ON THE DEVIATION OF ERGODIC AVERAGES FOR HOROCYCLE FLOWS

ANDREAS STRÖMBERGSSON

ABSTRACT. We give effective bounds on the deviation of ergodic averages for the horocycle flow on the unit tangent bundle of a non-compact hyperbolic surface of finite area. The bounds depend on the small eigenvalues of the Laplacian and on the rate of excursion into cusps for the geodesic corresponding to the given initial point. We also prove Ω-results which show that in a certain sense our bounds are essentially the best possible, for any given initial point.

1. INTRODUCTION

Let G denote the group $\mathrm{PSL}(2,\mathbb{R})$ and let Γ be a lattice in G. This means that Γ is a discrete subgroup of G and the measure ν on the quotient space $\Gamma\setminus G$ derived from the Haar measure on G is finite. We assume that ν is normalized, i.e. $\nu(\Gamma\setminus G)=1$.

The geodesic and the horocycle flows on $\Gamma\setminus G$ are defined by

$$g_t(\Gamma g)=\Gamma g\begin{pmatrix}e^{t/2}&0\\0&e^{-t/2}\end{pmatrix} \tag{1.1}$$

$$h_t(\Gamma g)=\Gamma g\begin{pmatrix}1&t\\0&1\end{pmatrix} \qquad (g\in G,\ t\in\mathbb{R}).$$

It was proved by Dani and Smillie [D, DS] that for each point $p=\Gamma g\in\Gamma\setminus G$ which does not belong to a closed orbit of the horocycle flow, the orbit $\{h_t(p)\mid 0\leqq t\leqq T\}$ becomes asymptotically equidistributed on $\Gamma\setminus G$ as $T\to\infty$. In other words, for any such p and any bounded continuous function $f:\Gamma\setminus G\to\mathbb{R}$, the ergodic averages satisfy

$$\frac{1}{T}\int_0^T f(h_t(p))\,dt\to\langle f\rangle:=\int_{\Gamma\setminus G}f\,d\nu,\qquad\text{as } T\to\infty. \tag{1.2}$$

This result has later been vastly generalized by Ratner [R2] to the case of an arbitrary unipotent flow on a general homogeneous space. As has been pointed out by Margulis and others (cf., e.g., [Ma]), an important (and difficult) open problem is to prove *effective* bounds on the rate of convergence in Ratner's general result.

For $G=\mathrm{PSL}(2,\mathbb{R})$ and Γ cocompact, such effective bounds were obtained by Burger in [Bur]. For Γ cocompact there are no closed horocycles on $\Gamma\setminus G$,

2000 *Mathematics Subject Classification.* Primary 37D40; Secondary 30F35.

and the bound proved in [Bur] for the rate of convergence in (1.2) is *uniform* with respect to $p \in \Gamma \backslash G$.

In the present paper, we treat the case of *non-cocompact* (but cofinite) $\Gamma \subset G = \mathrm{PSL}(2,\mathbb{R})$. In this case the dynamics of the horocycle flow is more complicated than for cocompact Γ, due to the presence of closed horocycles. Since the closed horocycles form a dense set in $\Gamma \backslash G$, it is clear that the rate of convergence in (1.2) must be highly sensitive to the choice of p.

Our main result is Theorem 1 below. This theorem gives an effective version of (1.2) for a restricted class of functions f; it gives a bound on the deviation of the ergodic average which depends on the small eigenvalues of the Laplacian and on the excursion rate of the geodesic $g_t(p)$ as $t \to \infty$. We also prove that in a certain sense this bound is essentially the best possible, for any given initial point (see Theorem 2).

To state Theorem 1, we let $\mathbb{H}$ be the Poincaré upper half-plane, with $G = \mathrm{PSL}(2,\mathbb{R})$ acting on $\mathbb{H}$ in the usual way, and let $\mathcal{M} = \Gamma\backslash\mathbb{H}$, a hyperbolic surface of finite area. If there exist small eigenvalues $\lambda \in (0, \frac{1}{4})$ in the discrete spectrum of the Laplacian on $\mathcal{M}$, then we let λ_1 be the smallest of these, and define $s_1 \in (\frac{1}{2}, 1)$ by $\lambda_1 = s_1(1-s_1)$; otherwise let $s_1 = \frac{1}{2}$. For each $j \in \{1, ..., \kappa\}$ we also let $\lambda_1^{(j)} \in [\lambda_1, \frac{1}{4})$ be the smallest positive eigenvalue for which there exists an eigenfunction which has non-zero constant term at the cusp η_j (i.e., $c_0^{(j)} \neq 0$ in (2.11) below), and define $s_1^{(j)} \in (\frac{1}{2}, 1)$ by $\lambda_1^{(j)} = s_1^{(j)}(1-s_1^{(j)})$; if no such $\lambda_1^{(j)}$ exists we let $s_1^{(j)} = \frac{1}{2}$. Note that by these definitions we have $\frac{1}{2} \leqq s_1^{(j)} \leqq s_1 < 1$ for each $j \in \{1, ..., \kappa\}$, and $s_1 = \frac{1}{2}$ if and only if $\Gamma \backslash \mathbb{H}$ admits no small eigenvalues.

We let $\pi : \Gamma \backslash G \to \mathcal{M}$ be the standard projection given by $\Gamma g \mapsto \Gamma g(i)$. (Then $\Gamma \backslash G$ is naturally identified with the unit tangent bundle of $\mathcal{M}$.) Let $\eta_1, ..., \eta_\kappa$ (where $\kappa \geqq 1$) be the inequivalent cusps of $\mathcal{M}$, and fix a neighborhood $\mathcal{C}_j \subset \Gamma \backslash G$ of each η_j in such a way that $\mathcal{C}_1, ..., \mathcal{C}_\kappa$ are pairwise disjoint. Fix a point $p_0 \in \mathcal{M}$. For $p \in \Gamma \backslash G$ we let $\mathrm{dist}(p)$ denote the hyperbolic distance from p_0 to $\pi(p)$ on $\mathcal{M}$.

For $f \in C^k(\Gamma \backslash G)$ we let $||f||_{W_k}$ be the Sobolev L^2 norm involving all the Lie derivatives of f up to the k:th order. We also introduce the following weighted supremum norm, for $f \in C(\Gamma \backslash G)$ and $\alpha \geqq 0$,

$$||f||_{N_\alpha} = \sup_{p \in \Gamma\backslash G} |f(p)| \cdot e^{-\alpha \cdot \mathrm{dist}(p)}. \tag{1.3}$$

Theorem 1. *Let $0 \leqq \alpha < \frac{1}{2}$. We then have, for all $p \in \Gamma \backslash G$, $T \geqq 10$, and all $f \in C^4(\Gamma \backslash G)$ such that $||f||_{W_4} < \infty$ and $||f||_{N_\alpha} < \infty$:*

$$\frac{1}{T}\int_0^T f(h_t(p))\,dt = \langle f \rangle + O\Big(||f||_{W_4}\Big)\Big\{r^{-\frac{1}{2}}\log^3(r+2) + r^{s_1^{(j)}-1} + T^{s_1-1}\Big\} + O\Big(||f||_{N_\alpha}\Big)\cdot r^{-\frac{1}{2}}, \tag{1.4}$$

where $r = r(p,T) = T \cdot e^{-\mathrm{dist}(g_{\log T}(p))}$, *and* $j = j(p,T)$ *is defined by* $g_{\log T}(p) \in \mathcal{C}_j$ *(if there is no such* j*, the term* $r^{s_1^{(j)}-1}$ *is omitted in the bound above). The implied constants depend only on* Γ, α *and* $p_0, \mathcal{C}_1, ..., \mathcal{C}_\kappa$.

We remark that the implied constants are *effective* in the sense that they can in principle be determined explicitly from the proof once Γ and the finite set of small eigenvalues $0 < \lambda < \frac{1}{4}$ on $\Gamma \setminus \mathbb{H}$ are known. Also, it is easy to see that for each $p \in \Gamma \setminus G$ for which the horocycle $\{h_t(p) \mid t \in \mathbb{R}\}$ is non-closed, one has $r = T \cdot e^{-\mathrm{dist}(g_{\log T}(p))} \to \infty$ as $T \to \infty$ (cf. (2.7) and Lemma 4.3 below). Hence Theorem 1 is indeed an effective version of (1.2).

The entity $r = T \cdot e^{-\mathrm{dist}(g_{\log T}(p))}$ is directly related to the asymptotic excursion rate of the geodesic $\{g_t(p)\}$, a concept which is well studied in the literature. For instance, let us define

$$\alpha_p = \limsup_{t \to \infty} \frac{\mathrm{dist}(g_t(p))}{t} \in [0,1]; \tag{1.5}$$

it then follows from Sullivan's logarithm law for geodesics [Su, §9], that $\alpha_p = 0$ holds for almost every point $p \in \Gamma \setminus G$ (with respect to the invariant volume measure). One even knows that for each $\alpha_0 \in [0,1]$ and each fiber $\pi^{-1}(q) \in \Gamma \setminus G$ (a one-dimensional circle), the subset $\{p \in \pi^{-1}(q) \mid \alpha_p \geqq \alpha_0\}$ has Hausdorff dimension $1 - \alpha_0$ (cf. [MP, Thm. 1]).

From these facts and Theorem 1 we see that for each point $p \in \Gamma \setminus G$ outside a "very small" set, the deviation of the ergodic average always decays like $O(T^{-\delta})$ as $T \to \infty$, for some $\delta > 0$. The next theorem shows that for each p, Theorem 1 gives the *optimal* exponent δ.

Let us define, for $j \in \{1, ..., \kappa\}$,

$$\alpha_{p,j} = \limsup_{\substack{t \to \infty \\ (g_t(p) \in \mathcal{C}_j)}} \frac{\mathrm{dist}(g_t(p))}{t}. \tag{1.6}$$

(We let $\alpha_{p,j} = 0$ if $g_t(p) \notin \mathcal{C}_j$ for all large t.) Note that $0 \leqq \alpha_{p,j} \leqq \alpha_p \leqq 1$.

Theorem 2. *Let* $p \in \Gamma \setminus G$ *be given, and let*

$$\delta_p = \min\big(1 - s_1,\ \min_j (1 - \alpha_{p,j})(1 - s_1^{(j)})\big).$$

Then for any fixed $\delta < \delta_p$*, and any fixed function* $f \in C^4(\Gamma \setminus G)$ *such that* $||f||_{W_4} < \infty$*, we have*

$$\frac{1}{T} \int_0^T f(h_t(p))\, dt = \langle f \rangle + O(T^{-\delta}), \qquad \text{as } T \to \infty. \tag{1.7}$$

On the other hand, there exists a function f *of the above type such that* (1.7) *does not hold for any* $\delta > \delta_p$.

In particular, by the logarithm law for geodesics, for almost every $p \in \Gamma \setminus G$ the exponent of optimal rate equals $\delta_p = 1 - s_1$. Also, in each fiber $\pi^{-1}(q)$, $\delta_p > 0$ holds for every $p \in \pi^{-1}(q)$ outside a set of Hausdorff dimension zero.

Remark 1.1. It is immediate from Theorem 1 that (1.7) holds for all $\delta < \delta_p$ and all functions $f \in C^4(\Gamma \setminus G)$ satisfying $||f||_{W_4} < \infty$ and $||f||_{N_\alpha} < \infty$ for some $\alpha < \frac{1}{2}$. Theorem 2 tells us that (1.7) even holds without the assumption $||f||_{N_\alpha} < \infty$. On the other hand, by studying special non-closed horocycles with $\delta_p = 0$ one can show that in Theorem 1 some assumption on f beyond $||f||_{W_k} < \infty$ *is* necessary, cf. Proposition 4.1.

The proof of Theorem 1 is based on an explicit identity for ergodic averages of the horocycle flow which was developed and used by Burger in [Bur]. However, we cannot use invariant norms on $L^2(\Gamma \setminus G)$ in the same direct way as was possible in [Bur]; instead we use Sobolev imbedding inequalities with explicit dependence on the point in $\Gamma \setminus G$. Extra care is required to treat initial points p with $\delta_p = 0$; for such points p we first make a careful splitting of the horocycle into several parts, and then deal with each part separately, cf. pp. 15–18. (This argument was inspired by Ratner, [R3, p. 20].)

The last statement in Theorem 2 is a consequence of more precise Ω-results which we prove in §4 and §5. The proofs in §5 involve use of the Fourier expansions of the individual eigenfunctions on $\Gamma \setminus \mathbb{H}$.

