# ON THE UNIFORM EQUIDISTRIBUTION OF LONG CLOSED HOROCYCLES

ANDREAS STRÖMBERGSSON

**Abstract**

*It is well known that on any given hyperbolic surface of finite area, a closed horocycle of length $\ell$ becomes asymptotically equidistributed as $\ell \to \infty$. In this paper we prove that any subsegment of length greater than $\ell^{1/2+\varepsilon}$ of such a closed horocycle also becomes equidistributed as $\ell \to \infty$. The exponent $1/2 + \varepsilon$ is the best possible and improves upon a recent result by Hejhal [He3]. We give two proofs of the above result; our second proof leads to explicit information on the rate of convergence. We also prove a result on the asymptotic joint equidistribution of a finite number of distinct subsegments having equal length proportional to $\ell$.*

**Contents**

1. Introduction . . . . . 507
2. Some preliminaries on $\Gamma \setminus \mathscr{H}$ and the distribution of cusps . . . . . 510
3. Proof of Theorem 1 using ergodic theory . . . . . 516
4. Spectral theory and the rate of convergence in (1.2) . . . . . 521
5. Bounding $\sum_{n=1}^{N} c_n e(n\alpha)$ for residual eigenfunctions . . . . . 531
6. Joint equidistribution of subsegments . . . . . 537
7. Some further results, applications, and comments . . . . . 542
References . . . . . 545

## 1. Introduction

Let $\Gamma$ be a cofinite Fuchsian group acting on the Poincaré upper half-plane $\mathscr{H}$ with metric $ds = |dz|/y$. We assume that $\Gamma \setminus \mathscr{H}$ has at least one cusp. By an auxiliary conjugation, we may then assume that $\Gamma \setminus \mathscr{H}$ has one cusp located at $\infty$ and that the isotropy group $\Gamma_\infty$ is generated by the translation $S(z) = z + 1$.

For any $y > 0$, the curve $\{x + iy \mid x \in [0, 1]\}$ is a *closed horocycle* of length $1/y$ on $\Gamma \setminus \mathscr{H}$. When $y \to 0$, this curve is known to become *equidistributed* on $\Gamma \setminus \mathscr{H}$ with

DUKE MATHEMATICAL JOURNAL
Vol. 123, No. 3, © 2004
Received 26 February 2003. Revision received 20 August 2003.
2000 *Mathematics Subject Classification*. Primary 37D40; Secondary 11F, 30F35.
Author's work partially supported by the European Commission under Research Training Network (Mathematical Aspects of Quantum Chaos) grant number HPRN-CT-2000-00103 under the Improving Human Potential Programme.

respect to the Poincaré area $\mu$. (Recall that $d\mu = y^{-2}\,dx\,dy$.) Investigations related to this fact have been carried out by a number of people over the years, including Selberg (unpublished), Zagier [Z], Sarnak [S], Hejhal [He2], [He3], and Flaminio and Forni [FF].

By elementary functional analysis, the equidistribution fact just mentioned is equivalent to the assertion that

$$\int_0^1 f(x+iy)\,dx \to \frac{1}{\mu(\Gamma\setminus\mathscr{H})}\int_{\Gamma\setminus\mathscr{H}} f(z)\,d\mu(z)$$

holds for every compactly supported function $f \in C(\Gamma\setminus\mathscr{H})$ as $y \to 0^+$.

In [He3], Hejhal asked the following question. To exactly what degree of *uniformity* does this equidistribution result hold? Specifically, for numbers $\alpha = \alpha(y) < \beta = \beta(y)$, under what conditions do we have

$$\frac{1}{\beta-\alpha}\int_\alpha^\beta f(x+iy)\,dx \to \frac{1}{\mu(\Gamma\setminus\mathscr{H})}\int_{\Gamma\setminus\mathscr{H}} f(z)\,d\mu(z) \tag{1.1}$$

as $y \to 0$?

The main result in [He3] is that there exists a positive constant $c(\Gamma) \leqq 1/3$, which depends only on the group $\Gamma$, such that (1.1) holds as $y \to 0$ so long as $\beta - \alpha$ is kept bigger than $y^{c(\Gamma)-\varepsilon}$. The proof is based on spectral-theoretic techniques. (The constant $c(\Gamma)$ obtained in [He3] depends on the smallest eigenvalue $\lambda_1 > 0$ of the Laplacian on $\Gamma\setminus\mathscr{H}$; in particular, [He3] gives $c(\Gamma) = 1/3$ if and only if $\lambda_1 \geqq 3/16$.)

For *fixed* $\alpha$ and $\beta$, a proof of (1.1) was outlined earlier in [He2, page 44], and the same assertion can also be obtained using ergodic-theoretic techniques (see [EM, Theorem 7.1]* as well as [Sh]).

In this paper, we improve the result from [He3] to show that, for *any* Fuchsian group $\Gamma$ as above, we may in fact take $c(\Gamma) = 1/2$. In other words, we prove the following.

THEOREM 1
*Let $\Gamma$ be a cofinite Fuchsian group such that $\Gamma\setminus\mathscr{H}$ has a cusp at $\infty$, with $\Gamma_\infty = [z \mapsto z+1]$. Let $\delta > 0$, and let $f : \mathscr{H} \to \mathbb{C}$ be any fixed, bounded, continuous, and $\Gamma$-invariant function. Then*

$$\frac{1}{\beta-\alpha}\int_\alpha^\beta f(x+iy)\,dx \to \frac{1}{\mu(\Gamma\setminus\mathscr{H})}\int_{\Gamma\setminus\mathscr{H}} f(z)\,d\mu(z) \tag{1.2}$$

*uniformly as $y \to 0$ so long as $\beta - \alpha$ remains bigger than $y^{1/2-\delta}$.*

*The proof in [EM] is for closed horocycles, that is, $\alpha = 0$, $\beta = 1$, but the proof can be adapted to work for any fixed $\alpha < \beta$.

The exponent $c(\Gamma) = 1/2$ is the best possible result because, as remarked in [He3, page 840], it is easily seen that there are numerous cases of $\alpha = \alpha(y)$, $\beta = \beta(y)$ with $\beta - \alpha = [\text{const}]\, y^{1/2}$ as $y \to 0$ such that all of the horocycle segments $[\alpha, \beta] + iy$ stay far out in one cusp of $\Gamma \setminus \mathscr{H}$.

We give two different proofs of Theorem 1. In §3 we give a proof using ergodic properties of the horocycle flow on $\Gamma \setminus \mathrm{PSL}(2, \mathbb{R})$. This approach leads to a more general version of Theorem 1 in that we obtain asymptotic equidistribution on the unit tangent bundle of $\Gamma \setminus \mathscr{H}$. We also replace the boundedness assumption on $f$ by a weaker condition on the growth of $f$ in each cusp.

Our second proof of Theorem 1 is given in §§4 and 5. This proof uses spectral theory and yields a stronger result than Theorem 1 as we also obtain explicit information on the *rate of convergence* in (1.2) (albeit for a more restricted class of test functions $f$). Such information does not seem possible to obtain using the ergodic methods of §3.

We remark that our approach in §§4 and 5 is different from the one used in [He3], although both are based on spectral-theoretic techniques in a completely classical style. The present proof uses the spectral expansion of the given test function $f$ in a direct way, and the argument ultimately relies on uniform Rankin-Selberg-type bounds on the Fourier coefficients $c_n$ of the eigenfunctions. In case there are small eigenvalues present ($0 < \lambda < 1/4$), we also need to use bounds on sums of the form $\sum_{n=1}^{N} c_n e(n\alpha)$. For a Maass cusp form, this sum is known to be bounded by $O(N^{1/2+\varepsilon})$; however, for a noncuspidal Maass waveform, the best possible uniform bound is $O(N^{1-\sqrt{1/4-\lambda}})$, which we prove in Proposition 5.1 (see also Remark 5.2).

We should mention a recent paper by Flaminio and Forni, [FF], in which a detailed analysis is made of the invariant distributions and the cohomological equation for the horocycle flow on $\Gamma \setminus \mathrm{SL}(2, \mathbb{R})$. One possible alternative approach to the question of the rate of convergence in (1.2) would be to build on the results in that paper, adapting the technique in [FF, §5].

We next turn to another natural question concerning the deeper properties of the asymptotic distribution of the closed horocycle $[0, 1] + iy$. To what extent are we able to assert that distinct subsegments $[\alpha_j, \alpha_j + \ell] + iy$ ($j = 1, \dots, N$) tend to become more and more decorrelated position-wise on $\Gamma \setminus \mathscr{H}$ as $y \to 0$? This question was raised by Hejhal in [He2] in connection with a heuristic and numerical study of the sum

$$S_{y,N}(x) = \sum_{j=0}^{N-1} F\Big(\frac{x+j}{N} + iy\Big).$$

Here $F$ is a test function on $\Gamma \setminus \mathscr{H}$, $x \in [0, 1]$, and $N \to \infty$, $y \to 0$ in such a way that $Ny \to 0$. The numerical studies in [He2] carried out on the nonarithmetic Hecke triangle groups $\Gamma = \mathbb{G}_5$ and $\Gamma = \mathbb{G}_7$ showed that for several choices of functions $F$

(of mean zero), the value distribution of $N^{-1/2} \cdot S_{y,N}(x)$ with respect to $x \in [0, 1]$ clearly approached a Gaussian curve. However, for the arithmetic group $\Gamma = \mathbb{G}_3 = \mathrm{PSL}(2, \mathbb{Z})$, this Gaussian behavior broke down completely, and an explanation of this phenomenon was given based on the existence of Hecke operators on $\mathrm{PSL}(2, \mathbb{Z})$.

In §6, we study this question for fixed $N$ by applying, on the group $\mathrm{PSL}(2, \mathbb{R})^N$, the topological rigidity of unipotent flows proved by Ratner in [R2] together with a theorem by Shah [Sh, Theorem 1.4] on the asymptotic equidistribution of expanding translates of certain orbits on homogeneous spaces. We prove in Theorem 4 that *unless* there are Hecke symmetries present to force correlations, distinct segments $[\alpha_j, \alpha_j + \ell] + iy$ indeed go decorrelated, and even jointly equidistributed, as $y \to 0$ (with $\ell$ fixed). As an application, using the central limit theorem for independent random variables, we show that on any nonarithmetic Hecke triangle group $\Gamma = \mathbb{G}_L$ (i.e., $L = 5$ or $L \geqq 7$), if $y = y(N)$ tends to zero *sufficiently rapidly* as $N \to \infty$, then the value distribution of $N^{-1/2} \cdot S_{y,N}(x)$ indeed has a Gaussian limit, as expected from [He2] (cf. Corollary 6.5 and Remark 6.6). We should stress, however, that we do not know of any way to make "sufficiently rapidly" effective in this statement, and we are still very far from being able to prove any result about $S_{y,N}(x)$ in the case that receives most attention in [He2], namely, $y = N^{-1-\varepsilon}$.

In the last section, §7, we state some further extensions and applications of Theorem 1; in particular, we present a new result on the value distribution of the generalized theta sum

$$\Theta_f(x + iy) = y^{1/4} \sum_{n\in\mathbb{Z}} f(ny^{1/2})e(n^2x),$$

which was studied by Marklof in [Mar1].

## 2. Some preliminaries on $\Gamma \setminus \mathscr{H}$ and the distribution of cusps

We start by introducing some notation that is in force throughout this paper. (To a large extent, our notation is the same as in [He1, page 268].)

We let $\Gamma$ be as in the introduction; that is, $\Gamma$ is a cofinite Fuchsian group having a normalized cusp at $\infty$. We let $\mathscr{F} \subset \mathscr{H}$ be a canonical (closed) fundamental domain for $\Gamma \setminus \mathscr{H}$, and we let $\eta_1 = \infty, \eta_2, \ldots, \eta_\kappa$ (where $\kappa \geqq 1$) be the vertices of $\mathscr{F}$ along $\partial\mathscr{H} = \mathbb{R} \cup \{\infty\}$. Since $\mathscr{F}$ is canonical, $\eta_1, \ldots, \eta_\kappa$ are $\Gamma$-inequivalent.

We denote

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

For each $k \in \{1, \ldots, \kappa\}$, we choose $N_k \in \mathrm{PSL}(2, \mathbb{R})$ such that $N_k(\eta_k) = \infty$ and such that the stabilizer $\Gamma_{\eta_k}$ is $[T_k]$, where $T_k := N_k^{-1} S^{-1} N_k$. We will always keep $N_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Since $\mathscr{F}$ is canonical, its intersection with $\{z \in \mathscr{H} \mid \operatorname{Im} z \geqq B\}$ for $B$ large is a vertical strip $[x, x+1] \times [B, \infty) \subset \mathscr{H}$; without loss of generality, we may assume that $\mathscr{F}$ was chosen so that $x = 0$. Then, by modifying $N_k$ for $k = 2, \ldots, \kappa$, we can ensure that

$$N_k(\mathscr{F}) \bigcap \left\{ z \in \mathscr{H} \,\middle|\, \operatorname{Im} z \geqq B \right\} = [0,1] \times [B, \infty) \tag{2.1}$$

holds for all $k \in \{1, \ldots, \kappa\}$ and for all $B \geqq B_0$, where $B_0 = B_0(\Gamma) > 1$ is a constant fixed once and for all.

For $B \geqq B_0$, the corresponding cuspidal region in $\mathscr{F}$ is called $\mathscr{C}_{kB}$:

$$\mathscr{C}_{kB} = N_k^{-1}\big([0,1] \times [B, \infty)\big) \subset \mathscr{F}. \tag{2.2}$$

We then define

$$\mathscr{F}_B = \mathscr{F} - \bigcup_{k=1}^{\kappa} \mathscr{C}_{kB}. \tag{2.3}$$

This is a bounded region.

Closely related to the above splitting of the fundamental domain is the *invariant height* function, $\mathscr{Y}_\Gamma(z)$. This is defined by

$$\mathscr{Y}_\Gamma(z) = \max_{k \in \{1,\ldots,\kappa\}} \max_{W \in \Gamma} \operatorname{Im} N_k W(z) \quad (z \in \mathscr{H}) \tag{2.4}$$

(cf. [I, (3.8)]). The function $\mathscr{Y}_\Gamma(z)$ is well known to be continuous and is $\Gamma$-invariant and is bounded from below by a positive constant that depends only on the group $\Gamma$. Notice that we have $\mathscr{Y}_\Gamma(z) \to \infty$ when $z \in \mathscr{F}$ approaches any of the cusps.

One important step in the ergodic-theoretic proof of Theorem 1 is carried out in this section. We show that it is not possible for a positive proportion of the horocycle segment $[\alpha, \beta] + iy$ to escape into some cusp as $y \to 0$, so long as we keep $\beta - \alpha \geqq y^{1/2-\delta}$. A precise form of this statement is given in the following proposition.

PROPOSITION 2.1
*Given any number $\varepsilon > 0$, there exists a continuous, $\Gamma$-invariant function $f : \mathscr{H} \to [0,1]$ which has* compact support *on $\Gamma \setminus \mathscr{H}$, such that the following holds. For any $\delta > 0$, there is a $y_0 > 0$ such that*

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta f(x + iy)\, dx \geqq 1 - \varepsilon$$

*for all $0 < y < y_0$ and all $\alpha, \beta \in \mathbb{R}$ such that $\beta - \alpha \geqq y^{1/2-\delta}$.*

Without much extra difficulty, we can in fact prove a stronger result, the use of which later on allows us to replace the boundedness assumption in Theorem 1 by the weaker assumption $f(z) = O\big(\sqrt{\mathscr{Y}_\Gamma(z)}\big)$.

Given any constant $M \geqq 0$, we define the cutoff function

$$\lfloor x \rfloor_M := \begin{cases} x & \text{if } x \geqq M, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

PROPOSITION 2.2

*Let $\delta > 0$ be given. We then have*

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta \sqrt{\mathscr{Y}_\Gamma(x+iy)}\, dx = O(1) \tag{2.6}$$

*for all $0 < y < 1$ and all $\alpha, \beta \in \mathbb{R}$ such that $\beta - \alpha \geqq y^{1/2-\delta}$. The implied constant depends only on $\Gamma$ and $\delta$. Furthermore, given any $M \geqq 10$, there is some $y_0 = y_0(M, \delta) > 0$ such that*

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta \sqrt{\lfloor \mathscr{Y}_\Gamma(x+iy) \rfloor_M}\, dx \leqq \frac{25\kappa}{\sqrt{M}} \tag{2.7}$$

*for all $0 < y < y_0$ and all $\alpha, \beta \in \mathbb{R}$ such that $\beta - \alpha \geqq y^{1/2-\delta}$.*

*Proof of Proposition 2.1 using Proposition 2.2*

Take $M \geqq 10$ so large that $25\kappa/\sqrt{M} < \varepsilon$, and let $F : \mathbb{R}^+ \to [0, 1]$ be a continuous function satisfying $F(y) = 1$ for $0 < y \leqq M$ and $F(y) = 0$ for $y \geqq M + 1$. Set $f(z) = F(\mathscr{Y}_\Gamma(z))$. Using $f(z) \geqq 1 - \sqrt{\lfloor \mathscr{Y}_\Gamma(z) \rfloor_M}$ ($\forall z \in \mathscr{H}$) and (2.7) in Proposition 2.2, we obtain the desired result. □

*Proof of Proposition 2.2*

We first prove the second assertion, (2.7). Using (2.4) and $N_k[T_k] = [S]N_k$, we see that the left-hand side in (2.7) is bounded from above by

$$\sum_{k=1}^{\kappa} \sum_{W_0 \in [T_k] \setminus \Gamma} \frac{1}{\beta - \alpha} \int_\alpha^\beta \sqrt{\lfloor \operatorname{Im} N_k W_0(x+iy) \rfloor_M}\, dx, \tag{2.8}$$

where $[T_k] \setminus \Gamma$ denotes a set of representatives of the cosets $\{[T_k]W \mid W \in \Gamma\}$. We temporarily fix some $M \geqq 10$, some $y \in (0, 1)$, and some $\alpha, \beta$ such that $\beta - \alpha \geqq y^{1/2-\delta}$.

