov [67] later analyzed the error term more carefully, and obtained that shown in (4.6).

Using the correspondence between the periodic points of F and $\tilde{F}$, we also obtain an analogous equidistribution result for the periodic points of $\tilde{F}$ as a corollary of Theorem 4. Define the function $h : [0,1]^2 \to \mathbb{R}$ by $h(x,y) = (x + y - xy)^2$ and the measure $\tilde{m}_T$ on $[0,1]^2$ by

$$\int_{[0,1]^2} f \, d\tilde{m}_T := \frac{\sum_{x \in Q_F(T)} h(x,\tilde{x}) f(x,\tilde{x})}{\sum_{x \in Q_F(T)} h(x,\tilde{x})}. \qquad (f \in C([0,1]^2))$$

We then have the following:

Corollary 4. *For all $f \in C([0,1]^2)$,*

$$\lim_{T \to \infty} \int_{[0,1]^2} f \, d\tilde{m}_T = \int_0^1 \int_0^1 f(x,y) \, dx \, dy,$$

that is, the weighted sequence of periodic points of $\tilde{F}$ given by $\tilde{m}_T$ equidistributes with respect to the Lebesgue measure on $[0,1]^2$.

Proof. We begin by following the reasoning of [45, Lemma 17] and showing the asymptotic formula

$$\lim_{T\to\infty} \frac{\sum_{x\in Q_F(T)} f(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} = \int_{[0,1]^2} f\, d\tilde{\mu} \tag{4.7}$$

for appropriate approximating functions f. We first verify (4.7) when f is an indicator function $1_{I_b\times I_{b'}}$, where $b=(b_1,\dots,b_n)$ and $b'=(b'_1,\dots,b'_{n'})$ are any tuples and $I_b = [\![b]\!]$ and $I_b = [\![b']\!]$ are the sets defined in Section 3.1. First note that if $B = b'_1+\cdots+b'_{n'}$, then $1_{I_b\times I_{b'}}\circ \tilde{F}^B = 1_{I_{b''}\times[0,1]}$, where

$$b'' = (1, b'_n, b'_{n-1}, \dots, b'_2, b'_1 + b_1 - 1, b_2, b_3, \dots, b_n).$$

Also, since $\tilde{F}$ forms a bijection of $\{(x,\tilde{x}) : x\in Q_F(T)\}$, we have

$$\sum_{x\in Q_F(T)} 1_{I_b\times I_{b'}}(x,\tilde{x}) = \sum_{x\in Q_F(T)} (1_{I_b\times I_{b'}}\circ \tilde{F}^j)(x,\tilde{x})$$

for any $j\in\mathbb{Z}$. Therefore, using Theorem 4, the equality (3.8), and the $\tilde{F}$-invariance of $\tilde{\mu}$, we have

$$\begin{aligned}
\lim_{T\to\infty} \frac{\sum_{x\in Q_F(T)} 1_{I_b\times I_{b'}}(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} &= \lim_{T\to\infty} \frac{\sum_{x\in Q_F(T)} (1_{I_b\times I_{b'}}\circ \tilde{F}^B)(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} \\
&= \lim_{T\to\infty} \frac{\sum_{x\in Q_F(T)} 1_{I_{b''}\times[0,1]}(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} = \lim_{T\to\infty} \frac{\sum_{x\in Q_F(T)} 1_{I_{b''}}(x)}{\sum_{x\in Q_F(T)} x} \\
&= \int_{[0,1]} 1_{I_{b''}}\, d\mu = \int_{[0,1]^2} 1_{I_{b''}\times[0,1]}\, d\tilde{\mu} = \int_{[0,1]^2} 1_{I_{b''}\times[0,1]}\circ \tilde{F}^B\, d\tilde{\mu} \\
&= \int_{[0,1]^2} 1_{I_b\times I_{b'}}\, d\tilde{\mu}.
\end{aligned}$$

We thus have (4.7) for f of the form $1_{I_b\times I_{b'}}$.

Next, (4.7) can be easily verified when $f(x,y)=x$. Also, we see in Section 4.2.3 that the set $\{(x,\tilde{x}) : x\in Q_F(T)\}$ is symmetric about the line $x=y$. As a result, (4.7) holds also when $f(x,y)=y$.

We now have a sufficient set of approximating functions, so let $f\in C([0,1]^2)$. By splitting f into its positive and negative parts, we may assume without loss of generality that $f\geq 0$. For a given $\epsilon > 0$, there exists a finite linear combination, which we denote by f_ϵ, of indicator functions of the form $1_{I_b \times I_{b'}}$ such that $f_\epsilon \le h \cdot f$ and

$$\int_{[0,1]^2} (h \cdot f - f_\epsilon)\, d\tilde{\mu} < \epsilon.$$

This is possible since the sets of the form $I_b \times I_{b'}$ generate the Borel σ-algebra of $[0,1]^2$ and

$$\int_{[0,1]^2} h \cdot f\, d\tilde{\mu} = \int_0^1 \int_0^1 f(x,y)\, dx\, dy < \infty.$$

We then have

$$\begin{aligned}\liminf_{T\to\infty} \frac{\sum_{x\in Q_F(T)} h(x,\tilde{x}) f(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} &\ge \lim_{T\to\infty} \frac{\sum_{x\in Q_F(T)} f_\epsilon(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} = \int_{[0,1]^2} f_\epsilon\, d\tilde{\mu} \\ &\ge \int_{[0,1]^2} h \cdot f\, d\tilde{\mu} - \epsilon.\end{aligned}$$

Letting $\epsilon \to 0$ yields

$$\liminf_{T\to\infty} \frac{\sum_{x\in Q_F(T)} h(x,\tilde{x}) f(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} \ge \int_{[0,1]^2} h \cdot f\, d\tilde{\mu}. \tag{4.8}$$

Now notice that $h(x,y) \cdot f(x,y) \le H(x,y) := \|f\|_\infty (x+y)$ for $(x,y) \in [0,1]^2$. Repeating the above process used to produce the inequality (4.8), while replacing the function $h \cdot f$ with $H - h \cdot f$, we find that

$$\liminf_{T\to\infty} \frac{\sum_{x\in Q_F(T)} (H - h\cdot f)(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} \ge \int_{[0,1]^2} (H - h \cdot f)\, d\tilde{\mu}.$$

Since (4.7) is satisfied when f is replaced by H, we can cancel the H in the above inequality to get

$$\limsup_{T\to\infty} \frac{\sum_{x\in Q_F(T)} h(x,\tilde{x}) f(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} \le \int_0^1 \int_0^1 f(x,y)\, dx\, dy,$$

and thus

$$\lim_{T\to\infty} \frac{\sum_{x\in Q_F(T)} h(x,\tilde{x}) f(x,\tilde{x})}{\sum_{x\in Q_F(T)} x} = \int_0^1 \int_0^1 f(x,y)\, dx\, dy.$$

Dividing this equality by the same equality, with f replaced by the constant function 1, yields the result. □

In an analogous manner to Theorem 4, one can view Corollary 4 as the equidistribution of the unweighted periodic points of $\tilde{F}$ according to the measure $\tilde{\mu}$. As before, one must restrict the functions over which to test the equidistribution and maintain the normalization $\sum_{x\in Q_F(T)} h(x,\tilde{x})$.

4.2 The modular surface and the Gauss and Farey maps

4.2.1 The modular surface and the geodesic flow

Let $\mathbb{H} = \{x+iy : x,y\in\mathbb{R}, y>0\}$ denote the upper half of the complex plane equipped with the hyperbolic metric $ds^2 = \frac{dx^2+dy^2}{y^2}$. The geodesics in $\mathbb{H}$ with this metric are vertical lines and the semicircles centered on the real line. The group $\mathrm{PSL}(2,\mathbb{R})$ acts isometrically on $\mathbb{H}$ by linear fractional transformation:

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

The modular surface is the quotient space $\mathcal{M} := \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$, whose geodesics are naturally projected from those of $\mathbb{H}$.

Let $T_1\mathbb{H}$ and $T_1\mathcal{M}$ be the unit tangent bundles of the upper half plane and modular surface, respectively. Then let $g_t : T_1\mathbb{H}\to T_1\mathbb{H}$ denote the geodesic flow on $T_1\mathbb{H}$ so that for $(z,v)\in T_1\mathbb{H}$, $g_t(z,v)$ is the tangent vector obtained by starting at the base point z, and moving a distance t along the geodesic tangent to the vector (z,v). Let $\{g_t : t\in\mathbb{R}\}$ also be defined on $T_1\mathcal{M}$ by natural projection. Next, define the coordinates (x,y,θ) on $T_1\mathbb{H}$ (and, locally, on $T_1\mathcal{M}$ by projection) corresponding to the vector with base point $x+iy$ and at an angle θ counterclockwise from the vertical upward-pointing vector. Then the g.-invariant Liouville measure $\tilde{\lambda}$, given by $d\tilde{\lambda} = \frac{dx\,dy\,d\theta}{y^2}$ is obtained by importing the Haar measure on $\mathrm{PSL}_2(\mathbb{R})$ to $T_1\mathbb{H}$ via the natural identification

$$\gamma\mapsto\gamma(i,v_0) = (\gamma(i),\gamma'(i)v_0) : \mathrm{PSL}_2(\mathbb{R}) \xrightarrow{\sim} T_1\mathbb{H}, \tag{4.9}$$

where v_0 is the upward-pointing vector at i. The Liouville measure naturally descends to $T_1\mathcal{M}$. Notice also that under the above map, $T_1\mathcal{M}$ is naturally identified with $\mathrm{PSL}(2,\mathbb{Z})\backslash\mathrm{PSL}(2,\mathbb{R}) \cong$

$\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$, and the geodesic flow g_t corresponds to right multiplication by

$$\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}.$$

Thus the functions $\tilde{g}_Q$ we considered in Section 2.6 can, in some cases, be viewed as the geodesic flow in a quotient of $T_1\mathbb{H}$.

We can use the alternative coordinates $(\alpha, \beta, t) \in \mathbb{R}^3$ on $T_1\mathbb{H}$, which correspond to the point $(z, v) \in T_1\mathbb{H}$ such that $\alpha = \alpha(z, v) := \lim_{t\to-\infty} g_t(z, v)$ is the endpoint of the geodesic $g_{(z,v)} := \{g_s(z, v) : s \in \mathbb{R}\}$ approached by (z, v) under the geodesic flow in the backward direction, $\beta = \beta(z, v) := \lim_{t\to\infty} g_t(z, v)$ is the endpoint of $g_{(z,v)}$ approached under the flow in the forward direction, and $t = t(z, v)$ is such that $g_{-t(z,v)}(z, v)$ is based at the apex of $g_{(z,v)}$. With respect to these coordinates, the Liouville measure is

$$d\tilde{\lambda} = \frac{d\alpha\, d\beta\, dt}{(\beta - \alpha)^2}.$$

4.2.2 The cross section of Series and $\tilde{G}$

The cross section of the geodesic flow considered by Series [63] is

$$X := \{(z, v) \in T_1\mathcal{M} : 0 < |\alpha(z, v)| \le 1, |\beta(z, v)| \ge 1, z \in i\mathbb{R}\}.$$

If $\beta(z, v) \ne \pm 1$, then the geodesic flow returns (z, v) to X. So the first return map P for X defined by

$$P(z, v) := g_{r(z,v)}(z, v), \quad \text{where} \quad r(z, v) := \min\{s > 0 : g_s(z, v) \in X\},$$

is well defined on $X^* := \{(z, v) \in X : |\beta(z, v)| > 1\}$. (Note that X^* is of full measure in X with respect to the measure $\frac{d\alpha\, d\beta}{(\beta-\alpha)^2}$ induced on X by $\tilde{\lambda}$.) Using the correspondence between the continued fraction expansions of $\alpha(z, v)$ and $\beta(z, v)$ and the cutting sequence of the geodesic $g_{(z,v)}$, Series proved that $\tilde{G}$ is a factor of P. Specifically, if we parameterize X by the coordinates $(U, V, \epsilon) \in (0, 1]^2 \times \{\pm 1\}$, where $U = |\beta|^{-1}$, $V = |\alpha|$, and $\epsilon = \mathrm{sign}(\beta)$, and abuse notation by identifying X with $(0,1]^2 \times \{\pm 1\}$, then

$$P(U,V,\epsilon) = (\tilde{G}(U,V), -\epsilon).$$

(Here we must also assume that $U^{-1} \notin \mathbb{Z}$.) So if we define $\pi_X : X \to [0,1]^2$ by $\pi_X(U,V,\epsilon) = (U,V)$, then we have the diagram

$$\begin{tikzcd} X^* \arrow[r, "P"] \arrow[d, "\pi_X"] & X \arrow[d, "\pi_X"] \\ {[0,1]^2} \arrow[r, "\tilde{G}"] & {[0,1]^2} \end{tikzcd}$$

which commutes a.e., and expresses $\tilde{G}$ as a factor of P.

The cross section X also allows us to see that there is a one-to-one correspondence between the closed geodesics in $T_1\mathcal{M}$ and periodic orbits of the return map P. Indeed, a closed geodesic is simply the g.-orbit in $T_1\mathcal{M}$ of one of points in a periodic orbit of P. A given periodic orbit of P is of the form

$$\{(\tilde{G}^j([\overline{a_1, a_2, \ldots, a_{2n}}], [\overline{a_{2n}, a_{2n-1}, \ldots, a_1}]), \pm(-1)^j) : j = 1, \ldots, 2n\}, \tag{4.10}$$

where $2n$ is the minimal even period length of the continued fraction $[\overline{a_1, a_2, \ldots, a_{2n}}]$. It follows from [63, Section 3.2] that the length of the geodesic corresponding to the orbit (4.10) is

$$-2 \sum_{j=1}^{2n} \log(G^j([\overline{a_1, a_2, \ldots, a_{2n}}])),$$

which inspired the definition (4.1) in [59]. Here, we wish to note that the closed geodesics corresponding to the orbits

$$\{(\tilde{G}^j([\overline{a_1, a_2, \ldots, a_{2n}}], [\overline{a_{2n}, a_{2n-1}, \ldots, a_1}]), \pm(-1)^j) : j = 1, \ldots, 2n\}, \text{ and}$$
$$\{(\tilde{G}^j([\overline{a_{2n}, a_{2n-1}, \ldots, a_1}], [\overline{a_1, a_2, \ldots, a_{2n}}]), \pm(-1)^j) : j = 1, \ldots, 2n\}$$

are permuted by the symmetry $(z,v) \mapsto (z,-v)$ on $T_1\mathcal{M}$. Indeed, the first orbit corresponds to the geodesic tangent to the element $(z, v) \in T_1\mathcal{M}$ with

$$\beta(z, v) = \pm[\overline{a_1; a_2, \ldots, a_{2n}}] \quad \text{and} \quad \alpha(z, v) = \mp[\overline{a_{2n}, a_{2n-1}, \ldots, a_1}],$$

whereas the second orbit corresponds to the geodesic tangent to the element $(z', v') \in T_1\mathcal{M}$ such that $z' \in i\mathbb{R}$,

$$\beta(z', v') = \pm[\overline{a_{2n}; a_{2n-1}, \ldots, a_1}], \quad \text{and} \quad \alpha(z', v') = \mp[\overline{a_1, a_2, \ldots, a_{2n}}].$$

Acting by the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in \mathrm{PSL}_2(\mathbb{Z})$, we see that this geodesic is the same as that tangent to $(z'', v'') \in T_1\mathbb{H}$, where $z'' = -(z')^{-1}$,

$$\beta(z'', v'') = \mp[\overline{a_{2n}, a_{2n-1}, \ldots, a_1}], \quad \text{and} \quad \alpha(z'', v'') = \pm[\overline{a_1; a_2, \ldots, a_{2n}}].$$

In other words, we have $(z'', v'') = (z, -v)$. Hence the two geodesics are the same length, i.e.,

$$\ell([\overline{a_1, a_2, \ldots, a_{2n}}]) = \ell([\overline{a_{2n}, a_{2n-1}, \ldots, a_1}]). \tag{4.11}$$

4.2.3 A new cross section for $\tilde{F}$

We now seek to find a cross section analogous to X whose return map under the geodesic flow is a double cover of $\tilde{F}$. The fact that $\tilde{G}$ is conjugate to the induced map (3.11) hints that we should seek to expand the range of the parameter V with respect to the coordinates (U, V, ϵ), or the range of α with respect to the coordinates (α, β). In fact, we define our new cross section $\bar{X}$ by

$$\bar{X} := \{(z, v) \in T_1\mathcal{M} : \alpha(z, v) \neq 0, |\beta(z, v)| \geq 1, z \in i\mathbb{R}\};$$

so we have just removed the restriction $|\alpha| \leq 1$ from the endpoint α of the geodesic determined by (z, v). One can also see that $\bar{X}$ is simply all the nonvertical tangent vectors with base point on the positive imaginary axis. We can parameterize $\bar{X}$ by the coordinates $(U, W, \epsilon) \in (0, 1] \times (0, 1) \times \{\pm 1\}$, where $W = \frac{1}{1+|\alpha|}$, and as before, $U = |\beta|^{-1}$ and $\epsilon = \mathrm{sign}(\beta)$. (We again abuse notation by identifying $\bar{X}$ with $(0, 1] \times (0, 1) \times \{\pm 1\}$.) Our definition of the coordinate W follows from the equality (3.12), which motivates us to change the second coordinates of our points in X according to the map

$$[b_1, b_2, \ldots] \mapsto [1, b_1, b_2, \ldots].$$

Now let $\bar{P} : \bar{X}^* \to \bar{X}$ be the first return map

$$\bar{P}(z,v) := g_{\bar{r}(z,v)}(z,v), \quad \text{where} \quad \bar{r}(z,v) := \min\{s > 0 : g_s(z,v) \in \bar{X}\}$$

and $\bar{X}^* = (0,1)^2 \times \{\pm 1\}$ is the set of points in $\bar{X}$ on which $\bar{P}$ is defined. Our main goal of this section is to prove the following:

Theorem 5. *The natural extension of the Farey map $\tilde{F}$ is a factor of the return map $\bar{P}$. Specifically, we have*

$$\bar{P}(U,W,\epsilon) = \begin{cases} (\tilde{F}(U,W), \epsilon) & \text{if } U \in \left(0, \frac{1}{2}\right] \\ (\tilde{F}(U,W), -\epsilon) & \text{if } U \in \left(\frac{1}{2}, 1\right) \end{cases} \tag{4.12}$$

so that if $\pi_{\bar{X}} : \bar{X} \to [0,1]^2$ is defined by $\pi_{\bar{X}}(U,W,\epsilon) = (U,W)$, we have the following commutative diagram:

$$\begin{tikzcd} \bar{X}^* \arrow[r, "\bar{P}"] \arrow[d, "\pi_{\bar{X}}"] & \bar{X} \arrow[d, "\pi_{\bar{X}}"] \\ {[0,1]^2} \arrow[r, "\tilde{F}"] & {[0,1]^2} \end{tikzcd}$$

Also, the first return time function $\bar{r}$ is given by

$$\bar{r}(U,W,\epsilon) = -\frac{1}{2}\log((1-U)(1-W)). \tag{4.13}$$

Proof. To begin, it is helpful to note that the set of base points of the vectors lifted from $\bar{X}$ to $T_1\mathbb{H}$ lie on the $\mathrm{PSL}_2(\mathbb{Z})$ translates of the positive imaginary axis, which make up the Farey tesselation. Each of these translates is either a vertical geodesic with integer real part, or a semicircle connecting rational numbers which are adjacent in a Farey sequence of some order. This is easy to see by evaluating $\lim_{t\to 0,\infty} \gamma(it)$ for $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$. The important thing to note is that all of the nonvertical translates lie below the semicircles connecting adjacent integers. Part of the Farey tessellation is depicted in Figures 4.1 and 4.2.