For Γ a congruence subgroup of $\mathrm{PSL}(2,\mathbb{Z})$, Theorem 2 allows a more explicit formulation, as we will now show. Recall that an irrational number $r \in \mathbb{R}$ is said to be of *(Diophantine) type* K if there exists a constant $C > 0$ such that $|r - m/n| > Cn^{-K}$ for all $m, n \in \mathbb{Z}$ with $(m,n) = 1$, $n > 0$. The smallest possible value of K is $K = 2$. Let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$ be a representative for a point $p \in \Gamma \setminus G$; then $\frac{a}{c} \in \mathbb{R} \cup \{\infty\} = \partial\mathbb{H}$ is the endpoint at infinity of the horocycle $\{h_t(g)\}$ (projected from G to $\mathbb{H}$). If Γ is any subgroup of finite index in $\mathrm{PSL}(2,\mathbb{Z})$ then the horocycle $\{h_t(p)\}$ is non-closed if and only if $c \neq 0$ and $\frac{a}{c}$ is irrational. For general Γ, there is a well-known correspondence between the excursion rate of the geodesic $g_t(p)$ and the well-approximability of $\frac{a}{c}$ by cusps of Γ, cf. [Su, MP, V2]; in particular, if Γ is any subgroup of finite index in $\mathrm{PSL}(2,\mathbb{Z})$ and $\{h_t(p)\}$ is non-closed, then

$$\alpha_p = 1 - 2/K_p,$$

where $K_p \in [2,\infty) \cup \{\infty\}$ is the infimum of all numbers $K \geqq 2$ such that $\frac{a}{c}$ is of type K. (This formula remains true if we define $K_p = \infty$ when $\{h_t(p)\}$ is closed.) Recall also that if Γ is a congruence subgroup of $\mathrm{PSL}(2,\mathbb{Z})$ then there is no residual spectrum on $\Gamma \setminus \mathbb{H}$, i.e. $s_1^{(j)} = \frac{1}{2}$ for all j. Hence:

Corollary 1. *If Γ is a congruence subgroup of PSL$(2,\mathbb{Z})$ then the optimal exponent δ_p in Theorem 2 is*

$$\delta_p = \min\bigl(1 - s_1,\ K_p^{-1}\bigr), \qquad \textit{for all } p \in \Gamma \setminus G.$$

In this connection, recall that the fundamental eigenvalue conjecture by Selberg is the statement that for each congruence subgroup Γ, there are no small eigenvalues $0 < \lambda < \frac{1}{4}$ on $\Gamma \setminus \mathbb{H}$. Hence, by the corollary, Selberg's conjecture is true if and only if, for each congruence subgroup Γ, there exists

at least one point $p \in \Gamma \setminus G$ such that $\delta_p = \frac{1}{2}$ (and if so, then $\delta_p = \frac{1}{2}$ holds for *almost all* $p \in \Gamma \setminus G$, and $\delta_p = K_p^{-1}$ holds for all $p \in \Gamma \setminus G$).

Finally, we should mention a closely related recent paper by Flaminio and Forni, [FF], in which a detailed study is made of the invariant distributions for the horocycle flow on $\Gamma \setminus G$, and the results are used to prove a more precise asymptotic version of Burger's result for cocompact Γ, and also, for non-cocompact Γ, an asymptotic formula for ergodic averages of the pull-backs under the geodesic flow of a given horocycle arc of finite length. It seems likely that an alternative proof of our Proposition 3.1 can be obtained based on [FF, §§5.2–5.4] together with Lemma 2.2, 2.3 below.

Another paper of related interest is Chernov [C], where an explicit bound (of the form $O\big(e^{-\alpha\sqrt{\log T}}\big)$ with $\alpha > 0$) is established for the deviation of ergodic averages for the horocycle flow on compact surfaces of (variable) negative curvature, as well as in more general cases.

This work was supported by a Swedish STINT Postdoctoral award. The method using Fourier series presented in §5 is based on earlier unpublished work supported by the European Commission under the Research Training Network (Mathematical Aspects of Quantum Chaos) HPRN-CT-2000-00103 of the IHP Programme. I would like to thank Alexander Bufetov, Livio Flaminio, Giovanni Forni, Dennis Hejhal, Jens Marklof, David Mieczkowiski, Hee Oh and Peter Sarnak for useful and inspiring discussions.

2. Decomposition of $L^2(\Gamma \setminus G)$ and Sobolev norms

We start by introducing necessary notation and recalling some basic facts regarding unitary representations of G, Sobolev norms, and the geometry of $\Gamma \setminus G$.

We let $G = \mathrm{PSL}(2,\mathbb{R})$, $\mathfrak{g} = \mathfrak{sl}(2,\mathbb{R})$. For $x \in \mathbb{R}$, $y > 0$, $\theta \in \mathbb{R}/\pi\mathbb{Z}$ we define the following elements in G:

$$\mathbf{n}(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \mathbf{a}(y) = \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}, \quad \mathbf{r}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Then each $g \in G$ has a unique factorization $g = \mathbf{n}(x)\,\mathbf{a}(y)\,\mathbf{r}(\theta)$ (Iwasawa decomposition). Using these coordinates, the bi-invariant Haar-measure dg on G is given by $dg = y^{-2}dxdyd\theta$.

We define $H, X_-, X_+ \in \mathfrak{g}$ by

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad X_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad X_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then the Casimir element in the universal enveloping algebra $U(\mathfrak{g})$ is

$$\Box = -\tfrac{1}{4}(H^2 + 2X_+X_- + 2X_-X_+). \tag{2.1}$$

If $(\mathcal{H}, \pi)$ is an irreducible unitary representation of $\mathrm{PSL}(2,\mathbb{R})$ then there is an orthonormal basis $\{\phi_n\}_{n\in\Sigma}$ in $\mathcal{H}$ such that Σ is a subset of the set $2\mathbb{Z}$

of even integers, each ϕ_n is a smooth vector, and $\pi(\mathbf{r}(\theta))\phi_n = e^{in\theta}\phi_n$ for all $n \in \Sigma$, $\theta \in \mathbb{R}$. The Casimir element acts as a scalar, $\pi(\Box) = \lambda \cdot \mathrm{Id}$, on the set of smooth vectors in $\mathcal{H}$. One knows that either $\lambda > 0$ and $\Sigma = 2\mathbb{Z}$ (then $(\mathcal{H}, \pi)$ is a representation of the principal series or the complementary series), or $\lambda = 0$ and $\Sigma = \{0\}$ (the trivial representation), or else $\lambda = \frac{m}{2}(1 - \frac{m}{2})$ for some even integer $m \geqq 2$, and $\Sigma = \{m, m+2, m+4, ...\}$ or $\Sigma = \{-m, -m-2, -m-4, ...\}$ (the discrete series).

Now let π be an arbitrary unitary representation of G on a (separable) Hilbert space $\mathcal{H}$. It is known that any such representation is unitarily equivalent to a direct integral

$$(\pi, \mathcal{H}) \cong \left(\int_{\mathsf{Z}}^{\oplus} \pi_\zeta \, d\nu(\zeta), \int_{\mathsf{Z}}^{\oplus} \mathcal{H}(\zeta) \, d\nu(\zeta) \right),$$

where Z is a locally compact Hausdorff space, ν is a positive Radon measure on Z, and for almost every $\zeta \in \mathsf{Z}$, π_ζ is an irreducible unitary representation of G in a separable Hilbert space $\mathcal{H}(\zeta)$. The Hilbert space $\int_{\mathsf{Z}}^{\oplus} \mathcal{H}(\zeta) \, d\nu(\zeta)$ is the L^2-space of all measurable functions f on Z with $f(\zeta) \in \mathcal{H}(\zeta)$ and $\int_{\mathsf{Z}} ||f(\zeta)||^2_{\mathcal{H}(\zeta)} \, d\nu(\zeta) < \infty$, and the action of $g \in G$ is given by $(\pi(g)f)(\zeta) = (\pi_\zeta(g))(f(\zeta))$. Cf., e.g., [M, §2.4] or [W, §§14.8, 14.9].

Using the notation introduced above for irreducible representations we define, for almost all ζ, $\lambda = \lambda(\zeta)$ so that $\pi_\zeta(\Box) = \lambda(\zeta) \cdot \mathrm{Id}$ in $\mathcal{H}(\zeta)$, $\Sigma = \Sigma(\zeta) \subset 2\mathbb{Z}$, and an orthonormal basis $\{\phi_n(\zeta)\}_{n \in \Sigma(\zeta)}$ in $\mathcal{H}(\zeta)$ consisting of smooth vectors $\phi_n = \phi_n(\zeta)$ satisfying $\pi_\zeta(\mathbf{r}(\theta))\phi_n = e^{in\theta}\phi_n$. We define $\phi_n(\zeta) = 0$ for all $n \in 2\mathbb{Z} - \Sigma(\zeta)$. Then $\lambda(\zeta)$ is a measurable function of ζ, and by [R1, Lemma 1.1] the basis elements ϕ_n may be chosen in such a way that $\phi_n(\zeta)$ is a measurable function of $\zeta \in \mathsf{Z}$ for each $n \in 2\mathbb{Z}$, and also so that a function f on Z with $f(\zeta) \in \mathcal{H}_\zeta$ is measurable if and only if the function $\mathsf{Z} \ni \zeta \mapsto \big\langle f(\zeta), \phi_n(\zeta) \big\rangle_{\mathcal{H}(\zeta)} \in \mathbb{C}$ is measurable for each $n \in 2\mathbb{Z}$.

We also define $s = s(\zeta)$ as the unique complex number such that $\lambda = s(1-s)$ and $\mathrm{Re}\, s \geqq \frac{1}{2}$, $\mathrm{Im}\, s \geqq 0$. We then have $s \in \frac{1}{2} + i\mathbb{R}_{\geqq 0}$ if $\lambda \geqq \frac{1}{4}$, $s \in (\frac{1}{2}, 1]$ if $0 \leqq \lambda < \frac{1}{4}$, and $s \in \mathbb{Z}^+$ if $\lambda \leqq 0$.

For $k \in \mathbb{Z}_{\geqq 0}$ we denote by $C^k(\pi)$ is the space of vectors $f \in \mathcal{H}$ such that the function $G \ni g \mapsto \pi(g)v \in \mathcal{H}$ is of class C^k. We let $\mathcal{H}^\infty = C^\infty(\pi)$ be the space of smooth vectors. This is a dense subspace in $\mathcal{H}$.

We next define the Sobolev norms which we will use. Fix any basis X_1, X_2, X_3 in $\mathfrak{g}$ and let $\Delta = -\sum_i X_i^2 \in U(\mathfrak{g})$. Then $\overline{\pi(\Delta)}$ (the closure of $\pi(\Delta)$) is a self-adjoint operator on $\mathcal{H}$ (cf. [N]). The (L^2) *Sobolev space* $W_k(\mathcal{H})$ of order $k \in \mathbb{R}^+$ is defined to be the domain of the operator $\big(I + \overline{\pi(\Delta)}\big)^{k/2}$. We define the *Sobolev norm* $||\cdot||_{W_k}$ on $W_k(\mathcal{H})$ by $||v||_{W_k} = ||\big(I + \overline{\pi(\Delta)}\big)^{k/2} v||$. The space $W_k(\mathcal{H})$ with norm $||\cdot||_{W_k}$ is in fact a Hilbert space, containing the space of smooth vectors $\mathcal{H}^\infty$ as a dense subspace.

A straightforward computation shows that the norm $||\cdot||_{W_k}$ is equivalent to the norm $||\cdot||_{W_k'}$ defined by

$$||v||_{W_k'}^2 = \int_{\mathsf{Z}} \sum_{n\in\Sigma(\zeta)} (1+n^2+|s(\zeta)|^2)^k \cdot \left|\left\langle v(\zeta), \phi_n(\zeta)\right\rangle_{\mathcal{H}(\zeta)}\right|^2 d\nu(\zeta), \tag{2.2}$$

and a vector $v \in \mathcal{H}$ belongs to $W_k(\mathcal{H})$ if and only if the right hand side above is finite. (The constants of equivalence between $||\cdot||_{W_k}$ and $||\cdot||_{W_k'}$ depend only on $k \in \mathbb{R}^+$ and the choice of basis X_1, X_2, X_3 in $\mathfrak{g}$.)

If k is an integer, then on the subspace $C^k(\pi) \subset W_k(\mathcal{H})$ the norm $||\cdot||_{W_k}$ is also equivalent to the norm $||\cdot||_{W_k''}$ defined by $||v||_{W_k''}^2 = \sum ||\pi(X_\alpha)v||^2$, where the sum runs over all monomials $X_\alpha = X_{i_1}X_{i_2}...X_{i_l} \in U(\mathfrak{g})$ of degree $\leqq k$. As usual, $\mathcal{H}^\infty$ is given the topology induced by the all the norms $||\cdot||_{W_k}$, $k \in \mathbb{Z}^+$. This makes $\mathcal{H}^\infty$ into a Fréchet space. (Cf., e.g., [W, Lemma 1.6.4].)

From now on we let $\Gamma \subset G = \mathrm{PSL}(2,\mathbb{R})$ be a cofinite Fuchsian group such that the hyperbolic surface $\mathcal{M} = \Gamma \setminus \mathbb{H}$ has at least one cusp.

Concerning the cusps and the fundamental domain, we will use the same notation as in [He, p. 268]. Specifically: we let $\mathcal{F} \subset \mathbb{H}$ be a canonical (closed) fundamental domain for $\Gamma \setminus \mathcal{H}$, and let $\eta_1, ..., \eta_\kappa$ (where $\kappa \geqq 1$) be the vertices of $\mathcal{F}$ along $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$. Since $\mathcal{F}$ is canonical, $\eta_1, ..., \eta_\kappa$ are $\Gamma-$inequivalent.