Take $k \in \{1, \ldots, \kappa\}$, and look at any $W_0 \in [T_k] \setminus \Gamma$ that gives a nonzero contribution to (2.8). Write $N_k W_0 = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then there is some $x \in [\alpha, \beta]$ for which

$$M \leqq \operatorname{Im} N_k W_0(x+iy) = \frac{y}{|c(x+iy)+d|^2}. \tag{2.9}$$

Hence $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \notin [S]$ since $M > y$. Using Lemma 2.3 below (with $j = 1$, $N_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$), we now get $|c| \geqq 1$. From (2.9) we also get $|c| \leqq (My)^{-1/2}$ and $|x + d/c| \leqq y^{1/2}|c|^{-1}M^{-1/2} \leqq (\beta - \alpha)M^{-1/2}$, and hence

$$-\frac{d}{c} \in [\mu, \nu], \quad \text{where } \mu = \alpha - \frac{\beta - \alpha}{\sqrt{M}}, \ \ \nu = \beta + \frac{\beta - \alpha}{\sqrt{M}}. \tag{2.10}$$

Now notice that

$$\begin{aligned} \int_\alpha^\beta \sqrt{\lfloor \operatorname{Im} N_k W_0(x + iy) \rfloor_M}\, dx &\leqq \int_{-\infty}^{\infty} \sqrt{\Big\lfloor \frac{y}{(cx + d)^2 + (cy)^2} \Big\rfloor_M}\, dx \\ &= \frac{\sqrt{y}}{|c|} \int_1^{(Mc^2y)^{-1}} \frac{du}{\sqrt{u}\sqrt{u - 1}}. \end{aligned}$$

(We substituted $x = -d/c \pm y\sqrt{u - 1}$.) Using $(u(u - 1))^{-1/2} \leqq (u - 1)^{-1/2}$ for $1 \leqq u \leqq 2$ and $(u(u - 1))^{-1/2} \leqq 2/u$ for $2 \leqq u$, we see that $\int_1^D (u(u - 1))^{-1/2}\, du \leqq 2(1 + \log D)$ for all $D \geqq 1$. Hence

$$\int_\alpha^\beta \sqrt{\lfloor \operatorname{Im} N_k W_0(x + iy) \rfloor_M}\, dx \leqq \frac{2\sqrt{y}}{|c|} \left(1 + 2 \log\left(\frac{1}{\sqrt{My}|c|}\right)\right).$$

In conclusion, we see that the left-hand side in (2.7) is bounded from above by

$$\frac{1}{\beta - \alpha} \sum_{k=1}^{\kappa} \sum_{W_0} \frac{2\sqrt{y}}{|c|} \left(1 + 2 \log\left(\frac{1}{\sqrt{My}|c|}\right)\right), \tag{2.11}$$

where the inner sum is taken over all $W_0 \in [T_k] \setminus \Gamma$ such that $1 \leqq |c| \leqq (My)^{-1/2}$ and $-d/c \in [\mu, \nu]$ for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = N_k W_0$.

Sums similar to the inner sum in (2.11) also occur in §5, where we prove a bound on sums of Fourier coefficients of Maass waveforms of residual type. The following three lemmas give useful inequalities related to these types of sums.

LEMMA 2.3

*Let $j, k \in \{1, \ldots, \kappa\}$ be given. Take $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in N_k \Gamma N_j^{-1}$. Then*

$$|c| \geqq 1, \quad \textit{unless } k = j \textit{ and } \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in [S].$$

*Proof*

We write $T = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = N_k W N_j^{-1}$ with $W \in \Gamma$. We then find that $TST^{-1} = \left(\begin{smallmatrix} * & * \\ -c^2 & * \end{smallmatrix}\right)$ and $TST^{-1} \in N_k \Gamma N_k^{-1}$. But $N_k \Gamma N_k^{-1}$ is a Fuchsian group having $\infty$ as a cusp, with stabilizer $(N_k \Gamma N_k^{-1})_\infty = [S]$. Hence by Shimizu's lemma (cf. [Shi, Lemma 4] or [Mi, Lemma 1.7.3]), we have either $|-c^2| \geqq 1$ or $TST^{-1} \in [S]$. In the second case, $T^{-1}(\infty)$ is a fix-point of $S$, and thus $T^{-1}(\infty) = \infty$, which gives $W(\eta_j) = \eta_k$ and hence $k = j$ and $W \in [T_k] = N_k^{-1}[S]N_k$, and $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = N_k W N_k^{-1} \in [S]$. □

We define

$$\mathfrak{C}^k_{\mu\nu}(X) := \sharp\Big\{W_0 = N_k^{-1}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in [T_k] \setminus \Gamma \;\Big|\; 0 < |c| \leqq X,\ -\frac{d}{c} \in [\mu, \nu]\Big\}. \tag{2.12}$$

LEMMA 2.4

*For any $k \in \{1, \ldots, \kappa\}$, $\mu < \nu$, and $X > 0$, we have*

$$\mathfrak{C}^k_{\mu\nu}(X) \leqq (\nu - \mu)X^2 + 1.$$

*Proof (Cf. [I, Proposition 2.8])*

Let $M$ be the set occurring in the right-hand side in the definition of $\mathfrak{C}^k_{\mu\nu}(X)$. If $W_0 = N_k^{-1}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $W_0' = N_k^{-1}\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$ are any two distinct elements of $M$, then $W_0 W_0'^{-1} \notin [T_k]$ and $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)^{-1} = N_k W_0 W_0'^{-1} N_k^{-1} \in N_k \Gamma N_k^{-1}$. Hence by Shimizu's lemma (cf. the proof of Lemma 2.3), we have $|cd' - dc'| \geqq 1$. But $0 < |c|, |c'| \leqq X$; hence

$$\left|\frac{d'}{c'} - \frac{d}{c}\right| \geqq |cc'|^{-1} \geqq X^{-2}.$$

The lemma now follows by ordering the elements in $M$ by increasing quotients $d/c$ and then adding the above inequality over all pairs of consecutive elements in $M$. □

LEMMA 2.5

*Let real numbers $\mu < \nu$ and $1 \leqq A \leqq B$ be given, and let $\sum_{W_0}$ refer to a sum over a set of representatives $W_0 = N_k^{-1}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in [T_k] \setminus \Gamma$ restricted by $A \leqq |c| \leqq B$ and $-d/c \in [\mu, \nu]$. We then have the following bounds:*

$$\sum_{W_0} \frac{1}{|c|^\delta} \leqq \frac{2}{2-\delta}(\nu - \mu)B^{2-\delta} + A^{-\delta} \quad \textit{for any } 0 < \delta < 2,$$

$$\sum_{W_0} \frac{1}{|c|^\delta} \leqq \frac{\delta}{\delta-2}(\nu - \mu)A^{2-\delta} + A^{-\delta} \quad \textit{for any } \delta > 2,$$

$$\sum_{W_0} \frac{1}{|c|} \log\left(\frac{B}{|c|}\right) \leqq 2(\nu - \mu)B + \frac{1}{A}\log\left(\frac{B}{A}\right).$$

(Notice that the second bound also holds for $B = \infty$, as follows by taking the limit $B \to \infty$.)

*Proof*

This is proved by standard integration by parts and by the use of Lemma 2.4. We give the details for only the third sum. We have, for any $\gamma \in (0,1)$,

$$\begin{aligned}\sum_{W_0} \frac{1}{|c|} \log\left(\frac{B}{|c|}\right) &\leqq \int_{A-\gamma}^{B} \frac{1}{x} \log\left(\frac{B}{x}\right) d\mathfrak{C}^k_{\mu\nu}(x)\\ &= \left[\frac{1}{x}\log\left(\frac{B}{x}\right)\mathfrak{C}^k_{\mu\nu}(x)\right]_{x=A-\gamma}^{x=B} + \int_{A-\gamma}^{B} \left(\frac{1}{x^2}\log\left(\frac{B}{x}\right) + \frac{1}{x^2}\right)\mathfrak{C}^k_{\mu\nu}(x)\,dx.\end{aligned}$$

Here the first term is at most zero, and thus by Lemma 2.4, the whole expression is

$$\leqq \int_{A-\gamma}^{B} \left(\frac{1}{x^2}\log\left(\frac{B}{x}\right) + \frac{1}{x^2}\right)\big((\nu-\mu)x^2 + 1\big)\,dx.$$

Substituting $x = B/u$, we obtain that the whole expression is

$$\begin{aligned}&= (\nu-\mu)B \int_1^{B/(A-\gamma)} \frac{\log(u)+1}{u^2}\,du + \frac{1}{B}\int_1^{B/(A-\gamma)} \big(\log(u)+1\big)\,du\\ &\leqq (\nu-\mu)B\left[\frac{-\log(u)-2}{u}\right]_{u=1}^{u=\infty} + \frac{1}{B}\big[u\log(u)\big]_{u=1}^{u=B/(A-\gamma)}\\ &= 2(\nu-\mu)B + \frac{1}{A-\gamma}\log\left(\frac{B}{A-\gamma}\right).\end{aligned}$$

The desired inequality follows by letting $\gamma \to 0$. □

We continue onward with the proof of Proposition 2.2. If $My > 1$, then the sum in (2.11) is empty; if $My \leqq 1$, Lemma 2.5 implies that the sum is bounded from above by

$$\kappa \frac{2\sqrt{y}}{\beta-\alpha}\left[\frac{2(\nu-\mu)}{\sqrt{My}} + 1 + \frac{4(\nu-\mu)}{\sqrt{My}} + 2\log\left(\frac{1}{\sqrt{My}}\right)\right].$$

From (2.10) and $M \geqq 10$, it follows that $\nu - \mu \leqq 2(\beta-\alpha)$. Using $\beta - \alpha \geqq y^{1/2-\delta}$, we finally conclude that

$$\begin{aligned}\frac{1}{\beta-\alpha}\int_\alpha^\beta \sqrt{\lfloor \mathscr{Y}_\Gamma(x+iy)\rfloor_M}\,dx \qquad (2.13)\\ \leqq \frac{24\kappa}{\sqrt{M}} + 2\kappa\, y^\delta\left(1 + 2\log^+\left(\frac{1}{\sqrt{My}}\right)\right).\end{aligned}$$

This holds for all $0 < y < 1$ and all $\alpha, \beta \in \mathbb{R}$ such that $\beta - \alpha \geqq y^{1/2-\delta}$. The second assertion in Proposition 2.2 follows immediately from this inequality.

The first assertion, (2.6), follows from (2.13) with $M = 10$ and the inequality $\sqrt{\mathscr{Y}_\Gamma(z)} \leqq \sqrt{10} + \sqrt{\lfloor \mathscr{Y}_\Gamma(z)\rfloor_{10}}$. □

## 3. Proof of Theorem 1 using ergodic theory

Let $G$ denote the group $\mathrm{PSL}(2, \mathbb{R})$. In this section we give a proof of Theorem 1 based on the ergodic properties of the horocycle flow on $\Gamma \setminus G$. We start by recalling some basic facts concerning the geodesic and horocycle flow and the standard identification of $G$ and $T_1\mathscr{H}$, the unit tangent bundle of $\mathscr{H}$ (cf. [S, §1], [M, §3.1], [Mar2, §2.3]).

We use the notation

$$S^x := \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{a}(y) := \begin{pmatrix} \sqrt{y} & 0 \\ 0 & 1/\sqrt{y} \end{pmatrix}, \qquad \boldsymbol{k}(\vartheta) = \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}.$$

We let $T_1\mathscr{H}$ denote the unit tangent bundle of $\mathscr{H}$. $T_1\mathscr{H}$ is parametrized by $(z, \theta) \in \mathscr{H} \times (\mathbb{R}/2\pi\mathbb{Z})$, where $\theta$ is an angular variable measured from the upward vertical counterclockwise. The action of $G$ on $T_1\mathscr{H}$ is given by

$$T(z, \theta) := \big(T(z), \theta - 2\vartheta_T(z)\big), \tag{3.1}$$

where

$$\vartheta_T(z) = \arg(cz + d) \quad \text{for} \quad T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We now identify $T_1\mathscr{H}$ and $G$ (as manifolds) through

$$G \to T_1\mathscr{H}, \qquad T \longmapsto T(i, 0) = \big(T(i), -2\vartheta_T(i)\big). \tag{3.2}$$

The inverse of this identification map is given by

$$T_1\mathscr{H} \to G, \qquad (x + iy, \theta) \longmapsto S^x \boldsymbol{a}(y)\boldsymbol{k}(-\theta/2). \tag{3.3}$$

Under this identification, the left- and right-invariant Haar measure on $G$ corresponds (up to a multiplicative constant) to the Liouville volume measure $\nu$ on $T_1\mathscr{H}$, given by

$$d\nu(z, \theta) := d\mu(z)\, d\theta = \frac{dx\, dy\, d\theta}{y^2}. \tag{3.4}$$

We let $g_t$ and $h_t$ denote the geodesic flow and the horocycle flow on $T_1\mathscr{H}$ (cf., e.g., [M, §3.1] for the intrinsic geometric definition of these flows). Under our identification $T_1\mathscr{H} \leftrightarrow G$, $g_t$ and $h_t$ are given by

$$g_t(T) = T\boldsymbol{a}(e^t), \qquad h_t(T) = TS^t \quad (\text{for } T \in G,\ t \in \mathbb{R}). \tag{3.5}$$

As before, we let $\Gamma$ be a cofinite Fuchsian group with a standard cusp at $\infty$, and we let $\boldsymbol{M} = \Gamma \setminus T_1\mathscr{H} = \Gamma \setminus G$. Clearly, a (closed) fundamental domain for the action of $\Gamma$ on $T_1\mathscr{H}$ is given by

$$\widetilde{\mathscr{F}} = \mathscr{F} \times (\mathbb{R}/2\pi\mathbb{Z}) \subset T_1\mathscr{H}.$$

Hence, in particular, $\nu(\boldsymbol{M}) = 2\pi\,\mu(\mathscr{F})$. The flows $g_t$ and $h_t$ induce well-defined flows on $\boldsymbol{M}$, which we again denote by $g_t$, $h_t$. It is well known that the volume measure $\nu$ on $\boldsymbol{M}$ is invariant under the horocycle flow $h_t$ and is ergodic. The same fact also holds for the geodesic flow $g_t$ (cf., e.g., [CFS, Chapter 4, §4]).

A point $p \in T_1\mathscr{H}$ belongs to a closed orbit of the horocycle flow on $\boldsymbol{M}$ if and only if it determines a closed horocycle encircling one of the $\kappa$ cusps, that is, if and only if

$$p \in \Gamma N_k^{-1}(z, 0) \quad \text{for some } k \in \{1, \ldots, \kappa\},\ z \in \mathscr{H}$$

(cf., e.g., [S, §1]).

Our goal in this section is to prove the following theorem. We write

$$\mathscr{Y}_\Gamma(p) := \mathscr{Y}_\Gamma(z) \quad \text{for } p = (z, \theta) \in T_1\mathscr{H}.$$

THEOREM 2

*Let $\delta > 0$, and let $f : T_1\mathscr{H} \to \mathbb{C}$ be any fixed, continuous, and $\Gamma$-invariant function satisfying the growth condition*

$$|f(p)| \leqq C\sqrt{\mathscr{Y}_\Gamma(p)}, \quad \forall p \in T_1\mathscr{H}, \tag{3.6}$$

*for some positive constant $C$. Then*

$$\frac{1}{\beta - \alpha}\int_\alpha^\beta f(x + iy, 0)\, dx \to \frac{1}{\nu(\boldsymbol{M})}\int_{\boldsymbol{M}} f(p)\, d\nu(p) \tag{3.7}$$

*uniformly as $y \to 0$ so long as $\beta - \alpha$ remains bigger than $y^{1/2-\delta}$.*

Clearly, this is a generalization of Theorem 1 stated in the introduction.

The fundamental result on which we build our proof of Theorem 2 is the fact that all the ergodic measures for the horocycle flow $h_t$ on $\boldsymbol{M}$ are explicitly known. If $\omega$ is a Borel probability measure on $\boldsymbol{M}$, invariant and ergodic under the flow $h_t$, then *either* $\omega$ is just the volume measure $\nu$ normalized by a factor $\nu(\boldsymbol{M})^{-1}$ *or* the support of $\omega$ is a closed orbit $\mathfrak{h}$ of the flow $h_t$ (and then $\omega$ has uniform mass along $\mathfrak{h}$). This is a special case of Dani's result in [D1], [D2]. (Dani's result was later vastly generalized by Ratner in [R1]; cf. also [R3].)