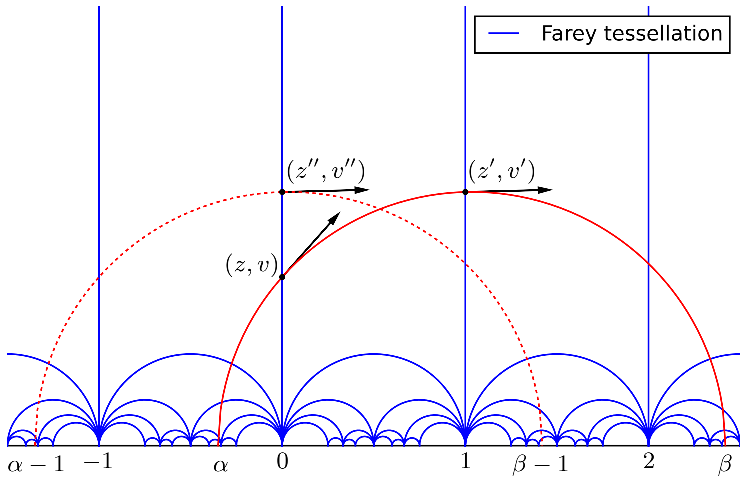

Figure 4.1: The case $U \leq \frac{1}{2}$

Now let $(z, v) \in \bar{X}^*$, which we identify with its coordinates $(U, W, \epsilon) \in (0,1)^2 \times \{\pm 1\}$, and assume for simplicity that $\epsilon = 1$. (The argument for $\epsilon = -1$ mirrors the following.) Let $U = [a_1, a_2, \ldots]$ and $W = [b_1, b_2, \ldots]$, either of which may terminate. We first consider the case when $U \in \left(0, \frac{1}{2}\right]$, and hence $a_1 \geq 2$. Then the geodesic $g_{(z,v)}$ has endpoints $\beta = \beta(z, v) = U^{-1} = [a_1; a_2, \ldots]$ and $\alpha = \alpha(z, v) = -(V^{-1} - 1) = -[b_1 - 1; b_2, \ldots]$. Since $\beta > 1$, the first point in $\bar{X}$ that (z, v) encounters under the geodesic flow is a point (z', v'), where $\mathrm{Re}(z') = 1$. The matrix $\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{PSL}(2, \mathbb{Z})$ identifies (z', v') with the point (z'', v''), where $\beta(z'', v'') = \beta - 1 = [a_1 - 1; a_2, \ldots]$ and $\alpha(z'', v'') = \alpha - 1 = -[b_1; b_2, \ldots]$. (See Figure 4.1.) The (U, W, ϵ) coordinates of this element are $U'' = [a_1 - 1, a_2, \ldots]$, $W'' = [b_1 + 1, b_2, b_3, \ldots]$, and $\epsilon'' = 1$. Thus (4.12) holds for $U \in \left(0, \frac{1}{2}\right]$.

Next, consider the case when $U \in \left(\frac{1}{2}, 1\right)$, i.e., $a_1 = 1$. As above, the corresponding geodesic $g_{(z,v)}$ has endpoints $\beta = [a_1; a_2, \ldots]$ and $\alpha = -[b_1 - 1; b_2, \ldots]$. Similar to the previous case, the first point in $\bar{X}$ that (z, v) passes through under the geodesic flow is a point (z', v'), where $\mathrm{Re}(z') = 1$. However, since $\beta(z, v) = [1; a_2, \ldots] < 2$, we must use the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right) \in \mathrm{PSL}(2, \mathbb{Z})$ to identify (z', v') with the point (z'', v'') satisfying

$$\beta(z'', v'') = \frac{-1}{\beta - 1} = \frac{-1}{[1; a_2, a_3, \ldots] - 1} = -[a_2; a_3, a_4, \ldots] \text{ and}$$
$$\alpha(z'', v'') = \frac{-1}{\alpha - 1} = \frac{-1}{-[b_1 - 1; b_2, b_3, \ldots] - 1} = [b_1, b_2, b_3, \ldots].$$

(See Figure 4.2.) The (U, W, ϵ) coordinates of (z'', v'') are $U'' = [a_2, a_3, \ldots]$, $W'' = [1, b_1, b_2, \ldots]$,

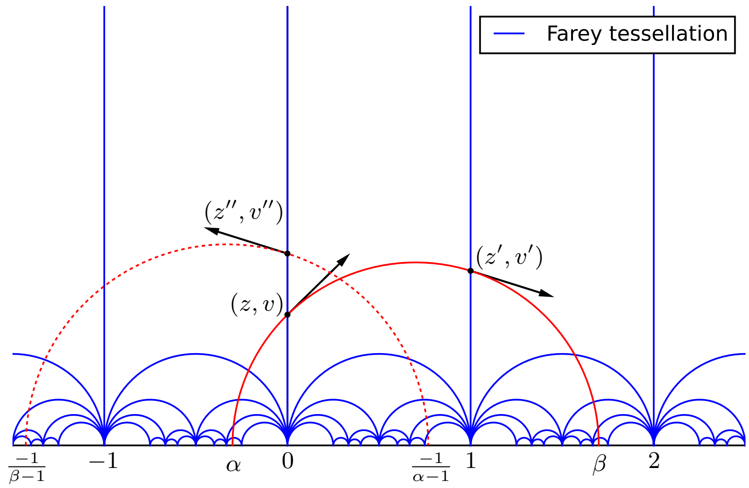

Figure 4.2: The case $U > \frac{1}{2}$

and $\epsilon'' = -1$, which shows (4.12) for $U \in \left(\frac{1}{2}, 1\right)$. This proves that $\tilde{F}$ is a factor of $\bar{P}$.

We now outline the calculation of the return time function $\bar{r}$. As above, let $(z, v) \in \bar{X}^*$ with coordinates $(U, V, \epsilon) \in (0, 1)^2 \times \{\pm 1\}$, and again assume that $\epsilon = 1$, since the case $\epsilon = -1$ is a mirror image. Also, let $z = iy$ for $y \in \mathbb{R}, y > 0$, and $\theta \in (0, \pi)$ be the angle v makes with the upward-pointing vector in the counterclockwise direction. Then under the identification (4.9), (z, v) is identified with

$$\begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}.$$

By the previous part of the proof, $\bar{r}(U, V, \epsilon)$ is the constant $t > 0$ such that the base point of $g_t(z, v) = (M(i), M'(i)v_0)$, where

$$M = \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix},$$

has real part equal to 1. By a straightforward calculation, this implies that

$$\bar{r}(U, V, \epsilon) = \frac{1}{2} \log \left(\frac{y \sin(\theta/2)\cos(\theta/2) + \cos^2(\theta/2)}{y \sin(\theta/2)\cos(\theta/2) - \sin^2(\theta/2)} \right).$$

Then, using the fact that $U = y^{-1} \tan\frac{\theta}{2}$ and $W = \frac{1}{1+y\tan(\theta/2)}$, we easily deduce (4.13). $\square$

Remark 2. See [3, Section 7] for a different way of relating the Farey map to a cross section of the geodesic flow in $T^1\mathcal{M}$.

Hence, we can see how a given periodic orbit

$$\left\{ \bar{P}^j([\overline{a_1, a_2, \ldots, a_{2n}}], [1, \overline{a_{2n}, a_{2n-1}, \ldots, a_1}], \pm 1) : j = 1, \ldots, \sum_{k=1}^{2n} a_k \right\} \tag{4.14}$$

of $\bar{P}$ naturally includes the periodic orbit (4.10) of P, and so corresponds to the same closed geodesic. We therefore extend our definition of the length function ℓ to the periodic points of F so that for all $k \in \mathbb{N}$, $\ell(F^k([\overline{a_1, \ldots, a_{2n}}]))$ is the length of the closed geodesic given by the $g_\cdot$-orbit of any point in (4.14), i.e.,

$$\ell(F^k([\overline{a_1, \ldots, a_{2n}}])) = -2 \sum_{j=1}^{2n} \log(G^j([\overline{a_1, \ldots, a_{2n}}])).$$

Note here that if $x = [\overline{a_1, \ldots, a_{2n}}]$ has minimal even periodic length $2n$ and $k \in \{0, \ldots, a_1 - 1\}$, then letting $y = F^k(x)$, we have

$$\begin{aligned} \ell(\tilde{y}) &= \ell([1 + k, \overline{a_{2n}, a_{2n-1}, \ldots, a_1}]) = \ell(F^{a_1 - 1 - k}([\overline{a_1, a_{2n}, a_{2n-1}, \ldots, a_2}])) \\ &= \ell([\overline{a_1, a_{2n}, a_{2n-1}, \ldots, a_2}]) = \ell([\overline{a_{2n}, a_{2n-1}, \ldots, a_1}]) = \ell([\overline{a_1, a_2, \ldots, a_{2n}}]) = \ell(y), \end{aligned}$$

where we used (4.11) for the penultimate equality. This shows that for any $T > 0$, the set $\{(x, \tilde{x}) : x \in Q_F(T)\}$ is symmetric about the line $x = y$, a fact we used in the proof of Corollary 4.

To conclude this section, we notice that the measure $\frac{d\alpha\, d\beta}{(\beta - \alpha)^2}$ on $\bar{X}$ induced by the Liouville measure is, in the coordinates (U, W) (with ϵ fixed as 1 or -1),

$$-\frac{d(U^{-1})\, d(-(W^{-1} - 1))}{(U^{-1} + (W^{-1} - 1))^2} = \frac{U^{-2} W^{-2} dU\, dV}{(U^{-1} + W^{-1} - 1)^2} = \frac{dU\, dW}{(U + W - UW)^2}.$$

Thus the Liouville measure naturally induces the invariant measure $\tilde{\mu}$ of $\tilde{F}$ on $[0, 1]^2$.

4.3 Proof of equidistribution

We now set out to prove Theorem 4. The result follows directly from the following asymptotic formula we aim to show, and which holds for all $f \in C([0,1])$:

$$\sum_{x \in Q_F(T)} xf(x) \sim \left(\frac{3}{\pi^2} \int_0^1 f(x)\, dx \right) e^T \qquad (T \to \infty) \tag{4.15}$$

In Section 4.3.1, we reduce the proof of (4.15) to showing a similar asymptotic formula for sums of certain analytic functions over the periodic points of the Gauss map. From that point we adapt the work of Pollicott [59] on a Ruelle-Perron-Frobenius operator of the Gauss map. We introduce this operator in Section 4.3.2 and establish its nuclearity and analyticity with respect to certain parameters in Section 4.3.3. Then in Section 4.3.4, we utilize the Fredholm determinants of the operator to construct a certain η function which is the Laplace transform of an approximation $\tilde{S}_f$ to a sum of the form

$$S_f(T) := \sum_{x \in Q_G(T)} f(x).$$

We then calculate the residue of a pole of the η function and apply the Wiener-Ikehara Tauberian theorem to determine the asymptotic growth rate of $\tilde{S}_f$. We conclude the proof in Section 4.3.5 by showing that $\tilde{S}_f$ is in fact asymptotically equivalent to S_f, which uses a fact due to Kelmer [45, Theorem 3] that the sum of f over the odd periodic continued fractions grows asymptotically slower than that over the even.

4.3.1 Reduction to equidistribution of $Q_G(T)$

First, let $f \in C([0,1])$, and assume without loss of generality that f is real valued and nonnegative. Then let $g : [0,1] \to \mathbb{R}$ be defined by $g(x) = xf(x)$. Notice that for any $T > 0$, we have

$$\begin{aligned}
\sum_{x \in Q_F(T)} xf(x) &= \sum_{[\overline{a_1,\ldots,a_{2n}}] \in Q_G(T)} \sum_{k=0}^{a_1-1} g([a_1 - k, \overline{a_2, \ldots, a_{2n}, a_1}]) \\
&= \sum_{[\overline{a_1,\ldots,a_{2n}}] \in Q_G(T)} \sum_{k=0}^{a_1-1} (g \circ F^k)([\overline{a_1, \ldots, a_{2n}}]) = \sum_{x \in Q_G(T)} \sum_{k=0}^{\lfloor 1/x \rfloor - 1} (g \circ F^k)(x).
\end{aligned}$$

So understanding the sum of g over $Q_F(T)$ is equivalent to understanding the sum of the function $\bar{g} : (0,1] \to \mathbb{R}$ defined by

$$\bar{g}(x) = \sum_{k=0}^{\lfloor 1/x \rfloor - 1} (g \circ F^k)(x)$$

over $Q_G(T)$. Notice also that

$$\begin{aligned}
\int_{[0,1]} \bar{g}\, d\nu &= \sum_{n=1}^{\infty} \int_{1/(n+1)}^{1/n} \sum_{k=0}^{n-1} (g \circ F^k)(x) \frac{dx}{(1+x)\log 2} = \sum_{k=0}^{\infty} \sum_{n=k+1}^{\infty} \int_{1/(n+1)}^{1/n} g\left(\frac{x}{1-kx}\right) \frac{dx}{(1+x)\log 2} \\
&= \sum_{k=0}^{\infty} \int_0^{1/(k+1)} g\left(\frac{x}{1-kx}\right) \frac{dx}{(1+x)\log 2} \\
&= \sum_{k=0}^{\infty} \int_0^1 g(y) \frac{dy}{(1+ky)^2(1+y/(1+ky))\log 2} \qquad\qquad \left(y = \frac{x}{1-kx}\right) \\
&= \frac{1}{\log 2} \int_0^1 g(y) \left(\sum_{k=0}^{\infty} \frac{1}{(1+ky)(1+(k+1)y)} \right) dy \\
&= \frac{1}{\log 2} \int_{[0,1]} g\, d\mu = \frac{1}{\log 2} \int_0^1 f(x)\, dx.
\end{aligned}$$

Thus the asymptotic formula (4.15) is equivalent to

$$\sum_{x \in Q_G(T)} \bar{g}(x) \sim \left(\frac{3 \log 2}{\pi^2} \int_{[0,1]} \bar{g}\, d\nu \right) e^T. \qquad (T \to \infty)$$

If $\bar{g}$ had an extension to a function in $C([0,1])$, this asymptotic formula would follow from the work of Pollicott. However, in general, $\bar{g}(x)$ has discontinuities at the points $x \in \{n^{-1} : n \in \mathbb{N}, n \geq 2\}$ and can grow without bound as $x \to 0^+$, and we must take these facts into account. On the other hand, the following lemma essentially shows that we can assume that $\bar{g}$ is an analytic function of a particular form.

Lemma 3. *The function $\bar{g}(x) = \sum_{k=0}^{\lfloor 1/x \rfloor - 1} (g \circ F^k)(x)$ can be approximated arbitrarily closely in $L^1(\nu)$ from above and below by functions of the form $p(x) = p_1(x)(1 - \log x)$, where p_1 is a polynomial.*

Proof. By the mutual absolute continuity between ν and the Lebesgue measure λ on $[0,1]$, it suffices to prove the approximation in the L^1 norm with respect to λ. It is clear that $\bar{\bar{g}}(x) := \frac{\bar{g}(x)}{1-\log x}$ is uniformly bounded on $(0,1]$, and it is also continuous except possibly at the points $\{n^{-1} : n \in$

$\mathbb{N}, n \geq 2\}$, where there could be jump discontinuities. For any $\epsilon > 0$, one can clearly find continuous functions $h_{1,\epsilon}, h_{2,\epsilon} : [0,1] \to \mathbb{R}$ such that $h_{1,\epsilon} \leq \bar{\bar{g}} \leq h_{2,\epsilon}$, $\|h_{1,\epsilon}\|_\infty, \|h_{2,\epsilon}\|_\infty \leq \|\bar{\bar{g}}\|_\infty$,

$$\int_0^1 (\bar{\bar{g}}(x) - h_{1,\epsilon}(x))\, dx \leq \epsilon, \quad \text{and} \quad \int_0^1 (h_{2,\epsilon}(x) - \bar{\bar{g}}(x))\, dx \leq \epsilon.$$

Then by the Stone-Weierstrass theorem, there exist polynomials $p_{1,\epsilon}, p_{2,\epsilon}$ such that $h_{1,\epsilon}(x) - \epsilon \leq p_{1,\epsilon}(x) \leq h_{1,\epsilon}(x)$ and $h_{2,\epsilon}(x) \leq p_{2,\epsilon}(x) \leq h_{2,\epsilon}(x) + \epsilon$ for $x \in [0,1]$. We then have

$$\begin{aligned}
\int_0^1 [\bar{g}(x) - p_{1,\epsilon}(x)(1-\log x)]\, dx &= \int_0^1 (\bar{\bar{g}}(x) - p_{1,\epsilon}(x))(1-\log x)\, dx \\
&= \int_0^1 (\bar{\bar{g}}(x) - h_{1,\epsilon}(x))(1-\log x)\, dx + \int_0^1 (h_{1,\epsilon}(x) - p_{1,\epsilon}(x))(1-\log x)\, dx \\
&\leq \int_0^\epsilon 2\|\bar{\bar{g}}\|_\infty (1-\log x)\, dx + (1+\log \epsilon^{-1}) \int_\epsilon^1 (\bar{\bar{g}}(x) - h_{1,\epsilon}(x))\, dx + \epsilon \int_0^1 (1-\log x)\, dx \\
&\leq 2\|\bar{\bar{g}}\|_\infty \epsilon(2-\log\epsilon) + \epsilon(1+\log\epsilon^{-1}) + 2\epsilon.
\end{aligned}$$

The last expression above approaches 0 as $\epsilon \to 0$. Thus, we can conclude that $\bar{g}(x)$ can be approximated arbitrarily closely in $L^1(\lambda)$ from below by functions of the desired form. Similarly, using $p_{\epsilon,2}$, one can show the approximation of $\bar{g}(x)$ by such functions from above. □

By this approximation result, we have reduced the proof to showing that

$$S_f(T) = \sum_{x \in Q_G(T)} f(x) \sim \left(\frac{3\log 2}{\pi^2} \int_{[0,1]} f\, d\nu \right) e^T. \qquad (T \to \infty)$$

for functions $f : (0,1] \to \mathbb{R}$ of the form $f(x) = p(x)(1-\log x)$ where p is a polynomial.