For each $j \in \{1, ..., \kappa\}$ we choose $\mathsf{N}_j \in G$ such that $\mathsf{N}_j(\eta_j) = \infty$ and such that the stabilizer Γ_{η_j} is $[\mathsf{T}_j]$, where $\mathsf{T}_j := \mathsf{N}_j^{-1} \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right) \mathsf{N}_j$. Since $\mathcal{F}$ is canonical, by modifying N_j we can also ensure that for all B large enough,

$$\begin{aligned} &\mathsf{N}_j(\mathcal{F}) \bigcap \{z \in \mathbb{H} \mid \mathrm{Im}\ z \geqq B\} \\ &= \{z \in \mathbb{H} \mid 0 \leqq \mathrm{Re}\ z \leqq 1,\ \mathrm{Im}\ z \geqq B\}. \end{aligned} \tag{2.3}$$

We recall the definition of the *invariant height function,* $\mathcal{Y}_\Gamma(z)$:

$$\mathcal{Y}_\Gamma(z) = \sup\Big\{ \mathrm{Im}\ \mathsf{N}_j W(z) \ \Big|\ j \in \{1, ..., \kappa\},\ W \in \Gamma \Big\}. \tag{2.4}$$

(Cf. [I, (3.8)].) This definition is in fact independent of the choice of $\mathcal{F}$ and of the maps N_j. One knows that the supremum in (2.4) is always attained for some j, W; we then write $j_\Gamma(z) = j$ (this makes $j_\Gamma(z)$ uniquely determined for each z with $\mathcal{Y}_\Gamma(z)$ large). The function $\mathcal{Y}_\Gamma(z)$ is well-known to be continuous and Γ-invariant; hence $\mathcal{Y}_\Gamma(\cdot)$ can be viewed as a function on $\mathcal{M}$.

In the lemmas below, we will also use $\mathcal{Y}_\Gamma(\cdot)$ and $j_\Gamma(\cdot)$ as functions on G or on $\Gamma \setminus G$, defined via composition with the standard projection maps $G \to \Gamma \setminus G \to \mathcal{M}$ (i.e. $g \mapsto \Gamma g$ and $\Gamma g \mapsto \Gamma g(i)$). Thus

$$\mathcal{Y}_\Gamma(g) = \sup\left\{ \frac{1}{c^2+d^2} \ \Big|\ \begin{pmatrix} * & * \\ c & d \end{pmatrix} = \mathsf{N}_j W g,\ j \in \{1, ..., \kappa\},\ W \in \Gamma \right\}.$$

We record the following inequalities for later use:

$$\mathcal{Y}_\Gamma(g\,\mathbf{a}(y)) \leqq \max(y,1/y)\cdot \mathcal{Y}_\Gamma(g) \qquad \forall y>0, \tag{2.5}$$

$$\mathcal{Y}_\Gamma(g\,\mathbf{n}(t)) \leqq (1+|t|)^2\cdot \mathcal{Y}_\Gamma(g) \qquad \forall t\in\mathbb{R}. \tag{2.6}$$

These follows from the inequalities $(c^2y+d^2/y)^{-1} \leqq \max(y,1/y)(c^2+d^2)^{-1}$ and $(c^2+(ct+d)^2)^{-1} \leqq (1+|t|)^2(c^2+d^2)^{-1}$, true for all $(c,d)\in\mathbb{R}^2\setminus\{(0,0)\}$. Furthermore, recalling the definition of the function $\mathrm{dist}(p)$ in the introduction, one easily checks that there are positive constants $C_1 = C_1(\Gamma,p_0)$, $C_2 = C_2(\Gamma,p_0)$ such that

$$C_1 e^{\mathrm{dist}(p)} \leqq \mathcal{Y}_\Gamma(p) \leqq C_2 e^{\mathrm{dist}(p)}, \qquad \forall p\in\mathcal{M}. \tag{2.7}$$

From now on in this paper, we will always let π denote the right regular representation of G on $\mathcal{H}=L^2(\Gamma\setminus G)$. This is a unitary representation. Since $\Gamma\setminus G$ is of finite volume, the direct integral decomposition $\mathcal{H}\cong\int_{\mathsf{Z}}^{\oplus}\pi_\zeta\,d\nu(\zeta)$ can in fact be constructed in such a way that Z is a disjoint union of three measurable subsets, $\mathsf{Z}=\mathsf{Z}_o\cup\mathsf{Z}_{ct}\cup\mathsf{Z}_{rs}$, such that the following holds (cf., e.g., [La], [Bo]):

(a) Writing $\mathcal{H}=\mathcal{H}_o\oplus\mathcal{H}_{ct}\oplus\mathcal{H}_{rs}$ for the corresponding decomposition of $\mathcal{H}$ as an orthogonal sum of closed subspaces, the space $\mathcal{H}_o$ coincides with the space ${}^oL^2(\Gamma\setminus G)$ of cuspidal elements in $L^2(\Gamma\setminus G)$.

(b) $\mathsf{Z}_o\cup\mathsf{Z}_{rs}$ is a discrete measure space, and we may thus assume $\nu(\{\zeta\})=1$ for all $\zeta\in\mathsf{Z}_o\cup\mathsf{Z}_{rs}$.

(c) For each $\zeta\in\mathsf{Z}_o$, π_ζ is nontrivial, and $\{\zeta\in\mathsf{Z}_o \mid |s(\zeta)|<S\}$ is finite for each $S>0$.

(d) Z_{rs} is finite. There is exactly one $\zeta\in\mathsf{Z}_{rs}$ such that π_ζ is the trivial representation, and for all other $\zeta\in\mathsf{Z}_{rs}$ we have $s(\zeta)\in(\frac12,1)$.

(e) For all $\zeta\in\mathsf{Z}_{ct}$, π_ζ is a principal series representation (and thus $s(\zeta)\in\frac12+i\mathbb{R}_{\geqq 0}$).

We can give more precise statements than (c) and (d) as follows: The set $\{\lambda(\zeta)\mid\zeta\in\mathsf{Z}_o, \mathrm{Re}\,s(\zeta)<1\}$ coincides (with multiplicities) with the set of cuspidal eigenvalues of the Laplace operator $-y^2(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2})$ on $\Gamma\setminus\mathbb{H}$; the set $\{\lambda(\zeta)\mid\zeta\in\mathsf{Z}_{rs}\}$ coincides with the set of residual eigenvalues of $-y^2(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2})$ on $\Gamma\setminus\mathbb{H}$; and for each $m\in 2\mathbb{Z}^+$ the number of elements $\zeta\in\mathsf{Z}_o$ with $s(\zeta)=m/2$ is equal to twice the dimension of the space of holomorphic cusp forms of weight m on $\Gamma\setminus\mathbb{H}$.

Note that by (b) above, we have $\mathcal{H}_o\cong\oplus_{\zeta\in\mathsf{Z}_o}\mathcal{H}(\zeta)$ and $\mathcal{H}_{rs}\cong\oplus_{\zeta\in\mathsf{Z}_{rs}}\mathcal{H}(\zeta)$, and in particular for each $\zeta\in\mathsf{Z}_o\cup\mathsf{Z}_{rs}$, $\mathcal{H}(\zeta)$ may be viewed as a closed subspace in $\mathcal{H}$. Note also that if $f\in\mathcal{H}^\infty$ and if f_o, f_{ct}, f_{rs} are the projections of f to $\mathcal{H}_o$, $\mathcal{H}_{ct}$ and $\mathcal{H}_{rs}$, then $f_o,f_{ct},f_{rs}\in\mathcal{H}^\infty$, and $||f||_{W_k}^2 = ||f_o||_{W_k}^2 + ||f_{ct}||_{W_k}^2 + ||f_{rs}||_{W_k}^2$ for any $k>0$. Furthermore, if $f\in\mathcal{H}^\infty$, then each projection of f to a subspace $\mathcal{H}(\zeta)$ ($\zeta\in\mathsf{Z}_o\cup\mathsf{Z}_{rs}$) remains in $\mathcal{H}^\infty$.

Lemma 2.1. *If $f \in W_2(\mathcal{H})$ then f is a continuous function on $\Gamma \setminus G$, and*

$$|f(p)| \ll_\Gamma ||f||_{W_2} \cdot \mathcal{Y}_\Gamma(p)^{\frac{1}{2}}, \qquad \forall p \in \Gamma \setminus G. \tag{2.8}$$

Proof. Cf. [FF, Lemma 5.3]. (Cf. also [BR, p. 349 Prop B.2]; notice that the function "$w(x)$" in [BR] is comparable with $\mathcal{Y}_\Gamma(x)$.) □

Lemma 2.2. *If $f \in W_3(\mathcal{H})$ and f is cuspidal then we also have the bound*

$$|f(p)| \ll_\Gamma ||f||_{W_3}, \qquad \forall p \in \Gamma \setminus G.$$

Proof. This is [BR, Prop 4.1]. As a preparation for the next lemma, we recall the proof from [BR, pp. 350–351]. In view of Lemma 2.1 and the density of $\mathcal{H}^\infty$ in $W_3(\mathcal{H})$ we may assume $f \in \mathcal{H}^\infty$. In particular, $d\pi(X)f$ and the pointwise (right) Lie derivative Xf are now the same, for all $X \in \mathfrak{g}$.

Also because of Lemma 2.1, we need only treat the case when p's projection onto $\Gamma\backslash\mathbb{H}$ lies in a cuspidal region $\mathcal{F} \cap \mathsf{N}_j^{-1}\{z \in \mathbb{H} \mid \text{Im } z \geqq B\}$, cf. (2.3). After an auxiliary conjugation, we may assume $\mathsf{N}_j = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, so that the cusp is $\eta_j = \infty$, and the stabilizer $\Gamma_\infty = [\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)]$. Then $\mathcal{Y}_\Gamma(p) = \text{Im } g(i) \geqq B$, for some representative $g \in G$ of p.

Let $g = \mathbf{n}(x)\,\mathbf{a}(y)\,\mathbf{r}(\theta)$ be the Iwasawa decomposition of g; then $y = \mathcal{Y}_\Gamma(p)$. Since f is cuspidal we have $\int_0^1 f(\mathbf{n}(u)g)\,du = 0$. Writing $\mathbf{n}_\theta(t) = \mathbf{r}(\theta)^{-1}\mathbf{n}(t)\,\mathbf{r}(\theta)$ and using $\mathbf{n}(u)\,\mathbf{a}(y) = \mathbf{a}(y)\,\mathbf{n}(u/y)$, we obtain

$$\int_0^{1/y} f\left(g\,\mathbf{n}_\theta(t)\right)\,dt = 0. \tag{2.9}$$

Now $\{\mathbf{n}_\theta(t) \mid t \in \mathbb{R}\}$ is a one-parameter subgroup in G generated by $X_\theta = \mathbf{r}(\theta)^{-1}X_+\,\mathbf{r}(\theta) \in \mathfrak{g}$. Clearly $||X_\theta f||_{W_2} \ll ||f||_{W_3}$, uniformly in θ. Notice also that $\mathcal{Y}_\Gamma(g\,\mathbf{n}_\theta(t)) = \mathcal{Y}_\Gamma(p)$, since $g\,\mathbf{n}_\theta(t) = \mathbf{n}(ty)g$ and $\text{Im } \mathbf{n}(ty)g(i) = \text{Im } g(i) \geqq B$. Hence by Lemma 2.1 applied to $X_\theta f$, we have for all $t \in \mathbb{R}$,

$$\left|\frac{d}{dt}f(g\mathbf{n}_\theta(t))\right| = \left|\left[X_\theta f\right](g\mathbf{n}_\theta(t))\right| \ll ||f||_{W_3} \cdot \mathcal{Y}_\Gamma(p)^{\frac{1}{2}}. \tag{2.10}$$

Clearly, the desired bound follows from $y = \mathcal{Y}_\Gamma(p)$ and (2.9), (2.10). (In fact, we even obtain the stronger bound $|f(p)| \ll_\Gamma ||f||_{W_3} \cdot \mathcal{Y}_\Gamma(p)^{-\frac{1}{2}}$.) □

In the next lemma we will prove a similar bound when $f \in \mathcal{H}_{rs}$. Assume $\zeta \in \mathsf{Z}_{rs}$ and $s = s(\zeta) \in (\frac{1}{2}, 1)$. Recall the definition on p. 6 of the orthonormal basis $\{\phi_n(\zeta)\}$ in $\mathbb{H}(\zeta)$. We write $\phi_n = \phi_n(\zeta)$. Now ϕ_0 can be viewed as a function ϕ on $\Gamma \setminus \mathbb{H}$ (since $\pi(\mathbf{r}(\theta))\phi_0 = \phi_0$ for all θ), and ϕ is an eigenfunction of the Laplace operator of eigenvalue $\lambda = s(1-s)$. Hence for each $j \in \{1, ..., \kappa\}$ we have a Fourier expansion [He, Ch. 6, §4]

$$\phi(z) = c_0^{(j)} y_j^{1-s} + \sum_{n \neq 0} c_n^{(j)} \sqrt{y_j} K_{s-\frac{1}{2}}(2\pi|n|y_j)\, e(nx_j), \tag{2.11}$$

where $x_j + iy_j := \mathsf{N}_j(z)$. (Since $\zeta \in \mathsf{Z}_{rs}$ we know that $c_0^{(j)} \neq 0$ for at least one j.) Note that if $c_0^{(j)} = 0$ for some j then $\int_0^1 \phi_0(\mathsf{N}_j^{-1}\mathbf{n}(u)g)\,du = 0$ for all $g \in G$, and then $\int_0^1 \big[\pi(X)\phi_0\big](\mathsf{N}_j^{-1}\mathbf{n}(u)g)\,du = 0$ for all $g \in G$ and all $X \in U(\mathfrak{g})$. Since the set $\big\{\pi(X)\phi_0\big\}$ spans $\mathcal{H}(\zeta)$ in Hilbert space sense, it follows that for *any* $f \in \mathcal{H}(\zeta)$, we have $\int_0^1 f(\mathsf{N}_j^{-1}\mathbf{n}(u)g)\,du = 0$ for almost all $g \in G$. (Cf. the proof of Prop. 8.2 in [Bo].)