We let $\mathscr{S}$ be the union of all closed orbits on $\boldsymbol{M}$; that is, we let

$$\mathscr{S} = \bigcup_{k=1}^{\kappa} \pi\left(N_k^{-1}(\mathscr{H} \times \{0\})\right) \subset \boldsymbol{M}, \tag{3.8}$$

where $\pi$ is the projection map $T_1\mathscr{H} \to \Gamma \setminus T_1\mathscr{H} = \boldsymbol{M}$. We call $\mathscr{S}$ the *singular set*. By an application of ergodic decomposition, the characterization of invariant measures stated above implies the following proposition.

PROPOSITION 3.1
*Let $\omega$ be a Borel probability measure on $\boldsymbol{M}$ invariant under the horocycle flow. Assume that $\omega(\mathscr{S}) = 0$. Then $\omega$ is equal to $\nu(\boldsymbol{M})^{-1}\nu$, the unit normalized volume measure on $\boldsymbol{M}$.*

We use the following notation for subsets of the singular set $\mathscr{S}$:

$$\mathscr{S}_{A,B} = \bigcup_{k=1}^{\kappa} \pi\left(N_k^{-1}(\mathbb{R} \times [A, B] \times \{0\})\right) \subset \mathscr{S}.$$

Here $A, B$ are any numbers such that $0 < A < B$.

In the proof of the next lemma, we use the same idea as in [R3, page 21 (bottom)].

LEMMA 3.2
*Let $\delta > 0$ be given. Then, for any numbers $0 < A < B$ and $\varepsilon > 0$, there exist a number $y_0 > 0$ and a continuous, $\Gamma$-invariant function $f : T_1\mathscr{H} \to [0, 1]$ which has compact support on $\boldsymbol{M} = \Gamma \setminus T_1\mathscr{H}$, such that*

$$f(p) = 1 \quad \textit{for all} \ \ p \in \mathscr{S}_{A,B} \tag{3.9}$$

*and*

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta f(x + iy, 0)\, dx \leqq \varepsilon \tag{3.10}$$

*for all $0 < y < y_0$ and all $\alpha, \beta \in \mathbb{R}$ satisfying $\beta - \alpha \geqq y^{1/2-\delta}$.*

*Proof*
Let $\delta, A, B, \varepsilon$ be given as in the lemma. Let $f_0(z)$ be a function as in Proposition 2.1 ($f_0(z)$ depends only on $\Gamma$ and $\varepsilon$). Since $f_0(z)$ has compact support in $\Gamma \setminus \mathscr{H}$, we can take $T > 0$ so large that $1 < e^T A$, and $\mathscr{Y}_\Gamma(z) < 0.5\, e^T A$ for all points $z$ in the support of $f_0$. Let $F(y)$ be a continuous function satisfying $\chi_{[e^T A, e^T B]} \leqq F \leqq \chi_{[0.5\, e^T A, 2e^T(B+1)]}$ on $\mathbb{R}^+$ (where $\chi$ denotes the characteristic function). Recall the definition of the geodesic flow $g_t$, (3.5). We define $f(p)$ for $p \in T_1\mathscr{H}$ by

$$f(p) = F\left(\mathscr{Y}_\Gamma(g_T(p))\right).$$

It is now clear that $f(p)$ is a continuous, $\Gamma$-invariant function on $T_1\mathscr{H}$ and that $f(p)$ has compact support on $\boldsymbol{M} = \Gamma \setminus T_1\mathscr{H}$.

To prove (3.9), let $p \in \mathscr{S}_{A,B}$. Then $p = \pi(N_j^{-1}(x + iy, 0))$ for some $x \in \mathbb{R}$, $y \in [A, B]$, $j \in \{1, \ldots, \kappa\}$, and $g_T(p) = \pi(N_j^{-1}(x + ie^T y, 0))$. Write $z' = N_j^{-1}(x + ie^T y)$. For $k = j$ and $W \in [T_j]$, we then have $\operatorname{Im} N_k W(z') = e^T y$. For all *other*

$\langle k, W\rangle \in \{1, \dots, \kappa\} \times \Gamma$, we have, writing $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = N_k W N_j^{-1}$,

$$\operatorname{Im} N_k W(z') = \frac{e^T y}{(cx+d)^2 + (ce^T y)^2} < 1 < e^T y$$

since $e^T y \geqq e^T A > 1$, and $c \geqq 1$ by Lemma 2.3. This shows that $\mathscr{Y}_\Gamma(z') = e^T y \in [e^T A, e^T B]$, and hence $f(p) = 1$.

To prove (3.10), notice that for all $x + iy \in \mathscr{H}$,

$$f(x+iy, 0) = F\big(\mathscr{Y}_\Gamma(x + ie^T y, 0)\big) \leqq 1 - f_0(x + ie^T y).$$

Also, for each sufficiently small $y$, we have $y^{1/2-\delta} > (e^T y)^{1/2-\delta/2}$. Hence by the property of $f_0(z)$ from Proposition 2.1 (with $\delta/2$ in place on $\delta$), we have

$$\frac{1}{\beta-\alpha}\int_\alpha^\beta f(x+iy,0)\,dx \leqq 1 - \frac{1}{\beta-\alpha}\int_\alpha^\beta f_0(x+ie^T y)\,dx \leqq \varepsilon$$

for all sufficiently small $y$ and all $\alpha, \beta \in \mathbb{R}$ satisfying $\beta - \alpha \geqq y^{1/2-\delta}$. □

*Proof of Theorem 2*

Let $\delta > 0$ be fixed. Let $C_0(\boldsymbol{M})$ denote the Banach space of all real continuous functions on $\boldsymbol{M}$ vanishing at infinity, with the norm being the supremum norm, $\|f\| = \sup_{p\in \boldsymbol{M}} |f(p)|$. Let $C_0^*(\boldsymbol{M})$ denote the dual of $C_0(\boldsymbol{M})$.

For $y > 0$ and $\alpha < \beta$, define $\Lambda_{y,\alpha,\beta} \in C_0^*(\boldsymbol{M})$ by

$$\Lambda_{y,\alpha,\beta}(f) = \frac{1}{\beta-\alpha}\int_\alpha^\beta f(x+iy,0)\,dx, \quad f \in C_0(\boldsymbol{M}).$$

We first prove that Theorem 2 holds for all functions $f \in C_0(\boldsymbol{M})$. Let $\mathscr{T}$ denote the set of all limit points in the weak *-topology on $C_0^*(\boldsymbol{M})$ of the set $\{\Lambda_{y,\alpha,\beta} \mid y > 0,\ \beta - \alpha \geqq y^{1/2-\delta}\}$ when $y \to 0$. It suffices to prove that $\mathscr{T} = \{\Lambda_\nu\}$, where $\Lambda_\nu$ is given by $\Lambda_\nu f = \nu(\boldsymbol{M})^{-1}\int_{\boldsymbol{M}} f\,d\nu$.

Take $\Lambda \in \mathscr{T}$. Then there are sequences $\{y_n\}_{n=1}^\infty$, $\{\alpha_n\}_{n=1}^\infty$, $\{\beta_n\}_{n=1}^\infty$ such that $y_1 > y_2 > \cdots$, $\lim_{n\to\infty} y_n = 0$, $\beta_n - \alpha_n \geqq y_n^{1/2-\delta}$ for all $n$, and

$$\lim_{n\to\infty} \Lambda_{y_n,\alpha_n,\beta_n}(f) = \Lambda f$$

for all $f \in C_0(\boldsymbol{M})$.

Let $\omega$ be the unique Borel measure on $\boldsymbol{M}$ such that

$$\Lambda f = \int_{\boldsymbol{M}} f\,d\omega, \quad f \in C_0(\boldsymbol{M}).$$

It is clear that $\omega(\boldsymbol{M}) \leqq 1$. Also, by Proposition 2.1, for any $\varepsilon > 0$ there is a function $f \in C_0(\boldsymbol{M})$ such that $\|f\| \leqq 1$ and $\Lambda f \geqq 1 - \varepsilon$, and thus $\omega(\boldsymbol{M}) \geqq 1 - \varepsilon$. Hence we have $\omega(\boldsymbol{M}) = 1$.

Furthermore, $\omega$ is $h_t$-invariant since, for any fixed $f, t$, we have

$$\begin{aligned}|\Lambda(f \circ h_t) - \Lambda f| &= \lim_{n\to\infty} \left|\Lambda_{y_n,\alpha_n,\beta_n}(f \circ h_t) - \Lambda_{y_n,\alpha_n,\beta_n}(f)\right| \\ &\leqq \limsup_{n\to\infty} \frac{1}{\beta_n - \alpha_n}\Big[\int_{\alpha_n}^{\alpha_n+ty_n} + \int_{\beta_n}^{\beta_n+ty_n}\Big] |f(x+iy_n,0)|\ |dx| \\ &\leqq \limsup_{n\to\infty} \frac{2|t|y_n\ \|f\|}{\beta_n - \alpha_n} = 0.\end{aligned}$$

Finally, Lemma 3.2 implies that $\omega(\mathscr{S}_{A,B}) \leqq \varepsilon$ for all $A < B$, $\varepsilon > 0$; hence $\omega(\mathscr{S}) = 0$.

Now Proposition 3.1 forces $\omega = \nu(\boldsymbol{M})^{-1}\nu$, as was to be shown.

To complete the proof of Theorem 2, we now use an approximation argument to treat the general case of functions $f \in C(\boldsymbol{M})$ restricted only by the growth condition (3.6). Given such a function $f$, and given any number $\varepsilon > 0$, we can find $M \geqq 10$ such that

$$\frac{C}{\nu(\boldsymbol{M})}\int_{\widetilde{\mathscr{F}}} \sqrt{\lfloor \mathscr{Y}_\Gamma(p)\rfloor_M}\, d\nu(p) \leqq \frac{\varepsilon}{10} \tag{3.11}$$

and

$$\frac{25\kappa C}{\sqrt{M}} \leqq \frac{\varepsilon}{10}, \tag{3.12}$$

where $C$ is as in (3.6). (To see that (3.11) can be achieved, one applies the decomposition of $\mathscr{F}$ into the bounded region $\mathscr{F}_B$ and the cuspidal regions $\mathscr{C}_{kB}$ (cf. (2.3)); one then uses the fact that $\mathscr{Y}_\Gamma(z)$ is bounded on $\mathscr{F}_B$ and that, for $B$ sufficiently large, $\mathscr{Y}_\Gamma(z) = \operatorname{Im} N_k(z)$ on $\mathscr{C}_{kB}$.) We let $F(y)$ be a continuous function on $\mathbb{R}^+$ satisfying $\chi_{(0,M]} \leqq F \leqq \chi_{(0,M+1]}$, and we define

$$f_1(p) = F\big(\mathscr{Y}_\Gamma(p)\big)\cdot f(p).$$

Clearly, $f_1 \in C_0(\boldsymbol{M})$, and by (3.6), we have

$$\big|f(p) - f_1(p)\big| \leqq C\sqrt{\lfloor \mathscr{Y}_\Gamma(p)\rfloor_M}, \quad \forall p \in T_1\mathscr{H}. \tag{3.13}$$

By what we have already proved, $f_1$ satisfies the conclusion (3.7) in Theorem 2. Using this fact together with (3.11), (3.12), (3.13), and (2.7) in Proposition 2.2, we now obtain

$$\left|\frac{1}{\beta-\alpha}\int_\alpha^\beta f(x+iy,0)\,dx - \frac{1}{\nu(\boldsymbol{M})}\int_{\boldsymbol{M}} f(p)\,d\nu(p)\right| \leqq \varepsilon$$

for all sufficiently small $y$ and all $\alpha, \beta \in \mathbb{R}$ such that $\beta - \alpha \geqq y^{1/2-\delta}$. Since $\varepsilon$ was arbitrary, this concludes the proof of Theorem 2. $\square$

## 4. Spectral theory and the rate of convergence in (1.2)

Let $D = y^2\big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\big)$ denote the non-Euclidean Laplacian, and let $\phi_0, \phi_1, \ldots$ be the discrete eigenfunctions of $-D$ on $\Gamma \setminus \mathscr{H}$, taken to be orthonormal and to have increasing eigenvalues $0 = \lambda_0 < \lambda_1 \leqq \lambda_2 \leqq \cdots$. In general, we do not know if the set $\{\phi_0, \phi_1, \ldots\}$ is infinite or not. We let $E_k(z, s)$ be the Eisenstein series associated to the cusp $\eta_k$ (cf. [He1, pages 280 (Definition 3.5), 130 (Definition 11.7, Theorem 11.8), 296–297]). We recall that whenever $\phi_m$ ($m \geqq 1$) is *not* a cusp form, we have $0 < \lambda_m < 1/4$, at least one of the $E_k(z, s)$ has a pole at $s = s_m = 1/2+\sqrt{1/4 - \lambda_m} \in (1/2, 1)$, and $\phi_m$ is a linear combination of the residues of $E_k(z, s)$ ($k = 1, \ldots, \kappa$) at $s = s_m$ (cf. [He1, pages 284–288]). We call such a $\phi_m$ a residual eigenfunction.

For $f \in L_2(\Gamma \setminus \mathscr{H})$, we let $\|f\|$ denote the $L_2$-norm, $\|f\| = \sqrt{\int_{\mathscr{F}} |f(z)|^2 \, d\mu(z)}$. Our goal in this section and in Section 5 is to prove the following theorem.

THEOREM 3

*Let $\varepsilon > 0$. We then have, for all $f \in C^2(\mathscr{H}) \cap L_2(\Gamma \setminus \mathscr{H})$ such that $Df \in L_2(\Gamma \setminus \mathscr{H})$ and all $y \in (0, 1)$, $\alpha, \beta$ such that $\sqrt{y} \leqq \beta - \alpha \leqq 1$:*

$$\begin{aligned} \frac{1}{\beta - \alpha} \int_\alpha^\beta f(x + iy)\, dx = {} & \frac{1}{\mu(\mathscr{F})} \int_{\mathscr{F}} f(z)\, d\mu(z) \\ & + O\big((\|f\| + \|Df\|)\, y^{1/2-\varepsilon} (\beta - \alpha)^{-1}\big) \\ & + O\big(\|f\|\, y^{1-s_1-\varepsilon} (\beta - \alpha)^{s_1 - 1}\big) \\ & + O\big(\|f\|\, y^{1-s_1'} (\beta - \alpha)^{2(s_1'-1)}\big). \end{aligned} \tag{4.1}$$

*(In each big $O$, the implied constant depends solely on $\Gamma$ and $\varepsilon$.) Here $s_1 \in (1/2, 1)$ is the largest number such that there is a cusp form on $\Gamma \setminus \mathscr{H}$ of eigenvalue $\lambda = s_1(1 - s_1)$, and $s_1' \in (1/2, 1)$ is the largest number such that there is a residual eigenfunction on $\Gamma \setminus \mathscr{H}$ of eigenvalue $\lambda = s_1'(1 - s_1')$. If there are no such cusp forms or residual eigenfunctions, it is understood that the corresponding error term in (4.1) is omitted.*

By a standard approximation argument, one can show that Theorem 3 implies Theorem 1, stated in the introduction.

The first step in the proof of Theorem 3 is to apply spectral decomposition to the given function $f(z)$. According to [He1, pages 317 (Proposition 5.3), 733 (note 5)], any function $f(z)$ as in Theorem 3 has a spectral expansion

$$f(z) = \sum_{m \geqq 0} d_m \phi_m(z) + \sum_{k=1}^{\kappa} \int_0^\infty g_k(R) E_k\Big(z, \frac{1}{2} + iR\Big)\, dR \tag{4.2}$$

with uniform and absolute convergence over $z \in \mathscr{H}$-compacta. Here $d_m = \langle f, \phi_m \rangle$, and "$g_k(R) = (2\pi)^{-1} \int_{\mathscr{F}} f(z)\overline{E_k(z, 1/2 + iR)}\, d\mu(z)$" (this has to be properly considered as a limit in the $L_2(0, \infty)$-norm); compare [He1, page 242 (Proposition 2.3), 317 (line 4)].

The proof in [He1] of the uniform and absolute convergence in (4.2) starts by considering the spectral expansion (in the $L_2$-sense) of the function $Df + a(1-a)f \in L_2(\Gamma \setminus \mathscr{H})$ for some fixed number $a > 1$; this is then integrated against the Green function $G_a(z, w)$. It is seen in this proof that

$$\sum_{m \geqq 0} |d_m|^2 \big(a(1-a) - \lambda_m\big)^2 + 2\pi \sum_{k=1}^{\kappa} \int_0^\infty |g_k(R)|^2 \Big(a(1-a) - \frac{1}{4} - R^2\Big)^2 \, dR = \int_{\mathscr{F}} |Df(z) + a(1-a)f(z)|^2 < \infty \quad (4.3)$$

(cf. [He1, pages 91 (9.36), 244–245, 291 (3.23)]).

Substituting (4.2) in $(1/(\beta - \alpha)) \int_\alpha^\beta f(x + iy)\, dx$ and changing order between summation and integration, we see that the contribution from the constant eigenfunction $\phi_0 \equiv \mu(\Gamma \setminus \mathscr{H})^{-1/2}$ gives exactly the main term in (4.1) since $d_0 = \langle f, \phi_0 \rangle$.