4.3.2 The Ruelle-Perron-Frobenius operator

We now begin following [59], as well as [22] and [45], in proving the above growth rate by analyzing a Ruelle-Perron-Frobenius operator of the Gauss map. For $r > 0$, let $D_r = \{z \in \mathbb{C} : |z-1| < r\}$ and $\overline{D}_r$ the closure of D_r, and let $a, b \in \mathbb{R}$ be any fixed constants satisfying $1 < a < b < \frac{3}{2}$. The Ruelle-Perron-Frobenius operator we define below acts on the disk algebra, which we denote by $\mathbb{A}$ and view as the set of continuous functions on $\overline{D}_a$ which are analytic in D_a, equipped with the supremum norm $\|\cdot\|_\infty$. Let f be a complex valued function of the form $f(z) = f_1(z)(1-\log z)$, where $\log z$ is the principal branch of the logarithm and f_1 is analytic in an open neighborhood of $\overline{D}_1$ and real valued and positive on $[0,1]$. Then f is analytic in an open neighborhood of $\overline{D}_1\backslash\{0\}$, and letting $K = \|f_1\|_\infty(1+\pi)$, we have $|f(z)| \le K(1-\log|z|)$ for all $z \in \overline{D}_1\backslash\{0\}$. Then letting $\chi_{s,\omega}(z) := z^{2s}e^{\omega f(z)}$ for $s,\omega \in \mathbb{C}$ (the $2s$ power coming from the principle branch of the logarithm), we define the Ruelle-Perron-Frobenius operator $L_{s,\omega} : \mathbb{A} \to \mathbb{A}$ by

$$(L_{s,\omega}g)(z) := \sum_{n=1}^{\infty} g\left(\frac{1}{z+n}\right)\chi_{s,\omega}\left(\frac{1}{z+n}\right).$$

Note that the particular operator $L_{1,0}$ is the transfer operator $\hat{G}_\lambda$ of the Gauss map. Letting $\mathcal{U} := \left\{(s,\omega) \in \mathbb{C}^2 : \mathrm{Re}(s) > \frac{1+K|\omega|}{2}\right\}$, we see that $L_{s,\omega}$ is a well defined and bounded operator for $(s,\omega) \in \mathcal{U}$ (with $s = \sigma + it$ for $\sigma, t \in \mathbb{R}$) by the calculation

$$\begin{aligned}\sum_{n=1}^{\infty}\left|g\left(\frac{1}{z+n}\right)\chi_{s,\omega}\left(\frac{1}{z+n}\right)\right| &\le \|g\|_\infty \sum_{n=1}^{\infty}\left|\frac{1}{(z+n)^{2s}}\right|\left|e^{\omega f(1/(z+n))}\right| \\ &= \|g\|_\infty \sum_{n=1}^{\infty} e^{\mathrm{Re}((2\sigma+2it)(-\log|z+n|-i\arg(z+n)))}e^{\mathrm{Re}(\omega f(1/(z+n)))} \\ &\le \|g\|_\infty \sum_{n=1}^{\infty} e^{-2\sigma\log|z+n|+2t\arg(z+n)}e^{|\omega|K(1+\log|z+n|)} \\ &\le \|g\|_\infty \sum_{n=1}^{\infty}\frac{e^{2\pi t+K|\omega|}}{|z+n|^{2\sigma-K|\omega|}} \le \|g\|_\infty \sum_{n=1}^{\infty}\frac{e^{2\pi t+K|\omega|}}{(n-\frac{1}{2})^{2\sigma-K|\omega|}}.\end{aligned} \tag{4.16}$$

4.3.3 $L_{s,\omega}$ is nuclear of order 0 and analytic

We now closely follow the arguments of Faivre [22] to show that for $(s,\omega) \in \mathcal{U}$, $L_{s,\omega}$ is a nuclear operator of order 0, and is analytic in (s,ω). We begin with nuclearity.

For a given $\epsilon > 0$, we wish to find a sequence $\{\Lambda_j \otimes e_j\}_{j=0}^{\infty} \subseteq \mathbb{A}^* \otimes \mathbb{A}$ such that

$$L_{s,\omega} = \sum_{j=0}^{\infty}\Lambda_j \otimes e_j, \quad \text{and} \quad \sum_{j=0}^{\infty}\|\Lambda_j\|^\epsilon\|e_j\|_\infty^\epsilon < \infty,$$

where $\Lambda_j \otimes e_j$ is defined as an operator on $\mathbb{A}$ by $((\Lambda_j \otimes e_j)g)(z) = (\Lambda_j g)e_j(z)$. Assume $(s,\omega) \in \mathcal{U}$ so that $\sigma > \frac{1+K|\omega|}{2}$, fix $g \in \mathbb{A}$, and for each $n \in \mathbb{N}$ let

$$(L_{n,s,\omega}g)(z) := h_{g,n,s,\omega}(z) := g\left(\frac{1}{z+n}\right)\chi_{s,\omega}\left(\frac{1}{z+n}\right).$$

It is easy to see that $h_{g,n,s,\omega}$ is an analytic function on $D_{3/2}$, and so for $z \in \overline{D}_a \subseteq D_b$,

$$h_{g,n,s,\omega}(z) = \sum_{j=0}^{\infty} \frac{h^{(j)}_{g,n,s,\omega}(1)}{j!}(z-1)^j.$$

Define the element $\Lambda_{n,j,s,\omega} \in \mathbb{A}^*$ by

$$\Lambda_{n,j,s,\omega} g := \frac{h^{(j)}_{g,n,s,\omega}(1)}{j!} = \frac{1}{2\pi i} \int\limits_{|\zeta-1|=b} \frac{h_{g,n,s,\omega}(\zeta)}{(\zeta-1)^{j+1}} d\zeta;$$

the latter equality follows from Cauchy's formula. Also define $e_j \in \mathbb{A}$ by $e_j(z) = (z-1)^j$ so that

$$h_{g,n,s,\omega} = \sum_{j=0}^{\infty} (\Lambda_{n,j,s,\omega} g) e_j.$$

By the calculation (4.16), we have

$$\|h_{g,n,s,\omega}\|_\infty \leq \frac{\|g\|_\infty e^{2\pi t + K|\omega|}}{(n-\frac{1}{2})^{2\sigma - K|\omega|}},$$

and therefore

$$|\Lambda_{n,j,s,\omega} g| \leq \frac{\|h_{g,n,s,\omega}\|_\infty}{2\pi} \int\limits_{|\zeta-1|=b} \frac{1}{|\zeta-1|^{j+1}} |d\zeta| = \frac{\|g\|_\infty e^{2\pi t + K|\omega|}}{b^j (n-\frac{1}{2})^{2\sigma - K|\omega|}}.$$

This implies that

$$\sum_{n=1}^{\infty} \|\Lambda_{n,j,s,\omega}\| \leq \sum_{n=1}^{\infty} \frac{e^{2\pi t + K|\omega|}}{b^j (n-\frac{1}{2})^{2\sigma - K|\omega|}},$$

which is a finite quantity, and we denote its product with b^j by $\kappa(s,\omega)$. Hence

$$\Lambda_{j,s,\omega} := \sum_{n=1}^{\infty} \Lambda_{n,j,s,\omega}$$

is a well defined element of $\mathbb{A}^*$.

Next, notice that

$$\sum_{j=0}^{\infty}\sum_{n=1}^{\infty}\|\Lambda_{n,j,s,\omega}\|\|e_j\|_\infty \leq \sum_{j=0}^{\infty}\sum_{n=1}^{\infty}\frac{e^{2\pi t+K|\omega|}}{b^j(n-\frac{1}{2})^{2\sigma-K|\omega|}}a^j = \kappa(s,\omega)\sum_{j=0}^{\infty}\left(\frac{a}{b}\right)^j = \frac{\kappa(s,\omega)}{1-(a/b)} < \infty.$$

As a result, we have

$$L_{s,\omega} = \sum_{n=1}^{\infty} L_{n,s,\omega} = \sum_{n=1}^{\infty}\sum_{j=0}^{\infty}\Lambda_{n,j,s,\omega}\otimes e_j = \sum_{j=0}^{\infty}\sum_{n=1}^{\infty}\Lambda_{n,j,s,\omega}\otimes e_j = \sum_{j=0}^{\infty}\Lambda_{j,s,\omega}\otimes e_j$$

Finally, note that for all $\epsilon > 0$,

$$\sum_{j=0}^{\infty}\|\Lambda_{j,s,\omega}\|^\epsilon\|e_j\|^\epsilon \leq \sum_{j=0}^{\infty}\frac{\kappa(s,\omega)^\epsilon}{b^{\epsilon j}}a^{\epsilon j} = \frac{\kappa(s,\omega)^\epsilon}{1-(a/b)^\epsilon} < \infty.$$

This completes the proof that $L_{s,\omega}$ is nuclear of order 0.

We now prove the analyticity of $L_{s,\omega}$. We specifically show that $(s,\omega)\mapsto L_{s,\omega}$ is analytic as a map from $\mathcal{U}$ to the order ϵ nuclear operators on $\mathbb{A}$ for any $\epsilon > 0$. Throughout the proof, we fix $(s_0,\omega_0)\in\mathcal{U}$ at which we prove the differentiability of $L_{s,\omega}$. Also, let $s_0 = \sigma_0 + it_0$ for $\sigma_0, t_0 \in \mathbb{R}$.

For our first step, we fix $n, j \in \mathbb{N}\cup\{0\}$ with $n \geq 1$, and show that $(s,\omega)\mapsto \Lambda_{n,j,s,\omega}$ is differentiable (as a map from $\mathcal{U}$ to $\mathbb{A}^*$) at (s_0,ω_0) with respect to s. Define the element $\theta^{[1]}_{n,j,s,\omega}\in\mathbb{A}^*$ by

$$\theta^{[1]}_{n,j,s,\omega}g = \frac{1}{2\pi i}\int\limits_{|\zeta-1|=b}\frac{-2h_{g,n,s,\omega}(\zeta)\log(\zeta+n)}{(\zeta-1)^{j+1}}d\zeta.$$

We aim to show that $\theta^{[1]}_{n,j,s,\omega} = \frac{\partial}{\partial s}\Lambda_{n,j,s,\omega}$. For $(s,\omega_0)\in\mathcal{U}$ and $g\in\mathbb{A}$, we have

$$\begin{aligned}
&\left|\left(\frac{\Lambda_{n,j,s,\omega_0}-\Lambda_{n,j,s_0,\omega_0}}{s-s_0} - \theta^{[1]}_{n,j,s_0,\omega_0}\right)g\right| \\
&\quad = \left|\frac{1}{2\pi i}\int\limits_{|\zeta-1|=b}\frac{g((\zeta+n)^{-1})e^{\omega_0 f((\zeta+n)^{-1})}}{(\zeta-1)^{j+1}}\left(\frac{(\zeta+n)^{-2s}-(\zeta+n)^{-2s_0}}{s-s_0}+\frac{2\log(\zeta+n)}{(\zeta+n)^{2s_0}}\right)d\zeta\right| \\
&\quad = \left|\frac{1}{2\pi i}\int\limits_{|\zeta-1|=b}\frac{g((\zeta+n)^{-1})e^{\omega_0 f((\zeta+n)^{-1})}}{(\zeta-1)^{j+1}(\zeta+n)^{2s_0}}\left(\frac{(\zeta+n)^{-2(s-s_0)}-1}{s-s_0}+2\log(\zeta+n)\right)d\zeta\right|. \qquad (4.17)
\end{aligned}$$

Notice that for a function ψ that is analytic in an open neighborhood N of 0, and for $h\in\mathbb{C}$ such that the straight line segment $[0,h]$ connecting 0 and h is in N, we have

$$\begin{aligned}
|\psi(h)-\psi(0)-h\psi'(0)| &= \left|\int_{[0,h]}\psi(u)\,du - h\psi'(0)\right| = \left|\int_{[0,h]}(\psi'(u)-\psi'(0))du\right| \\
&= \left|\int_{[0,h]}\frac{\psi'(u)-\psi'(0)}{u}u\,du\right| \le \int_{[0,h]}\max_{u'\in[0,u]}|\psi''(u')||u||du| \\
&\le \left(\max_{u\in[0,h]}|\psi''(u)|\right)\int_0^{|h|}x\,dx = \frac{|h|^2}{2}\max_{u\in[0,h]}|\psi''(u)|. \qquad (4.18)
\end{aligned}$$

This implies that

$$\begin{aligned}
\left|\frac{(\zeta+n)^{-2(s-s_0)}-1}{s-s_0}+2\log(\zeta+n)\right| &\le \frac{|s-s_0|}{2}\max_{u\in[0,s-s_0]}\left|\frac{4(\log(\zeta+n))^2}{(\zeta+n)^{2u}}\right| \\
&\le 2|s-s_0|(\pi+\log(n+1+b))^2 e^{2\pi|\operatorname{Im}(s-s_0)|}(n+1+b)^{2|\operatorname{Re}(s-s_0)|}.
\end{aligned}$$

Hence, (4.17) is at most

$$\frac{\|g\|_\infty e^{2\pi t_0+K|\omega_0|}}{b^j(n-\frac{1}{2})^{2\sigma_0-K|\omega_0|}}2|s-s_0|(\pi+\log(n+1+b))^2 e^{2\pi|\operatorname{Im}(s-s_0)|}(n+1+b)^{2|\operatorname{Re}(s-s_0)|},$$

and therefore

$$\begin{aligned}
&\left\|\frac{\Lambda_{n,j,s,\omega_0}-\Lambda_{n,j,s_0,\omega_0}}{s-s_0}-\theta^{[1]}_{n,j,s_0,\omega_0}\right\| \\
&\qquad\le \frac{2(\pi+\log(n+1+b))^2 e^{2\pi t_0+K|\omega_0|+2\pi|\operatorname{Im}(s-s_0)|}(n+1+b)^{2|\operatorname{Re}(s-s_0)|}}{b^j(n-\frac{1}{2})^{2\sigma_0-K|\omega_0|}}|s-s_0|,
\end{aligned}$$

which approaches 0 as $s\to s_0$. This proves that $\theta^{[1]}_{n,j,s,\omega}=\frac{\partial}{\partial s}\Lambda_{n,j,s,\omega}$.

Next, notice that

$$\sum_{n=1}^{\infty}\|\theta^{[1]}_{n,j,s,\omega}\| \le \sum_{n=1}^{\infty}\frac{e^{2\pi t+K|\omega|}2(\pi+\log(n+1+b))}{b^j(n-\frac{1}{2})^{2\sigma-K|\omega|}},$$

which is finite for $(s,\omega)\in\mathcal{U}$; and hence

$$\theta^{[1]}_{j,s,\omega} := \sum_{n=1}^{\infty}\theta^{[1]}_{n,j,s,\omega}$$

is a well defined element of $\mathbb{A}^*$ for all $j \in \mathbb{N} \cup \{0\}$. We then have

$$\left\| \frac{\Lambda_{j,s,\omega_0} - \Lambda_{j,s_0,\omega_0}}{s - s_0} - \theta^{[1]}_{j,s_0,\omega_0} \right\| \\ \leq \sum_{n=1}^{\infty} \frac{2(\pi + \log(n+1+b))^2 e^{2\pi t_0 + K|\omega_0| + 2\pi|\operatorname{Im}(s-s_0)|}(n+1+b)^{2|\operatorname{Re}(s-s_0)|}}{b^j (n - \frac{1}{2})^{2\sigma_0 - K|\omega_0|}} |s - s_0|,$$

which is finite as long as s is close enough to s_0 so that $2\sigma_0 - K|\omega_0| - 2|\operatorname{Re}(s - s_0)| > 1$. Such s comprise an open neighborhood of s_0 since $2\sigma_0 - K|\omega_0| > 1$. It is thus clear that as $s \to s_0$, the left side above approaches 0. This proves that $\frac{\partial}{\partial s}\Lambda_{j,s,\omega}$ exists and equals $\theta^{[1]}_{j,s,\omega}$.