Recall the definition of $j_\Gamma(\cdot)$, just below (2.4).

Lemma 2.3. *Let $\zeta \in \mathsf{Z}_{rs}$ and $s = s(\zeta)$ be as above. We then have for each $f \in W_3(\mathcal{H}) \cap \mathcal{H}(\zeta)$,*

$$|f(p)| \ll_\Gamma ||f||_{W_3} \cdot \mathcal{Y}_\Gamma(p)^{1-s}, \qquad \forall p \in \Gamma \setminus G.$$

Furthermore, if $c_0^{(j)} = 0$ for some $j \in \{1, ..., \kappa\}$ then for all $p \in \Gamma \setminus G$ with $j_\Gamma(p) = j$ we also have the stronger bound $\big|f(p)\big| \ll_\Gamma ||f||_{W_3}$.

Proof. The second statement follows from the proof of Lemma 2.2, in view of our remarks above. We now prove the first statement. Let $\phi_n = \phi_n(\zeta)$ be as above.

As before, we may assume $f \in \mathcal{H}^\infty \cap \mathcal{H}(\zeta)$, $\mathsf{N}_j = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and and $\mathcal{Y}_\Gamma(p) = \mathrm{Im}\, g_0(i) \geqq B$, where $g_0 \in G$ is a representative for p. Assume $f = \sum_{n\in 2\mathbb{Z}} d_n\phi_n$. Each ϕ_n belongs to $C^\infty(G)\cap L_2(\Gamma\backslash G)$ and satisfies $\phi_n(g\,\mathbf{r}(\theta)) = \phi_n(g)e^{in\theta}$ and $\square\phi_n = \lambda\phi_n$, where $\lambda = s(1-s) \in (0, \frac{1}{4})$. Now let $F_n(g) = \int_0^1 \phi_n(\mathbf{n}(u)\,g)\,du$. Then $\square\mathcal{F}_n = \lambda F_n$, and since $F_n(\mathbf{n}(x)\,\mathbf{a}(y)\,\mathbf{r}(\theta)) = F_n(\mathbf{a}(y))e^{in\theta}$ for all x, y, θ, and $\square = -y^2\big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\big) + y\frac{\partial^2}{\partial x\partial\theta}$ in these coordinates, we obtain $-y^2\frac{\partial^2}{\partial y^2}F_n(\mathbf{a}(y)) = \lambda F_n(\mathbf{a}(y))$, that is, $\frac{\partial}{\partial y}y^{2s}\frac{\partial}{\partial y}y^{-s}F_n(\mathbf{a}(y)) = 0$. Hence $F_n(\mathbf{a}(y)) = A_n y^s + A_n' y^{1-s}$ for some constants $A_n, A_n' \in \mathbb{C}$.

Using $\mathsf{N}_j = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, (2.3) and Cauchy's inequality, we now have

$$\begin{aligned} 1 = \int_{\Gamma\backslash G} |\phi_n(g)|^2\,dg &= \int_{\mathcal{F}}\int_{\mathbb{R}/\pi\mathbb{Z}} |\phi_n(\mathbf{n}(x)\,\mathbf{a}(y)\,\mathbf{r}(\theta))|^2\,d\theta\,\frac{dxdy}{y^2} \\ &\geqq \pi\int_B^\infty\int_0^1 |\phi_n(\mathbf{n}(x)\mathbf{a}(y))|^2\,dx\,\frac{dy}{y^2} \geqq \pi\int_B^\infty |F_n(\mathbf{a}(y))|^2\,\frac{dy}{y^2}. \end{aligned}$$

This forces $A_n = 0$ and $A_n' \ll_\Gamma 1$ (uniformly in n).

Now write $g_0 = \mathbf{n}(x)\mathbf{a}(y)\mathbf{r}(\theta)$ so that $y = \mathcal{Y}_\Gamma(p)$ and $F_n(g_0) = A_n' y^{1-s}e^{in\theta}$, and let $\mathbf{n}_\theta(t)$ be as in the proof of Lemma 2.2. We then have

$$\begin{aligned} y\Big|\int_0^{1/y} f(g_0\mathbf{n}_\theta(t))\,dt\Big| = \Big|\int_0^1 f(\mathbf{n}(u)g_0)\,du\Big| &= \Big|\sum_{n\in 2\mathbb{Z}} d_n F_n(g_0)\Big| \\ &\ll \sum_{n\in 2\mathbb{Z}} |d_n|\cdot y^{1-s} \ll ||f||_{W_1}\cdot y^{1-s} \end{aligned} \tag{2.12}$$

by Cauchy's inequality, since $\sum_{n\in 2\mathbb{Z}}(1+n^2)|d_n|^2 \ll ||f||_{W_1}^2$ by (2.2).

The proof is now completed as the proof of Lemma 2.2, using (2.10) and (2.12). □

3. Bounding the deviation of ergodic averages

As before, we let $\Gamma \subset G = \mathrm{PSL}(2,\mathbb{R})$ be a cofinite Fuchsian group such that $\mathcal{M} = \Gamma\backslash\mathbb{H}$ has at least one cusp, and we let π denote the representation of G on $\mathcal{H} = L^2(\Gamma \setminus G)$ given by right translations. For any $f \in \mathcal{H}$ we write (noting $\mathcal{H} \subset L^1(\Gamma \setminus G)$)

$$\langle f \rangle = \frac{1}{\mathrm{vol}(\Gamma \setminus G)} \int_{\Gamma\setminus G} f(g)\, dg.$$

This agrees with the definition in (1.2). Also recall the definition of s_1 and $s_1^{(j)}$ given in the introduction.

Proposition 3.1. *For all $f \in W_4(\mathcal{H})$, $p \in \Gamma \setminus G$ and $T \geqq 10$ we have, if $j = j_\Gamma(p\,\mathbf{a}(T))$,*

$$\frac{1}{T}\int_0^T f(p\mathbf{n}(t))\, dt = \langle f\rangle + O\Big(||f||_{W_4}\Big)\Big\{\Big(\frac{T}{\mathcal{Y}_\Gamma\big(p\,\mathbf{a}(T)\big)}\Big)^{-\frac{1}{2}} \cdot (\log T)^2 + \Big(\frac{T}{\mathcal{Y}_\Gamma\big(p\,\mathbf{a}(T)\big)}\Big)^{s_1^{(j)}-1} + T^{s_1-1}\Big\}. \tag{3.1}$$

The implied constant depends only on Γ.

Proof. For *fixed* p and T, notice that $\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\, dt$ depends continuously on $f \in W_4(\mathcal{H})$ with respect to the norm $||\cdot||_{W_4}$, by Lemma 2.1. The same is clearly true for $\langle f\rangle$, since $||\cdot||_{W_4}$ is stronger than $||\cdot||$. Hence by the density of $\mathcal{H}^\infty$ in $W_4(\mathcal{H})$, we may assume from start that $f \in \mathcal{H}^\infty$.

Using $\mathcal{H} = \mathcal{H}_o \oplus \mathcal{H}_{ct} \oplus \mathcal{H}_{rs}$ we may assume from start that $f \in \mathcal{H}_o$, $f \in \mathcal{H}_{ct}$ or $f \in \mathcal{H}_{rs}$. Let $\mathsf{Z}_o^+ = \{\zeta \in \mathsf{Z}_o \mid s(\zeta) \in (\frac{1}{2}, 1)\}$ and $\mathcal{H}_o^+ = \oplus_{\zeta\in\mathsf{Z}_o^+}\mathcal{H}(\zeta) \subset \mathcal{H}_o$, and let $\mathcal{H}_o^-$ be the orthogonal complement of $\mathcal{H}_o^+$ in $\mathcal{H}_o$. Since Z_o^+ and Z_{rs} are finite, we may in fact assume that one of the following holds: $f \in \mathcal{H}(\zeta)$ for some $\zeta \in \mathsf{Z}_o^+ \cup \mathsf{Z}_{rs}$, or $f \in \mathcal{H}_o^-$, or $f \in \mathcal{H}_{ct}$. (We still have $f \in \mathcal{H}^\infty$.)

In the case when f is a constant function (i.e., $f \in \mathcal{H}(\zeta)$ for the unique $\zeta \in \mathsf{Z}_{rs}$ with $s(\zeta) = 1$), (3.1) is trivial, and the error term vanishes. Hence, from now on, we may assume that f is orthogonal to the space of constant functions.

We will now recall the integral formula in [Bur] which lies at the heart of the proof of [Bur, Theorem 2][1]. We first have to recall the definition of some auxiliary intertwining operators. Fix a number $Y \geqq 1$ (we will later take $Y = T$). For each $y \geqq 1$ we define F_y and S_y to be the intertwining

[1]We modify Burger's formula to "$T^{-1}\int_0^T$" instead of "$(2T)^{-1}\int_{-T}^T$". Also, we correct a minor mistake in the case "$\alpha = \frac{1}{2}$" in [Bur, pp. 788(**), 790(2),(3), etc.].

operators $\mathcal{H} \to \mathcal{H}$ which are determined, via the integral decomposition of $\mathcal{H}$, by the following functions $\mathsf{Z} \to \mathbb{C}$ (we write $s = s(\zeta)$):

$$f_y(\zeta) = \begin{cases} \frac{sy^{s-1}-(1-s)y^{-s}}{2s-1}, & \text{if Re } s < 1, s \neq \frac{1}{2} \\ \frac{2+\log y}{2\sqrt{y}}, & \text{if } s = \frac{1}{2} \\ y^{-s}, & \text{if } s \in \mathbb{Z}^+, \end{cases}$$

$$s_y(\zeta) = \begin{cases} \frac{y^{s-1}-y^{-s}}{2s-1}, & \text{if Re } s < 1, s \neq \frac{1}{2} \\ \frac{\log y}{\sqrt{y}}, & \text{if } s = \frac{1}{2} \\ 0, & \text{if } s \in \mathbb{Z}^+. \end{cases}$$

Also, for each $y > 0$ we define T_y to be the intertwining operator $\mathcal{H} \to \mathcal{H}$ determined by

$$t_y(\zeta) = \begin{cases} s_y(\zeta), & \text{if Re } s < 1, y \geqq 1 \\ 0, & \text{if Re } s < 1, y < 1 \\ y^{s-1}\big(\frac{\max(1,y)^{1-2s} - Y^{1-2s}}{1-2s}\big), & \text{if } s \in \mathbb{Z}^+. \end{cases} \tag{3.2}$$

We record the following bounds (write $\mathsf{Z}_o^- = \mathsf{Z}_o - \mathsf{Z}_o^+$):

$$\sup_{\zeta \in \mathsf{Z}_o^- \cup \mathsf{Z}_{ct}} |f_y(\zeta)| \leqq \frac{2+\log y}{2\sqrt{y}}, \qquad \sup_{\zeta \in \mathsf{Z}_o^- \cup \mathsf{Z}_{ct}} |s_y(\zeta)| \leqq \frac{\log y}{\sqrt{y}} \quad (\forall y \geqq 1); \tag{3.3}$$

$$\sup_{\zeta \in \mathsf{Z}_o^- \cup \mathsf{Z}_{ct}} |t_y(\zeta)| \leqq \frac{2\log(y+10)}{\sqrt{y+1}} \qquad (\forall y \in (0, Y]).$$

These bounds are easy to verify from the definitions, using the fact that $\zeta \in \mathsf{Z}_o^- \cup \mathsf{Z}_{ct}$ implies either $s = \frac{1}{2} + it$ ($t \geqq 0$) or $s \in \mathbb{Z}^+$, and for $s = \frac{1}{2} + it$ ($t > 0$) we have $f_y(\zeta) = \frac{\cos(t \log y)}{\sqrt{y}} + \frac{\sin(t \log y)}{2t\sqrt{y}}$ and $s_y(\zeta) = y^{-s} \int_1^y x^{2s-2}\, dx$.

It follows from the bounds (3.3), and the fact that Z_o^+ and Z_{rs} are finite, that F_y, S_y and T_y are bounded operators $\mathcal{H} \to \mathcal{H}$ for each y, and also bounded operators $W_k(\mathcal{H}) \to W_k(\mathcal{H})$ for each $k > 0$, as well as continuous operators $\mathcal{H}^\infty \to \mathcal{H}^\infty$. One easily verifies that $t_y(\zeta)$ is continuous in y, uniformly with respect to ζ, viz., for each fixed $y_0 > 0$ we have $\sup_{\zeta \in \mathsf{Z}} |t_y(\zeta) - t_{y_0}(\zeta)| \to 0$ as $y \to y_0$. It follows from this that T_y is continuous in y with respect to the operator norm in each space $W_k(\mathcal{H})$ ($k > 0$): $||T_y - T_{y_0}||_{W_k} \to 0$ as $y \to y_0$.