To treat the other contributions from (4.2), we use the Fourier expansions of the eigenfunctions $\phi_m(z)$ $(m \geqq 1)$ and $E_k(z, 1/2 + iR)$ at the cusp $\eta_1 = \infty$. Recall that for $E_k(z, 1/2 + iR)$, this expansion is

$$E_k\Big(x + iy, \frac{1}{2} + iR\Big) = \delta_{k1} y^{1/2+iR} + \varphi_{k1}\Big(\frac{1}{2} + iR\Big) y^{1/2-iR} + \sum_{n \neq 0} c_n \sqrt{y} K_{iR}(2\pi |n| y)\, e(nx), \quad (4.4)$$

where $\varphi_{k1}(s)$ is an element in the scattering matrix $\Phi(s) = (\varphi_{ij}(s))$ (cf. [He1, Chapter 8]). Of course, the coefficients $c_n$ depend on $R$.

Integrating (4.4) over the horocycle segment $[\alpha, \beta] + iy$, we get

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta E_k\Big(x + iy, \frac{1}{2} + iR\Big)\, dx = \delta_{k1} y^{1/2+iR} + \varphi_{k1}\Big(\frac{1}{2} + iR\Big) y^{1/2-iR} + \frac{1}{\beta - \alpha} \sum_{n \neq 0} c_n \sqrt{y} K_{iR}(2\pi |n| y)\, \frac{e(n\beta) - e(n\alpha)}{2\pi i n}. \quad (4.5)$$

In order to obtain an upper bound on this expression, we prove a Rankin-Selberg-type bound on the Fourier coefficients $c_n$.

We need our bounds to be uniform over $R \geqq 0$. A valuable tool in this regard is the spectral majorant function $\omega(R)$, which is defined in [He1, pages 161, 299 (line

14)]. This function depends only on $\Gamma$, and it satisfies $\omega(-R) = \omega(R) \geqq 1$ and

$$\operatorname{Tr}\Big[\Phi'\Big(\frac{1}{2} + iR\Big)\Phi\Big(\frac{1}{2} + iR\Big)^{-1}\Big] = O\big(\omega(R)\big) \tag{4.6}$$

for all $R \in \mathbb{R}$. One also has

$$\int_0^T \omega(R)\,dR = O(T^2) \quad \text{as } T \to \infty \tag{4.7}$$

(cf. [He1, page 315 (i) – (iii)]). The implied constants in (4.6) and (4.7) depend on $\Gamma$.

PROPOSITION 4.1
*In the Fourier expansion (4.4) we have, uniformly over $N \geqq 1$ and $R \geqq 0$,*

$$\sum_{1 \leqq |n| \leqq N} |c_n|^2 = O\big(e^{\pi R}(N + R)\big)\Big\{\omega(R) + \log\Big(\frac{2N}{R+1} + R\Big)\Big\}.$$

*(The implied constant depends only on $\Gamma$.)*

*Proof*
The method of proof is similar to [Wo1, Proposition 5.1], where the case of $\Gamma = \mathrm{PSL}(2, \mathbb{Z})$ was treated.

We keep $0 < Y < H$ and try to find a bound from above for the integral

$$J = \int_{\mathscr{D}} \Big|E_k\Big(z, \frac{1}{2} + iR\Big)\Big|^2 d\mu(z), \quad \text{where } \mathscr{D} = (0, 1) \times (Y, H). \tag{4.8}$$

We tessellate $\mathscr{D}$ by translates of the fundamental region; that is, we write $\mathscr{D} = \bigcup_{T \in \Gamma}(\mathscr{D} \cap T(\mathscr{F}))$, an essentially disjoint union. Using the automorphy of $E_k(z, 1/2 + iR)$, we then get

$$J = \sum_{T \in \Gamma} \int_{\mathscr{F}} I[T(z) \in \mathscr{D}] \cdot \Big|E_k\Big(z, \frac{1}{2} + iR\Big)\Big|^2 d\mu(z), \tag{4.9}$$

where $I[\,\cdot\,]$ is the indicator function.

Recall our definition of $B_0$ just below (2.1), and recall relations (2.2) and (2.3) defining $\mathscr{C}_{jB}$ and $\mathscr{F}_B$. We take

$$B = \max\Big(B_0, H, \frac{1}{Y}\Big). \tag{4.10}$$

We then claim that in (4.9), the integrand is zero for all $z \in \mathscr{F} - \mathscr{F}_B$ and all $T \in \Gamma$. Indeed, given $z \in \mathscr{F} - \mathscr{F}_B$, there is some $j \in \{1, \dots, \kappa\}$ such that $z \in \mathscr{C}_{jB}$, that is, $\operatorname{Im} N_j(z) \geqq B$. Recall $N_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. For each $T \in \Gamma$, writing $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = N_1 T N_j^{-1}$ and using Lemma 2.3, we now have *either* $j = 1$, $T \in [S]$, and $\operatorname{Im} T(z) = \operatorname{Im} z \geqq B \geqq H$, *or else* $|c| \geqq 1$, and then

$$\begin{aligned}\operatorname{Im} T(z) = \operatorname{Im} N_1 T(z) = \operatorname{Im} \frac{aN_j(z)+b}{cN_j(z)+d} &= \frac{\operatorname{Im} N_j(z)}{|cN_j(z)+d|^2} \\ &\leqq \big(\operatorname{Im} N_j(z)\big)^{-1} \leqq B^{-1} \leqq Y.\end{aligned}$$

In both cases, we get $T(z) \notin \mathscr{D}$, that is, $I[T(z) \in \mathscr{D}] = 0$. This proves our claim, and as a result we see that in (4.9), we may replace $\mathscr{F}$ by $\mathscr{F}_B$. Changing order between summation and integration, we get

$$J = \int_{\mathscr{F}_B} \#\big\{T \in \Gamma \bigm| T(z) \in \mathscr{D}\big\} \cdot \Big| E_k\Big(z, \frac{1}{2} + iR\Big)\Big|^2 \, d\mu(z).$$

Next, we have, by [I, Lemma 2.10],

$$\begin{aligned}\#\big\{T \in \Gamma \bigm| T(z) \in \mathscr{D}\big\} &\leqq \#\big\{W_0 \in [S] \setminus \Gamma \bigm| \operatorname{Im} W_0(z) > Y\big\} \\ &\leqq 1 + O(Y^{-1}),\end{aligned}$$

where the implied constant depends only on $\Gamma$; that is, the bound is uniform over all $z \in \mathscr{H}$ and $Y > 0$. This implies

$$J \leqq O(1 + Y^{-1}) \int_{\mathscr{F}_B} \Big|E_k\Big(z, \frac{1}{2} + iR\Big)\Big|^2 \, d\mu(z). \tag{4.11}$$

We define $E_k^B(z,s)$ for $z \in \mathscr{F}$ by

$$E_k^B(z,s) = \begin{cases} E_k(z,s) & \text{if } z \in \mathscr{F}_B, \\ E_k(z,s) - \delta_{jk} \cdot (\operatorname{Im} N_j z)^s - \varphi_{kj}(s) \cdot (\operatorname{Im} N_j z)^{1-s} & \text{if } z \in \mathscr{C}_{jB}. \end{cases}$$

Using the appropriate Maass-Selberg identity, we have for all $R > 0$ (cf., e.g., [He1, pages 301 (3.43), 311 – 312, 315 (i)]),

$$\begin{aligned}&\sum_{k'=1}^{\kappa} \int_{\mathscr{F}_B} \Big|E_{k'}\Big(z, \frac{1}{2} + iR\Big)\Big|^2 d\mu(z) \leqq \sum_{k'=1}^{\kappa} \int_{\mathscr{F}} \Big|E_{k'}^B\Big(z, \frac{1}{2} + iR\Big)\Big|^2 d\mu(z) \\ &= 2\kappa \log B - \operatorname{Tr}\Big[\Phi'\Big(\frac{1}{2} + iR\Big)\Phi\Big(\frac{1}{2} + iR\Big)^{-1}\Big] + \sum_{k'=1}^{\kappa} \operatorname{Re}\bigg(\overline{\varphi_{k'k'}\Big(\frac{1}{2} + iR\Big)\frac{B^{2iR}}{iR}}\bigg).\end{aligned}$$

Recall that the scattering matrix $\Phi(s)$ is unitary for $s = 1/2 + iR$. Because of (4.6), we now obtain, for all $R \geqq 1$,

$$J \leqq O\big(1 + Y^{-1}\big)\big\{\log(B) + \omega(R)\big\}. \tag{4.12}$$

The implied constant depends only on $\Gamma$. The inequality (4.12) also holds for $0 \leqq R \leqq 1$, as follows directly from (4.11) by using [He1, page 301 (d)] and decomposing $\mathscr{F}_B$ as a union of $\mathscr{F}_{B_0}$ and $(\mathscr{C}_{jB_0} - \mathscr{C}_{jB})$ for $j = 1, \ldots, \kappa$.

On the other hand, substituting (4.4) directly in the definition of $J$, (4.8), and then applying Parseval's formula, we get

$$J \geqq \sum_{n \neq 0} |c_n|^2 \int_{2\pi|n|Y}^{2\pi|n|H} K_{iR}(y)^2 \, \frac{dy}{y}.$$

We now take $Y = (R+1)/(8\pi N)$, $H = (R+1)/(4\pi)$. With this choice we have

$$\Big[\frac{1}{4}(R+1), \frac{1}{2}(R+1)\Big] \subset \big[2\pi|n|Y, 2\pi|n|H\big] \quad \text{whenever } 1 \leqq |n| \leqq N,$$

and hence

$$\sum_{1 \leqq |n| \leqq N} |c_n|^2 \leqq C^{-1} J, \quad \text{where } C = \int_{(R+1)/4}^{(R+1)/2} K_{iR}(y)^2 \, \frac{dy}{y}. \tag{4.13}$$

However, from the asymptotic formula for $K_{iR}(y)$ for $R$ large and $y < (1-\varepsilon)R$ (cf. [B1], [B2], and also [EMOT, page 88 (19)]), it follows that $C^{-1} \leqq O((R+1)e^{\pi R})$ holds for all sufficiently large $R$. By a simple continuity argument, this bound then actually holds for all $R \geqq 0$. Using this fact and (4.10), (4.12), and (4.13) (remembering $\omega(R) \geqq 1$), we obtain the desired result. □

PROPOSITION 4.2
*Let $\varepsilon > 0$ and $k \in \{1, \ldots, \kappa\}$, and keep $0 < y < 1$, $R \geqq 0$, $\alpha < \beta$. We then have*

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta E_k\Big(x + iy, \frac{1}{2} + iR\Big)\, dx = O\big(y^{1/2-\varepsilon}\big)\Big\{1 + \frac{(R+1)^{1/6+\varepsilon}\sqrt{\omega(R)}}{\beta - \alpha}\Big\}. \tag{4.14}$$

*(The implied constant depends only on $\Gamma$ and $\varepsilon$.)*

*Proof*
This result follows from relation (4.5) and Proposition 4.1 by use of partial summation. The details are as follows.

We keep $0 < \varepsilon < 1$ and write $\varepsilon' = \varepsilon/2$. Notice that $|\varphi_{k1}(1/2 + iR)| \leqq 1$ since $\Phi(s)$ is unitary at $s = 1/2 + iR$. A convenient bound on the $K$-Bessel function is given by

$$K_{iR}(y_0) = O\big(e^{-(\pi/2)R}(R+1)^{-1/3+\varepsilon'} y_0^{-\varepsilon'} \cdot \min(1, e^{(\pi/2)R - y_0})\big). \tag{4.15}$$

This holds uniformly for all $R \geqq 0$ and $y_0 > 0$. To prove (4.15), we first notice that by the integral representation of $K_{iR}(y_0)$ (see [W, p. 181 (5)]), we have for all $R \geqq 0$ and $y_0 > 0$,

$$|K_{iR}(y_0)| \leqq \int_0^\infty e^{-y_0(1+t^2/2)}\,dt = \sqrt{\frac{\pi}{2y_0}}e^{-y_0}. \tag{4.16}$$

Now keep $R$ restricted to an arbitrary *compact* interval $[0, R_0]$. Then for $1 \leqq y_0$, (4.15) follows directly from (4.16). For $0 < y_0 \leqq 1$, we use [W, pages 77 (2), 78 (6)] to show $|K_\upsilon(y_0)| \leqq C y_0^{-\varepsilon'}$ for all $\upsilon$ on the boundary of the rectangle $[-\varepsilon', \varepsilon'] \times [-R_0, R_0]$ in the complex plane (where $C > 0$ depends only on $R_0$, $\varepsilon'$), and (4.15) follows by applying the modulus principle in the $\upsilon$-variable. In the remaining case, that is, $R$ large, (4.15) follows from the asymptotic formula for $K_{iR}(y_0)$ given in [B1], [B2]. (In the case of $R \geqq 1$ and $y_0 \geqq (\pi/2)R$, (4.15) also follows more directly from (4.16).)

We now get the following upper bound on the expression in (4.5):

$$O\big(\sqrt{y}\big) + O\Big(e^{-(\pi/2)R}(R+1)^{-1/3+\varepsilon'}\frac{y^{1/2-\varepsilon'}}{\beta-\alpha}\Big)\sum_{n=1}^\infty(|c_n|+|c_{-n}|)\cdot f(n), \tag{4.17}$$

where $f(X) := X^{-1-\varepsilon'}\cdot\min\big(1, e^{(\pi/2)R-2\pi yX}\big)$.

We define

$$S(X) := \sum_{1\leqq|n|\leqq X}|c_n|.$$

Then, by a weak form of Proposition 4.1 and Cauchy's inequality,

$$S(X) = O\big(e^{\pi R/2}(R+1)^{\varepsilon'}\sqrt{\omega(R)}X^{1/2+\varepsilon'}\sqrt{X+R}\big), \quad \forall X \geqq \frac{1}{2}.$$

Notice that for given $R \geqq 0$, $0 < y < 1$, the function $f(X)$ is continuous and piecewise smooth, and $f(X)S(X) \to 0$ as $X \to \infty$. We now get

$$\sum_{n=1}^\infty(|c_n|+|c_{-n}|)\cdot f(n) = \int_{1/2}^\infty f(X)\,dS(X) = -\int_{1/2}^\infty f'(X)S(X)\,dX. \tag{4.18}$$

But we have $f'(X) = O(X^{-2-\varepsilon'})$ for $X < R/4y$, while $f'(X) = O((1+yX)X^{-2-\varepsilon'}e^{(\pi/2)R-2\pi yX})$ and $\sqrt{X+R} = O(\sqrt{X})$ for $X > R/4y$. Using these facts, we see that (4.18) is bounded by

$$O\big(e^{(\pi/2)R}(R+1)^{\varepsilon'}\sqrt{\omega(R)}\big)\Big[\int_{1/2}^{\max(1/2,R/4y)}\big(X^{-1}+\sqrt{R}X^{-3/2}\big)\,dX \\ + \int_{\max(1/2,R/4y)}^\infty(1+yX)e^{(\pi/2)R-2\pi yX}\,\frac{dX}{X}\Big]. \tag{4.19}$$

Here the second integral is equal to $\int_{\max(y/2,R/4)}^{\infty}(1+u)e^{(\pi/2)R-2\pi u}\,\frac{du}{u}$, which is clearly bounded by $O(\log(2/y))$ if $0 \leqq R \leqq 1$ and by $O(1)$ if $R \geqq 1$. Hence (4.19) is

$$\begin{aligned}&= O\Big(e^{(\pi/2)R}(R+1)^{\varepsilon'}\sqrt{\omega(R)}\Big)\Big[\log^{+}\Big(\frac{R}{2y}\Big)+\sqrt{R}+\log\Big(\frac{2}{y}\Big)\Big]\\ &= O\Big(e^{(\pi/2)R}(R+1)^{1/2+\varepsilon'}\sqrt{\omega(R)}y^{-\varepsilon'}\Big).\end{aligned}$$

Using this bound in (4.17), we obtain the desired result. □

PROPOSITION 4.3

*Let $\varepsilon > 0$ and $k \in \{1, \ldots, \kappa\}$. Take $f \in C^2(\mathscr{H}) \cap L_2(\Gamma \setminus \mathscr{H})$ such that $Df \in L_2(\Gamma \setminus \mathscr{H})$, and let the spectral expansion of $f$ be as in (4.2). We then have, for all $0 < y < 1$ and all $\alpha, \beta$ such that $0 < \beta - \alpha \leqq 1$,*

$$\frac{1}{\beta-\alpha}\int_{\alpha}^{\beta}\Big\{\int_0^{\infty} g_k(R)E_k\Big(x+iy,\frac{1}{2}+iR\Big)\,dR\Big\}\,dx \\ = O\big((\|f\|+\|Df\|)\,y^{1/2-\varepsilon}(\beta-\alpha)^{-1}\big). \quad (4.20)$$

*(The implied constant depends only on $\Gamma$ and $\varepsilon$.)*

*Proof*

We keep $\varepsilon < 1/10$. Changing order of integration and applying Cauchy's inequality, we find that the absolute value of the left-hand side in (4.20) is less than or equal to

$$\sqrt{\int_0^{\infty}(R+1)^4|g_k(R)|^2\,dR} \\ \times\sqrt{\int_0^{\infty}(R+1)^{-4}\Big|\frac{1}{\beta-\alpha}\int_{\alpha}^{\beta}E_k\Big(x+iy,\frac{1}{2}+iR\Big)\,dx\Big|^2\,dR}.$$

The first factor is bounded by $O(\|f\|+\|Df\|)$ because of (4.3). Also, by Proposition 4.2 (and $\omega(R) \geqq 1$, $\beta - \alpha \leqq 1$), the second factor is bounded by

$$O\big(y^{1/2-\varepsilon}(\beta-\alpha)^{-1}\big)\sqrt{\int_0^{\infty}(R+1)^{-11/3+2\varepsilon}\omega(R)\,dR}.$$

Here the integral is convergent, as follows from (4.7) using integration by parts. This concludes the proof. □

The above treatment of the Eisenstein series can easily be carried over to the discrete eigenfunctions $\phi_m(z)$, except those that have small eigenvalues $\lambda_m < 1/4$.