Next fix $\epsilon > 0$. Then notice that

$$\sum_{j=0}^{\infty} \|\theta^{[1]}_{j,s,\omega}\|^{\epsilon} \|e_j\|^{\epsilon} \leq \sum_{j=0}^{\infty} \left(\sum_{n=1}^{\infty} \frac{e^{2\pi t + K|\omega|} 2(\pi + \log(n+1+b))}{b^j (n - \frac{1}{2})^{2\sigma - K|\omega|}} \right)^{\epsilon} a^{j\epsilon} \\ = \left(\sum_{n=1}^{\infty} \frac{e^{2\pi t + K|\omega|} 2(\pi + \log(n+1+b))}{(n - \frac{1}{2})^{2\sigma - K|\omega|}} \right)^{\epsilon} \frac{1}{1 - (a/b)^{\epsilon}} < \infty,$$

for $(s, \omega) \in \mathcal{U}$, implying that

$$\theta^{[1]}_{s,\omega} := \sum_{j=0}^{\infty} \theta^{[1]}_{j,s,\omega} \otimes e_j$$

is a nuclear operator of order ϵ. We also have

$$\sum_{j=0}^{\infty} \left\| \frac{\Lambda_{j,s,\omega_0} - \Lambda_{j,s_0,\omega_0}}{s - s_0} - \theta^{[1]}_{j,s_0,\omega_0} \right\|^{\epsilon} \|e_j\|^{\epsilon} \\ \leq \sum_{j=0}^{\infty} \left(\sum_{n=1}^{\infty} \frac{2(\pi + \log(n+1+b))^2 e^{2\pi t_0 + K|\omega_0| + 2\pi|\operatorname{Im}(s-s_0)|}(n+1+b)^{2|\operatorname{Re}(s-s_0)|}}{b^j (n - \frac{1}{2})^{2\sigma_0 - K|\omega_0|}} |s - s_0| \right)^{\epsilon} a^{j\epsilon} \\ = |s - s_0|^{\epsilon} \left(\sum_{n=1}^{\infty} \frac{2(\pi + \log(n+1+b))^2 e^{2\pi t_0 + K|\omega_0| + 2\pi|\operatorname{Im}(s-s_0)|}(n+1+b)^{2|\operatorname{Re}(s-s_0)|}}{(n - \frac{1}{2})^{2\sigma_0 - K|\omega_0|}} \right)^{\epsilon} \frac{1}{1 - (a/b)^{\epsilon}},$$

which again is finite as long as $2\sigma_0 - K|\omega_0| - 2|\operatorname{Re}(s - s_0)| > 1$, and approaches 0 as $s \to s_0$. This means that

$$\frac{L_{s,\omega_0} - L_{s_0,\omega_0}}{s - s_0} - \theta^{[1]}_{s_0,\omega_0} = \sum_{j=0}^{\infty} \left(\frac{\Lambda_{j,s,\omega_0} - \Lambda_{j,s_0,\omega_0}}{s - s_0} - \theta^{[1]}_{j,s_0,\omega_0} \right) \otimes e_j$$

approaches 0 in the norm on order ϵ nuclear operators. This completes the proof that $(s, \omega) \mapsto L_{s,\omega}$ is analytic with respect to s as a map to the order ϵ nuclear operators ($\forall \epsilon > 0$) and $\frac{\partial}{\partial s} L_{s,\omega} = \theta^{[1]}_{s,\omega}$.

We will now essentially repeat the above argument to show that

$$\frac{\partial}{\partial \omega} L_{s,\omega} = \sum_{j=0}^{\infty} \sum_{n=1}^{\infty} \theta^{[2]}_{n,j,s,\omega} \otimes e_j,$$

where

$$\theta^{[2]}_{n,j,s,\omega} g = \frac{1}{2\pi i} \int\limits_{|\zeta-1|=b} \frac{h_{g,n,s,\omega}(\zeta) f((\zeta+n)^{-1})}{(\zeta-1)^{j+1}}\, d\zeta.$$

For $(s_0, \omega) \in \mathcal{U}$ and $g \in \mathbb{A}$, we have

$$\begin{aligned}
&\left| \left(\frac{\Lambda_{n,j,s_0,\omega} - \Lambda_{n,j,s_0,\omega_0}}{\omega - \omega_0} - \theta^{[2]}_{n,j,s_0,\omega_0} \right) g \right| \\
&= \left| \frac{1}{2\pi i} \int\limits_{|\zeta-1|=b} \frac{g((\zeta+n)^{-1})}{(\zeta-1)^{j+1}(\zeta+n)^{2s_0}} \left(\frac{e^{\omega f((\zeta+n)^{-1})} - e^{\omega_0 f((\zeta+n)^{-1})}}{\omega - \omega_0} - e^{\omega_0 f((\zeta+n)^{-1})} f\left(\frac{1}{\zeta+n}\right) \right) d\zeta \right| \\
&= \left| \frac{1}{2\pi i} \int\limits_{|\zeta-1|=b} \frac{g((\zeta+n)^{-1}) e^{\omega_0 f((\zeta+n)^{-1})}}{(\zeta-1)^{j+1}(\zeta+n)^{2s_0}} \left(\frac{e^{(\omega-\omega_0) f((\zeta+n)^{-1})} - 1}{\omega - \omega_0} - f\left(\frac{1}{\zeta+n}\right) \right) d\zeta \right| \qquad (4.19)
\end{aligned}$$

Using (4.18), we have

$$\begin{aligned}
\left| \frac{e^{(\omega-\omega_0) f((\zeta+n)^{-1})} - 1}{\omega - \omega_0} - f\left(\frac{1}{\zeta+n}\right) \right| &\leq \frac{|\omega-\omega_0|}{2} \max_{u \in [0, \omega-\omega_0]} \left| e^{u f((\zeta+n)^{-1})} f\left(\frac{1}{\zeta+n}\right)^2 \right| \\
&\leq \frac{1}{2} |\omega-\omega_0| K^2 (1 + \log(n+1+b))^2 e^{|\omega-\omega_0| K (1+\log(n+1+b))}.
\end{aligned}$$

Therefore, (4.19) is at most

$$\frac{\|g\|_\infty e^{2\pi t_0 + K|\omega_0|}}{2b^j (n-\frac{1}{2})^{2\sigma_0 - K|\omega_0|}} |\omega-\omega_0| K^2 (1+\log(n+1+b))^2 e^{K|\omega-\omega_0|} (n+1+b)^{K|\omega-\omega_0|},$$

and so

$$\begin{aligned}
&\left\| \frac{\Lambda_{n,j,s_0,\omega} - \Lambda_{n,j,s_0,\omega_0}}{\omega - \omega_0} - \theta^{[2]}_{n,j,s_0,\omega_0} \right\| \\
&\qquad = \frac{K^2 (1+\log(n+1+b))^2 e^{2\pi t_0 + K|\omega_0| + K|\omega-\omega_0|} (n+1+b)^{K|\omega-\omega_0|}}{2b^j (n-\frac{1}{2})^{2\sigma_0 - K|\omega_0|}} |\omega-\omega_0|,
\end{aligned}$$

which approaches 0 as $\omega \to \omega_0$. We have thus shown that $\theta^{[2]}_{n,j,s,\omega} = \frac{\partial}{\partial\omega} \Lambda_{n,j,s,\omega}$.

Now notice that

$$\sum_{n=1}^{\infty} \|\theta^{[2]}_{n,j,s,\omega}\| \leq \sum_{n=1}^{\infty} \frac{e^{2\pi t+K|\omega|}K^2(1+\log(n+1+b))}{b^j(n-\frac{1}{2})^{2\sigma-K|\omega|}},$$

which is finite for $(s,\omega)\in\mathcal{U}$. So

$$\theta^{[2]}_{j,s,\omega} := \sum_{n=1}^{\infty} \theta^{[2]}_{n,j,s,\omega}$$

is a well defined element of $\mathbb{A}^*$ for all $j\in\mathbb{N}\cup\{0\}$. Then

$$\left\|\frac{\Lambda_{j,s_0,\omega}-\Lambda_{j,s_0,\omega_0}}{\omega-\omega_0} - \theta^{[2]}_{j,s_0,\omega_0}\right\| \\ \leq \sum_{n=1}^{\infty} \frac{K^2(1+\log(n+1+b))^2 e^{2\pi t_0+K|\omega_0|+K|\omega-\omega_0|}(n+1+b)^{K|\omega-\omega_0|}}{2b^j(n-\frac{1}{2})^{2\sigma_0-K|\omega_0|}}|\omega-\omega_0|,$$

which is finite as long as ω is close enough to ω_0. Thus, if $\omega\to\omega_0$, the left side above approaches 0. This proves that $\frac{\partial}{\partial\omega}\Lambda_{j,s,\omega}$ exists and equals $\theta^{[2]}_{j,s,\omega}$.

Again fix $\epsilon>0$ and notice that

$$\sum_{j=0}^{\infty} \|\theta^{[2]}_{j,s,\omega}\|^{\epsilon}\|e_j\|^{\epsilon} \leq \sum_{j=0}^{\infty}\left(\sum_{n=1}^{\infty} \frac{e^{2\pi t+K|\omega|}K^2(1+\log(n+1+b))}{b^j(n-\frac{1}{2})^{2\sigma-K|\omega|}}\right)^{\epsilon} a^{j\epsilon} \\ = \left(\sum_{n=1}^{\infty} \frac{e^{2\pi t+K|\omega|}K^2(1+\log(n+1+b))}{(n-\frac{1}{2})^{2\sigma-K|\omega|}}\right)^{\epsilon} \frac{1}{1-(a/b)^{\epsilon}} < \infty,$$

for $(s,\omega)\in\mathcal{U}$, implying that

$$\theta^{[2]}_{s,\omega} := \sum_{j=0}^{\infty} \theta^{[2]}_{j,s,\omega}\otimes e_j$$

is a nuclear operator of order ϵ. We also have

$$\sum_{j=0}^{\infty}\left\|\frac{\Lambda_{j,s,\omega_0}-\Lambda_{j,s_0,\omega_0}}{s-s_0} - \theta^{[1]}_{j,s_0,\omega_0}\right\|^{\epsilon}\|e_j\|^{\epsilon} \\ \leq \sum_{j=0}^{\infty}\left(\sum_{n=1}^{\infty} \frac{K^2(1+\log(n+1+b))^2 e^{2\pi t_0+K|\omega_0|+K|\omega-\omega_0|}(n+1+b)^{K|\omega-\omega_0|}}{2b^j(n-\frac{1}{2})^{2\sigma_0-K|\omega_0|}}|\omega-\omega_0|\right)^{\epsilon} a^{j\epsilon} \\ = |\omega-\omega_0|^{\epsilon}\left(\sum_{n=1}^{\infty} \frac{K^2(1+\log(n+1+b))^2 e^{2\pi t_0+K|\omega_0|+K|\omega-\omega_0|}(n+1+b)^{K|\omega-\omega_0|}}{2(n-\frac{1}{2})^{2\sigma_0-K|\omega_0|}}\right)^{\epsilon} \frac{1}{1-(a/b)^{\epsilon}},$$

which is finite as long as $2\sigma_0 - K|\omega_0| - K|\omega-\omega_0| > 1$, and approaches 0 as $\omega\to\omega_0$. This implies that

$$\frac{L_{s_0,\omega} - L_{s_0,\omega_0}}{\omega - \omega_0} - \theta^{[2]}_{s_0,\omega_0} = \sum_{j=0}^{\infty} \left(\frac{\Lambda_{j,s_0,\omega} - \Lambda_{j,s_0,\omega_0}}{\omega - \omega_0} - \theta^{[2]}_{j,s_0,\omega_0} \right) \otimes e_j$$

approaches 0 in the norm on order ϵ nuclear operators. Thus $(s,\omega) \mapsto L_{s,\omega}$ is analytic with respect to ω, and hence with respect to (s,ω), as a map to the order ϵ nuclear operators ($\forall \epsilon > 0$) and $\frac{\partial}{\partial \omega} L_{s,\omega} = \theta^{[2]}_{s,\omega}$.

4.3.4 The η-function

By the work of Grothendieck [28, 29] on the theory of Fredholm determinants of nuclear operators on Banach spaces, the functions

$$Z_{\pm}(s,\omega) := \det(I \pm L_{s,\omega}) = \prod_{\lambda_{s,\omega} \in \mathrm{spec}(L_{s,\omega})} (1 \pm \lambda_{s,\omega}),$$

where $\mathrm{spec}(L_{s,\omega})$ is the set of eigenvalues of $L_{s,\omega}$ (counted with multiplicity), are well defined and analytic on $\mathcal{U}$. Furthermore, the product over the eigenvalues converges absolutely. By [22, Proposition 3.4], if $\mathrm{Re}(s) > 1$ or $s = 1 + it$ with $t \in \mathbb{R}\backslash\{0\}$, then the spectral radius of $L_{s,0}$ is less than 1, and hence there is an open neighborhood $\mathcal{V}$ of $\{(s,0) \in \mathbb{C}^2 : \mathrm{Re}(s) \geq 1, s \neq 1\}$ in $\mathcal{U}$ such that the spectral radius of $L_{s,\omega}$ is less than 1 for all $(s,\omega) \in \mathcal{V}$. Thus, for $(s,\omega) \in \mathcal{V}$, $Z_{\pm}(s,\omega) \neq 0$ and

$$\begin{aligned} Z_{\pm}(s,\omega) &= \exp\left(\sum_{\lambda_{s,\omega} \in \mathrm{spec}(L_{s,\omega})} \log(1 \pm \lambda_{s,\omega}) \right) = \exp\left(\sum_{\lambda_{s,\omega} \in \mathrm{spec}(L_{s,\omega})} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} (\pm \lambda_{s,\omega})^n \right) \\ &= \exp\left(-\sum_{n=1}^{\infty} \frac{(\mp 1)^n}{n} \mathrm{Tr}(L^n_{s,\omega}) \right). \end{aligned}$$

Assuming that $\mathrm{Re}(s) > 1$ and following [56] (see also [22, Theorem 3.3] for a detailed argument which can readily be applied to our situation), we have

$$\mathrm{Tr}(L^n_{s,\omega}) = \sum_{a_1,\ldots,a_n=1}^{\infty} \frac{\prod_{j=1}^{n} \chi_{s,\omega}(G^j([\overline{a_1,\ldots,a_n}]))}{1 - (-1)^n \prod_{j=1}^{n} G^j([\overline{a_1,\ldots,a_n}])^2}.$$

It then follows that

$$\zeta_+(s,\omega) := \frac{Z_+(s+1,\omega)}{Z_-(s,\omega)} = \exp\left(\sum_{n=1}^{\infty}\frac{1}{n}\sum_{a_1,\ldots,a_n=1}^{\infty}\prod_{j=1}^{n}\chi_{s,\omega}(G^j([\overline{a_1,\ldots,a_n}]))\right),$$

$$\zeta_-(s,\omega) := \frac{Z_-(s+1,\omega)}{Z_+(s,\omega)} = \exp\left(\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sum_{a_1,\ldots,a_n=1}^{\infty}\prod_{j=1}^{n}\chi_{s,\omega}([G^j(\overline{a_1,\ldots,a_n}]))\right),$$

and hence

$$\eta(s,\omega) := \frac{1}{2}\log(\zeta_+(s,\omega)\zeta_-(s,\omega)) = \sum_{n=1}^{\infty}\frac{1}{2n}\sum_{a_1,\ldots,a_{2n}=1}^{\infty}\prod_{j=1}^{2n}\chi_{s,\omega}(G^j[\overline{a_1,\ldots,a_{2n}}]).$$

Taking the derivative of η with respect to ω and setting $\omega = 0$ yields the function

$$\hat{\eta}(s) := \sum_{n=1}^{\infty}\frac{1}{2n}\sum_{a_1,\ldots,a_{2n}=1}^{\infty}\left(\sum_{j=1}^{2n} f(G^j([\overline{a_1,\ldots,a_{2n}}]))\right) e^{-s\ell(a_1,\ldots,a_{2n})}$$

For a given tuple $a = (a_1,\ldots,a_n) \in \mathbb{N}^n$ of any length, we define $\mathrm{per}(a)$ be the length of the minimal period in the periodic continued fraction $[\overline{a}]$, which is the number of distinct tuples of length n that one can cyclically permute to obtain a. Then the term $(2n)^{-1}f([\overline{a}])e^{-s\ell(a)}$ appears $\mathrm{per}(a)$ times in the sum defining $\hat{\eta}$, and so we can rearrange terms to get

$$\hat{\eta}(s) = \sum_{n=1}^{\infty}\sum_{a\in\mathbb{N}^{2n}}\frac{\mathrm{per}(a)}{2n}f([\overline{a}])e^{-s\ell(a)}.$$

This establishes $\hat{\eta}$ as the Laplace transform

$$\int_0^{\infty} e^{-st}\, d\tilde{S}_f(t)$$

of the function

$$\tilde{S}_f(T) = \sum_{n=1}^{\infty}\sum_{\substack{a\in\mathbb{N}^{2n}\\ \ell(a)\le T}}\frac{\mathrm{per}(a)}{2n}f([\overline{a}]).$$

Now since $\hat{\eta}(s)$ is the ω-derivative of $\eta(s,\omega)$ at $\omega = 0$, $\hat{\eta}$ extends analytically to a neighborhood of $\{s \in \mathbb{C} : \mathrm{Re}(s) \ge 1\}\backslash\{1\}$. In this section, we see that $s = 1$ is a simple pole of $\hat{\eta}$ and calculate its corresponding residue, which allows us to determine the asymptotic growth rate of $\tilde{S}_f$.

Recall that as an operator on $C^1([0,1])$, $\hat{G}_\nu$ has maximal eigenvalue 1, and all other eigenvalues have modulus less than 0.31. Hence, the same is true for $\hat{G}_\lambda = L_{1,0}$ as an operator on the smaller space $\mathbb{A}$. By analytic perturbation theory (see [44]), there is a neighborhood $\mathcal{W}$ of $(1,0)$ such that for $(s,\omega) \in \mathcal{W}$, the maximal eigenvalue $\lambda_1(s,\omega)$ of $L_{s,\omega}$ is analytic in (s,ω) and the lesser eigenvalues of $L_{s,\omega}$ are of modulus less than 1. So for $(s,\omega) \in \mathcal{W}$ such that $\lambda_1(s,\omega) \neq 1$,

$$\eta(s,\omega) = -\frac{1}{2}\log(1-\lambda_1(s,\omega)^2) + \Phi(s,\omega),$$

where Φ is analytic in $\mathcal{V}$. Hence

$$\hat{\eta}(s) = \frac{\lambda_1(s,0)}{1-(\lambda_1(s,0))^2}\frac{\partial\lambda_1}{\partial\omega}(s,0) + \frac{\partial\Phi}{\partial\omega}(s,0),$$

which extends the domain of $\hat{\eta}(s)$ to the set of $s \in \mathbb{C}$ such that $(s,0) \in \mathcal{W}$ and $\lambda_1(s,0) \neq \pm 1$. So if $\frac{\partial\lambda_1}{\partial s}(1,0) \neq 0$, then $\hat{\eta}(s)$ has a simple pole at $s=1$ with residue

$$\frac{-\frac{\partial\lambda_1}{\partial\omega}(1,0)}{2\frac{\partial\lambda_1}{\partial s}(1,0)}.$$

From the proof of [59, Proposition 2] and [22, Theorem 3.6], it follows that $-\frac{\partial\lambda_1}{\partial s}(1,0)$ is the entropy of the Gauss map, which is $\frac{\pi^2}{6\log 2}$; and [59] also establishes that $\frac{\partial\lambda_1}{\partial\omega}(1,0) = \int_{[0,1]} f\,d\nu$.