Since our function $f \in \mathcal{H}^\infty$ is orthogonal to the constants, the integral formula from [Bur, Lemma 1 and pp. 790–791] now applies as follows:

$$\begin{aligned} \frac{1}{T}\int_0^T \pi(\mathbf{n}(t))f\, dt &= \frac{1}{T}\int_0^T \pi(\mathbf{n}(t)\mathbf{a}(Y))F_Y(f)\, dt \\ &\quad - \frac{1}{2T}\int_0^T \pi(\mathbf{n}(t)\mathbf{a}(Y))S_Y\big(d\pi(H)f\big)\, dt \\ &\quad + \frac{1}{T}\int_0^Y \big[1 - \pi(\mathbf{n}(T))\big]\pi(\mathbf{a}(y))T_y\big(d\pi(X_-)f\big)\, dy. \end{aligned} \tag{3.4}$$

The first three integrals are well-defined as integrals of $\mathcal{H}^\infty$-valued functions, since the integrands therein are continuous functions from $[0,T]$ to $\mathcal{H}^\infty$. Similarly, the last integrand is a continuous function from $(0,Y]$ to $\mathcal{H}^\infty$, and hence if we replace $\int_0^Y$ by $\int_\epsilon^Y$ for any $\epsilon > 0$, the last integral is well-defined in $\mathcal{H}^\infty$. Also, as $\epsilon \to 0$, the last integral certainly converges *in* $\mathcal{H}$, since the $||\cdot||$-norm of the integrand is uniformly bounded for $y \in (0,Y]$.

We first treat the case $f \in \mathcal{H}_{ct}$. Since $s \in \frac{1}{2} + i\mathbb{R}$ for all $\zeta \in Z_{ct}$, we may here replace $\int_0^Y$ by $\int_1^Y$ in the last line of (3.4), and all the integrals then converge in $\mathcal{H}^\infty$. For each fixed point $p \in \Gamma \setminus G$ the map $v \mapsto v(p)$ is a continuous linear functional on $\mathcal{H}^\infty$, by Lemma 2.1. Applying this functional to both sides of (3.4) we obtain

$$\begin{aligned}\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt &= \frac{1}{T}\int_0^T \big[F_Y f\big](p\,\mathbf{n}(t)\mathbf{a}(Y))\,dt \\ &- \frac{1}{2T}\int_0^T \big[S_Y d\pi(H) f\big](p\,\mathbf{n}(t)\mathbf{a}(Y))\,dt \\ &+ \frac{1}{T}\int_1^Y \Big(\big[T_y d\pi(X_-) f\big](p\,\mathbf{a}(y)) - \big[T_y d\pi(X_-) f\big](p\,\mathbf{n}(T)\mathbf{a}(y))\Big)\,dy.\end{aligned}$$

By Lemma 2.1 we have $|v(x)| \ll ||v||_{W_3} \mathcal{Y}_\Gamma(x)^{\frac{1}{2}}$ for all $v \in \mathcal{H}^\infty$ and all $x \in \Gamma \setminus G$. Using (3.3) we also have $||F_Y f||_{W_3} \leqq \frac{2+\log Y}{2\sqrt{Y}} ||f||_{W_3}$. Similarly, $||S_Y d\pi(H) f||_{W_3} \ll \frac{\log Y}{\sqrt{Y}} ||f||_{W_4}$ and $||T_y d\pi(X_-) f||_{W_3} \ll \frac{\log y}{\sqrt{y}} ||f||_{W_4}$ for all $y \geqq 1$ (cf. (3.2)).

Let us now take $Y = T$. Using the relation $\mathbf{n}(u)\mathbf{a}(y) = \mathbf{a}(y)\mathbf{n}(u/y)$ and (2.5), (2.6), we see that for all $t \in [0,T]$ and all $y \in [1,T]$:

$$\begin{aligned}&\mathcal{Y}_\Gamma(p\,\mathbf{n}(t)\mathbf{a}(Y)) = \mathcal{Y}_\Gamma(p\mathbf{a}(T)\mathbf{n}(t/T)) \leqq 4 \cdot \mathcal{Y}_\Gamma(p\,\mathbf{a}(T)); \\ &\mathcal{Y}_\Gamma(p\,\mathbf{n}(T)\mathbf{a}(y)) = \mathcal{Y}_\Gamma(p\,\mathbf{a}(T)\mathbf{n}(1)\mathbf{a}(y/T)) \leqq \frac{4T}{y} \cdot \mathcal{Y}_\Gamma(p\,\mathbf{a}(T)); \\ &\mathcal{Y}_\Gamma(p\,\mathbf{a}(y)) \leqq \frac{T}{y} \cdot \mathcal{Y}_\Gamma(p\,\mathbf{a}(T)).\end{aligned}$$

Using these inequalities we obtain

$$\begin{aligned}\left|\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt\right| &\ll \frac{1}{T}\int_0^T \frac{1+\log T}{\sqrt{T}} ||f||_{W_4} \cdot \sqrt{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}\,dt \\ &\quad + \frac{1}{T}\int_1^T \frac{\log y}{\sqrt{y}} \cdot ||f||_{W_4} \cdot \sqrt{\frac{T}{y} \cdot \mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}\,dy \\ &\ll ||f||_{W_4} \cdot \Big(\frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}\Big)^{-\frac{1}{2}} (\log T)^2.\end{aligned}$$

This completes the proof in the case $f \in \mathcal{H}_{ct}$.

Now assume $f \in \mathcal{H}(\zeta)$ for some $\zeta \in \mathsf{Z}_o^+ \cup \mathsf{Z}_{rs}$, and that f is orthogonal to the constant functions. Let $s = s(\zeta) \in (\frac{1}{2}, 1)$. In (3.4), the intertwining operator F_Y is now simply multiplication with $f_y(\zeta)$, and similarly for S_Y

and T_y. We may apply the above argument, with the only differences that now by Lemma 2.2 and Lemma 2.3, $|v(x)| \ll ||v||_{W_3} \mathcal{Y}_\Gamma(x)^{1-s}$ for all $v \in \mathcal{H}(\zeta) \cap \mathcal{H}^\infty$ and all $x \in \Gamma \setminus G$, and furthermore $||F_Y f||_{W_3} \ll_\Gamma Y^{s-1}||f||_{W_3}$, $||S_Y d\pi(H) f||_{W_3} \ll_\Gamma Y^{s-1}||f||_{W_4}$, and $||T_y d\pi(X_-) f||_{W_3} \ll_\Gamma y^{s-1}||f||_{W_4}$ for all $y \geqq 1$. We obtain

$$\left| \frac{1}{T} \int_0^T f(p\,\mathbf{n}(t))\, dt \right| \ll_\Gamma ||f||_{W_4} \Big(\frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))} \Big)^{s-1}. \tag{3.5}$$

This implies the desired bound whenever $\frac{1}{2} < s \leqq s_1^{(j)}$, since, writing $r = T/\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))$, we then have $r^{s-1} \leqq \max(r^{-1/2}, r^{s_1^{(j)}-1})$. Clearly, (3.5) also implies the desired bound whenever $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) \ll 1$, since $s \leqq s_1$.

Next, assume $s_1^{(j)} < s < 1$ and $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))$ large. By the definition of $s_1^{(j)}$ and the second bound in Lemma 2.3 (or by Lemma 2.2, if $\zeta \in \mathsf{Z}_o^+$) we have $v(x) \ll_\Gamma ||v||_{W_3}$ for all $v \in \mathcal{H}(\zeta) \cap \mathcal{H}^\infty$ and all $x \in \Gamma \setminus G$ with $j_\Gamma(x) = j = j_\Gamma(p\,\mathbf{a}(T))$.

Let us take $B > 1$ as in (2.3) so large that that the cuspidal regions $\mathsf{N}_k^{-1}\big([0,1] \times [B,\infty)\big) \subset \mathcal{F}$ are pairwise disjoint; clearly then $j_\Gamma(z) = j$ for all $z \in \mathbb{H}$ with $\operatorname{Im} \mathsf{N}_j(z) \geqq B$. We may assume $\operatorname{Im} \mathsf{N}_j g\,\mathbf{a}(T)(i) \geqq 4B$ for some representative $g \in G$ for p, since $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))$ is large and $j_\Gamma(p\,\mathbf{a}(T)) = j$. By the proof of (2.6) we then have $\operatorname{Im} \mathsf{N}_j g\,\mathbf{n}(t)\mathbf{a}(T)(i) = \operatorname{Im} \mathsf{N}_j g\,\mathbf{a}(T)\mathbf{n}(t/T)(i) \geqq B$ for all $t \in [0,T]$, and thus $v(p\,\mathbf{n}(t)\mathbf{a}(T)) \ll_\Gamma ||v||_{W_3}$ for all $t \in [0,T]$ and all $v \in \mathcal{H}(\zeta) \cap \mathcal{H}^\infty$.

Furthermore, if $\operatorname{Im} \mathsf{N}_j g\,\mathbf{n}(T)\mathbf{a}(y)(i) < B$ for some $y \in [1,T]$ then $y < T$ by what we have just noted, and we may find $y_0 \in (y,T]$ such that $\mathcal{Y}_\Gamma(g\,\mathbf{n}(T)\mathbf{a}(y_0)) = \operatorname{Im} \mathsf{N}_j g\,\mathbf{n}(T)\mathbf{a}(y_0)(i) = B$, and thus by (2.5), $\mathcal{Y}_\Gamma(g\,\mathbf{n}(T)\mathbf{a}(y)) \leqq B y_0/y \ll_\Gamma T/y$. Hence we conclude that $v(p\,\mathbf{n}(T)\mathbf{a}(y)) \ll_\Gamma (T/y)^{1-s}||v||_{W_3}$ for all $y \in [0,T]$ and $v \in \mathcal{H}(\zeta) \cap \mathcal{H}^\infty$. Similarly, $v(p\,\mathbf{a}(y)) \ll_\Gamma (T/y)^{1-s}||v||_{W_3}$. Using these inequalities and computing as above, we obtain

$$\left| \frac{1}{T} \int_0^T f(p\mathbf{n}(t))\, dt \right| \ll_\Gamma ||f||_{W_4} T^{s-1}. \tag{3.6}$$

This implies the desired bound, since $s \leqq s_1$.

Finally we treat the case $f \in \mathcal{H}_o^-$, by a similar argument as in [Bur, p. 791]: It follows from Lemma 2.2 that the supremum norm $N(f) = \sup_{p \in \Gamma\backslash G} |f(p)|$ is a well defined and continuous function on $\mathcal{H}_o^- \cap \mathcal{H}^\infty$. Write

$$v_\varepsilon = \frac{1}{T} \int_\varepsilon^Y \big[1 - \pi(\mathbf{n}(T))\big] \pi(\mathbf{a}(y)) T_y\big(d\pi(X_-) f\big)\, dy \in \mathcal{H}_o^- \cap \mathcal{H}^\infty;$$

then the last line in (3.4) is the same as $v_0 = \lim_{\varepsilon \to 0+} v_\varepsilon$ (limit in the norm $||\cdot||$). The norm N is clearly invariant under $\pi(g)$, for all $g \in G$, and hence

by Lemma 2.2 and (3.3),

$$N(v_\varepsilon) \leqq \frac{2}{T}\int_\varepsilon^Y N\big(T_y d\pi(X_-)f\big)\,dy \ll_\Gamma \frac{2}{T}\int_\varepsilon^Y ||T_y d\pi(X_-)f||_{W_3}\,dy$$
$$\ll \frac{\sqrt{Y}\log(Y+1)}{T}\cdot||f||_{W_4}.$$

Similar estimates also show that $\{v_{j^{-1}}\}_{j=1}^\infty$ is a Cauchy sequence with respect to the norm $N(\cdot)$. Hence there exists a function w in $C_b(\Gamma\setminus G)$, the space of bounded continuous functions on $\Gamma\setminus G$, such that $N(v_{j^{-1}}-w)\to 0$ as $j\to\infty$. But using the fact that $||v||\leqq\sqrt{\mathrm{vol}(\Gamma\setminus G)}\cdot N(v)$ for all $v\in C_b(\Gamma\setminus G)$, we see that we must have $w=v_0$, and hence $N(v_0)=\lim_{j\to\infty}N(v_{j^{-1}})\ll T^{-1}\sqrt{Y}\log(Y+1)\cdot||f||_{W_4}$.

The other integrals in (3.4) are dealt with more easily, since they are convergent in $\mathcal{H}^\infty$. We obtain

$$\left|\frac{1}{T}\int_0^T f(p\mathbf{n}(t))\,dt\right| \leqq N\Big(\frac{1}{T}\int_0^T \pi(\mathbf{n}(t))f\,dt\Big) \tag{3.7}$$
$$\ll \frac{1+\log Y}{\sqrt{Y}}\cdot||f||_{W_4}+\frac{\sqrt{Y}\log(Y+1)}{T}\cdot||f||_{W_4}.$$

Taking $Y=T$ again gives the desired result, and the proof is complete. □

Note that Proposition 3.1 does not imply the fact that *each* non-closed horocycle goes asymptotically equidistributed. The problem is that there exist non-closed horocycles for which $\big(T/\mathcal{Y}_\Gamma(p\mathbf{a}(T))\big)^{-\frac{1}{2}}(\log T)^2$ does not tend to 0. We will now prove Theorem 1 (cf. p. 2), which rectifies this problem, at the price of also having to use the weighted supremum norm $||\cdot||_{N_\alpha}$ in the bounds. The proof is carried out by splitting the long horocycle into several pieces, and applying Proposition 3.1 to each piece except possibly one. On the exceptional piece we instead use a supremum bound. We first prove a simple lemma.

Lemma 3.2. *For any given $g\in G$, $j\in\{1,...,\kappa\}$ and $W\in\Gamma$ we have*

$$\mathcal{Y}_\Gamma(g)\leqq\max\big(\mathrm{Im}\ \mathsf{N}_jWg(i),(\mathrm{Im}\ \mathsf{N}_jWg(i))^{-1}\big),$$

with equality whenever $\mathrm{Im}\ \mathsf{N}_jWg(i)\geqq 1$.