PROPOSITION 4.4
*Let $\varepsilon > 0$, take $m \geqq 1$ such that $\lambda_m \geqq 1/4$, and define $R \geqq 0$ through $\lambda_m = 1/4 + R^2$. We then have, for all $0 < y < 1$ and all $\alpha < \beta$,*

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta \phi_m(x + iy)\, dx = O\Big((R+1)^{1/6+\varepsilon} \frac{y^{1/2-\varepsilon}}{\beta - \alpha}\Big). \tag{4.21}$$

*(The implied constant depends only on $\Gamma$ and $\varepsilon$.)*

*Proof*
This is very similar to the proof of Proposition 4.2. One uses the Fourier expansion of $\phi_m(z)$ at the cusp $\eta_1 = \infty$,

$$\phi_m(x + iy) = \sum_{n \neq 0} c_n \sqrt{y} K_{iR}\,(2\pi\,|n|y)\, e(nx) \tag{4.22}$$

($\phi_m$ is certainly a cusp form since $\lambda_m \geqq 1/4$). Since we are assuming $\|\phi_m\| = 1$, we have the following bound on the coefficients $c_n$:

$$\sum_{|n| \leqq N} |c_n|^2 = O\big(e^{\pi R}(N + R)\big), \quad \forall N \geqq 1$$

(cf. [I, Theorem 3.2, the first bound]).* □

PROPOSITION 4.5
*Let $\varepsilon > 0$. Take $f \in C^2(\mathscr{H}) \cap L_2(\Gamma \setminus \mathscr{H})$ such that $Df \in L_2(\Gamma \setminus \mathscr{H})$, and let the spectral expansion of $f$ be as in (4.2). We then have, for all $0 < y < 1$ and all $\alpha, \beta$ such that $0 < \beta - \alpha \leqq 1$,*

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta \Big\{ \sum_{\lambda_m \geqq 1/4} d_m \phi_m(x + iy) \Big\}\, dx = O\big((\|f\| + \|Df\|)\, y^{1/2-\varepsilon} (\beta - \alpha)^{-1}\big).$$

*(The implied constant depends only on $\Gamma$ and $\varepsilon$.)*

*Proof*
Mimic the proof of Proposition 4.3. One uses the fact that $\sum_{\lambda_m \geqq 1/4} (R_m + 1)^{-11/3+2\varepsilon} < \infty$ for $R_m = \sqrt{\lambda_m - 1/4}$ and $\varepsilon$ small; this fact can be deduced from [He1, page 315 (ii)]. □

*Notice that some minor revisions are called for in the proof given in [I] for, in fact, one has $\int_R^\infty K_{iR}(y)^2 \frac{dy}{y} = O(R^{-4/3} e^{-\pi R})$ as $R \to \infty$; that is, [I, p. 61 (line 7)] is false (cf. the proof of Proposition 4.1 and the choice of $Y$ therein).

Recall that $0 = \lambda_0 < \lambda_1 \leqq \lambda_2 \leqq \cdots$. Notice that if we have $\lambda_1 \geqq 1/4$, then the statement in Theorem 3 follows from (4.2), Proposition 4.3, and Proposition 4.5 (and our remark just below (4.3)). However, when $\lambda_m < 1/4$, the best bound on $(1/(\beta-\alpha))\int_\alpha^\beta \phi_m(x+iy)\,dx$ that can be obtained by use of the Rankin-Selberg bound on $\sum_{|n|\leqq N}|c_n|^2$ as in the above proofs is*

$$\frac{1}{\beta-\alpha}\int_\alpha^\beta \phi_m(x+iy)\,dx = O\big(y^{1-s}(\beta-\alpha)^{s-3/2}\big),$$
$$\text{where } s = \frac{1}{2}+\sqrt{\frac{1}{4}-\lambda_m} \in \Big(\frac{1}{2},1\Big). \tag{4.23}$$

This holds uniformly over all $y \in (0,1)$ and all $\alpha, \beta$ such that $y \leqq \beta-\alpha \leqq 1$. Clearly, the bound (4.23) is not sufficient for our purposes. The largest constant $c$ for which $y^{1-s}(\beta-\alpha)^{s-3/2} \to 0$ holds anytime $\beta-\alpha \geqq y^{c-\varepsilon}$ and $y \to 0$ is

$$c = 1-\frac{1}{3-2s} < \frac{1}{2}, \tag{4.24}$$

whereas our goal is to reach $c = 1/2$. (One may notice that when $s \geqq 3/4$, the constant $c$ in (4.24) is the same as "$c(\Gamma)$" in [He3, Theorem A].)

If $\phi_m$ is a *cusp form*, we improve (4.23) by using the following bound from Hafner [H, Theorem 3]:†

$$S_\alpha(X) = \sum_{1\leqq n\leqq X} c_n e(n\alpha) = O(X^{1/2+\varepsilon}), \quad \forall X \geqq 0. \tag{4.25}$$

This holds for any cusp form $\phi_m$ having Fourier expansion as in (4.22). The same bound also holds for the sum $\sum_{-X\leqq n\leqq -1} c_n e(n\alpha)$. The implied constant in (4.25) depends only on $\Gamma$, $\varepsilon$, and $\phi_m$; in particular, the bound is uniform over all $\alpha \in \mathbb{R}$.

PROPOSITION 4.6
*Let $\varepsilon > 0$, and take $m \geqq 1$ such that $\phi_m$ is a cusp form. Define $s$ through $\lambda_m = s(1-s)$, $s \in [1/2,1) \cup [1/2, 1/2+i\infty)$. We then have, uniformly over all $y \in (0,1)$ and all $\alpha, \beta$ such that $y \leqq \beta-\alpha \leqq 1$,*

$$\frac{1}{\beta-\alpha}\int_\alpha^\beta \phi_m(x+iy)\,dx = O\big(y^{1-\operatorname{Re}s-\varepsilon}(\beta-\alpha)^{\operatorname{Re}s-1}\big). \tag{4.26}$$

*The implied constant depends on $\Gamma$, $\varepsilon$, and $\phi_m(z)$.*

*In the proof of (4.23), one uses $(e(n\beta)-e(n\alpha))/n = O(\min(\beta-\alpha, n^{-1}))$ in place of the coarser $O(n^{-1})$ that we used in the proofs of Propositions 4.2 and 4.4 below.
†We state here a slightly modified form of the result in [H]. The proof is completely analogous (cf. the proof of Proposition 5.1 below).

*Proof*
Since $\phi_m(z)$ is a cusp form, it has a Fourier expansion as in (4.22), wherein $iR = s - 1/2$. The analog of (4.5) reads

$$\frac{1}{\beta-\alpha}\int_\alpha^\beta \phi_m(x+iy)\,dx = \frac{1}{\beta-\alpha}\sum_{n\neq 0} c_n\sqrt{y}K_{s-1/2}(2\pi|n|y)\frac{e(n\beta)-e(n\alpha)}{2\pi i n}. \tag{4.27}$$

We define $\delta = \beta - \alpha$, and

$$f(X) := K_{s-1/2}\big(2\pi yX\big)\frac{e(X\delta)-1}{X}, \qquad g(X) = K_{s-1/2}(2\pi yX)\frac{1}{X}.$$

Also, recall the definition of $S_\alpha(X)$ in (4.25). We can now express the part corresponding to $n > 0$ in the right-hand side of (4.27) as

$$\begin{aligned}
&\frac{\sqrt{y}}{2\pi i\delta}\Big\{\int_{1/2}^{\delta^{-1}} f(X)\,dS_\alpha(X) + \int_{\delta^{-1}}^\infty g(X)\,dS_\beta(X) - \int_{\delta^{-1}}^\infty g(X)\,dS_\alpha(X)\Big\}\\
&\quad= \frac{\sqrt{y}}{2\pi i\delta}\Big\{f(\delta^{-1})S_\alpha(\delta^{-1}) - g(\delta^{-1})S_\beta(\delta^{-1}) + g(\delta^{-1})S_\alpha(\delta^{-1})\\
&\qquad\qquad - \int_{1/2}^{\delta^{-1}} f'(X)S_\alpha(X)\,dX - \int_{\delta^{-1}}^\infty g'(X)S_\beta(X)\,dX\\
&\qquad\qquad + \int_{\delta^{-1}}^\infty g'(X)S_\alpha(X)\,dX\Big\}.
\end{aligned} \tag{4.28}$$

(Convergence follows easily from (4.25) and the exponential decay of $g(X)$ and $g'(X)$ as $X \to \infty$.) Let us write $\sigma = \operatorname{Re} s$. One knows that

$$\begin{aligned}
&K_{s-1/2}(u) = O(u^{1/2-\sigma-\varepsilon}), \qquad K'_{s-1/2}(u) = O(u^{-1/2-\sigma}) \quad \text{for } 0 \leqq u \leqq 2\pi,\\
&K_{s-1/2}(u) = O(u^{-1/2}e^{-u}), \qquad K'_{s-1/2}(u) = O(u^{-1/2}e^{-u}) \quad \text{for } 2\pi \leqq u
\end{aligned} \tag{4.29}$$

(cf. [W, pages 77 (2), 78 (6), 79 (2), 80 (14), 202 (1)]; the implied constants depend on $s$ and $\varepsilon$). Recall that $y \leqq \delta \leqq 1$. It is now easy to verify that

$$\begin{aligned}
f(\delta^{-1}),\ g(\delta^{-1}) &= O(\delta^{1/2+\sigma+\varepsilon}y^{1/2-\sigma-\varepsilon}),\\
f'(X) &= O(\delta y^{1/2-\sigma-\varepsilon}X^{-1/2-\sigma-\varepsilon}) \quad \text{for } \frac{1}{2} \leqq X \leqq \delta^{-1},\\
g'(X) &= O(y^{1/2-\sigma-\varepsilon}X^{-3/2-\sigma-\varepsilon}) \quad \text{for } \delta^{-1} \leqq X \leqq y^{-1},\\
g'(X) &= O(\sqrt{y}X^{-3/2}e^{-2\pi yX}) \quad \text{for } y^{-1} \leqq X.
\end{aligned}$$

It follows by a short computation, using the above bounds and (4.25), that the whole expression in (4.28) is bounded by $O\big(y^{1-\sigma-\varepsilon}\delta^{\sigma-1}\big)$. The part corresponding to $n < 0$ in (4.27) can be treated in an entirely similar way. This concludes the proof. □

*Remark 4.7*
For $\phi_m$ fixed, the error term in Proposition 4.6 is much stronger than what we obtained in Proposition 4.4 (for comments related to this fact, see §7(I)).

Notice that in view of Proposition 4.6, in the case when there exist no residual eigenfunctions on $\Gamma \backslash \mathscr{H}$, the proof of Theorem 3 is now complete since $|d_m| = |\langle f, \phi_m \rangle| \leqq \| f \|$ for each $m$, and there are at most finitely many $m$ such that $0 < \lambda_m < 1/4$.

## 5. Bounding $\sum_{n=1}^{N} c_n e(n\alpha)$ for residual eigenfunctions

If $\phi_m$ is a residual eigenfunction, the bound (4.25) is not true for all $\alpha$ (cf. Remark 5.2). The best possible uniform bound is given in the following proposition.

PROPOSITION 5.1
*Let $m \geqq 1$, and assume that $\phi = \phi_m$ is a residual eigenfunction (hence $0 < \lambda_m < 1/4$). Define $s$ through $\lambda_m = s(1-s)$, $s \in (1/2, 1)$. Let the Fourier expansion of $\phi(z)$ at the cusp $\eta_1 = \infty$ be*

$$\phi(z) = c_0 y^{1-s} + \sum_{n \neq 0} c_n \sqrt{y} K_{s-1/2}(2\pi |n| y) e(nx). \tag{5.1}$$

*We then have, uniformly over all $N \geqq 1$ and $\alpha \in \mathbb{R}$,*

$$\sum_{n=1}^{N} c_n e(n\alpha) = O(N^{3/2-s}).$$

*(The implied constant depends on $\Gamma$ and $\phi(z)$.) The same bound holds for the sum $\sum_{n=-N}^{-1} c_n e(n\alpha)$.*

*Proof*
The basic idea is the same as in [H, Theorem 3], but the computations in the present case are much more involved. We fix a number $\delta \in (s, 1)$, and we let

$$I = \int_0^\infty \int_{\alpha-1/2}^{\alpha+1/2} \phi(x+iy) \Big( \sum_{m=1}^{N} e\big(m(\alpha - x)\big) \Big) \, dx \, \frac{dy}{y^\delta}. \tag{5.2}$$

We remark that the double integral is not absolutely convergent, but we have

$$\int_0^\infty \Big| \int_{\alpha-1/2}^{\alpha+1/2} \cdots \, dx \Big| \, y^{-\delta} dy < \infty.$$

In fact, for each fixed $y > 0$, using (5.1), we see that the inner integral in (5.2) equals $\sum_{m=1}^{N} c_m \sqrt{y} K_{s-1/2}(2\pi m y) e(m\alpha)$, and the absolute convergence follows at once (cf.

(4.29)) since the sum is finite. In fact, using [W, page 388 (8)], we find that

$$I = \sum_{m=1}^{N} c_m e(m\alpha)(2\pi m)^{\delta-3/2} 2^{-1/2-\delta}\, \Gamma\Big(\frac{2-s-\delta}{2}\Big)\Gamma\Big(\frac{1+s-\delta}{2}\Big). \tag{5.3}$$

We now estimate $|I|$ from above. For each $k \in \{1, \dots, \kappa\}$, we know from the Fourier expansion of $\phi(z)$ at the cusp $\eta_k$ that

$$\phi(z) = c_0^{(k)}\big(\operatorname{Im} N_k(z)\big)^{1-s} + O(e^{-2\pi \operatorname{Im} N_k(z)}) \quad \text{as } \operatorname{Im} N_k(z) \to \infty \tag{5.4}$$

(cf. (5.1); of course, $c_0^{(1)} = c_0$). Also, $\phi(z)$ is bounded on any bounded region $\mathscr{F}_B$. Using $\Gamma$-invariance, it now follows that

$$|\phi(z)| = O\big(\mathscr{Y}_\Gamma(z)^{1-s}\big) = O\big(1 + \lfloor \mathscr{Y}_\Gamma(z)\rfloor_1^{1-s}\big), \quad \forall z \in \mathscr{H} \tag{5.5}$$

(cf. (2.4), (2.5)). Here, and in all big $O$ estimates in the rest of this proof, the implied constant depends solely on $\Gamma$, $\phi(z)$, and $\delta$.