We prove the latter assertion in detail, taking into account the fact that $f(z)$ can have an asymptote at $z=0$. Note first that if $\omega \in \mathbb{R}$, then it is clear that $\mathrm{Tr}(L^n_{1,\omega}) > 0$ for all $n \in \mathbb{N}$, and hence $\lambda_1(1,\omega)$ is real and positive. Therefore, $\lambda_1(1,\omega)$ can be calculated as the spectral radius of $L^n_{1,\omega}$ for $\omega \in \mathbb{R}$. So letting $\omega \in \mathbb{R}$ and following [22, Theorem 3.6], we have

$$\begin{aligned}
\log\lambda_1(1,\omega) &= \lim_{n\to\infty}\frac{1}{n}\log\|L^n_{1,\omega}\| \\
&\geq \lim_{n\to\infty}\frac{1}{n}\log\left(\max_{z\in D_a}\left|\sum_{a_1,\ldots,a_n=1}^{\infty}\left(\prod_{j=1}^{n}[a_j,\ldots,a_n+z]^2\right)\exp\left(\omega\sum_{j=1}^{n}f([a_j,\ldots,a_n+z])\right)\right|\right) \\
&\geq \lim_{n\to\infty}\frac{1}{n}\log\left(\sum_{a_1,\ldots,a_n=1}^{\infty}\left(\prod_{j=1}^{n}[a_j,\ldots,a_n]^2\right)\exp\left(\omega\sum_{j=1}^{n}f([a_j,\ldots,a_n])\right)\right)
\end{aligned}$$

By the properties (3.2) and (3.3), we have

$$\prod_{j=1}^{n}[a_j,\dots,a_n]^2 = \frac{1}{q_{1,n}^2} \geq \frac{1}{q_{1,n}(q_{1,n}+q_{1,n-1})} = \lambda([\![a_1,\dots,a_n]\!]).$$

These inequalities and the concavity of log yields

$$\begin{aligned}\log\lambda_1(1,\omega) &\geq \lim_{n\to\infty}\frac{1}{n}\log\left(\sum_{a_1,\dots,a_n=1}^{\infty}\lambda([\![a_1,\dots,a_n]\!])\exp\left(\omega\sum_{j=1}^{n}f([a_j,\dots,a_n])\right)\right)\\ &\geq \omega\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\sum_{a_1,\dots,a_n=1}^{\infty}\lambda([\![a_1,\dots,a_n]\!])\cdot f([a_j,\dots,a_n]).\end{aligned}$$

Next, notice that

$$\begin{aligned}f([a_j,\dots,a_n]) &= \frac{1}{\lambda([\![a_1,\dots,a_n]\!])}\int_{[\![a_1,\dots,a_n]\!]}(f\circ G^{j-1})\,d\lambda + O\left(\left\|f'|_{[\![a_j,\dots,a_n]\!]}\right\|_{\infty}\cdot\lambda([\![a_j,\dots,a_n]\!])\right)\\ &= \frac{1}{\lambda([\![a_1,\dots,a_n]\!])}\int_{[\![a_1,\dots,a_n]\!]}(f\circ G^{j-1})\,d\lambda + O\left(\frac{\|f'|_{[\![a_j,\dots,a_n]\!]}\|_{\infty}}{q_{j,n}(q_{j,n}+q_{j,n-1})}\right);\end{aligned}$$

and hence

$$\begin{aligned}\log\lambda_1(1,\omega) &\geq \omega\lim_{n\to\infty}\left(\int_{[0,1]}\frac{1}{n}\sum_{j=1}^{n}(f\circ G^{j-1})\,d\lambda + O\left(\sup_{a_1,\dots,a_n}\frac{1}{n}\sum_{j=1}^{n}\frac{\|f'|_{[\![a_j,\dots,a_n]\!]}\|_{\infty}}{q_{j,n}(q_{j,n}+q_{j,n-1})}\right)\right)\\ &= \omega\int_{[0,1]}f\,d\nu + O\left(\omega\lim_{n\to\infty}\frac{1}{n}\sup_{a_1,\dots,a_n}\sum_{j=1}^{n}\frac{\|f'|_{[\![a_j,\dots,a_n]\!]}\|_{\infty}}{q_{j,n}(q_{j,n}+q_{j,n-1})}\right),\end{aligned}\tag{4.20}$$

where we used the von Neumann ergodic theorem [58] and the ergodicity of G to derive the equality.

We wish to prove that the error term in (4.20) is equal to 0. To do so, note that by the definition of f, there exists $C>0$ such that $|f'(z)| \leq \frac{C}{|z|}$ for $z\in\overline{D}_1\backslash\{0\}$, and so $\|f'|_{[\![a_j,\dots,a_n]\!]}\|_{\infty} \leq C(a_j+1)$. From Section 3.1, we have the property $q_{j,n}=a_jq_{j+1,n}+q_{j+2,n}$, and as a result,

$$\sum_{j=1}^{n}\frac{\|f'|_{[\![a_j,\dots,a_n]\!]}\|_{\infty}}{q_{j,n}(q_{j,n}+q_{j,n-1})} \leq \sum_{j=1}^{n}\frac{2C}{q_{j,n}}.$$

Now for any $a_1,\dots,a_n\in\mathbb{N}$, $q_{j,n}\geq F_{n-j+1}$, where $\{F_j\}_{j=1}^{\infty}$ is the Fibonacci sequence with $F_1=$

$F_2 = 1$; and since the Fibonacci sequence grows at an exponential rate, we can say that

$$\sup_n \sup_{a_1,\ldots,a_n} \sum_{j=1}^{n} \frac{2C}{q_{j,n}} \le \sum_{j=1}^{\infty} \frac{2C}{F_j} < \infty.$$

This proves that

$$\lim_{n\to\infty} \frac{1}{n} \sup_{a_1,\ldots,a_n} \sum_{j=1}^{n} \frac{\|f'|_{[\![a_j,\ldots,a_n]\!]}\|_\infty}{q_{j,n}(q_{j,n}+q_{j,n-1})} = 0,$$

and hence

$$\log \lambda_1(1,\omega) - \omega \int_{[0,1]} f \, d\nu \ge 0.$$

for $(1,\omega) \in \mathcal{U}$. Since $\mathcal{U}$ contains an open neighborhood of $(1,0)$, and thus the above holds for ω in an open neighborhood of 0, and the expression on the left is equal to 0 when $\omega = 0$, the expression's derivative at $\omega = 0$ must vanish. So

$$\frac{\frac{\partial \lambda_1}{\partial \omega}(1,0)}{\lambda_1(1,0)} - \int_{[0,1]} f \, d\nu = 0;$$

and since $\lambda_1(1,0) = 1$, we have

$$\frac{\partial \lambda_1}{\partial \omega}(1,0) = \int_{[0,1]} f \, d\nu.$$

We have therefore established that the function $\hat{\eta}(s)$ has a pole at $s = 1$ with residue $\frac{3\log 2}{\pi^2} \int_{[0,1]} f \, d\nu$. Then by the Wiener-Ikehara Tauberian theorem [52, Section III, Theorem 4.2],

$$\tilde{S}_f(T) \sim \left(\frac{3\log 2}{\pi^2} \int_{[0,1]} f \, d\nu \right) e^T. \qquad (T \to \infty)$$

4.3.5 $S_f(T)$ and $\tilde{S}_f(T)$ are asymptotically equivalent

We first rewrite the sum defining $\tilde{S}_f(T)$ in terms of the periodic points of G. For this we need to distinguish between periodic continued fractions of even and odd period by defining the sets

$$Q_{G,\text{even}}(T) = \{[\overline{a}] : a \in \mathbb{N}^{2n}, n \in \mathbb{N}, \text{per}(a) = 2n, \ell([\overline{a}]) \le T\},$$

$$Q_{G,\text{odd}}(T) = \{[\overline{a}] : a \in \mathbb{N}^{2n}, n \in \mathbb{N} \text{ is odd}, \text{per}(a) = n, \ell([\overline{a}]) \le T\}.$$

Now let $a \in \mathbb{N}^{2n}$, with $\ell(a) \leq T$, be a tuple represented in the sum defining $\tilde{S}_f(T)$. Then by the definition of $\mathrm{per}(a)$, a is the concatenation of the tuple $(a_1, \ldots, a_{\mathrm{per}(a)})$ with itself $\frac{2n}{\mathrm{per}(a)}$ times. If $\mathrm{per}(a)$ is even, let $k = \frac{2n}{\mathrm{per}(a)}$, and if $\mathrm{per}(a)$ is odd, let $k = \frac{n}{\mathrm{per}(a)}$. Then letting $x = [\overline{a}]$, we have $\ell(a) = k\ell(x)$ whether $\mathrm{per}(a)$ is even or odd. Thus, for a given tuple a in the sum defining $\tilde{S}_f(T)$, we have associated corresponding elements $x \in Q_G$ and $k \in \mathbb{N}$ such that $\ell(a) = k\ell(x)$. So we can rewrite $\tilde{S}_f(T)$ as follows. First define

$$\bar{S}_f(T) = \sum_{x \in Q_{G,\mathrm{even}}(T)} f(x) + \frac{1}{2} \sum_{x \in Q_{G,\mathrm{odd}}(T)} f(x). \tag{4.21}$$

We then have

$$\tilde{S}_f(T) = \sum_{k=1}^{\lfloor T/\ell_0 \rfloor} \frac{1}{k} \bar{S}_f\left(\frac{T}{k}\right),$$

where ℓ_0 is the length of the shortest closed geodesic in $T_1\mathcal{M}$. Noting that f is real valued and positive on $(0,1]$, we see that $\bar{S}_f(T) \ll \tilde{S}_f(T) \ll e^T$ as $T \to \infty$, and hence

$$\sum_{k=2}^{\lfloor T/\ell_0 \rfloor} \bar{S}_f\left(\frac{T}{k}\right) \ll T e^{T/2}. \qquad (T \to \infty)$$

This yields

$$\bar{S}_f(T) \sim \tilde{S}_f(T) \sim \left(\frac{3\log 2}{\pi^2} \int_{[0,1]} f\, d\nu\right) e^T. \qquad (T \to \infty) \tag{4.22}$$

To complete the proof, by (4.21) it suffices to establish that $\sum_{x \in Q_{G,\mathrm{odd}}(T)} f(x) \ll e^{T/2}$. To show this, note that

$$\frac{\partial}{\partial \omega} \left(\log \zeta_+(s,\omega)\right)\Big|_{\omega=0} = \sum_{n=1}^{\infty} \sum_{a \in \mathbb{N}^n} \frac{\mathrm{per}(a)}{n} f([\overline{a}]) e^{-s\ell(a)}$$

is the Laplace transform of the function

$$\bar{\bar{S}}_f(T) = \sum_{n=1}^{\infty} \sum_{\substack{a \in \mathbb{N}^n \\ \ell(a) \leq T}} \frac{\mathrm{per}(a)}{n} f([\overline{a}]),$$

and has a simple pole at $s = 1$, though is otherwise analytic on $\{s \in \mathbb{C} : \mathrm{Re}(s) \geq 1\}$. So by the

Wiener-Ikehara Tauberian theorem, $\bar{\bar{S}}_f(T) \ll e^T$. Note also that

$$\sum_{x \in Q_{G,\mathrm{odd}}(T)} f(x) \leq \bar{\bar{S}}\left(\frac{T}{2}\right),$$

since if $x = [\overline{a_1, \ldots, a_n}]$, where n is odd and the minimal period length in the continued fraction expansion of x, then $\ell(x) = 2\ell(a_1, \ldots, a_n)$ so that the inequalities $\ell(x) \leq T$ and $\ell(a_1, \ldots, a_n) \leq \frac{T}{2}$ are equivalent. We thus have

$$\sum_{x \in Q_{G,\mathrm{odd}}(T)} f(x) \ll e^{T/2}.$$

(One can likely adapt the work of Kelmer [45, Theorem 3] to prove a more precise estimate.) This imples that

$$S_f(T) \sim \bar{S}_f(T) \sim \left(\frac{3\log 2}{\pi^2} \int_{[0,1]} f\, d\nu\right) e^T, \qquad (T \to \infty)$$

and therefore the proof of Theorem 4 is complete.

Chapter 5

Effective equidistribution of preimages of iterates of the Farey map

5.1 Introduction

In this chapter, we consider the equidistribution of preimages of the Farey map, building upon the author's work in [34]. We begin in this section by stating our main results and drawing out their dynamical and number theoretical implications.

5.1.1 The preimages $F^{-n}([\alpha,\beta])$

As noted in Chapter 1, a motivation for the work of Kesseböhmer and Stratmann on the preimages of F was the problem of estimating the Lebesgue measure of the sum-level set for continued fractions

$$\mathscr{C}_n = \left\{ [a_1, a_2, \ldots] \in [0,1] : \sum_{i=1}^{k} a_i = n \text{ for some } k \in \mathbb{N} \right\}$$

as $n \to \infty$. Using the fact that $\mathscr{C}_1 = \left[\frac{1}{2}, 1\right]$ and that $F^{-1}(\mathscr{C}_n) = \mathscr{C}_{n+1}$ (following from (3.9)), we clearly have $\mathscr{C}_n = F^{-(n-1)}(\mathscr{C}_1) = F^{-(n-1)}\left(\left[\frac{1}{2}, 1\right]\right)$ (see [49, Lemma 2.1]). This is illustrated in Figure 5.1 when $n = 4$. Kesseböhmer and Stratmann proved the asymptotic equivalence [49, Theorem 1.3]

$$\lambda(\mathscr{C}_n) \sim \frac{1}{\log_2 n}. \qquad (n \to \infty)$$

Then in [48], they examined the more general preimages $(F^{-n}([\alpha,\beta]))_n$, with $[\alpha,\beta] \subseteq (0,1]$, which can be interpreted as

$$F^{-n}([\alpha,\beta]) = \left\{ [a_1, a_2, \ldots] \in [0,1] : \left[\sum_{i=1}^{\iota_n(a)} a_i - n, a_{\iota_n(a)+1}, a_{\iota_n(a)+2}, \ldots \right] \in [\alpha,\beta] \right\},$$

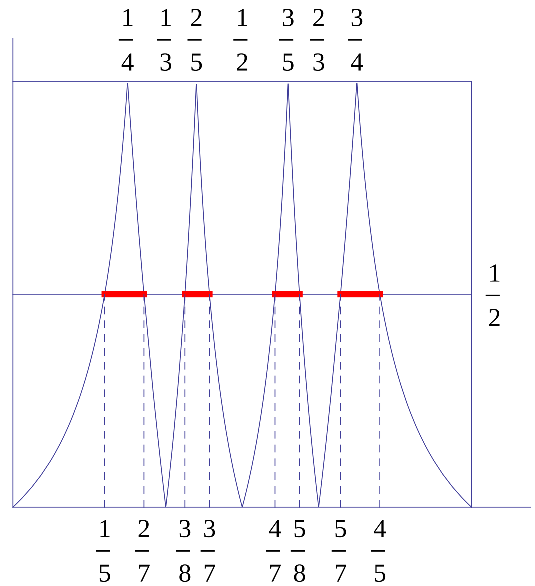

Figure 5.1: The graph of F^3 and $\mathscr{C}_4$ shown as the preimage $F^{-3}\left(\left[\frac{1}{2}, 1\right]\right)$.

where $\iota_n(a) = \iota_n(a_1, a_2, \ldots)$ is the least index such that $\sum_{i=1}^{\iota_n(a)} a_i > n$. Specifically, they showed that

$$\lim_{n\to\infty} \frac{\log n}{\log(\beta/\alpha)} \int_{F^{-n}([\alpha,\beta])} f\, d\lambda = \int_{[0,1]} f\, d\lambda, \tag{5.1}$$

which establishes the decay rate of the Lebesgue measures and the equidistribution of the preimages $(F^{-n}([\alpha,\beta]))_n$. Their proofs applied results in infinite ergodic theory following from the work of Aaronson [1, 2], and Kesseböhmer and Slassi [46, 47], to the Farey map. The goal of this chapter is to prove the following result, which provides an effective version of (5.1) when f is a C^2 function. We can think of this result as an analogue for the Farey map of the mixing rate of the Gauss map because of the similarity to the characterization of mixing (3.7).

Theorem 6. *For any interval $[\alpha,\beta] \subseteq (0,1]$ and $f \in C^2([0,1])$, we have*

$$\frac{\log n}{\log(\beta/\alpha)} \int_{F^{-n}([\alpha,\beta])} f\, d\lambda = \int_{[0,1]} f\, d\lambda + O_{\alpha,\beta}\left(\frac{\|f\|_{C^2}}{\log n}\right), \qquad (n \to \infty) \tag{5.2}$$

where $\|f\|_{C^2} = \|f\|_\infty + \|f'\|_\infty + \|f''\|_\infty$.

We prove Theorem 6 as a corollary of the following result.

Theorem 7. *Let $u \in (0,1)$ and $\varphi \in L^1(\mu) \cap \{f \in C^2([0,1]) : h, h' \geq 0, h'' \leq 0\}$, and let*

$$\mu_{\varphi,n}^{(u)} := \int_{F^{-n}([u,1])} \varphi \, d\mu.$$

Then:

(a) $\left(\mu_{\varphi,n}^{(u)}\right)_n$ *is a nonincreasing sequence;*

(b) $\displaystyle\sum_{k=0}^{n} \mu_{\varphi,k}^{(u)} = \frac{n \log(1/u)}{\log n} \left(\int_{[0,1]} \varphi \, d\mu + O_u\left(\frac{\|\varphi'\|_\infty}{\log n} \right) \right). \qquad (n \to \infty)$

Establishing Theorem 6 from Theorem 7 involves finding, for each $f \in C^2([0,1])$, an appropriate way of writing the function $x \mapsto xf(x)$ as the difference of two functions having the properties of φ in Theorem 7. We detail how this is done in Section 5.2.1.