Proof. Writing $z=\mathsf{N}_jWg(i)$ we need to prove $\mathrm{Im}\ \mathsf{N}_{j'}W'g(i)\leqq\max(\mathrm{Im}\ z,(\mathrm{Im}\ z)^{-1})$, for any given $j'\in\{1,...,\kappa\}$ and $W'\in\Gamma$. Now if $U=\left(\begin{smallmatrix}*&*\\c&d\end{smallmatrix}\right)=\mathsf{N}_{j'}W'W^{-1}\mathsf{N}_j^{-1}$ we have either $|c|\geqq 1$ or $U=\left(\begin{smallmatrix}1&*\\0&1\end{smallmatrix}\right)$ (cf., e.g., [St, Lemma 2.3]) and hence either $\mathrm{Im}\ \mathsf{N}_{j'}W'g(i)=\mathrm{Im}\ U(z)=\frac{\mathrm{Im}\ z}{|cz+d|^2}\leqq|c|^{-2}(\mathrm{Im}\ z)^{-1}\leqq(\mathrm{Im}\ z)^{-1}$ or $\mathrm{Im}\ \mathsf{N}_{j'}W'g(i)=\mathrm{Im}\ U(z)=\mathrm{Im}\ z$. □

Proof of Theorem 1. Without loss of generalization we may take α close to $\frac{1}{2}$; in particular, we may assume that $\alpha>1-s(\zeta)$ for all $\zeta\in\mathsf{Z}_{rs}$.

Notice that by (1.1), (2.7), the variable r in (1.4) satisfies

$$C_1 \frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))} \leqq r \leqq C_2 \frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))} \tag{3.8}$$

for some constants $C_1 = C_1(\Gamma, p_0)$, $C_2 = C_2(\Gamma, p_0)$. By the proof of Proposition 3.1 (cf. (3.5), (3.6), (3.7)), it now follows that (1.4) holds whenever $f \in \mathcal{H}_o \oplus \mathcal{H}_{rs}$. Notice also that if f_0 denotes the projection of f to $\mathcal{H}_o \oplus \mathcal{H}_{rs}$, then $||f_0||_{N_\alpha} \ll_{\Gamma,\alpha} ||f||_{W_4}$ (and thus $||f - f_0||_{N_\alpha} \ll_{\Gamma,\alpha} ||f||_{W_4} + ||f||_{N_\alpha}$). This follows from Lemma 2.2, Lemma 2.3 and our assumption $\alpha > 1 - s(\zeta)$ for all $\zeta \in \mathsf{Z}_{rs}$. Because of these facts we may from now on assume that $f \in \mathcal{H}_{ct}$.

Let $p \in \Gamma \setminus G$ and $T \geqq 10$ be given. In view of Proposition 3.1, we may assume $T \geqq 10^{10}$ and $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) \geqq T^{9/10}$ from start. Now there is a representative $g \in G$ for p and some $j = j_\Gamma(p\,\mathbf{a}(T)) \in \{1, ..., \kappa\}$ such that $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) = \operatorname{Im} \mathsf{N}_j g \mathbf{a}(T)(i)$. Then

$$\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) = \frac{1}{c^2 T + d^2/T}, \qquad \text{where } \mathsf{N}_j g = \begin{pmatrix} * & * \\ c & d \end{pmatrix}. \tag{3.9}$$

We choose signs so that $c \geqq 0$. It then follows from $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) \geqq T^{9/10}$ that $c^2 T \leqq T^{-9/10}$, $d^2/T \leqq T^{-9/10}$ and thus $0 \leqq c \leqq T^{-19/20} < 10^{-9}$ and $|d| \leqq T^{1/20}$.

Below, we will make a specific choice of points $0 = \tau_0 < \tau_1 < ... < \tau_n = T$, for some $n \in \mathbb{Z}^+$. Writing $T_k = \tau_{k+1} - \tau_k$ we then have

$$\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt = \sum_{k=0}^{n-1} \frac{T_k}{T}\Big(\frac{1}{T_k}\int_0^{T_k} f(p\,\mathbf{n}(\tau_k + t))\,dt\Big). \tag{3.10}$$

Assuming $T_k \geqq 10$ for each k we may apply Proposition 3.1 to each term, obtaining

$$\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt = O\Big(||f||_{W_4}\Big) \sum_{k=0}^{n-1} \frac{T_k}{T} \cdot \sqrt{\frac{\mathcal{Y}_\Gamma\big(p\,\mathbf{n}(\tau_k)\mathbf{a}(T_k)\big)}{T_k}} \cdot (\log T_k)^2. \tag{3.11}$$

(Notice that $\langle f \rangle = 0$, and that the last two terms in (3.1) may be ignored, since $f \in \mathcal{H}_{ct}$.) Let us define $d_k = d + c\tau_k$, so that $\mathsf{N}_j g\,\mathbf{n}(\tau_k) = \left(\begin{smallmatrix} * & * \\ c & d_k \end{smallmatrix}\right)$. Then $\operatorname{Im} \mathsf{N}_j g \mathbf{n}(\tau_k)\mathbf{a}(T_k)(i) = (c^2 T_k + d_k^2/T_k)^{-1}$ and $c^2 T_k \leqq c^2 T \leqq 10^{-9}$. We will choose the sequence $\tau_0, ..., \tau_n$ in such a way that for all $k \in \{0, 1, ..., n-1\}$ except at most one k, we have

$$\tfrac{1}{5} \leqq d_k^2/T_k \leqq \tfrac{1}{2}. \tag{3.12}$$

Let us call the exceptional index k_0, if it exists. Write $\mathsf{M} = \{0, 1, ..., n-1\} \setminus \{k_0\}$ if k_0 exists, and otherwise $\mathsf{M} = \{0, 1, ..., n-1\}$. It then follows from Lemma 3.2 that $\mathcal{Y}_\Gamma(p\mathbf{n}(\tau_k)\mathbf{a}(T_k)) \leqq 5$ for all $k \in \mathsf{M}$. Notice also that for all

$k \in \mathsf{M}$ we have $T_k \leqq 5d_k^2 \leqq 5(|d| + cT)^2$. Hence we obtain from (3.10) and (3.11),

$$\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt = O\Big(||f||_{W_4}\Big)\cdot\frac{\log^2(|d|+cT+2)}{T}\cdot\sum_{k\in\mathsf{M}}|d_k|$$

$$+\Big[\text{If } k_0 \text{ exists: } \frac{1}{T}\int_0^{T_{k_0}} f(p\,\mathbf{n}(\tau_{k_0}+t))\,dt\Big]. \tag{3.13}$$

The last conditions which we impose on the sequence $0 = \tau_0 < \tau_1 < ... < \tau_n = T$ are the following:

$$\forall k \in \mathsf{M}: \qquad [d_k, d_{k+1}] \cap (-100, 100) = \emptyset; \tag{3.14}$$

$$\text{If } k_0 \text{ exists:} \qquad [d_{k_0}, d_{k_0+1}] \subset [-200, 200]. \tag{3.15}$$

Before giving the detailed verification that a sequence $\{\tau_k\}$ satisfying our conditions does indeed exist, we will show how to prove (1.4) using all our assumptions.

Notice that (3.12) implies $2cd_k^2 \leqq d_{k+1} - d_k \leqq 5cd_k^2$ for all $k \in \mathsf{M}$, and here $c|d_k| \leqq c(|d| + cT) \leqq 2T^{-9/10} < 10^{-8}$. Hence for all $k \in \mathsf{M}$, the numbers d_k and d_{k+1} have the same sign, and $|x| < 2|d_k|$ for all $x \in [d_k, d_{k+1}]$. Hence $\int_{d_k}^{d_{k+1}} |x|^{-1}\,dx > (d_{k+1} - d_k)/2|d_k| \geqq c|d_k|$ for all $k \in \mathsf{M}$. If $c > 0$ we may now conclude that the first error term in (3.13) is

$$\leqq O\Big(||f||_{W_4}\Big)\cdot\frac{\log^2(|d|+cT+2)}{cT}\cdot\int_{[d_0,d_n]\setminus(-100,100)}\frac{dx}{|x|}.$$

But $d_0 = d$ and $d_n = d + cT$. Notice that

$$\int_{[d,d+cT]\setminus(-100,100)}\frac{dx}{|x|} \ll \begin{cases} cT/|d| & \text{if } |d| > 2cT \\ \log(cT+2) & \text{if } |d| \leqq 2cT. \end{cases}$$

Hence, using (3.9), we see that the first error term in (3.13) is

$$\leqq O\Big(||f||_{W_4}\Big)\cdot\sqrt{\frac{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}{T}}\cdot\log^3\Big(\frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}+2\Big). \tag{3.16}$$

In the remaining case, $c = 0$, we have $d_k = d$ for all k, $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) = T/d^2$ by (3.9), and by adding $2d^2 \leqq T_k$ (cf. (3.12)) over $k \in \mathsf{M}$ we see that $|\mathsf{M}| \leqq T/2d^2$. Hence the first error term in (3.13) is bounded by (3.16) also when $c = 0$.

We now turn to the k_0-term in (3.13). Assuming that k_0 exists, we have by (3.15)

$$-200 \leqq d_{k_0} + ct \leqq 200 \qquad \text{for all } \ t \in [0, T_{k_0}].$$

Hence $\operatorname{Im}\,\mathsf{N}_j g\mathbf{n}(\tau_{k_0}+t)(i) = (c^2+(d_{k_0}+ct)^2)^{-1} \gg 1$ (for recall $0 \leqq c < 10^{-9}$), and by Lemma 3.2, $\mathcal{Y}_\Gamma(p\,\mathbf{n}(\tau_{k_0}+t)) \ll (c^2+(d_{k_0}+ct)^2)^{-1}$ for all $t \in [0, T_{k_0}]$. But by (1.3) and (2.7) we have $\big|f(p\,\mathbf{n}(\tau_{k_0}+t))\big| \ll_\Gamma ||f||_{N_\alpha}\cdot\mathcal{Y}_\Gamma(p\,\mathbf{n}(\tau_{k_0}+t))^\alpha$.

If $c>0$, then it follows that the k_0-term in (3.13) is

$$\begin{aligned}&\ll ||f||_{N_\alpha}\cdot\frac{1}{T}\int_0^{T_{k_0}}(c^2+(d_{k_0}+ct)^2)^{-\alpha}\,dt\\&\leqq\frac{||f||_{N_\alpha}}{T}\int_{-200/c}^{200/c}(c^2+(ct)^2)^{-\alpha}\,dt\ll\frac{||f||_{N_\alpha}}{cT},\end{aligned}$$

where we used $\alpha<\frac{1}{2}$. If $|d|<10^4cT$ then by (3.9) we obtain

$$\ll||f||_{N_\alpha}\sqrt{\frac{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}{T}},$$

and hence (1.4) holds.

The remaining case, $|d|\geqq 10^4cT$, is easy: In this case $d_{k_0}\in[d,d+cT]$ and (3.15) imply $|d+ct|\leqq 201$ for all $t\in[0,T]$ (in particular $|d|\leqq 201$). This gives $\mathcal{Y}_\Gamma(p\,\mathbf{n}(t))\ll|d|^{-2}$, by the same argument as above, and hence

$$|f(p\,\mathbf{n}(t))|\ll||f||_{N_\alpha}\cdot|d|^{-2\alpha}\ll||f||_{N_\alpha}\cdot|d|^{-1},\qquad\forall t\in[0,T].$$

We also have $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))\geqq T/2d^2$ by (3.9); hence (1.4) is true by simple inspection (without using (3.13)).

We now conclude the proof by showing that it is indeed possible to choose a sequence satisfying all the assumptions made above.

If $c=0$ and $|d|>100$ then one easily checks that we may take n as the largest integer $\leqq T/2d^2$ (thus $n\geqq 10^8$) and $\tau_k=kT/n$ for $k=0,1,...,n$. If $c=0$ and $|d|\leqq 100$ we may clearly take simply $n=1$, $\tau_0=0$, $\tau_1=T$. Otherwise, if $c>0$, it suffices to construct a sequence $\{d_k\}_{k=0}^n$ satisfying $d=d_0<d_1<...<d_n=d+cT$ and (3.14), (3.15) and

$$\forall k\in\mathsf{M}:\qquad 2cd_k^2\leqq d_{k+1}-d_k\leqq 5cd_k^2. \tag{3.17}$$

(For if we then define $\tau_k=(d_k-d)/c$ and $T_k=\tau_{k+1}-\tau_k$, we will have $\frac{1}{5}\leqq d_k^2/T_k\leqq\frac{1}{2}$, $\forall k\in\mathsf{M}$, because of (3.17), and also $T_k\geqq 10$, $\forall k\in\mathsf{M}$, because of (3.14), (3.17).)

The existence of such a sequence $\{d_k\}_{k=0}^n$ is now quite obvious, once we observe (in connection with (3.17)) that for any $x\in[d,d+cT]$ we have $5cx^2\leqq 5c(|d|+cT)|x|\leqq 10^{-8}|x|$.