For $y \geqq 1$, we substitute (5.4) (with $k = 1$) directly in (5.2); the $c_0$-term is then killed in the inner integral. For $y \leqq 1$, we use (5.5). We then get

$$\begin{aligned}|I| \leqq &\int_1^\infty \int_{\alpha-1/2}^{\alpha+1/2} O(e^{-2\pi y})\Big|\sum_{m=1}^{N} e\big(m(\alpha-x)\big)\Big| \frac{dx\,dy}{y^\delta}\\ &+ \int_0^1 \int_{\alpha-1/2}^{\alpha+1/2} O\big(1 + \lfloor \mathscr{Y}_\Gamma(z)\rfloor_1^{1-s}\big)\Big|\sum_{m=1}^{N} e\big(m(\alpha-x)\big)\Big| \frac{dx\,dy}{y^\delta}.\end{aligned}$$

We have, by direct evaluation, $\big|\sum_{m=1}^N e(m(\alpha-x))\big| = O(\min(N, |x-\alpha|^{-1}))$ for all $x \in [\alpha - 1/2, \alpha + 1/2]$, and thus $\int_{\alpha-1/2}^{\alpha+1/2} \big|\sum_{m=1}^N e(m(\alpha-x))\big|\,dx = O(\log 2N)$. Hence

$$|I| \leqq O(\log 2N) + O(1)\int_0^1 \int_{\alpha-1/2}^{\alpha+1/2} \lfloor \mathscr{Y}_\Gamma(z)\rfloor_1^{1-s} \min(N, |x-\alpha|^{-1})\,\frac{dx\,dy}{y^\delta}.$$

Using here the definition of $\mathscr{Y}_\Gamma(z)$, we see that the double integral is bounded from above by

$$\sum_{k=1}^{\kappa} \sum_{W_0 \in [T_k]\backslash\Gamma} \int_0^1 \int_{\alpha-1/2}^{\alpha+1/2} \lfloor \operatorname{Im} N_k W_0(z)\rfloor_1^{1-s} \min(N, |x-\alpha|^{-1})\,\frac{dx\,dy}{y^\delta}. \tag{5.6}$$

We now temporarily fix some $k \in \{1, \dots, \kappa\}$ and $W_0 \in [T_k] \setminus \Gamma$ which give a nonzero contribution in (5.6), and we write $\left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) = N_k W_0$. Then $\operatorname{Im} N_k W_0(z) \geqq 1$ for some $z = x + iy \in (\alpha - 1/2, \alpha + 1/2) \times (0, 1)$. Hence $N_k W_0 \notin [S]$, and by Lemma 2.3 with $j = 1$, $N_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1\end{smallmatrix}\right)$, we have $|c| \geqq 1$. Notice that $\operatorname{Im} N_k W_0(z) \geqq 1$

means that $z$ belongs to the horoball tangent to $\mathbb{R}$ at $-d/c$ with Euclidean radius $r = (1/2)c^{-2} \leqq 1/2$. In particular, $|x + d/c| \leqq r \leqq 1/2$, and combining this with $|x - \alpha| < 1/2$, we see that $-d/c \in (\alpha - 1, \alpha + 1)$. We let

$$\alpha' = \alpha + \frac{d}{c}. \tag{5.7}$$

CLAIM 1

*The contribution from each $\langle k, W_0\rangle$ in (5.6) is bounded by*

$$O\big[|c|^{2s-2} \min(N, |\alpha'|^{-1})^{\delta-s} \cdot (1 + \log^+(N|\alpha'|))\big]. \tag{5.8}$$

*Proof*

As noticed above, the integrand in (5.6) vanishes outside the vertical strip $|x + d/c| \leqq r$. Using this and $\operatorname{Im} W_0(z) = y/|cz + d|^2$, we get the following upper bound on the double integral in (5.6):

$$|c|^{2s-2} \int_{-r}^{r} \left( \int_0^1 \left( \frac{y}{x^2 + y^2} \right)^{1-s} \frac{dy}{y^\delta} \right) \min(N, |x - \alpha'|^{-1})\, dx. \tag{5.9}$$

For any $x \neq 0$, the inner integral is less than

$$\int_0^\infty \left( \frac{y}{x^2 + y^2} \right)^{1-s} \frac{dy}{y^\delta} \leqq \int_0^{|x|} \left( \frac{y}{x^2} \right)^{1-s} \frac{dy}{y^\delta} + \int_{|x|}^\infty \left( \frac{y}{y^2} \right)^{1-s} \frac{dy}{y^\delta} = O(|x|^{s-\delta}) \tag{5.10}$$

(we used $1/2 < s < \delta < 1$). Using this fact and $|x - \alpha'| \geqq \big||x| - |\alpha'|\big|$, we find that (5.9) is bounded from above by

$$O(|c|^{2s-2}) \int_0^\infty x^{s-\delta} \min(N, |x - |\alpha'||^{-1})\, dx. \tag{5.11}$$

If $\left|\alpha'\right| \leqq 100/N$, we get that (5.11) is

$$= O(|c|^{2s-2}) \int_0^{200/N} x^{s-\delta} N\, dx + O(|c|^{2s-2}) \int_{200/N}^\infty x^{s-\delta-1}\, dx = O(|c|^{2s-2} N^{\delta-s})$$

(again using $1/2 < s < \delta < 1$). The bound thus obtained is the same as (5.8) since $\left|\alpha'\right| \leqq 100/N$.

On the other hand, if $|\alpha'| > 100/N$, then by splitting the integral at $x = |\alpha'|/2$ and $x = 3|\alpha'|/2$ and using obvious inequalities, we see that (5.11) is less than

$$\begin{aligned} O(|c|^{2s-2}) \Big[ &\int_0^{|\alpha'|/2} x^{s-\delta} |\alpha'|^{-1}\, dx + \int_{3|\alpha'|/2}^\infty x^{s-\delta} x^{-1}\, dx \\ &+ \int_{|\alpha'|/2}^{3|\alpha'|/2} |\alpha'|^{s-\delta} \min(N, |x - |\alpha'||^{-1})\, dx \Big] \\ = O\big(&|c|^{2s-2} |\alpha'|^{s-\delta} (1 + \log(N|\alpha'|))\big). \end{aligned}$$

This is, again, the same as (5.8). □

CLAIM 2

*In the case* $c^2 \geqq \min(N, |\alpha'|^{-1})$, *we also have the following upper bound on the contribution from* $\langle k, W_0 \rangle$ *in (5.6):*

$$O\Big[|c|^{2\delta-4}\min(N, |\alpha'|^{-1})\Big]. \tag{5.12}$$

*Proof*

Assume that $c^2 \geqq \min(N, |\alpha'|^{-1})$. We now maintain that

$$\min(N, |x-\alpha'|^{-1}) \leqq 2\min(N, |\alpha'|^{-1}), \quad \forall x \in [-r, r]. \tag{5.13}$$

Indeed, if $N \leqq |\alpha'|^{-1}$, then (5.13) is trivial since the right-hand side equals $2N$. In the other case, $N > |\alpha'|^{-1}$, we prove (5.13) by first noticing that

$$(2r)^{-1} = c^2 \geqq \min(N, |\alpha'|^{-1}) = |\alpha'|^{-1},$$

and thus $r \leqq |\alpha'|/2$. We then get, for all $x \in [-r, r]$,

$$|x-\alpha'| \geqq |\alpha'| - r \geqq \frac{|\alpha'|}{2},$$

and hence $|x-\alpha'|^{-1} \leqq 2|\alpha'|^{-1} = 2\min\big(N, |\alpha'|^{-1}\big)$, and (5.13) is proved.

Using (5.13), (5.10), and $\int_{-r}^{r} |x|^{s-\delta}\,dx = O(r^{s-\delta+1}) = O(|c|^{-2(s-\delta+1)})$, we find that (5.9) is less than (5.12). □

We continue onward with our proof of Proposition 5.1. We add up all contributions to (5.6) for one fixed $k \in \{1, \ldots, \kappa\}$. As we have already noticed, we get a nonzero contribution from $W_0 = N_k^{-1}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in [T_k] \setminus \Gamma$ only if $|c| \geqq 1$ and $-d/c \in (\alpha - 1, \alpha + 1)$. We split the analysis into the following cases:

$$\begin{aligned}
&-\frac{d}{c} \in [\alpha - N^{-1}, \alpha + N^{-1}],\\
&-\frac{d}{c} \in [\alpha + 2^{\ell}N^{-1}, \alpha + 2^{\ell+1}N^{-1}] \quad \big(\ell \in \{0, 1, 2, \ldots, [\log_2 N]\}\big),\\
&-\frac{d}{c} \in [\alpha - 2^{\ell+1}N^{-1}, \alpha - 2^{\ell}N^{-1}] \quad \big(\ell \in \{0, 1, 2, \ldots, [\log_2 N]\}\big).
\end{aligned}$$

Clearly, each $W_0$ giving nonzero contribution to (5.6) is then counted at least once.

Let us consider the second case in detail; that is, we take $\ell \in \{0, 1, 2, \ldots, [\log_2 N]\}$ and consider all $W_0$ satisfying $-d/c \in [\alpha + 2^{\ell}N^{-1}, \alpha + 2^{\ell+1}N^{-1}]$. We then have $\min(N, |\alpha'|^{-1}) = |\alpha'|^{-1} \leqq 2^{-\ell}N$ (cf. (5.7)), and hence by Claims 1 and 2, we get the following bound on the total contribution from these $W_0$'s:

$$O\Big( \sum_{\substack{W_0 \\ 1 \leqq |c| < \sqrt{2^{-\ell} N}}} |c|^{2s-2} (2^{-\ell} N)^{\delta - s} \big(1 + \log(2^{\ell+1})\big) + \sum_{\substack{W_0 \\ |c| \geqq \sqrt{2^{-\ell} N}}} |c|^{2\delta - 4} (2^{-\ell} N) \Big).$$

(Of course, $2^{-\ell} N \geqq 1$ since $\ell \leqq [\log_2 N]$.) The sums are taken over a set of representatives $W_0 = N_k^{-1} \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in [T_k] \setminus \Gamma$ restricted by $-d/c \in [\alpha + 2^{\ell} N^{-1}, \alpha + 2^{\ell+1} N^{-1}]$ together with the stated bounds on $|c|$. By a quick computation using Lemma 2.5, we now get

$$= O\big[(\ell + 1)\big((2^{-\ell} N)^{\delta - s} + (2^{-\ell} N)^{\delta - 1}\big)\big] = O\big((\ell + 1)(2^{-\ell} N)^{\delta - s}\big).$$

We get exactly the same bound for $-d/c \in [\alpha - 2^{\ell+1} N^{-1}, \alpha - 2^{\ell} N^{-1}]$. Also, by a similar computation using Claim 1, Claim 2, and Lemma 2.5, we find that the total contribution from all $W_0$ satisfying $-d/c \in [\alpha - N^{-1}, \alpha + N^{-1}]$ is bounded by $O(N^{\delta - s})$.

We now add up *all* these contributions to (5.6), for each $k \in \{1, \dots, \kappa\}$. Since $\sum_{\ell=0}^{\infty} (\ell + 1) 2^{-\ell(\delta - s)} < \infty$, we finally obtain

$$|I| = O(\log 2N + N^{\delta - s}) = O(N^{\delta - s}).$$

Hence by (5.3), $\sum_{m=1}^{N} c_m e(m\alpha) m^{\delta - 3/2} = O(N^{\delta - s})$, and by partial summation, we obtain

$$\sum_{n=1}^{N} c_n e(n\alpha) = O(N^{3/2 - s}),$$

and we are done. The proof of $\sum_{n=-N}^{-1} c_n e(n\alpha) = O(N^{3/2 - s})$ is entirely similar. □

*Remark 5.2*

Proposition 5.1 is complemented by the following result, which shows that the exponent $3/2 - s$ therein is the best possible.

PROPOSITION 5.1′

*Let $\phi(z)$ be as in Proposition 5.1. Take $k \in \{1, \dots, \kappa\}$ such that $c_0^{(k)} \neq 0$ in (5.4) (such $k$ always exists since $\phi$ is a residual eigenfunction), and let $\alpha \in \mathbb{R}$ be any cusp equivalent to $\eta_k$.* Then at least one of*

$$\sum_{n=1}^{N} c_n e(n\alpha) = \Omega(N^{3/2 - s}), \qquad \sum_{n=-N}^{-1} c_n e(n\alpha) = \Omega(N^{3/2 - s})$$

*It is well known that the set of such points $\alpha$ is dense in $\mathbb{R}$. A more precise statement of this nature is provided in §7(II).

*must hold as* $N \to \infty$.

*Proof*
Assume that this is not true; that is, assume that $S_+(X) = \sum_{1 \leqq n \leqq X} c_n e(n\alpha) = o(X^{3/2-s})$ and $S_-(X) = \sum_{1 \leqq n \leqq X} c_{-n} e(-n\alpha) = o(X^{3/2-s})$ as $X \to \infty$. (Here $\alpha$ is fixed.) We now study $\phi(\alpha + iy)$ as $y \to 0$.

It follows from (5.1), using partial summation (treating $n > 0$ and $n < 0$ separately), that

$$\phi(\alpha + iy) = c_0 y^{1-s} - 2\pi y^{3/2} \int_{1/2}^{\infty} K'_{s-1/2}(2\pi X y)\big(S_+(X) + S_-(X)\big)\, dX.$$

Convergence follows easily using our bounds on $S_\pm(X)$ and the exponential decay of $K'_{s-1/2}(u)$ (cf. (4.29)).

Now let $\varepsilon > 0$ be given. It follows from our assumption on $S_\pm(X)$ that there is some $X_0 = X_0(\varepsilon) > 1$ such that $|S_+(X) + S_-(X)| \leqq \varepsilon X^{3/2-s}$ for all $X \geqq X_0$. Let $M = \sup_{X \in [1/2, X_0]} |S_+(X) + S_-(X)|$. We then get, for all $y < 1/X_0$,

$$\begin{aligned}
|\phi(\alpha + iy)| &\leqq |c_0| y^{1-s} + 2\pi y^{3/2} \int_{1/2}^{X_0} C(Xy)^{-1/2-s} M\, dX \\
&\quad + 2\pi y^{3/2} \int_{X_0}^{1/y} C(Xy)^{-1/2-s} \varepsilon X^{3/2-s}\, dX \\
&\quad + 2\pi y^{3/2} \int_{1/y}^{\infty} C e^{-Xy} \varepsilon X^{3/2-s}\, dX \\
&< \Big(|c_0| + 2\pi C M \int_{1/2}^{X_0} X^{-1/2-s}\, dX\Big) y^{1-s} \\
&\quad + \varepsilon \Big(\frac{2\pi C}{2-2s} + 2\pi C \int_1^{\infty} e^{-u} u^{3/2-s}\, du\Big) y^{s-1}.
\end{aligned}$$

Recall that $s \in (1/2, 1)$; hence $s - 1 < 0 < 1 - s$. Since $\varepsilon$ was arbitrary, and since the expression inside the parentheses in front of $y^{s-1}$ above does not depend on $\varepsilon$, it now follows that

$$\phi(\alpha + iy) = o(y^{s-1}) \quad \text{as } y \to 0. \tag{5.14}$$

On the other hand, take $T \in \Gamma$ such that $\alpha = T(\eta_k)$, and write $N_k T^{-1} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. We then have $N_k T^{-1}(\alpha) = \infty$, and thus $c\alpha + d = 0$. Hence $c \neq 0$, and by a quick computation, $\operatorname{Im} N_k T^{-1}(\alpha + iy) = |c|^{-2} y^{-1}$ for all $y > 0$. Using (5.4) and $\Gamma$-invariance, we now get

$$\phi(\alpha + iy) \sim c_0^{(k)} \big(\operatorname{Im} N_k T^{-1}(\alpha + iy)\big)^{1-s} = c_0^{(k)} |c|^{2(s-1)} y^{s-1}$$

as $y \to 0$. This is a contradiction to (5.14). □

PROPOSITION 5.3
*Let $m \geqq 1$ be such that $\phi_m$ is a residual eigenfunction (hence $0 < \lambda_m < 1/4$). Define $s$ by $\lambda_m = s(1-s)$, $s \in (1/2, 1)$. We then have, uniformly over all $y \in (0,1)$ and all $\alpha, \beta$ such that $y \leqq \beta - \alpha \leqq 1$,*

$$\frac{1}{\beta-\alpha}\int_\alpha^\beta \phi_m(x+iy)\,dx = O\big(y^{1-s}(\beta-\alpha)^{2s-2}\big).$$

*The implied constant depends on $\Gamma$ and $\phi_m$.*

*Proof*
The proof is almost identical to the proof of Proposition 4.6, except that we use the bound from Proposition 5.1 instead of (4.25). Notice that since $1/2 < s < 1$, (4.29) holds with $\varepsilon = 0$. □

The proof of Theorem 3 is now complete in view of Propositions 4.3, 4.5, 4.6, and 5.3.

## 6. Joint equidistribution of subsegments

In this section we prove a result on the joint distribution of several subsegments of the closed horocycle $\{x+iy \mid x \in [0,1]\}$.

As before, we let $\Gamma$ be a cofinite Fuchsian group with a standard cusp at infinity. We use the same notation as in §3, in particular, $G = \mathrm{PSL}(2,\mathbb{R})$. Let us write

$$\mathrm{Comm}(\Gamma) = \big\{g \in G \,\big|\, g\Gamma g^{-1} \text{ and } \Gamma \text{ are commensurable}\big\}.$$

(Two subgroups of a group are called *commensurable* if their intersection has finite index in both of them.) For basic information concerning $\mathrm{Comm}(\Gamma)$ and its use for the construction of Hecke operators on $\Gamma \setminus \mathscr{H}$, the reader is referred to [Shim, Chapter 3]. We also remark that if $\Gamma$ is a nonarithmetic group, then $\Gamma$ is of finite index in $\mathrm{Comm}(\Gamma)$ (cf. [Ma, Chapter IX, Theorem 1.16]).

THEOREM 4
*Let $n \geqq 2$, let $\ell > 0$, and let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be real numbers such that*

$$\begin{pmatrix} 1 & \alpha_j - \alpha_k \\ 0 & 1 \end{pmatrix} \notin \mathrm{Comm}(\Gamma) \quad \textit{for all } \; j \neq k. \tag{6.1}$$

*We then have, for any bounded continuous function $f$ on $\boldsymbol{M}^n = (\Gamma \setminus T_1\mathscr{H})^n$,*

$$\lim_{y\to 0}\frac{1}{\ell}\int_0^\ell f\big((\alpha_1+x+iy,0),(\alpha_2+x+iy,0),\ldots,(\alpha_n+x+iy,0)\big)\,dx$$
$$= \frac{1}{\nu(\boldsymbol{M})^n}\int_{\boldsymbol{M}^n} f(p_1,\ldots,p_n)\,d\nu(p_n)\cdots d\nu(p_1). \tag{6.2}$$

*Remark 6.1*
The condition (6.1) is also a *necessary* condition, for if $S^{\alpha_j-\alpha_k} \in \mathrm{Comm}(\Gamma)$ for some $j \neq k$, then one easily checks that the whole subset $\{(S^{\alpha_j}p, S^{\alpha_k}p) \mid p \in T_1\mathscr{H}\}$ of $(T_1\mathscr{H})^2$ projects onto a closed submanifold of codimension 3 in $(\Gamma \setminus T_1\mathscr{H})^2$, so that (6.2) cannot hold for all $f$.