5.1.2 The Stern-Brocot sequence

The sets $\mathcal{S}_n = \left\{ \frac{s_{n,k}}{t_{n,k}} : k = 1, \ldots, 2^n + 1 \right\}$ in the Stern-Brocot sequence $(\mathcal{S}_n)_n$ are defined recursively as follows:

- $s_{0,1} := 0$ and $s_{0,2} := t_{0,1} := t_{0,2} := 1$;
- $s_{n+1,2k-1} := s_{n,k}$ and $t_{n+1,2k-1} := t_{n,k}$ for $k = 1, \ldots, 2^n + 1$;
- $s_{n+1,2k} := s_{n,k} + s_{n,k+1}$ and $t_{n+1,2k} := t_{n,k} + t_{n,k+1}$ for $k = 1, \ldots, 2^n + 1$.

In other words, similar to how the Farey sequence can be generated, we have $\mathcal{S}_0 = \left\{\frac{0}{1}, \frac{1}{1}\right\}$, and $\mathcal{S}_{n+1}$ is the union of $\mathcal{S}_n$ together with the mediants of its consecutive elements. However, unlike the Farey sequence, there are no restrictions on the mediants of $\mathcal{S}_n$ to include in $\mathcal{S}_{n+1}$.

It is elementary to show that $\mathcal{S}_n = F^{-(n+1)}(0)$ and $\mathcal{S}_{n+1} \backslash \mathcal{S}_n = F^{-(n+1)}(1) = F^{-n}\left(\frac{1}{2}\right)$ for $n \geq 0$. Additionally, the sum-level sets $\mathscr{C}_n = F^{-(n-1)}\left[\frac{1}{2}, 1\right]$ can be written as gaps in elements of the Stern-Brocot sequence. Specifically, we have $\mathscr{C}_1 = \left[\frac{1}{2}, 1\right] = \left[\frac{s_{1,2}}{t_{1,2}}, \frac{s_{1,3}}{t_{1,3}}\right]$, and for $n \geq 2$,

$$\mathscr{C}_n = \bigcup_{k=1}^{2^{n-2}} \left[\frac{s_{n,4k-2}}{t_{n,4k-2}}, \frac{s_{n,4k}}{t_{n,4k}} \right].$$

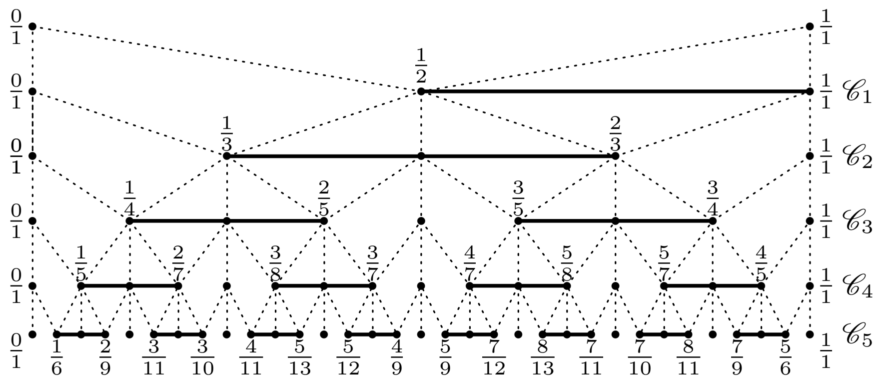

Figure 5.2: The sum-level sets as Stern-Brocot intervals

(See Figure 5.2.) This was the characterization of $\mathscr{C}_n$ considered by Fiala and Kleban [25] motivated by their study of spin chain models.

Now all rational numbers in $[0, 1]$ are contained in $\bigcup_{n=0}^{\infty} \mathcal{S}_n$. So for a given $\frac{v}{w} \in \mathbb{Q} \cap (0, 1)$, there exists $n_0 \in \mathbb{N}$ such that $\frac{v}{w} \in \mathcal{S}_{n_0} = F^{-(n_0+1)}(0)$; and as a result, for every $n \in \mathbb{N} \cup \{0\}$, $F^{-n}\left(\frac{v}{w}\right) \subseteq F^{-(n+n_0+1)}(0) = \mathcal{S}_{n+n_0}$. So the preimages $(F^{-n}\left(\frac{v}{w}\right))_n$ form a sequence of subsets of the Stern-Brocot sequence. By finding a way to shrink the interval $[\alpha, \beta]$ to a single point in (5.1), Kesseböhmer and Stratmann [48, Theorem 1.2] were able to prove that for all $\frac{v}{w} \in \mathbb{Q} \cap (0, 1)$ with $\gcd(v, w) = 1$,

$$\lim_{n\to\infty} \log(n^{vw}) \sum_{\substack{p/q \in F^{-n}(v/w) \\ \gcd(p,q)=1}} q^{-2} f\left(\frac{p}{q}\right) = \int_{[0,1]} f \, d\lambda \quad \text{for } f \in C([0, 1]). \tag{5.3}$$

So the weighted subsets $\left(F^{-n}\left(\frac{v}{w}\right)\right)_n$ equidistribute with respect to the Lebesgue measure when $\frac{p}{q} \in F^{-n}\left(\frac{v}{w}\right)$ is weighted by q^{-2}. In a similar way, we obtain the following effective version of this result as a reasonably straightforward corollary of Theorem 6.

Theorem 8. *Let $\frac{v}{w} \in \mathbb{Q} \cap (0, 1)$ with $\gcd(v, w) = 1$ and $f \in C^2([0, 1])$. Then*

$$\log(n^{vw}) \sum_{\substack{p/q \in F^{-n}(v/w) \\ \gcd(p,q)=1}} q^{-2} f\left(\frac{p}{q}\right) = \int_{[0,1]} f \, d\lambda + O_{v/w}\left(\frac{\|f\|_{C^2}}{\log n}\right). \qquad (n \to \infty)$$

We prove our results in the following section. As in the work of Kesseböhmer and Stratmann, we analyze the transfer operator $\hat{F} : L^1(\mu) \to L^1(\mu)$ of F, which we recall satisfies

$$\int_B \hat{F} f \, d\mu = \int_{F^{-1}(B)} f \, d\mu, \quad \text{for all } B \subseteq [0,1] \text{ Borel and } f \in L^1(\mu),$$

and

$$\hat{F} f(x) = \frac{f\left(\frac{x}{1+x}\right) + x f\left(\frac{1}{1+x}\right)}{1+x}. \tag{5.4}$$

We begin in Section 5.2.1 with reducing Theorem 6 to Theorem 7. Then in Section 5.2.2, we prove part (a) of Theorem 7 in a straightforward manner using previously known elementary properties of $\hat{F}$. We then prove part (b) in Section 5.2.3. To do so, we establish estimates involving sums of the iterates $\hat{F}^k \varphi$, and make careful applications of the equality (5.7) following from [1, Lemma 3.8.4] and Karamata's Tauberian theorem [42], which are important results underlying much of the machinery used in [48, 49], so as to obtain error terms. In particular, we make an application of Freud's effective version of Karamata's theorem [27] in establishing an asymptotic estimate of a certain weighted sum of the values $\mu^{(u)}_{\varphi,n}$ from an estimate of its Laplace transform derived from (5.7). We can then remove the weights to prove (b) by a standard analytic number theory argument. We conclude by proving Theorem 8 in Section 5.2.4. See [57] for other asymptotic results derived from operator renewal theory involving the iterates of transfer operators of infinite measure preserving systems.

5.2 Proofs

5.2.1 Reduction of Theorem 6 to Theorem 7

We begin by establishing from Theorem 7 the special case of Theorem 6 in which the function $f_0 \in C^2([0,1])$ defined by $f_0(x) := x f(x)$ is in $L^1(\mu) \cap \{h \in C^2([0,1]) : h, h' \geq 0, h'' \leq 0\}$. First define $\lambda^{(u)}_{f,n}$ by

$$\lambda^{(u)}_{f,n} := \int_{F^{-n}([u,1])} f \, d\lambda.$$

Then note that (a) and (b) imply that for all $u \in (0,1)$,

$$\lambda_{f,n}^{(u)} = \mu_{f_0,n}^{(u)} \leq \frac{1}{n}\sum_{k=0}^{n} \mu_{f_0,k}^{(u)} = \frac{\log(1/u)}{\log n}\left(\int_{[0,1]} f\,d\lambda + O_u\left(\frac{\|f_0'\|_\infty}{\log n}\right)\right), \text{ and}$$

$$\begin{aligned}
\lambda_{f,n}^{(u)} = \mu_{f_0,n}^{(u)} &\geq \frac{1}{n}\sum_{k=n+1}^{2n} \mu_{f_0,k}^{(u)} = \frac{1}{n}\left(\frac{2n\log(1/u)}{\log 2n} - \frac{n\log(1/u)}{\log n}\right)\int_{[0,1]} f\,d\lambda + O_u\left(\frac{\|f_0'\|_\infty}{\log^2 n}\right) \\
&= \frac{\log(1/u)}{\log n}\left(1 + O\left(\frac{1}{\log n}\right)\right)\int_{[0,1]} f\,d\lambda + O_u\left(\frac{\|f_0'\|_\infty}{\log^2 n}\right) \\
&= \frac{\log(1/u)}{\log n}\left(\int_{[0,1]} f\,d\lambda + O_u\left(\frac{\|f\|_\infty + \|f'\|_\infty}{\log n}\right)\right).
\end{aligned}$$

Hence, we have

$$\lambda_{f,n}^{(u)} = \frac{\log(1/u)}{\log n}\left(\int_{[0,1]} f\,d\lambda + O_u\left(\frac{\|f\|_\infty + \|f'\|_\infty}{\log n}\right)\right). \qquad (n \to \infty) \tag{5.5}$$

Then subtracting this expression for $u = \beta$ from that for $u = \alpha$ yields (5.2) with the error term $O_{\alpha,\beta}\left(\frac{\|f\|_\infty + \|f'\|_\infty}{\log n}\right)$.

To prove Theorem 6 for general $f \in C^2([0,1])$, we wish to write f as the difference $f_1 - f_2$ of functions $f_1, f_2 \in C^2([0,1])$ such that for $j = 1,2$ $f_j \geq \|f_j'\|_\infty$, $f_j', f_j'' \leq 0$, and $\|f_j\|_\infty \leq 2\|f\|_{C^2}$. We have

$$\begin{aligned}
f(x) &= f(0) + f'(0)x + \int_0^x \int_0^t f''(s)\,ds\,dt \\
&= f(0) + f'(0)x + \left(\|f''\|_\infty - \int_0^x \int_0^t f_-''(s)\,ds\,dt\right) - \left(\|f''\|_\infty - \int_0^x \int_0^t f_+''(s)\,ds\,dt\right),
\end{aligned}$$

where f_+'' and f_-' are the positive and negative parts of f'', respectively. Let $\tilde{f}_1$ be $\|f''\|_\infty - \int_0^x \int_0^t f_-''(s)\,ds\,dt$, added to $f(0)$ if $f(0) \geq 0$, and to $f'(0)x$ if $f'(0) \leq 0$; and let $\tilde{f}_2$ be $\|f''\|_\infty - \int_0^x \int_0^t f_+''(s)\,ds\,dt$, added to $-f(0)$ if $f(0) < 0$, and to $-f'(0)x$ if $f'(0) > 0$. Then $\tilde{f}_1$ and $\tilde{f}_2$ are functions such that $f = \tilde{f}_1 - \tilde{f}_2$, and for $j = 1,2$, $\tilde{f}_j{}', \tilde{f}_j{}'' \leq 0$, $\|\tilde{f}_j{}'\|_\infty \leq \|f'\|_\infty + \|f''\|_\infty$, and $-\|f'\|_\infty \leq \tilde{f}_j \leq \|f\|_\infty + \|f''\|_\infty$. Thus the functions $f_j = \tilde{f}_j + 2\|f'\|_\infty + \|f''\|_\infty$, $j = 1,2$, satisfy all our desired properties.

As a result, the functions $x \mapsto xf_1(x)$ and $x \mapsto xf_2(x)$ are in $\{h \in C^2([0,1]) : h, h' \geq 0, h'' \leq 0\}$. Therefore (5.2) is valid for f_1 and f_2, with the error terms being $O_{\alpha,\beta}\left(\frac{\|f_j\|_\infty + \|f_j'\|_\infty}{\log n}\right)$, $j = 1,2$.

Subtracting these two asymptotic formulas yields Theorem 6 for the function f, since $\|f_j\|_\infty + \|f_j'\|_\infty \le 3\|f\|_{C^2}$, $j = 1, 2$.

5.2.2 Proof of Theorem 7(a)

We now set out to prove Theorem 7, and we begin by fixing $\varphi \in L^1(\mu) \cap \{h \in C^2([0,1]) : h, h' \ge 0, h'' \le 0\}$, and proving that $\left(\mu_{\varphi,n}^{(u)}\right)_n$ is a nonincreasing sequence. Note that by [46, Lemma 3.2], $\hat{F}$ maps the set $\{h \in C^2([0,1]) : h, h' \ge 0, h'' \le 0\}$ to itself. Thus if $x \in \left[\frac{1}{2}, 1\right]$ and $n \in \mathbb{N} \cup \{0\}$, we have $\hat{F}^n\varphi(x) \ge \hat{F}^n\varphi\left(\frac{1}{2}\right) = \hat{F}^{n+1}\varphi(1) \ge \hat{F}^{n+1}\varphi(x)$. Thus (a) holds for $u \ge \frac{1}{2}$. (See the proof of [49, Theorem 1.1].) Now assume that $u \in \left(0, \frac{1}{2}\right)$. By [46, Lemma 3.2], it suffices to show that

$$\int_u^1 \hat{F} f \, d\mu \le \int_u^1 f \, d\mu$$

whenever $f' \ge 0$. This follows from

$$\begin{aligned}\int_u^1 f \, d\mu - \int_u^1 \hat{F} f \, d\mu &= \int_u^1 f \, d\mu - \int_{F^{-1}([u,1])} f \, d\mu = \int_u^1 f \, d\mu - \int_{u/(1+u)}^{1/(1+u)} f \, d\mu \\ &= \int_{1/(1+u)}^1 f \, d\mu - \int_{u/(1+u)}^{u} f \, d\mu \ge \left(f\left(\frac{1}{1+u}\right) - f(u)\right) \log(1+u) \ge 0.\end{aligned}$$

5.2.3 Proof of Theorem 7(b)

We first consider the case where $u = \frac{1}{N}$, with $N \in \mathbb{N}, N \ge 2$. Define the function $a : \mathbb{R} \to \mathbb{R}$ by

$$a(\sigma) = \frac{1}{\log N} \sum_{k=0}^{\lfloor \sigma \rfloor} \mu_{\varphi,k}^{(u)},$$

which is the μ-average of the function $\hat{F}_\sigma\varphi := \sum_{k=0}^{\lfloor\sigma\rfloor} \hat{F}^k\varphi$ on $\mathscr{C}_1^u = \left[\frac{1}{N}, 1\right]$ by

$$\frac{1}{\mu(\mathscr{C}_1^u)} \int_{\mathscr{C}_1^u} \sum_{k=0}^{\lfloor\sigma\rfloor} \hat{F}^k \varphi \, d\mu = \frac{1}{\log N} \sum_{k=0}^{\lfloor\sigma\rfloor} \int_{F^{-k}(\mathscr{C}_1^u)} \varphi \, d\mu = \frac{1}{\log N} \sum_{k=0}^{\lfloor\sigma\rfloor} \mu_{\varphi,k}^{(u)}.$$

We have the following bound between $\hat{F}_\sigma\varphi$ and $a(\sigma)$.

Lemma 4. *For all $\sigma \in \mathbb{R}$ and $x \in \mathscr{C}_1^u$,*

$$|\hat{F}_\sigma \varphi(x) - a(\sigma)| \leq \|\varphi\|_\infty \frac{N(N-1)}{2}.$$

Proof. Without loss of generality, we assume that $\sigma = n \in \mathbb{N} \cup \{0\}$. By [46, Lemma 3.2], we know that $\hat{F}_n\varphi$ is nondecreasing. So the difference between $\hat{F}_n\varphi(x)$ and $a(n)$ is at most $\hat{F}_n\varphi(1) - \hat{F}_n\varphi\left(\frac{1}{N}\right)$. Using the equality

$$\hat{F}^k\varphi\left(\frac{1}{j}\right) = \frac{j}{j-1}\hat{F}^{k+1}\varphi\left(\frac{1}{j-1}\right) - \frac{1}{j-1}\hat{F}^k\varphi\left(\frac{j-1}{j}\right) \qquad (j, k \in \mathbb{N} \cup \{0\}, j \geq 2)$$

following from (5.4) and the fact that $\hat{F}^k\varphi$ is nondecreasing, we have

$$\begin{aligned}
\hat{F}_n\varphi(1) - \hat{F}_n\varphi\left(\frac{1}{N}\right) &= \sum_{k=0}^{n}\left(\hat{F}^k\varphi(1) - \hat{F}^k\varphi\left(\frac{1}{N}\right)\right) \\
&= \sum_{k=0}^{n}\left(\hat{F}^k\varphi(1) - \frac{N}{N-1}\hat{F}^{k+1}\varphi\left(\frac{1}{N-1}\right) + \frac{1}{N-1}\hat{F}^k\varphi\left(\frac{N-1}{N}\right)\right) \\
&\leq \frac{N}{N-1}\sum_{k=0}^{n}\left(\hat{F}^k\varphi(1) - \hat{F}^{k+1}\varphi\left(\frac{1}{N-1}\right)\right).
\end{aligned}$$

Using this inequality recursively, and also the equality $\hat{F}f(1) = f\left(\frac{1}{2}\right)$, yields

$$\begin{aligned}
\hat{F}_n\varphi(1) - \hat{F}_n\varphi\left(\frac{1}{N}\right) &\leq \frac{N}{2}\sum_{k=0}^{n}\left(\hat{F}^k\varphi(1) - \hat{F}^{k+N-2}\varphi\left(\frac{1}{2}\right)\right) \\
&= \frac{N}{2}\sum_{k=0}^{n}(\hat{F}^k\varphi(1) - \hat{F}^{k+N-1}\varphi(1)) \leq \varphi(1)\frac{N(N-1)}{2} = \|\varphi\|_\infty\frac{N(N-1)}{2}.
\end{aligned}$$

$\square$

Next, we let $S : (0, \infty) \to \mathbb{R}$ be the Laplace transform of a given by

$$S(\sigma) = \int_{0-}^{\infty} e^{-t/\sigma}\, da(t) = \frac{1}{\log N}\sum_{n=0}^{\infty} e^{-n/\sigma}\mu_{\varphi,n}^{(u)}$$

and prove the following bound similar to Lemma 4.