For example, if $d\leqq -200$ and $d+cT\geqq 200$ we may define d_k recursively by $d_0=d$ and $d_{k+1}=d_k+2cd_k^2$ for $k=0,1,...$ until we get $d_k\in[-200,-199]$ for some k; then set $k_0=k$, $d_{k_0+1}=100$. We continue by letting (again) $d_{k+1}=d_k+2cd_k^2$ for $k=k_0+1,k_0+2,...$ until we obtain $d_k\leqq d+cT<d_{k+1}$ for some $k=n$. We may then redefine d_n as $d_n=d+cT$ (leaving $d_0,d_1,...,d_{n-1}$ intact); it is easy to verify that (3.17) remains true also for $k=n-1$, and that the sequence $\{d_k\}_{k=0}^n$ has all the desired properties. On the other hand, if $d\leqq -200$ and $-100\leqq d+cT<200$, we apply the above construction up until the definition of $d_{k_0}\in[-200,-199]$, and then simply let $d_{k_0+1}=d+cT$ and $n=k_0+1$, again obtaining a valid sequence $\{d_k\}_{k=0}^n$. Similar constructions can be made in all the remaining cases. $\square$

Remark 3.3. Note that Proposition 3.1 remains true if the left side in (3.1) is replaced by $\frac{1}{T}\int_{-T}^{0} f(p\,\mathbf{n}(t))\,dt$ and the right side is left unchanged. To see this we need merely apply Proposition 3.1 to the point $q = p\,\mathbf{n}(-T)$, and observe that $\mathcal{Y}_\Gamma(q\,\mathbf{a}(T)) = \mathcal{Y}_\Gamma(p\,\mathbf{a}(T)\mathbf{n}(-1)) = c\cdot\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))$ for some $\frac{1}{4} \leqq c \leqq 4$, by (2.6). A similar remark holds for Theorem 1.

Remark 3.4. Note that Proposition 3.1 and Theorem 1 apply also when $\{h_t(p)\}$ is a closed horocycle. In particular they can be used to derive stronger versions of the main theorem in [St], concerning subsegments of long closed horocycles. To see this, let us assume that $N_1 = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$, so that ∞ is a cusp with $\Gamma_\infty = \left[\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)\right]$, and let $p \in \Gamma\backslash G$ be the point represented by the element $\mathbf{n}(x)\mathbf{a}(y) \in G$, for some $x \in \mathbb{R}$, $0 < y < 0.1$. Then $\{h_t(p)\}$ is a closed horocycle of length $1/y$. For any $10 \leqq T \leqq 1/y$ we now have $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) = \mathcal{Y}_\Gamma(\mathbf{n}(x)\mathbf{a}(yT)) \leqq (yT)^{-1}$, by Lemma 3.2, and hence Theorem 1 gives (using (3.8))

$$\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt = \langle f\rangle + O\Big(||f||_{W_4}\Big)\Big\{T^{-1}y^{-\frac{1}{2}}\log^3(T^2y+2) + T^{2s_1'-2}y^{s_1'-1} + T^{s_1-1}\Big\} + O\Big(||f||_{N_\alpha}\Big)\cdot T^{-1}y^{-\frac{1}{2}},$$

and Proposition 3.1 gives

$$\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt = \langle f\rangle + O\Big(||f||_{W_4}\Big)\Big\{T^{-1}y^{-\frac{1}{2}}(\log T)^2 + T^{2s_1'-2}y^{s_1'-1} + T^{s_1-1}\Big\}.$$

Here $s_1' = \max_j s_1^{(j)}$. Both these estimates give effective versions of [St, Thm. 1], for given any $\delta > 0$ the error terms tend uniformly to 0 as $y \to 0$ if we keep $y^{-\frac{1}{2}-\delta} \leqq T \leqq y^{-1}$.

4. Ω-results due to cuspidal excursions

In this section we will prove various Ω-results for the deviation of ergodic averages for the horocycle flow, using comparatively elementary observations on the cuspidal excursions made by the horocycle orbit.

First, we give an example which shows that the norm $||\cdot||_{N_\alpha}$ in the right hand side of (1.4) in Theorem 1 cannot be replaced by a Sobolev norm of any order. For convenience, let us make a specific choice of the operator Δ in the definition of the Sobolev norm $||\cdot||_{W_k}$ on p. 6:

$$\Delta = -\tfrac{1}{4}\big(H^2 + 2(X_+)^2 + 2(X_-)^2\big) = \Box - \tfrac{1}{2}(X_+ - X_-)^2,$$

cf. (2.1). This operator acts as the Laplace operator $-y^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$ on any function $G \to \mathbb{C}$ which factors through the standard projection $G \to \mathbb{H}$.

Now fix a constant $\delta \in (\frac{1}{2}, 1)$, and let $F \in C^\infty(\mathbb{R}^+)$ be a smooth non-negative function such that $F(y) = 0$ for $0 < y \leqq 2$ and

$$F(y) = \sqrt{y}\,(\log y)^{-\delta} \qquad \text{for } y \geqq 3.$$

We then define a function $f \in C(\Gamma \setminus G)$ by

$$f(g) = F(\mathcal{Y}_\Gamma(g)).$$

Given any $g_0 \in G$ for which $f(g_0) > 0$, we have $\mathrm{Im}\, \mathsf{N}_j W g_0(i) = \mathcal{Y}_\Gamma(g_0) > 2$ for some $j \in \{1, ..., \kappa\}$ and $W \in \Gamma$. Hence by Lemma 3.2, $\mathcal{Y}_\Gamma(g) = \mathrm{Im}\, \mathsf{N}_j W g(i)$ holds for all g in some neighbourhood of g_0. It follows from this that f is smooth, and that $\Delta^k f(g) = F_k(\mathcal{Y}_\Gamma(g))$ for $k = 1, 2, ...$, where $F_k(y) = (-y^2 \frac{\partial^2}{\partial y^2})^k F(y)$. One checks by a quick computation that for each k we have $|F_k(y)| \ll_k \sqrt{y}\,(\log y)^{-\delta}$ for all $y \geqq 2$, and of course $F_k(y) = 0$ for $0 < y \leqq 2$. It now follows that

$$\int_{\Gamma\setminus G} |\Delta^k f(g)|^2\, dg < \infty \qquad \text{for each } \ k \in \mathbb{Z}^+,$$

as one verifies by splitting the fundamental region $\mathcal{F}$ into a compact part and κ cuspidal regions (cf. (2.3)), and using $\int_2^\infty y(\log y)^{-2\delta}\, dy/y^2 < \infty$. Hence $||f||_{W_k} < \infty$ for each k, and it also follows that $f \in \mathcal{H}^\infty$.

Note, however, that $||f||_{N_\alpha} = \infty$ for each $\alpha \in (0, \frac{1}{2})$.

Proposition 4.1. *Let $f \in C^\infty(G) \cap C(\Gamma \setminus G)$ be as above. Then there exists a point $p \in \Gamma \setminus G$ for which*

$$\limsup_{T\to\infty} \frac{1}{T} \int_0^T f(p\,\mathbf{n}(t))\, dt = \infty. \tag{4.1}$$

In particular, (4.1) implies that $\{p\,\mathbf{n}(t) \mid t \in \mathbb{R}\}$ is not a closed horocycle on $\Gamma \setminus G$, and hence by Theorem 1, for any function $f_1 \in C(\Gamma \setminus G) \cap C^4(G)$ such that $||f_1||_{W_4} < \infty$ and $||f_1||_{N_\alpha} < \infty$ for some $\alpha < \frac{1}{2}$, we have $\frac{1}{T}\int_0^T f_1(p\,\mathbf{n}(t))\, dt \to \langle f_1 \rangle$ as $T \to \infty$. Proposition 4.1 shows that the corresponding statement for f does *not* hold.

Proof. After an auxiliary conjugation we may assume that $\mathsf{N}_j = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ for some j, so that $\Gamma \setminus G$ has a cusp at ∞, and $\Gamma_\infty = \left[\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)\right]$.

It is well-known that the cusps equivalent to ∞ are dense on the real line (cf., e.g., [P]). In other words one knows that for any non-empty open interval $I \subset \mathbb{R}$ there is some $g \in \Gamma$ such that $g^{-1}(\infty) \in I$. Notice that $g^{-1}(\infty) = -\frac{d}{c}$ if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and that by Shimizu's lemma, $|c| \geqq 1$ (cf. [Sh, Lemma 4], or [Mi, Lemma 1.7.3]). We will use these facts in the construction below.

By definition $F(y) \geqq 0$ and $\int_4^\infty F(y)\, dy/y^{3/2} = \infty$. Hence there exists a decreasing function $h : \mathbb{R}^+ \to (0, \frac{1}{2})$ such that

$$\forall B > 0 :\ 0 < \ell \leqq h(B) \Longrightarrow \int_4^{\ell^{-2}} F(y)\, \frac{dy}{y^{3/2}} > B.$$

Of course, we necessarily have $\lim_{B\to\infty} h(B) = 0$.

We will now make a recursive definition of a sequence of elements $g_k = \left(\begin{smallmatrix} * & * \\ c_k & d_k \end{smallmatrix}\right) \in \Gamma$ and open non-empty intervals $(1, 2) \supseteq I_1 \supseteq I_2 \supseteq ....$ We first

take $g_1 = \left(\begin{smallmatrix} * & * \\ c_1 & d_1 \end{smallmatrix}\right) \in \Gamma$ arbitrary with $1 < g_1^{-1}(\infty) < 2$, and let

$$I_1 = \Big(1,2\Big) \bigcap \Big(g_1^{-1}(\infty),\, g_1^{-1}(\infty) + |c_1|^{-1} h(|d_1|)\Big).$$

Clearly then $(1,2) \supseteq I_1 \neq \emptyset$. For $k \geqq 2$, assuming that $g_1, ...g_{k-1}$ and $(1,2) \supseteq I_1 \supseteq ... \supseteq I_{k-1} \neq \emptyset$ have already been defined, we take $g_k = \left(\begin{smallmatrix} * & * \\ c_k & d_k \end{smallmatrix}\right) \in \Gamma$ arbitrary with $g_k^{-1}(\infty) \in I_{k-1}$, and then let

$$I_k = I_{k-1} \bigcap \Big(g_k^{-1}(\infty),\, g_k^{-1}(\infty) + |c_k|^{-1} h(k|d_k|)\Big).$$

Clearly then $(1,2) \supseteq I_1 \supseteq ... \supseteq I_{k-1} \supseteq I_k \neq \emptyset$, and the definition may be iterated indefinitely.

We have $|c_k| \geqq 1$ by Shimizu's lemma, for all k, and hence $|d_k| > 1$, since $-\frac{d_k}{c_k} = g_k^{-1}(\infty) \subseteq (1,2)$. We choose signs so that $c_k \geqq 1$ and $d_k < -1$. Hence $|I_k| = c_k^{-1} h(k|d_k|) \to 0$ as $k \to \infty$.

It follows that there is a unique point $\alpha \in [1,2]$ which belongs to the closure of each interval I_k. Now let $p \in \Gamma \setminus G$ be the point given by $g = \left(\begin{smallmatrix} \alpha & 0 \\ 1 & \alpha^{-1} \end{smallmatrix}\right) \in G$. We will prove that (4.1) holds for this point p.

Since f is Γ-invariant we have, for any k and any $T > 0$,

$$\frac{1}{T} \int_0^T f(p\,\mathbf{n}(t))\, dt = \frac{1}{T} \int_0^T f(g_k g\,\mathbf{n}(t))\, dt.$$

Let us define $\gamma_k = c_k \alpha + d_k$ and $\delta_k = d_k/\alpha$, so that

$$g_k g\,\mathbf{n}(t) = \begin{pmatrix} * & * \\ \gamma_k & \gamma_k t + \delta_k \end{pmatrix}.$$

Then $d_k \leqq \delta_k \leqq d_k/2 < -\frac{1}{2}$. It follows from our construction that the sequence of lower endpoints of the intervals $I_1, I_2, ...$ is strictly increasing, and hence α is larger than each of these. Hence by the definition of I_k we must have $g_k^{-1}(\infty) < \alpha \leqq g_k^{-1}(\infty) + c_k^{-1} h(k|d_k|)$ for all k, and thus $0 < \gamma_k \leqq h(k|d_k|) < \frac{1}{2}$. This implies that

$$\int_4^{\gamma_k^{-2}} F(y)\, \frac{dy}{y^{3/2}} > k|d_k|.$$

Taking $T = (-\delta_k + 1/2)/\gamma_k > 0$ and writing $T' = (-\delta_k - 1/2)/\gamma_k \in \big(0, T\big)$, we find by using Lemma 3.2 with $\mathsf{N}_j = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$,

$$\begin{aligned}
&\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt \geqq \frac{1}{T}\int_{T'}^T F(\mathcal{Y}_\Gamma(\left(\begin{smallmatrix} * & * \\ \gamma_k & \gamma_k t+\delta_k \end{smallmatrix}\right)))\,dt \\
&\quad = \frac{2}{T}\int_0^{1/2\gamma_k} F((\gamma_k^2 + (\gamma_k t)^2)^{-1})\,dt = \frac{1}{T\gamma_k}\int_{\gamma_k^2}^{\gamma_k^2+1/4} F(u^{-1})\,\frac{du}{\sqrt{u-\gamma_k^2}} \\
&\quad \geqq \frac{1}{T\gamma_k}\int_{\gamma_k^2}^{\gamma_k^2+1/4} F(u^{-1})\,\frac{du}{\sqrt{u}} \geqq \frac{1}{T\gamma_k}\int_4^{\gamma_k^{-2}} F(y)\frac{dy}{y^{3/2}} \\
&\quad > \frac{k|d_k|}{T\gamma_k} = \frac{k|d_k|}{|\delta_k|+1/2} \geqq \frac{k|d_k|}{|d_k|+1/2} > \frac{2k}{3} \to \infty,
\end{aligned}$$

as $k \to \infty$. We also have $T \to \infty$ as $k \to \infty$. Hence p satisfies (4.1). □

The next proposition shows that at least if there are no small eigenvalues present, then the error terms in Theorem 1 (and in Proposition 3.1) are in a certain sense close to being optimal, at least for one of the limits $T \to \infty$ and $T \to -\infty$ (recall Remark 3.3).