*Proof of Theorem 4*
We apply Shah, [Sh, Theorem 1.4], for the group $L = G^n = G \times \cdots \times G$, the lattice $\Lambda = \Gamma^n \subset L$, the expanding horospherical subgroup $U^+ = \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \subset G$, and the probability measure $\lambda$ on $U^+$ defined by $\lambda(A) = \ell^{-1} \cdot m\big([0,\ell] \cap \big\{x \bigm| \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) \in A\big\}\big)$ for any Borel set $A \subset U^+$, where $m(\cdot)$ is the Lebesgue measure on $\mathbb{R}$. Furthermore, we take $G$ to be imbedded as a closed Lie subgroup of $L$ by the map

$$G \ni g \mapsto \big(S^{\alpha_1} g S^{-\alpha_1},\ S^{\alpha_2} g S^{-\alpha_2},\ \ldots,\ S^{\alpha_n} g S^{-\alpha_n}\big) \in L.$$

Let $\pi$ be the projection $L \to \Lambda\backslash L$. We will prove below the following proposition.

PROPOSITION 6.2
*Under assumption (6.1), $\pi(G)$ is dense in $\Lambda\backslash L$.*

Using this proposition, it is now easy to check that all assumptions in Shah's Theorem 1.4 are fulfilled. The conclusion from Shah's theorem is that, for any sequence of positive real numbers $y_j$ with $\lim_{j\to\infty} y_j = 0$ and for any bounded continuous function $f_0$ on $\Gamma^n\backslash L$, we have

$$\lim_{j\to\infty} \frac{1}{\ell} \int_0^{\ell} f_0\big(S^{x+\alpha_1}\boldsymbol{a}(y_j)S^{-\alpha_1},\ \ldots,\ S^{x+\alpha_n}\boldsymbol{a}(y_j)S^{-\alpha_n}\big)\, dx = \int_{\Gamma^n\backslash L} f_0\, d\mu_L. \quad (6.3)$$

Here $\mu_L$ is the unique $L$-invariant probability measure on $\Gamma^n\backslash L$.

We let $T$ denote right multiplication by $(S^{\alpha_1}, S^{\alpha_2}, \ldots, S^{\alpha_n})$ on $\Gamma^n\backslash L$; then $T$ is a homeomorphism of $\Gamma^n\backslash L$ onto itself preserving the measure $\mu_L$. Given any bounded continuous function $f$ on $\Gamma^n\backslash L$, we now apply (6.3) to $f_0 = f \circ T$. Reinterpreting the result via the identifications $\Gamma^n\backslash L = (\Gamma\backslash G)^n = \boldsymbol{M}^n$, we obtain (6.2). □

It remains to prove Proposition 6.2. We first prove the following lemma.

LEMMA 6.3
*Let $\mathfrak{g}$ be the Lie algebra of $G = \mathrm{PSL}(2,\mathbb{R})$, and let $\mathfrak{h}$ be a Lie subalgebra of the direct sum $\mathfrak{l} = \mathfrak{g} \oplus \mathfrak{g} \oplus \cdots \oplus \mathfrak{g}$ ($n$ copies). Assume that $\mathfrak{h}$ contains the diagonal $\{(X, X, \ldots, X) \mid X \in \mathfrak{g}\}$. Then there exists an equivalence relation $\sim$ on the set $\{1, \ldots, n\}$ such that*

$$\mathfrak{h} = \big\{(X_1, X_2, \ldots, X_n) \in \mathfrak{l} \bigm| X_i = X_j \text{ whenever } i \sim j\big\}.$$

*Proof*
Given $\mathfrak{h}$ as above, we define a relation $\sim$ on $\{1, \ldots, n\}$ by letting $i \sim j$ hold if and only if $X_i = X_j$ holds for all vectors $(X_1, \ldots, X_n) \in \mathfrak{h}$. Clearly, $\sim$ is an equivalence relation, and

$$\mathfrak{h} \subset \left\{(X_1, X_2, \ldots, X_n) \in \mathfrak{l} \,\middle|\, X_i = X_j \text{ whenever } i \sim j\right\}. \tag{6.4}$$

We have to prove that the opposite inclusion holds as well. Clearly, the right-hand side of (6.4) is isomorphic to a direct sum of $e$ copies of $\mathfrak{g}$, where $e$ is the number of equivalence classes of $\sim$. Using this isomorphism, we may reduce to the case when $n = e$; that is, we may assume that $i \not\sim j$ holds for all $i \neq j$. Our task is now to prove that $\mathfrak{h} = \mathfrak{l}$.

We let $H, R, L \in \mathfrak{g}$ be a standard basis with $[R, L] = H$, $[H, L] = -2L$, $[H, R] = 2R$. Also, we let $d : \mathfrak{g} \to \mathfrak{l}$ be the diagonal map $X \mapsto (X, X, \ldots, X)$, so that, by our assumption, $d(X) \in \mathfrak{h}$ for all $X \in \mathfrak{g}$.

*Step 1.* For any given $i \neq j$ in $\{1, \ldots, n\}$, there exists a vector $(t_1 H, t_2 H, \ldots, t_n H) \in \mathfrak{h}$ ($t_k \in \mathbb{R}$) such that $t_i \neq t_j$. To prove this claim, first note that since $i \not\sim j$, there is a vector $(X_1, \ldots, X_n) \in \mathfrak{h}$ with $X_i \neq X_j$. Thus, writing $X_k = a_k H + b_k R + c_k L$, at least one of $a_i \neq a_j$, $b_i \neq b_j$, $c_i \neq c_j$ must hold. Now the claim follows since the following vectors belong to $\mathfrak{h}$:

$$\begin{aligned}
\operatorname{ad} d(L) \operatorname{ad} d(H) \operatorname{ad} d(R)\, (X_1, X_2, \ldots, X_n) &= (4a_1 H, 4a_2 H, \ldots, 4a_n H), \\
\operatorname{ad} d(L) \operatorname{ad} d(H)\, (X_1, X_2, \ldots, X_n) &= (-2b_1 H, -2b_2 H, \ldots, -2b_n H), \\
\operatorname{ad} d(R) \operatorname{ad} d(H)\, (X_1, X_2, \ldots, X_n) &= (-2c_1 H, -2c_2 H, \ldots, -2c_n H).
\end{aligned}$$

*Step 2.* There is a vector $(t_1 H, t_2 H, \ldots, t_n H) \in \mathfrak{h}$ such that all the numbers $t_1, \ldots, t_n$ are nonzero and pairwise distinct. This follows easily by constructing a suitable linear combination of the vector $d(H)$ and the vectors that we obtained in Step 1.

*Step 3.* The Lie algebra $\mathfrak{h}$ contains *each* vector $(x_1 H, x_2 H, \ldots, x_n H)$ ($x_j \in \mathbb{R}$). To prove this, let $X^{(1)} = (t_1 H, t_2 H, \ldots, t_n H) \in \mathfrak{h}$ be a vector as in Step 2. Define the vectors $X^{(2)}, X^{(3)}, \ldots \in \mathfrak{h}$ recursively by

$$X^{(k)} = \frac{1}{4}\big[[d(L), X^{(1)}], [d(R), X^{(k-1)}]\big].$$

We then find that

$$X^{(k)} = (t_1^k H, t_2^k H, \ldots, t_n^k H) \quad \text{for each } k \geqq 1.$$

Since all the $t_j$'s are nonzero and pairwise distinct, it follows using the Vandermonde determinant that the vectors $X^{(1)}, \ldots, X^{(n)}$ span the space $\{(x_1 H, x_2 H, \ldots, x_n H) \mid x_j \in \mathbb{R}\}$, and the claim is proved.

Applying now $\operatorname{ad} d(L)$ and $\operatorname{ad} d(R)$ to the $\mathfrak{h}$-vectors obtained in Step 3, we find that $\mathfrak{h}$ contains all vectors $(x_1 L, x_2 L, \dots, x_n L)$ and $(x_1 R, x_2 R, \dots, x_n R)$ as well. Hence $\mathfrak{h} = \mathfrak{l}$, and the lemma is proved. $\square$

*Proof of Proposition 6.2*

It is convenient to alter the notation by a conjugation so as to make $G$ imbedded in $L$ by the diagonal map $G \ni g \mapsto (g, g, \dots, g) \in L$, and $\Lambda = S^{-\alpha_1}\Gamma S^{\alpha_1} \times \cdots \times S^{-\alpha_n}\Gamma S^{\alpha_n}$.

Notice that $G$ is generated by (Ad-)unipotent one-parameter subgroups of $L$ contained in $G$. Hence Ratner's result [R2, Corollary B] applies, and it follows that there is a closed subgroup $H \subset L$ such that $G \subset H$, $H \cap \Lambda$ is a lattice in $H$, and $\overline{\pi(G)} = \pi(H)$ in $\Lambda\backslash L$. By Lemma 6.3 applied to the Lie algebra of $H$, the identity component $H^0$ has the following explicit form for some fixed equivalence relation $\sim$ on $\{1, \dots, n\}$:

$$H^0 = \big\{(g_1, g_2, \dots, g_n) \in L \,\big|\, g_i = g_j \text{ whenever } i \sim j\big\}.$$

Clearly, if $C_1, \dots, C_e \subseteq \{1, \dots, n\}$ are the distinct equivalence classes of $\sim$, then there is a natural isomorphism $H^0 \cong \prod_{m=1}^{e} G$ under which $H^0 \cap \Lambda$ corresponds to $\prod_{m=1}^{e}\big(\bigcap_{j\in C_m} S^{-\alpha_j}\Gamma S^{\alpha_j}\big)$. Since $H^0 \cap \Lambda$ is a lattice in $H^0$, it follows that $\bigcap_{j\in C_m} S^{-\alpha_j}\Gamma S^{\alpha_j}$ must be a lattice in $G$ for each $m$. Hence $|C_m| = 1$ for each $m$ since, by assumption, $S^{\alpha_j-\alpha_k} \notin \operatorname{Comm}(\Gamma)$ whenever $j \neq k$. Hence $H^0 = L$, and thus $\overline{\pi(G)} = \Lambda\backslash L$. $\square$

Now Theorem 4 is completely proved. We give two corollaries concerning the sum $S_{y,N}(x)$.

As in the introduction, we define

$$S_{y,N}(x) = \sum_{j=0}^{N-1} F\Big(\frac{x+j}{N} + iy\Big),$$

where $F : \mathscr{H} \to \mathbb{R}$ is a fixed, bounded, continuous, and $\Gamma$-invariant function. For each $y > 0$ and $N \in \mathbb{Z}^+$, we view $S_{y,N}$ as a random variable by taking $x$ in the probability space $([0,1], m)$ with $m =$ Lebesgue measure. We also let $Y_1, Y_2, \dots$ be independent, identically distributed random variables with distribution given by

$$\operatorname{Prob}(Y_n \in A) = \frac{\mu\big\{z \in \Gamma \setminus \mathscr{H} \mid F(z) \in A\big\}}{\mu(\Gamma \setminus \mathscr{H})} \quad \text{(for each Borel set } A \subset \mathbb{R}).$$

Given $D \in \mathbb{Z}^+$, we define

$$\mathbb{Z}_D^+ = \big\{N \in \mathbb{Z}^+ \,\big|\, \gcd(N, D) = 1\big\}.$$

COROLLARY 6.4
*If $\Gamma$ is nonarithmetic, then there exists a $D = D(\Gamma) \in \mathbb{Z}^+$ such that for each fixed $N \in \mathbb{Z}_D^+$ and each fixed bounded continuous function $F : \Gamma\backslash\mathscr{H} \to \mathbb{R}$, $S_{y,N}$ converges in distribution to $Y_1 + Y_2 + \cdots + Y_N$ as $y \to 0$.*

*Proof*
Since $\Gamma$ is nonarithmetic, $\Gamma$ is of finite index in $\mathrm{Comm}(\Gamma)$ (cf. [Ma, Chapter IX, Theorem 1.16]). In particular, since $S^{x_1}\Gamma = S^{x_2}\Gamma \iff S^{x_1-x_2} \in \Gamma \iff x_1 - x_2 \in \mathbb{Z}$, there exist at most finitely many rational numbers $q \in (0,1)$ for which $S^q \in \mathrm{Comm}(\Gamma)$. Let $D = D(\Gamma)$ be the product of the denominators of all these numbers $q$. (Let $D = 1$ if there are no such numbers $q$.)

We now fix some $N \in \mathbb{Z}_D^+$ and some function $F$ as above. We then have $S^{(j-k)/N} \notin \mathrm{Comm}(\Gamma)$ for all $j \neq k \in \{1, \ldots, N\}$.

Let $k$ be a given positive integer, and define $f(z_1, \ldots, z_N) = \left(\sum_{j=1}^N F(z_j)\right)^k$; then $f$ is a bounded continuous function $(\Gamma \setminus \mathscr{H})^N \to \mathbb{R}$. The moments $\mathbf{E}\ S_{y,N}^k$ and $\mathbf{E}\ (Y_1 + \cdots + Y_N)^k$ can now be expressed as follows:

$$\mathbf{E}\ S_{y,N}^k = \int_0^1 f\Big(\frac{x}{N} + iy, \frac{x+1}{N} + iy, \ldots, \frac{x+N-1}{N} + iy\Big)\, dx;$$
$$\mathbf{E}\ (Y_1 + \cdots + Y_N)^k = \mu(\Gamma \setminus \mathscr{H})^{-N} \int_{(\Gamma\setminus\mathscr{H})^N} f(z_1, \ldots, z_N)\, d\mu(z_N) \cdots d\mu(z_1).$$

Hence by Theorem 4 (if $N = 1$, Theorem 1) and our remarks above regarding $\mathrm{Comm}(\Gamma)$, we have $\mathbf{E}\ S_{y,N}^k \to \mathbf{E}\ (Y_1 + \cdots + Y_N)^k$ as $y \to 0$. This holds for each fixed $k \in \mathbb{Z}^+$, and hence $S_{y,N}$ converges in distribution to $Y_1 + \cdots + Y_N$ as $y \to 0$ (cf., e.g., [F, Example VIII.1(d)]). $\square$

COROLLARY 6.5
*Let $\Gamma$, $D(\Gamma)$, $F$ be as in Corollary 6.4, and assume furthermore that $\int_{\Gamma\setminus\mathscr{H}} F\, d\mu = 0$. Then there exists a sequence $\{h_N\}_{N\in\mathbb{Z}_D^+}$ of positive numbers with $\lim_{N\to\infty} h_N = 0$, such that, for any sequence $\{y_N\}_{N\in\mathbb{Z}_D^+}$ of numbers satisfying $0 < y_N \leqq h_N$, the random variable $N^{-1/2} S_{y_N,N}$ converges in distribution to a Gaussian with mean zero and variance $\sigma_F^2 = \mu(\Gamma \setminus \mathscr{H})^{-1} \int_{\Gamma\setminus\mathscr{H}} F^2\, d\mu$ as $N \in \mathbb{Z}_D^+$, $N \to \infty$. (The sequence $\{h_N\}_{N\in\mathbb{Z}_D^+}$ may depend on $\Gamma$ and $F$.)*

*Proof*
For each $N \in \mathbb{Z}_D^+$, Corollary 6.4 shows that we may take $h_N > 0$ so small that, for all $y \in (0, h_N]$,
$$d\big[N^{-1/2} S_{y,N};\, N^{-1/2}(Y_1 + \cdots + Y_N)\big] \leqq N^{-1},$$
where $d[\cdot\ ;\ \cdot]$ is the Lévy distance between the corresponding distribution functions.

But by the central limit theorem for sums of independent, identically distributed random variables, we have

$$d\big[N^{-1/2}(Y_1 + \cdots + Y_N);\ \mathfrak{N}(0, \sigma_F^2)\big] \to 0 \quad \text{as } N \to \infty,$$

where $\mathfrak{N}(0, \sigma_F^2)$ denotes the normal distribution with mean zero and variance $\sigma_F^2$. The corollary follows from this. □

*Remark 6.6*

The numerical investigations of $S_{y,N}$ in [He2, §5] were carried out on Hecke triangle groups $\mathbb{G}_L$ for $L = 3, 5, 7$. We let $\mathbb{G}'_L$ denote a conjugated version of $\mathbb{G}_L$, normalized to have a cusp of standard width 1 at $\infty$, as in [He1, p. 569 (7.2)]. We remark that *if* $\Gamma$ *is a nonarithmetic Hecke triangle group* $\Gamma = \mathbb{G}'_L$ (i.e., $L = 5$ or $L \geqq 7$), then $\mathrm{Comm}(\Gamma) = \Gamma$ by [L], and hence *Corollaries 6.4 and 6.5 hold with* $D = D(\Gamma) = 1$.

## 7. Some further results, applications, and comments

(I) In the special case of $f = \phi$ a cusp form, Proposition 4.6 shows that the exponent $c(\Gamma) = 1/2$ in Theorem 1 can be improved all the way up to $c(\Gamma) = 1$. In [St, §4], we used this fact together with methods involving the incomplete Eisenstein series to show that, for arbitrary $f$, (1.2) holds with $c(\Gamma) = 1$, *so long as* $\alpha$ *(or* $\beta$*) is kept generic.* In precise terms, we have the following.