Lemma 5. *For all $x \in \mathscr{C}_1^u$ and all $\sigma > 0$,*

$$\left|\sum_{n=0}^{\infty} e^{-n/\sigma}\hat{F}^n\varphi(x) - S(\sigma)\right| \leq \|\varphi\|_\infty \frac{N(N-1)}{2}.$$

Proof. We first note the equality

$$\sum_{n=0}^{\infty} a_n e^{-n/\sigma} = (1-e^{-1/\sigma})\sum_{n=0}^{\infty} e^{-n/\sigma}\left(\sum_{k=0}^{n} a_k\right), \tag{5.6}$$

which holds for all sequences $(a_n)_n$ satisfying $\sum_{k=0}^{n} a_k = O(n)$ as $n \to \infty$ and all $\sigma > 0$.

Let $x \in \mathscr{C}_1^u$, $\delta_n(x) = \hat{F}_n\varphi_0(x) - a(n)$, and $\sigma > 0$. Using (5.6) twice, we have

$$\begin{aligned}\sum_{n=0}^{\infty} e^{-n/\sigma}\hat{F}^n\varphi(x) = (1-e^{-1/\sigma})\sum_{n=0}^{\infty} e^{-n/\sigma}\hat{F}_n\varphi(x) &= (1-e^{-1/\sigma})\sum_{n=0}^{\infty} e^{-n/\sigma}(a(n)+\delta_n(x)) \\ &= S(\sigma) + (1-e^{-1/\sigma})\sum_{n=0}^{\infty} e^{-n/\sigma}\delta_n(x).\end{aligned}$$

Since $|\delta_n(x)| \leq \|\varphi\|_\infty \frac{N(N-1)}{2}$ for all $n \geq 0$, we have

$$\left|(1-e^{-1/\sigma})\sum_{n=0}^{\infty} e^{-n/\sigma}\delta_n(x)\right| \leq (1-e^{-1/\sigma})\sum_{n=0}^{\infty} e^{-n/\sigma}\|\varphi\|_\infty\frac{N(N-1)}{2} = \|\varphi\|_\infty\frac{N(N-1)}{2}.$$

□

To continue the proof, we will make use of the following equality given by [1, Lemma 3.8.4].

$$\int_A\left(\sum_{n=0}^{\infty} e^{-n/\sigma}\hat{F}^n f\right)(1-e^{-\phi_A/\sigma})\,d\mu = \sum_{n=0}^{\infty} e^{-n/\sigma}\int_{A_n} f\,d\mu \tag{5.7}$$

Here, f is any function in $L^1(\mu)$, σ is any positive real number, $A \subseteq [0,1]$ is any subset such that $\mu(A) < \infty$, $A_0 = A$, and $A_n = F^{-n}(A)\setminus\bigcup_{k=0}^{n-1} F^{-k}(A)$ for $n \geq 1$. Also, $\phi_A : A \to \mathbb{N}$ is the return time function on A defined by $\phi_A(x) = \min\{n \in \mathbb{N} : F^n(x) \in A\}$.

Letting $A = \mathscr{C}_1^u$ and $f = \varphi$ in (5.7), and noting that $A_n = [\frac{1}{n+N}, \frac{1}{n+N-1})$ for $n \geq 1$, we have

$$\int_{1/N}^{1}\left(\sum_{n=0}^{\infty} e^{-n/\sigma}\hat{F}^k\varphi\right)(1-e^{-\phi_A/\sigma})\,d\mu = \int_{1/N}^{1}\varphi\,d\mu + \sum_{n=1}^{\infty} e^{-n/\sigma}\int_{1/(n+N)}^{1/(n+N-1)}\varphi\,d\mu.$$

On the other hand, using Lemma 5, we see that the left side of the above is also equal to

$$\begin{aligned}
&(S(\sigma)+O_N(\|\varphi\|_\infty))(1-e^{-1/\sigma})\int_{1/N}^{1}\left(\sum_{n=0}^{\infty}e^{-n/\sigma}\hat{F}^k 1\right)(1-e^{-\phi_A/\sigma})\,d\mu\\
&\qquad=(S(\sigma)+O_N(\|\varphi\|_\infty))(1-e^{-1/\sigma})\left(\mu(\mathscr{C}_1^u)+\sum_{n=1}^{\infty}e^{-n/\sigma}\int_{1/(n+N)}^{1/(n+N-1)}d\mu\right)\\
&\qquad=(S(\sigma)+O_N(\|\varphi\|_\infty))(1-e^{-1/\sigma})\left(\log N+\sum_{n=1}^{\infty}e^{-n/\sigma}\log\left(\frac{n+N}{n+N-1}\right)\right)
\end{aligned}$$

as $\sigma\to\infty$. (Note that (5.7) holds for the constant function $f(x)=1$ in spite of the fact that $1\notin L^1(\mu)$ since $\sum_{n=0}^{\infty}e^{-n/\sigma}\hat{F}^k 1$ has finite integral over $\mathscr{C}_1^u$.) For our next step, we determine the asymptotic behavior of

$$\int_{1/N}^{1}\varphi\,d\mu+\sum_{n=1}^{\infty}e^{-n/\sigma}\int_{1/(n+N)}^{1/(n+N-1)}\varphi\,d\mu\quad\text{and}\quad\log N+\sum_{n=1}^{\infty}e^{-n/\sigma}\log\left(\frac{n+N}{n+N-1}\right).$$

Lemma 6.

$$\int_{1/N}^{1}\varphi\,d\mu+\sum_{n=1}^{\infty}e^{-n/\sigma}\int_{1/(n+N)}^{1/(n+N-1)}\varphi\,d\mu=\int_{[0,1]}\varphi\,d\mu+O\left(\|\varphi'\|_\infty\frac{\log\sigma}{\sigma}\right)\qquad(\sigma\to\infty)$$

Proof. Let $S_1:\mathbb{R}\to\mathbb{R}$ be defined by

$$S_1(t)=\left(\int_{1/N}^{1}\varphi\,d\mu\right)1_{[0,\infty)}(t)+\sum_{n=1}^{\lfloor t\rfloor}\int_{1/(n+N)}^{1/(n+N-1)}\varphi\,d\mu=\begin{cases}\int_{1/(\lfloor t\rfloor+N)}^{1}\varphi\,d\mu & \text{if } t\geq 0\\ 0 & \text{if } t<0.\end{cases}$$

Then for $\sigma>0$,

$$\begin{aligned}
\int_{1/N}^{1}\varphi\,d\mu+\sum_{n=1}^{\infty}e^{-n/\sigma}\int_{1/(n+N)}^{1/(n+N-1)}\varphi\,d\mu&=\int_{0-}^{\infty}e^{-t/\sigma}\,dS_1(t)=\frac{1}{\sigma}\int_0^\infty\left(\int_{1/(\lfloor t\rfloor+N)}^{1}\varphi\,d\mu\right)e^{-t/\sigma}\,dt\\
&=\int_{[0,1]}\varphi\,d\mu-\int_0^\infty\left(\int_0^{1/(\lfloor\sigma x\rfloor+N)}\varphi\,d\mu\right)e^{-x}\,dx.
\end{aligned}$$

Next, letting $\bar{\varphi}:(0,1]\to\mathbb{R}$ be defined by $\bar{\varphi}(x)=\frac{\varphi(x)}{x}$, we have

$$0\leq\int_0^\infty\left(\int_0^{1/(\lfloor\sigma x\rfloor+N)}\varphi\,d\mu\right)e^{-x}\,dx\leq\|\bar{\varphi}\|_\infty\int_0^\infty\frac{e^{-x}\,dx}{\lfloor\sigma x\rfloor+N}.$$

Note that since $\varphi \in L^1(\mu) \cap C^2([0,1])$, $\varphi(0) = 0$. So for each $x \in (0,1]$,

$$\bar{\varphi}(x) = \frac{\varphi(x) - \varphi(0)}{x - 0} = \varphi'(y)$$

for some $y \in (0, x)$ by the mean value theorem, which shows that $\|\bar{\varphi}\|_\infty \le \|\varphi'\|_\infty$. Also, since the inequality $\lfloor t \rfloor + N \ge \frac{1}{2}(t+2)$ holds for $t \ge 0$,

$$\int_0^\infty \frac{e^{-x}\,dx}{\lfloor \sigma x \rfloor + N} \le 2\int_0^\infty \frac{e^{-x}\,dx}{\sigma x + 2} \le \int_0^1 \frac{2\,dx}{\sigma x + 2} + \int_1^\infty \frac{2e^{-x}}{\sigma}\,dx \ll \frac{\log \sigma}{\sigma}, \qquad (\sigma \to \infty)$$

which completes the proof. $\square$

Lemma 7. *We have*

$$\log N + \sum_{n=1}^\infty e^{-n/\sigma} \log\left(\frac{n+N}{n+N-1}\right) = \log(\sigma + N) + C + O_N\left(\frac{\log \sigma}{\sigma}\right), \qquad (\sigma \to \infty)$$

where

$$C = \int_0^1 \frac{e^{-x} - 1}{x}\,dx + \int_1^\infty \frac{e^{-x}}{x}\,dx.$$

Proof. Let $S_2 : \mathbb{R} \to \mathbb{R}$ be defined by

$$S_2(t) = (\log N) 1_{[0,\infty)}(t) + \sum_{n=1}^{\lfloor t \rfloor} \log\left(\frac{n+N}{n+N-1}\right) = \begin{cases} \log(\lfloor t \rfloor + N) & \text{if } t \ge 0 \\ 0 & \text{if } t < 0. \end{cases}$$

Then for $\sigma > 0$,

$$\begin{aligned} \log N + \sum_{n=1}^\infty e^{-n/\sigma} \log\left(\frac{n+N}{n+N-1}\right) &= \int_{0-}^\infty e^{-t/\sigma}\,dS_2(t) = \frac{1}{\sigma}\int_0^\infty e^{-t/\sigma} \log(\lfloor t \rfloor + N)\,dt \\ &= \int_0^\infty e^{-x} \log(\lfloor \sigma x \rfloor + N)\,dx \\ &= \int_0^\infty e^{-x} \log(\sigma x + N)\,dx - \int_0^\infty e^{-x} \log\left(\frac{\sigma x + N}{\lfloor \sigma x \rfloor + N}\right) dx. \end{aligned}$$

Using the inequality $\log(1+x) \le x$, we have

$$\int_0^\infty e^{-x} \log\left(\frac{\sigma x + N}{\lfloor \sigma x \rfloor + N}\right) dx = \int_0^\infty e^{-x} \log\left(1 + \frac{\{\sigma x\}}{\lfloor \sigma x \rfloor + N}\right) dx \ll \int_0^\infty \frac{e^{-x}\, dx}{\lfloor \sigma x \rfloor + N},$$

which is $O\left(\frac{\log \sigma}{\sigma}\right)$ as $\sigma \to \infty$ by the proof of Lemma 6.

Next, integration by parts yields

$$\int_0^\infty e^{-x} \log(\sigma x + N)\, dx = \log N + \int_0^\infty \frac{\sigma e^{-x}\, dx}{\sigma x + N}. \tag{5.8}$$

To continue, we consider the integral on the right over $[0,1]$ by writing

$$\int_0^1 \frac{\sigma e^{-x}\, dx}{\sigma x + N} = \int_0^1 \frac{\sigma\, dx}{\sigma x + N} + \int_0^1 \frac{\sigma(e^{-x} - 1)}{\sigma x + N}\, dx.$$

The first integral on the right equals $\log(\sigma + N) - \log N$, while the second equals

$$\begin{aligned} \int_0^1 \frac{e^{-x} - 1}{x}\, dx - N \int_0^1 \frac{e^{-x} - 1}{x(\sigma x + N)}\, dx &= \int_0^1 \frac{e^{-x} - 1}{x}\, dx + O\left(\int_0^1 \frac{N\, dx}{\sigma x + N}\right) \\ &= \int_0^1 \frac{e^{-x} - 1}{x}\, dx + O_N\left(\frac{\log \sigma}{\sigma}\right). \qquad (\sigma \to \infty) \end{aligned}$$

Now considering the integral in (5.8) over $[1,\infty)$, we write

$$\int_1^\infty \frac{\sigma e^{-x}\, dx}{\sigma x + N} = \int_1^\infty \frac{e^{-x}}{x}\, dx - N \int_1^\infty \frac{e^{-x}\, dx}{x(\sigma x + N)} = \int_1^\infty \frac{e^{-x}}{x}\, dx + O_N\left(\frac{1}{\sigma}\right).$$

Putting these results together proves the lemma. $\square$

Lemma 6 and the preceding equalities give

$$\begin{aligned} (S(\sigma) + O_N(\|\varphi\|_\infty))(1 - e^{-1/\sigma}) \left(\log N + \sum_{n=1}^\infty e^{-n/\sigma} \log\left(\frac{n+N}{n+N-1}\right)\right) \\ = \int_{[0,1]} \varphi\, d\mu + O\left(\|\varphi'\|_\infty \frac{\log \sigma}{\sigma}\right) \end{aligned} \tag{5.9}$$

as $\sigma \to \infty$. Using Lemma 7 and noting that $\|\varphi\|_\infty \le \|\bar{\varphi}\|_\infty \le \|\varphi'\|_\infty$, we have

$$S(\sigma) = \frac{\sigma}{\log \sigma + C} \int_{[0,1]} \varphi \, d\mu + O_N(\|\varphi'\|_\infty). \qquad (\sigma \to \infty) \tag{5.10}$$

At this point, an application of Karamata's Tauberian theorem [42] yields

$$\sum_{k=0}^{n} \mu_{\varphi,k}^{(u)} \sim \frac{n}{\log_N n} \int_{[0,1]} \varphi \, d\mu. \qquad (n \to \infty)$$

Furthermore, one can apply an adaptation of Freud's effective version of Karamata's theorem [27] (see also [66, Section 7.4]) accommodating logarithms to (5.10) in order to prove

$$\sum_{k=0}^{n} \mu_{\varphi,k}^{(u)} = \frac{n}{\log_N n} \left(\int_{[0,1]} \varphi \, d\mu + O_N\left(\frac{1 + \|\varphi'\|_\infty}{\log \log n} \right) \right). \qquad (n \to \infty)$$

This, together with the fact that $\left(\mu_{\varphi,n+1}^{(u)}\right)_n$ is a nonincreasing sequence, implies that

$$\mu_{\varphi,k}^{(u)} = \frac{1}{\log_N n} \left(\int_{[0,1]} \varphi \, d\mu + O_N\left(\frac{1 + \|\varphi'\|_\infty}{\log \log n} \right) \right) \qquad (n \to \infty)$$

by the reasoning of Section 5.2.1. To obtain an error term of $O_N\left(\frac{1}{\log n}\right)$, we evaluate the equality (5.9) more precisely. Instead of directly establishing an asymptotic equality for $S(\sigma)$, we divide by $1 - e^{-1/\sigma}$ and multiply the series expression for $S(\sigma)$ together with the other series on the left side. Together with Lemma 7, this process yields

$$\frac{1}{\log N} \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} \mu_{\varphi,k}^{(u)} \ell_N(n-k) \right) e^{-n/\sigma} = \sigma \left(\int_{[0,1]} \varphi \, d\mu + O_N\left(\|\varphi'\|_\infty \frac{\log \sigma}{\sigma} \right) \right), \qquad (\sigma \to \infty)$$

where we let $\ell_N(0) = \log N$ and $\ell_N(n) = \log\left(\frac{n+N}{n+N-1}\right)$ for $n > 0$. Now a direct application of Freud's effective Tauberian theorem yields

$$\sum_{k=0}^{n} \sum_{j=0}^{k} \mu_{\varphi,j}^{(u)} \ell_N(k-j) = n \log N \left(\int_{[0,1]} \varphi \, d\mu + O_N\left(\frac{\|\varphi'\|_\infty}{\log n} \right) \right). \qquad (n \to \infty)$$

The left side of this expression is equal to

$$\begin{aligned}\sum_{k=0}^{n}\left(\mu_{\varphi,k}^{(u)}\log N+\sum_{j=0}^{k-1}\mu_{\varphi,j}^{(u)}\ell_N(k-j)\right) &= \log N\sum_{k=0}^{n}\mu_{\varphi,k}^{(u)}+\sum_{j=0}^{n-1}\mu_{\varphi,j}^{(u)}\sum_{k=j+1}^{n}\ell_N(k-j)\\ &= \log N\sum_{k=0}^{n}\mu_{\varphi,k}^{(u)}+\sum_{j=0}^{n-1}\mu_{\varphi,j}^{(u)}\log\left(\frac{n-j+N}{N}\right)\\ &= \sum_{k=0}^{n}\mu_{\varphi,k}^{(u)}\log(n-k+N),\end{aligned}$$

where the second equality follows from the definition of ℓ_N and telescoping. We can rewrite the last expression above as

$$\log(n+N)\sum_{k=0}^{n}\mu_{\varphi,k}^{(u)}+\sum_{k=1}^{n}\mu_{\varphi,k}^{(u)}\log\left(1-\frac{k}{n+N}\right).$$

So if we can show that

$$\sum_{k=1}^{n}\mu_{\varphi,k}^{(u)}\log\left(1-\frac{k}{n+N}\right)=O_N\left(\|\varphi'\|_\infty\frac{n}{\log n}\right),\qquad (n\to\infty)\tag{5.11}$$

then

$$\sum_{k=0}^{n}\mu_{\varphi,k}^{(u)}=\frac{n}{\log_N n}\left(\int_{[0,1]}\varphi\,d\mu+O_N\left(\frac{\|\varphi'\|_\infty}{\log n}\right)\right).\qquad (n\to\infty)$$

First note that since $\left(\mu_{\varphi,n}^{(u)}\right)_n$ is nonincreasing, we have

$$\mu_{\varphi,n}^{(u)}\sum_{k=0}^{n}\log(n-k+N)\le\sum_{k=0}^{n}\mu_{\varphi,k}^{(u)}\log(n-k+N)=n\log N\left(\int_{[0,1]}\varphi\,d\mu+O_N\left(\frac{\|\varphi'\|_\infty}{\log n}\right)\right),$$

from which it is easy to see that $\mu_{\varphi,n}^{(u)}=O_N\left(\frac{\|\varphi'\|_\infty}{\log n}\right)$ as $n\to\infty$. Now since individual terms on the left side of (5.11) decay to 0 as $n \to \infty$, we can consider the sum starting from $k = 3$. We have

$$\begin{aligned}
\left|\sum_{k=3}^{n} \mu_{\varphi,k}^{(u)} \log\left(1 - \frac{k}{n+N}\right)\right| &= \sum_{k=3}^{n} \mu_{\varphi,k}^{(u)} \sum_{j=1}^{\infty} \frac{1}{j}\left(\frac{k}{n+N}\right)^{j} = \sum_{j=1}^{\infty} \frac{1}{j(n+N)^{j}} \sum_{k=3}^{n} k^{j} \mu_{\varphi,k}^{(u)} \\
&\ll_N \sum_{j=1}^{\infty} \frac{\|\varphi'\|_\infty}{j(n+N)^{j}} \sum_{k=3}^{n} \frac{k^{j}}{\log k} \ll \sum_{j=1}^{\infty} \frac{\|\varphi'\|_\infty}{j(n+N)^{j}} \int_{3}^{n+1} \frac{x^{j}\,dx}{\log x} \\
&\ll \sum_{j=1}^{\infty} \frac{\|\varphi'\|_\infty}{j(n+N)^{j}} \left(\frac{(n+1)^{j+1}}{(j+1)\log(n+1)}\right) \\
&\ll \frac{n}{\log n} \sum_{j=1}^{\infty} \frac{\|\varphi'\|_\infty}{j(j+1)} \left(\frac{n+1}{n+N}\right)^{j} \\
&\ll \|\varphi'\|_\infty \frac{n}{\log n}. \qquad (n \to \infty)
\end{aligned}$$

This proves Theorem 7 in the case that $u = \frac{1}{N}$.

For the general case $u \in (0,1)$, let $N = \left\lceil \frac{1}{u} \right\rceil$ so that $[u,1] \subseteq \left[\frac{1}{N}, 1\right]$. Then for $x \in \left[\frac{1}{N}, 1\right]$, we have

$$\sum_{k=0}^{n} \hat{F}^{k}\varphi(x) = \frac{1}{\log N} \sum_{k=0}^{n} \mu_{\varphi,k}^{(1/N)} + O_N(\|\varphi\|_\infty) = \frac{n}{\log n}\left(\int_{[0,1]} \varphi\,d\mu + O_N\left(\frac{\|\varphi'\|_\infty}{\log n}\right)\right). \qquad (n \to \infty)$$

Integrating the first and last expressions over $[u,1]$ yields

$$\sum_{k=0}^{n} \mu_{\varphi,k}^{(u)} = \frac{n\log(1/u)}{\log n}\left(\int_{[0,1]} \varphi\,d\mu + O_N\left(\frac{\|\varphi'\|_\infty}{\log n}\right)\right), \qquad (n \to \infty)$$

completing the proof of Theorem 7.

5.2.4 The weighted Stern-Brocot subsets

Finally, we prove Theorem 8. Let $\gamma = \frac{v}{w} \in \mathbb{Q} \cap (0,1)$ with $\gcd(v,w) = 1$, $\epsilon \in (0, 1-\gamma)$, and $f \in C^2([0,1])$. By Theorem 6, we have

$$\frac{\log n}{\log((\gamma+\epsilon)/\gamma)} \int_{F^{-n}([\gamma,\gamma+\epsilon])} f\,d\lambda = \int_{[0,1]} f\,d\lambda + O_\gamma\left(\frac{\|f\|_{C^2}}{\log n}\right). \qquad (n \to \infty) \qquad (5.12)$$

(Note that the constants associated with the error term can be chosen large enough so they remain valid independent of ϵ, since these constants tend to grow as γ decreases, not as γ increases.) Let

$\mathscr{F}^n$ be the set of continuous functions $g : [0,1] \to [0,1]$ making up the inverse branches of F^n. Then

$$\begin{aligned}\lim_{\epsilon\to 0^+} \frac{\log n}{\log((\gamma+\epsilon)/\gamma)} \int_{F^{-n}([\gamma,\gamma+\epsilon])} f\,d\lambda &= \lim_{\epsilon\to 0^+} \frac{\gamma \log n}{\epsilon} \sum_{g\in\mathscr{F}^n} \pm \int_{g(\gamma)}^{g(\gamma+\epsilon)} f(x)\,dx \\ &= \gamma \log n \sum_{g\in\mathscr{F}^n} |g'(\gamma)| f(g(\gamma)), \end{aligned} \tag{5.13}$$

where the $\pm$ is chosen to be $+$ if $g(\gamma+\epsilon) \geq g(\gamma)$, and $-$ otherwise. Now the inverse branches of F are the maps $x \mapsto \frac{x}{x+1}$ and $x \mapsto \frac{1}{x+1}$, i.e., the linear fractional transformations determined by the matrices $M_1 = \left(\begin{smallmatrix}1 & 0\\ 1 & 1\end{smallmatrix}\right)$ and $M_2 = \left(\begin{smallmatrix}0 & 1\\ 1 & 1\end{smallmatrix}\right)$. Thus any $g \in \mathscr{F}^n$ is the linear fractional transformation determined by a product of M_1 and M_2, which is of the form $\left(\begin{smallmatrix}p & p'\\ q & q'\end{smallmatrix}\right)$, where $0 \leq p \leq q$, $0 \leq p' \leq q'$, and $pq' - p'q = \pm 1$; hence $g(x) = \frac{px+p'}{qx+q'}$. Now $g(\gamma) = \frac{pv+p'w}{qv+q'w}$ is a fraction in $F^{-n}(\gamma)$ such that $(pv+p'w, qv+q'w) = 1$ since $pq' - p'q = \pm 1$ and $\gcd(v,w) = 1$. We also have $|g'(\gamma)| = \frac{w^2}{(qv+q'w)^2}$. So we can rewrite (5.13) as

$$\log(n^{vw}) \sum_{\substack{p/q\in F^{-n}(v/w)\\ \gcd(p,q)=1}} q^{-2} f\left(\frac{p}{q}\right).$$

Thus, letting $\epsilon \to 0^+$ in (5.12) yields

$$\log(n^{vw}) \sum_{\substack{p/q\in F^{-n}(v/w)\\ \gcd(p,q)=1}} q^{-2} f\left(\frac{p}{q}\right) = \int_{[0,1]} f\,d\lambda + O_\gamma\left(\frac{\|f\|_{C^2}}{\log n}\right). \qquad (n\to\infty)$$

References

[1] J. Aaronson. *An introduction to infinite ergodic theory.* Mathematical Surveys and Monographs 50, A.M.S., Providence, RI, 1997.

[2] J. Aaronson. *Random f-expansions.* Ann. Probab. 14(3) (1986), 1037–1057.

[3] P. Arnoux. *Le codage du flot géodésique sur la surface modulaire.* Enseign. Math. (2) 40(1-2) (1994), 29–48.

[4] E. Artin. *Ein mechanisches system mit quasiergodischen bahnen.* Abh. Math. Sem. Univ. Hamburg 3(1) (1924), 170–175.

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

[6] J. S. Athreya, J. Chaika, S. Leliévre. *The gap distribution of slopes on the golden L.* In: Recent trends in ergodic theory and dynamical systems, Contemp. Math. 631, 47–62. A.M.S., Providence, RI, 2015.

[7] J. S. Athreya, Y. Cheung. *A Poincaré section for the horocycle flow on the space of lattices.* Int. Math. Res. Not. 10 (2014), 2643–2690.

[8] J. S. Athreya, A. Ghosh. *The Erdős-Szüsz-Turán distribution for equivariant processes.* Preprint arXiv:1508.01886 math.DS.

[9] V. Augustin, F. P. Boca, C. Cobeli, A. Zaharescu. *The h-spacing distribution between Farey points.* Math. Proc. Cambridge Philos. Soc. 131(1) (2001), 23–38.

[10] D. A. Badziahin, A. K. Haynes. *A note on Farey fractions with denominators in arithmetic progressions.* Acta Arith. 147(3) (2011), 205–215.

[11] P. Billingsley. Convergence of probability measures. Second ed. Wiley Series in Probability and Statistics, John Wiley and Sons, Inc., New York, 1999.

[12] F. P. Boca. *A problem of Erdős, Szüsz, and Turán concerning Diophantine approximations.* Int. J. Number Theory 4(4) (2008), 691–708.

[13] F. P. Boca. *Products of matrices $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$ and the distribution of reduced quadratic irrationals.* J. Reine Angew. Math. 606 (2007), 149–165.

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

[15] F. P. Boca, R. N. Gologan, A. Zaharescu. *The average length of a trajectory in a certain billiard in a flat two-torus.* New York J. Math. 9 (2003), 303–330.

[16] F. P. Boca, B. Heersink, P. Spiegelhalter. *Gap distribution of Farey fractions under some divisibility constraints.* Integers 13 (2013), A#44.

[17] F. P. Boca, A. Zaharescu. *The correlations of Farey fractions.* J. London Math. Soc. (2) 72(1) (2005), 25–39.

[18] G. Brown, Q. Yin. *Metrical theory for Farey continued fractions.* Osaka J. Math. 33(4) (1996), 951–970.

[19] N. Elkies, C. McMullen. *Gaps in* $\sqrt{n}$ mod 1 *and ergodic theory.* Duke Math. J. 123(1) (2004), 95–139.

[20] P. Erdős, P. Szüsz, P. Turán. *Remarks on the theory of Diophantine approximation.* Colloq. Math. 6 (1958), 119–126.

[21] A. Eskin, C. McMullen. *Mixing, counting, and equidistribution in Lie groups.* Duke Math. J. 71(1) (1993), 181–209.

[22] C. Faivre. *Distribution of Lévy constants for quadratic numbers.* Acta Arith. 61(1) (1992), 13–34.

[23] M. Feigenbaum. *Presentation functions, fixed points, and a theory of scaling function dynamics.* J. Statist. Phys. 52(3-4) (1988), 527–569.

[24] M. Feigenbaum, I. Procaccia, T. Tél. *Scaling properties of multifractals as an eigenvalue problem.* Phys. Rev. A (3) 39(10) (1989), 5359–5372.

[25] J. Fiala, P. Kleban. *Intervals between Farey fractions in the limit of infinite level.* Ann. Sci. Math. Québec 34(1) (2010), 63–71.

[26] A. M. Fisher, T. A. Schmidt. *Distribution of approximants and geodesic flows.* Ergodic Theory Dynam. Systems 34(6) (2014), 1832–1848.

[27] G. Freud. *Restglied eines Tauberschen Satzes I.* Acta Math. Acad. Sci. Hung. 2 (1951), 299–308.

[28] A. Grothendieck. *Produits tensoriels topologiques et espaces nucléaires.* Mem. Amer. Math. Soc. 16 (1955).

[29] A. Grothendieck. *La théorie de Fredholm.* Bull. Soc. Math. France 84 (1956), 319–384.

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

[31] R. R. Hall, G. Tenenbaum. *On consecutive Farey arcs.* Acta Arith. 44(4) (1984), 397–405.

[32] G. H. Hardy, E. M. Wright. *An introduction to the theory of numbers.* Sixth ed. Revised by D. R. Heath-Brown, J. H. Silverman. Oxford Univ. Press, Oxford, 2008.

[33] G. Hedlund. *Fuchsian groups and transitive horocycles.* Duke Math. J. 2(3) (1936), 530–542.

[34] B. Heersink. *An effective estimate for the Lebesgue measure of preimages of iterates of the Farey map.* Adv. Math. 291 (2016), 621–634.

[35] B. Heersink. *Poincaré sections for the horocycle flow in covers of* $\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$ *and applications to Farey fraction statistics*. Monatsh. Math. 179(3) (2016), 389–420.

[36] 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, 839–853. Int. Press, Somerville, MA, 2000.

[37] R. Howe, C. C. Moore. *Asymptotic properties of unitary representation*. J. Funct. Anal. 32(1) (1979), 72–96.

[38] M. Iosifescu, C. Kraaikamp. *Metrical theory of continued fractions*. Mathematics and Its Applications 547, Kluwer Academic Publishers, Dordrecht, 2002.

[39] S. Isola. *On the spectrum of Farey and Gauss maps*. Nonlinearity 15(5) (2002), 1521–1539.

[40] S. Ito. *Algorithms with mediant convergents and their metrical theory*. Osaka J. Math. 26(3) (1989), 557–578.

[41] J. Kallies, A. Özlük, M. Peter, C. Snyder. *On asymptotic properties of a number theoretic function arising out of a spin chain model in statistical mechanics*. Comm. Math. Phys. 222(1) (2001), 9–43.

[42] J. Karamata. *Neuer Beweis und Verallgemeinerung der Tauberschen Sätze, welche die Laplacesche und Stieltjessche Transformation betreffen*. J. Reine Angew. Math. 164 (1931) 27–39.

[43] P. Kargaev, A. Zhigljavsky. *Asymptotic distribution of the distance function to the Farey points*. J. Number Theory 65(1) (1997), 130–149.

[44] T. Kato. *Perturbation theory for linear operators*. Classics in Mathematics, Springer-Verlag, Berlin, 1995.

[45] D. Kelmer. *Quadratic irrationals and linking numbers of modular knots*. J. Mod. Dyn. 6(4) (2012), 539–561.

[46] M. Kesseböhmer, M. Slassi. *A distributional limit law for the continued fraction digit sum*. Math. Nachr. 281(9) (2008), 1294–1306.

[47] M. Kesseböhmer, M. Slassi. *Limit laws for distorted critical return time processes in infinite ergodic theory*. Stoch. Dyn. 7(1) (2007), 103–121.

[48] M. Kesseböhmer, B. Stratmann. *A dichotomy between uniform distributions of the Stern-Brocot and the Farey sequence*. Unif. Distrib. Theory 7(2) (2012), 21–33.

[49] M. Kesseböhmer, B. Stratmann. *On the asymptotic behavior of the Lebesgue measure of sum-level sets for continued fractions*. Discrete Contin. Dyn. Sys. 32(7) (2012), 2437–2451.

[50] H. Kesten. *Some probabilistic theorems on Diophantine approximations*. Trans. Amer. Math. Soc. 103 (1962), 189–217.

[51] H. Kesten, V. T. Sós. *On two problems of Erdős, Szüsz and Turán concerning diophantine approximations*. Acta Arith. 12 (1966), 183–192.

[52] J. Korevaar. *Tauberian Theory: A Century of Developments*. Grundlehren der Mathematischen Wissenschaften 329, Springer-Verlag, Berlin, 2004.

[53] J. Marklof. *The asymptotic distribution of Frobenius numbers*. Invent. Math. 181(1) (2010), 179–207.

[54] J. Marklof. *Fine-scale statistics for the multidimensional Farey sequence*. In: Limit Theorems in Probability, Statistics and Number Theory, Springer Proc. Math. Stat. 42, 49–57. Springer, Heidelberg, 2013.

[55] J. Marklof, A. Strömbergsson. *The distribution of free path lengths in the periodic Lorentz gas and related lattice point problems*. Ann. Math. 172(3) (2010), 1949–2033.

[56] D. Mayer. *On a ζ function related to the continued fraction transformation*. Bull. Soc. Math. France 104(2) (1976), 195–203.

[57] I. Melbourne, D. Terhesiu. *Operator renewal theory and mixing rates for dynamical systems with infinite measure*. Invent. Math. 189(1) (2012), 61–110.

[58] J. von Neumann. *Proof of the quasi-ergodic hypothesis*. Proc. Nat. Acad. Sci. U.S.A. 18(1) (1932), 70–82.

[59] M. Pollicott. *Distribution of closed geodesics on the modular surface and quadratic irrationals*. Bull. Soc. Math. France 114(4) (1986), 431–446.

[60] T. Prellberg. *Towards a complete determination of the spectrum of a transfer operator associated with intermittency*. J. Phys. A 36(10) (2003), 2455–2461.

[61] V. A. Rohlin. *Exact endomorphisms of a Lebesgue space*. Izv. Akad. Nauk SSSR Ser. Mat. 25 (1961), 499–530. English translation: Amer. Math. Soc. Transl. (2) 39 (1964), 1–36.

[62] P. Sarnak. *Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series*. Comm. Pure Appl. Math. 34(6) (1981), 719–739.

[63] C. Series. *The modular surface and continued fractions*. J. London Math. Soc. (2) 31(1) (1985), 69–80.

[64] C. Silva. *On μ-recurrent nonsingular endomorphisms*. Israel J. Math. 61(1) (1988), 1–13.

[65] A. Strömbergsson. *On the uniform equidistribution of long closed horocycles*. Duke Math. J. 123(3) (2004), 507–547.

[66] G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*. Cambridge Studies in Advanced Mathematics 46, Cambridge Univ. Press, Cambridge, 1995.

[67] A. V. Ustinov. *Spin chains and Arnold's problem on the Gauss-Kuz'min statistics for quadratic irrationals*. Mat. Sb. 204(5) (2013), 143–160. English translation: Sb. Math. 204(5) (2013), 762–779.

[68] C. Uyanik, G. Work. *The distribution of gaps for saddle connections on the octagon*. Int. Math. Res. Not. 18 (2016), 5569–5602.

[69] M. Xiong, A. Zaharescu. *Correlation of fractions with divisibility constraints*. Math. Nachr. 284(2-3) (2011), 393–407.

[70] M. Xiong, A. Zaharescu. *A problem of Erdős-Szüsz-Turán on Diophantine approximation.* Acta Arith. 125(2) (2006), 163–177.

[71] 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, 275–301. Tata Inst. Fund. Res., Bombay, 1981.