THEOREM 7.1

*Let* $\Gamma$ *be given as above, and fix a number* $1/2 < \gamma < 1$. *Then there exists a family of subsets* $\mathcal{G}(y) \subseteq \mathbb{R}$ *such that for any fixed* $\mu < \nu$,

$$\lim_{y \to 0^+} m\big([\mu, \nu] \cap \mathcal{G}(y)\big) = \nu - \mu$$

*(*$m =$ *Lebesgue measure), and for any bounded, continuous, and* $\Gamma$*-invariant function* $f : \mathcal{H} \to \mathbb{C}$,

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta f(x + iy)\, dx \to \frac{1}{\mu(\Gamma \setminus \mathcal{H})} \int_{\Gamma \setminus \mathcal{H}} f(z)\, d\mu(z)$$

*uniformly as* $y \to 0^+$ *so long as* $\alpha \in \mathcal{G}(y)$ *(or* $\beta \in \mathcal{G}(y)$*) and* $\beta - \alpha \geqq y^\gamma$.

A result of similar nature can also be deduced from [DM, Theorem 3] (cf. [St, Remark 5.2.5]).

(II) Our result on the uniform equidistribution of horocycles can be applied to obtain an asymptotic formula for the counting function $\mathfrak{C}^k_{\mu\nu}(X)$ (cf. (2.12)) as $X \to \infty$.

THEOREM 7.2
*Let $\Gamma$ be given as above; let $k \in \{1, \ldots, \kappa\}$ and $\delta > 0$. We then have*

$$\frac{1}{X^2(\nu - \mu)} \mathfrak{C}^k_{\mu\nu}(X) \longrightarrow \frac{1}{\pi\ \mu(\Gamma \setminus \mathscr{H})}$$

*uniformly as $X \to \infty$ so long as $\nu - \mu \geqq X^{-1+\delta}$.*

This was proved in [St, §4.6] as a consequence of our main theorem, Theorem 1.

We remark that Theorem 7.2 can be interpreted as a statement concerning the distribution of the cusps of $\Gamma$ along the boundary $\partial\mathscr{H} = \mathbb{R} \cup \{\infty\}$ of $\mathscr{H}$. To see this, we associate to each cusp $\eta \in \partial\mathscr{H}$ the unique horoball $\mathfrak{B} = \mathfrak{B}_\eta$ which is tangent to $\partial\mathscr{H}$ at $\eta$ and for which $\Gamma_\eta \backslash \mathfrak{B}$ has hyperbolic area 1. (This means that $\mathfrak{B} = UN_k^{-1}\{z \in \mathscr{H} \mid \operatorname{Im} z \geqq 1\}$ for any $U \in \Gamma$, $k \in \{1, \ldots, \kappa\}$ such that $\eta = U(\eta_k)$.) We then have

$$\mathfrak{C}^k_{\mu\nu}(X) = \sharp\big\{\eta \in [\mu, \nu] \,\big|\, \eta \text{ is a cusp equivalent to } \eta_k, \\ \text{and } \mathfrak{B}_\eta \text{ intersects the line } \operatorname{Im} z = X^{-2}\big\}.$$

(III) Another application concerns the value distribution of the generalized theta sum

$$\Theta_f(x + iy) = y^{1/4} \sum_{n \in \mathbb{Z}} f(ny^{1/2}) e(n^2 x),$$

which was studied by Marklof in [Mar1]. Using our results on subsegments of closed horocycles, we are able to give a more uniform version of one of the main theorems in [Mar1], Theorem 7.1, as follows.

Let $f$ be a function from $\mathbb{R}$ to $\mathbb{C}$ which satisfies $f(x) = O((|x| + 1)^{-\eta})$ for some $\eta > 1$ and which is Riemann-integrable on every bounded interval. Let $\mathscr{B} \subset \mathbb{C}$ be an open convex set containing zero and with smooth boundary, and let, for $w \in \mathbb{C}$, $R > 0$,

$$\mathscr{B}(w, R) = \big\{Rz + w \,\big|\, z \in \mathscr{B}\big\}.$$

Let $\Psi(R) = \Psi_{\mathscr{B}}(w, R)$ be defined as in [Mar1, Theorem 7.1]. Then $\Psi$ is an increasing function from $\mathbb{R}^+$ to $[0, 1]$ with $\lim_{R\to\infty} \Psi(R) = 1$, and $\Psi$ is uniquely determined by $f, \mathscr{B}, w$. In the following, we keep $f, \mathscr{B}, w$ fixed.

THEOREM 7.3
*There exists a countable subset $\mathfrak{E} \subset \mathbb{R}^+$ such that for each $\delta > 0$, and for each $R > 0$ outside $\mathfrak{E}$, we have (with $m$ = Lebesgue measure)*

$$\frac{1}{\beta - \alpha} m\Big(\big\{x \in [\alpha, \beta] \,\big|\, \Theta_f(x + iy) \in \mathscr{B}(w, R)\big\}\Big) \to \Psi(R) \tag{7.1}$$

*uniformly as $y \to 0^+$ so long as $\beta - \alpha \geqq y^{1/2-\delta}$.*

Notice here that [Mar1, Theorem 7.1] corresponds to keeping $\alpha = 0$, $\beta = 1$ in Theorem 7.3. Notice also that Theorem 7.3 implies that (7.1) actually holds for each $R > 0$ which is a point of continuity of $\Psi(R)$.

Theorem 7.3 may in particular be applied to the classical theta sum $S_N(x) = \sum_{n=1}^{N} e(n^2 x)$ (cf. [Mar1, page 152]). We then obtain a strengthened version of the uniform central limit theorem for $S_N(x)$ which was proved by Jurkat and van Horne in [JV]: *There is a decreasing function* $\Phi : [0, \infty) \to [0, 1]$ *such that for each* $\lambda \geqq 0$ *which is a point of continuity of* $\Phi$*, we have*

$$\frac{1}{\beta - \alpha}\, m\big(\{x \in [\alpha, \beta] \,\big|\, N^{-1/2} |S_N(x)| \geqq \lambda\}\big) \to \Phi(\lambda)$$

*uniformly as* $N \to \infty$ *so long as* $\beta - \alpha \geqq N^{-1+2\delta}$. (In [JV] this was proved in the case of fixed $\alpha, \beta$.)

*Sketch of the proof of Theorem 7.3*

In [Mar1], $\Theta_f$ is identified (for functions $f$ of Schwarz class) as the restriction of a function living on a space $\mathscr{M}$ which is a 4-fold cover of $\Gamma_1(4)\backslash T_1\mathscr{H}$. Our proof of Theorem 2 in §3 can easily be carried over to the case of $[\alpha, \beta]$-segments of closed horocycles in $\mathscr{M}$. Hence we obtain $[\alpha, \beta]$-versions of [Mar1, Proposition 4.3, Corollary 4.4, Theorem 5.3]. (We have to keep $\sigma \leqq 1/2$ in the $[\alpha, \beta]$-version of [Mar1, Proposition 4.3]). Our goal is to prove an $[\alpha, \beta]$-version of [Mar1, Theorem 7.1], and the only step in the proof in [Mar1] which does not carry over immediately is the following. (Without loss of generalization, we take $f$ even.) Given $\varepsilon > 0$ and an even Schwartz function $f_\varepsilon$ on $\mathbb{R}$ such that $\int_{-\infty}^{\infty} |f(t) - f_\varepsilon(t)|^2\, dt < \varepsilon$, we need to show

$$J = \frac{1}{\beta - \alpha} \int_\alpha^\beta \big|\Theta_f(x + iy) - \Theta_{f_\varepsilon}(x + iy)\big|^2\, dx < 3\varepsilon \tag{7.2}$$

for all sufficiently small $y$ and all $\alpha, \beta$ such that $\beta - \alpha \geqq y^{1/2-\delta}$ (cf. [Mar1, (86)]). But $J$ can be expanded as follows (writing $g = f - f_\varepsilon$):

$$\begin{aligned} J = {} & y^{1/2}\Big(|g(0)|^2 + 2\sum_{n \neq 0} |g(ny^{1/2})|^2\Big) \\ & + \frac{y^{1/2}}{\beta - \alpha} \sum_{\substack{n,m \in \mathbb{Z} \\ n \neq \pm m}} g(ny^{1/2})\overline{g(my^{1/2})} \int_\alpha^\beta e\big((n^2 - m^2)x\big)\, dx. \end{aligned} \tag{7.3}$$

Here the expression on the first line converges to $2\int_{-\infty}^{\infty} |f(t) - f_\varepsilon(t)|^2\, dt < 2\varepsilon$ as $y \to 0$. Using $|g(x)| \leqq O_{f,f_\varepsilon}((|x|+1)^{-\eta})$ (with $\eta > 1$) and $\big|\int_\alpha^\beta e((n^2 - m^2)x)\, dx\big| \leqq \min(\beta - \alpha, |n^2 - m^2|^{-1})$, and careful summation, the expression in the second line of (7.3) is seen to be, for $0 < y < 1$ and $\alpha < \beta$,

$$O_{f,f_\varepsilon}\Big(\frac{y^{1/2}}{\beta-\alpha}\Big)\bigg(\Big(\log\frac{2}{y}\Big)^2+\log^+\Big(\frac{1}{\beta-\alpha}\Big)\bigg).$$

This tends to zero as $y \to 0$, $\beta - \alpha \geqq y^{1/2-\delta}$, and we thus obtain (7.2). □

*Acknowledgment.* Theorem 1 in this paper is the main result in Chapter 5 of my Ph.D. thesis [St]. I am grateful to my advisor Dennis Hejhal for suggesting this topic and for many helpful and inspiring discussions. I also thank Jens Marklof for valuable discussions and Nimish Shah and Scott A. Wolpert for insightful comments about the paper.

## References

[A] A. ALVAREZ-PARRILLA, *Asymptotic relations among Fourier coefficients of real-analytic Eisenstein series*, Trans. Amer. Math. Soc. **352** (2000), 5563 – 5582. MR 1675233

[B1] C. B. BALOGH, *Uniform asymptotic expansions of the modified Bessel function of the third kind of large imaginary order*, Bull. Amer. Math. Soc. **72** (1966), 40 – 43. MR 0188504

[B2] ———, *Asymptotic expansions of the modified Bessel function of the third kind of imaginary order*, SIAM J. Appl. Math. **15** (1967), 1315 – 1323. MR 0222354

[CFS] I. P. CORNFELD, S. V. FOMIN, and YA. G. SINAĬ, *Ergodic Theory*, Grundlehren Math. Wiss. **245**, Springer, New York, 1982. MR 0832433

[D1] S. G. DANI, *Invariant measures of horospherical flows on noncompact homogeneous spaces*, Invent. Math. **47** (1978), 101 – 138. MR 0578655

[D2] ———, *Invariant measures and minimal sets of horospherical flows*, Invent. Math. **64** (1981), 357 – 385. MR 0629475

[DM] S. G. DANI and G. A. MARGULIS, "Limit distributions of orbits of unipotent flows and values of quadratic forms" in *I. M. Gel'fand Seminar*, Adv. Soviet Math. **16**, Part I, Amer. Math. Soc., Providence, 1993, 91 – 137. MR 1237827

[EMOT] A. ERDÉLYI, W. MAGNUS, F. OBERHETTINGER, and F. G. TRICOMI, *Higher Transcendental Functions, Vol. II*, McGraw-Hill, New York, 1953. MR 0058756

[EM] A. ESKIN and C. MCMULLEN, *Mixing, counting, and equidistribution in Lie groups*, Duke Math. J. **71** (1993), 181 – 209. MR 1230290

[F] W. FELLER, *An Introduction to Probability Theory and Its Applications, Vol. II*, Wiley, New York, 1966. MR 0210154

[FF] L. FLAMINIO and G. FORNI, *Invariant distributions and time averages for horocycle flows*, Duke Math. J. **119** (2003), 465 – 526. MR 2003124

[H] J. L. HAFNER, *Some remarks on odd Maass wave forms (and a correction to "Zeros of L-functions attached to Maass forms"* [by C. Epstein, J. L. Hafner, and P. Sarnak, Math. Z. **190** (1985), 113 – 128; MR 0793354]), Math. Z. **196** (1987), 129 – 132. MR 0907415

[He1] D. A. HEJHAL, *The Selberg Trace Formula for* PSL(2, $\mathbb{R}$)*, Vol. 2*, Lecture Notes in Math. **1001**, Springer, Berlin, 1983. MR 0711197

[He2] ———, "On value distribution properties of automorphic functions along closed horocycles" in *XVIth Rolf Nevanlinna Colloquium (Joensuu, Finland, 1995)*, de Gruyter, Berlin, 1996, 39 – 52. MR 1427069

[He3] ———, "On the uniform equidistribution of long closed horocycles" in *Loo-Keng Hua: A Great Mathematician of the Twentieth Century*, Asian J. Math. **4**, Int. Press, Somerville, Mass., 2000, 839 – 853. MR 1870662

[I] H. IWANIEC, *Introduction to the Spectral Theory of Automorphic Forms,* Bibl. Rev. Mat. Iberoamericana, Rev. Mat. Iberoamericana, Madrid, 1995. MR 1325466

[JV] W. B. JURKAT and J. W. VAN HORNE, *The uniform central limit theorem for theta sums*, Duke Math. J. **50** (1983), 649 – 666. MR 0714822

[L] A. LEUTBECHER, *Über die Heckeschen Gruppen* $\mathbb{G}(\lambda)$, Abh. Math. Sem. Univ. Hamburg **31** (1967), 199 – 205. MR 0228438

[M] A. MANNING, "Dynamics of geodesic and horocycle flows on surfaces of constant negative curvature" in *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces (Trieste, Italy, 1989)*, Oxford Sci. Publ., Oxford Univ. Press, New York, 1991, 71 – 91. MR 1130173

[Ma] G. A. MARGULIS, *Discrete Subgroups of Semisimple Lie Groups*, Ergeb. Math. Grenzgeb (3) **17**, Springer, Berlin, 1991. MR 1090825

[Mar1] J. MARKLOF, *Limit theorems for theta sums*, Duke Math. J. **97** (1999), 127 – 153. MR 1682276

[Mar2] ———, "Theta sums, Eisenstein series, and the semiclassical dynamics of a precessing spin" in *Emerging Applications of Number Theory (Minneapolis, 1996)*, IMA Vol. Math. Appl. **109**, Springer, New York, 1999, 405 – 450. MR 1691543

[Mi] T. MIYAKE, *Modular Forms*, Springer, Berlin, 1989. MR 1021004

[P] S. J. PATTERSON, *Diophantine approximation in Fuchsian groups*, Philos. Trans. Roy. Soc. London Ser. A **282** (1976), 527 – 563. MR 0568140

[R1] M. RATNER, *On Raghunathan's measure conjecture*, Ann. of Math. (2) **134** (1991), 545 – 607. MR 1135878

[R2] ———, *Raghunathan's topological conjecture and distributions of unipotent flows*, Duke Math. J. **63** (1991), 235 – 280. MR 1106945

[R3] ———, *Raghunathan's conjectures for* SL(2, $\mathbb{R}$), Israel J. Math. **80** (1992), 1 – 31. MR 1248925

[S] P. SARNAK, *Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series*, Comm. Pure Appl. Math. **34** (1981), 719 – 739. MR 0634284

[Sh] N. SHAH, *Limit distributions of expanding translates of certain orbits on homogeneous spaces*, Proc. Indian Acad. Sci. Math. Sci. **106** (1996), 105 – 125. MR 1403756

[Shi] H. SHIMIZU, *On discontinuous groups operating on the product of the upper half planes*, Ann. of Math. (2) **77** (1963), 33 – 71. MR 0145106

[Shim] G. SHIMURA, *Introduction to the Arithmetic Theory of Automorphic Forms*, Kanô Memorial Lectures **1**, Publ. Math. Soc. Japan **11**, Iwanami Shoten, Tokyo; Princeton Univ. Press, 1971. MR 0314766

[St] A. STRÖMBERGSSON, "Some results on the uniform equidistribution of long closed horocycles" in *Studies in the Analytic and Spectral Theory of Automorphic Forms*, Ph.D. thesis, Uppsala University, Uppsala, Sweden, 2001, 137 – 226. http://www.math.uu.se/∼astrombe/papers.html

[W] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge Univ. Press, Cambridge, 1944. MR 0010746

[Wo1] S. A. WOLPERT, *Semiclassical limits for the hyperbolic plane*, Duke Math. J. **108** (2001), 449 – 509. MR 1838659

[Wo2] ———, *Asymptotic relations among Fourier coefficients of automorphic eigenfunctions*, Trans. Amer. Math. Soc. **356** (2004), 427 – 456. MR 2022706

[Z] D. ZAGIER, "Eisenstein series and the Riemann zeta function" in *Automorphic Forms, Representation Theory and Arithmetic (Bombay, 1979)*, Tata Inst. Fund. Res. Studies in Math. **10**, Tata Inst. Fundamental Res., Bombay, 1981, 275 – 301. MR 0633666

Department of Mathematics, Uppsala University, Box 480, S-75106 Uppsala, Sweden; astrombe@math.uu.se