Proposition 4.2. *Given a continuous function $f \in C(\Gamma \setminus G)$ of compact support and with $\langle f\rangle \neq 0$, there exist positive constants $C_1 = C_1(\Gamma, f)$ and $C_2 = C_2(\Gamma, f)$ such that the following holds. For any $p \in \Gamma\setminus G$ and any $u_0 > C_2$ such that $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \leqq u_0$ and such that the function $u \mapsto \mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum at $u = u_0$, there exists some $T \in [u_0, 2u_0]$ such that either*

$$\Big|\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt - \langle f\rangle\Big| \geqq C_1\Big(\frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}\Big)^{-\frac{1}{2}} \tag{4.2}$$

or

$$\Big|\frac{1}{T}\int_{-T}^0 f(p\,\mathbf{n}(t))\,dt - \langle f\rangle\Big| \geqq C_1\Big(\frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}\Big)^{-\frac{1}{2}} \tag{4.3}$$

Before proving this proposition, let us clarify its role by noting that for any non-closed horocycle there exist arbitrarily large numbers u_0 satisfying the stated assumptions:

Lemma 4.3. *Given $p \in \Gamma \setminus G$ such that the horocycle $\{p\,\mathbf{n}(t)\}$ is non-closed, we have $\lim_{u\to\infty} u/\mathcal{Y}_\Gamma(p\,\mathbf{a}(u)) = \infty$, and there exist arbitrarily large numbers $u_0 > 0$ such that the function $u \mapsto \mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum at $u = u_0$.*

Proof. By [R3, Prop. 1.1]) there exists a compact subset $K \subset \Gamma \setminus G$ such that for any p for which the horocycle $\{p\,\mathbf{n}(t)\}$ is non-closed, the geodesic $\{p\,\mathbf{a}(u)\}$ keeps returning to K as $u \to \infty$. Hence $\lim_{u\to\infty} u/\mathcal{Y}_\Gamma(p\,\mathbf{a}(u)) = \infty$ (using (2.5)).

To prove the second statement, we may assume that $\mathcal{Y}_\Gamma(p\mathbf{a}(u))$ stays bounded for all $u > 0$ (for otherwise the desired statement follows directly using continuity and the fact that $\{p\,\mathbf{a}(u)\}$ keeps returning to K). Then

$$K_1 = \bigcap_{u_1>0} \overline{\{p\,\mathbf{a}(u) \mid u \geqq u_1\}}$$

is a compact non-empty subset of $\Gamma \setminus G$, and hence there is a point $q \in K_1$ such that $\mathcal{Y}_\Gamma(q) \geqq \mathcal{Y}_\Gamma(q')$ for all $q' \in K_1$. Clearly $q\,\mathbf{a}(t) \in K_1$ for all $t > 0$. Also, the function $t \mapsto \mathcal{Y}_\Gamma(q\,\mathbf{a}(t))$ is easily seen to be non-constant on any non-empty open interval $t \in I \subset \mathbb{R}^+$, for if I is of finite length then there exists a *finite* subset $\mathsf{M} \subset \mathrm{PSL}(2,\mathbb{R})$ (depending on Γ, q, I) such that

$$\mathcal{Y}_\Gamma(q\,\mathbf{a}(t)) = \max\bigl\{(c^2t + d^2/t)^{-1} \bigm| \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathsf{M}\bigr\}, \qquad \forall t \in I.$$

Hence we may fix some numbers $0 < t_1 < 1 < t_2$ such that $\mathcal{Y}_\Gamma(q\,\mathbf{a}(t_\ell)) < \mathcal{Y}_\Gamma(q)$ for $\ell = 1, 2$. By continuity, we now have

$$\mathcal{Y}_\Gamma(q'\,\mathbf{a}(t_\ell)) < \mathcal{Y}_\Gamma(q') \qquad (\ell = 1, 2) \tag{4.4}$$

for all points $q' \in \Gamma \setminus G$ lying sufficiently close to q. Hence by the definition of K_1 we see that for each $U > 0$ there exists some number $u_2 > U$ such that (4.4) holds for $q' = p\,\mathbf{a}(u_2)$. It then follows that $u \mapsto \mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum for some $u_0 \in [u_2t_1, u_2t_2]$. □

Proof of Proposition 4.2 Let $\mathsf{m}_\Gamma = \inf_{g\in G} \mathcal{Y}_\Gamma(g) > 0$. Since f has compact support on $\Gamma \setminus G$, we may fix $C_3 > 1$ such that $f(g) = 0$ whenever $\mathcal{Y}_\Gamma(g) \geqq C_3$. Let $C_2 = C_3/\mathsf{m}_\Gamma$.

Take p, u_0 such that the assumptions hold. Now we can find a representative $g \in G$ for p and some $j \in \{1, ..., \kappa\}$ such that

$$\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) = \mathrm{Im}\ \mathsf{N}_j g\mathbf{a}(u_0)(i) = \frac{1}{c^2u_0 + d^2/u_0}, \quad \mathsf{N}_j g = \begin{pmatrix} * & * \\ c & d \end{pmatrix},\ c \geqq 0.$$

By (2.4) we also have $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u)) \geqq (c^2u + d^2/u)^{-1}$ for *all* $u > 0$. Hence since $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum at $u = u_0$, we necessarily have

$$c > 0, \quad d \neq 0, \quad u_0 = \frac{|d|}{c}, \quad \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) = \frac{1}{2c|d|}.$$

Now $\mathsf{m}_\Gamma \leqq \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \leqq u_0$ and $u_0 > C_2 = C_3/\mathsf{m}_\Gamma$ imply $c < 1/\sqrt{2C_3}$ and $|d| \geqq 1/\sqrt{2}$. Notice that $\mathrm{Im}\ \mathsf{N}_j g\mathbf{n}(t)(i) = (c^2 + (ct + d)^2)^{-1} > ((2C_3)^{-1} + (ct + d)^2)^{-1}$, and hence for each $t \in \mathbb{R}$ with $|t + d/c| \leqq (c\sqrt{2C_3})^{-1}$ we have $\mathrm{Im}\ \mathsf{N}_j g\mathbf{n}(t)(i) > C_3$, so that $\mathcal{Y}_\Gamma(g\mathbf{n}(t)) > C_3$, and thus $f(g\mathbf{n}(t)) = 0$. Hence, writing

$$T_1 = u_0 = \frac{|d|}{c} \qquad \text{and} \qquad T_2 = u_0 + \frac{1}{c\sqrt{2C_3}},$$

we have $f(g\mathbf{n}(t)) = 0$ either for all $t \in [T_1, T_2]$ (if $d < 0$), or for all $t \in [-T_2, -T_1]$ (if $d > 0$). Let us assume $d < 0$; we then conclude $\int_0^{T_1} f(g\,\mathbf{n}(t))\,dt =$

$\int_0^{T_2} f(g\,\mathbf{n}(t))\,dt$, and hence

$$\frac{\langle f\rangle}{c\sqrt{2C_3}} = \Big(\int_0^{T_1} f(g\,\mathbf{n}(t))\,dt - T_1\langle f\rangle\Big) - \Big(\int_0^{T_2} f(g\,\mathbf{n}(t))\,dt - T_2\langle f\rangle\Big).$$

It follows that

$$\Big|\int_0^{T_\ell} f(g\,\mathbf{n}(t))\,dt - T_\ell\langle f\rangle\Big| \geqq \frac{|\langle f\rangle|}{2c\sqrt{2C_3}} \tag{4.5}$$

for at least one $\ell \in \{1,2\}$. But using $C_3 > 1$ and $|d| \geqq 1/\sqrt{2}$ one checks that $u_0 = T_1 < T_2 < 2u_0$. Hence by (2.5), we also have $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T_2)) \leqq 2\mathcal{Y}_\Gamma(p\,\mathbf{a}(T_1)) = (c|d|)^{-1}$. Hence, for the same ℓ as in (4.5),

$$\Big|\frac{1}{T_\ell}\int_0^{T_\ell} f(g\,\mathbf{n}(t))\,dt - \langle f\rangle\Big| \geqq \frac{\big|\langle f\rangle\big|}{4|d|\sqrt{2C_3}} \geqq \frac{\big|\langle f\rangle\big|}{4\sqrt{2C_3}}\cdot\Big(\frac{T_\ell}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T_\ell))}\Big)^{-\frac12}.$$

This means that (4.2) holds with $C_1 = \frac{\big|\langle f\rangle\big|}{4\sqrt{2C_3}}$ and $T = T_\ell \in [u_0, 2u_0]$. In the other case, $d > 0$, exactly the same argument leads to (4.3). □

It seems that these elementary methods do not allow us to prove a similar lower bound *separately* for the two cases $T \to \infty$ and $T \to -\infty$; such a lower bound (only slightly weaker by a logarithm factor) will be obtained in Proposition 5.1 below using more difficult methods.

However, studying only cuspidal excursions we can at least obtain a $\Omega(T^{-\frac12})$-result for $T \to \infty$, as is seen in the next proposition.

Proposition 4.4. *Given a continuous function $f \in C(\Gamma\backslash G)$ of compact support and with $\langle f\rangle \neq 0$, there exists a positive constant $C_1 = C_1(\Gamma, f)$ such that for any $p \in \Gamma\backslash G$ for which $\{p\,\mathbf{n}(t)\}$ is a non-closed horocycle, there is a sequence $1 < T_1 < T_2 < \ldots$ with $\lim_{k\to\infty} T_k = \infty$ such that*

$$\Big|\frac{1}{T}\int_0^{T} f(p\,\mathbf{n}(t))\,dt - \langle f\rangle\Big| \geqq C_1 T^{-\frac12} \tag{4.6}$$

for each $T = T_k$.

Proof. We first introduce some new notation. We write $\mu(g) := a^2+b^2+c^2+d^2 \geqq 2$ for $g = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in G$. We also write $\mu_j(g) := \inf_{n\in\mathbb{Z}} \mu(g\mathsf{T}_j^n)$ (recall $\Gamma_{\eta_j} = [\mathsf{T}_j]$, cf. p. 7). Let A be the linear fractional map $A(z) = (z-i)/(z+i)$ which maps $\mathbb{H}$ onto the unit disk. We define a metric ρ on $\partial\mathbb{H} = \mathbb{R}\cup\{\infty\}$ by $\rho(z,w) := |A(z) - A(w)|$ for $z, w \in \partial\mathbb{H}$ (here $A(z), A(w)$ lie on the unit circle). Note that $\rho(g(z), g(w)) \ll_g \rho(z,w)$ for all $g \in G$, $z, w \in \partial\mathbb{H}$. We agree to write $z \prec w$ to indicate that $z, w \in \partial\mathbb{H}$, $\rho(z,w) < \frac{1}{10}$ and $A(w)$ is obtained from $A(z)$ by a short rotation in the *positive* direction along the unit circle. Note that if $w \in \mathbb{R}$ lies sufficiently close to a fixed point $z \in \mathbb{R}$, then $z \prec w \iff z < w$.

By Patterson, [P, p. 545, Thm. 1], there exists a constant $C_\Gamma > 0$ such that for all $\alpha \in \partial\mathbb{H}$ and all $X \geqq 2$ there are some $j \in \{1, \ldots, \kappa\}$ and $\gamma \in \Gamma$

such that

$$\mu_j(\gamma) \leqq X \quad \text{and} \quad \rho(\gamma(\eta_j), \alpha) \leqq \frac{C_\Gamma}{\sqrt{\mu_j(\gamma) X}}. \tag{4.7}$$

By the same theorem there also exists a constant $C'_\Gamma > 0$ such that for all $j, j' \in \{1, ..., \kappa\}$ and $\gamma, \gamma' \in \Gamma$ one has

$$\gamma'(\eta_{j'}) \neq \gamma(\eta_j) \Longrightarrow \rho(\gamma'(\eta_{j'}), \gamma(\eta_j)) > \frac{C'_\Gamma}{\sqrt{\mu_{j'}(\gamma')\mu_j(\gamma)}}. \tag{4.8}$$

(In [P, p. 545, Thm. 1] these two facts are stated with μ in place of μ_j, $\mu_{j'}$, but this is clearly equivalent to (4.7), (4.8), since $\gamma \mathsf{T}_j^n(\eta_j) = \gamma(\eta_j)$ for all n.)

Now fix $p \in \Gamma \setminus G$, let $g_0 \in G$ be a representative for p, and let $\alpha = g_0(\infty) \in \partial\mathbb{H}$. We assume that $\{p\,\mathbf{n}(t) \mid t \in \mathbb{R}\}$ is not a closed horocycle. This means that α is not a cusp, i.e. $\gamma(\eta_j) \neq \alpha$ for all $\gamma \in \Gamma$, $j \in \{1, ..., \kappa\}$. We wish to prove that there exist good approximations $\gamma(\eta_j)$ to α with $\gamma(\eta_j) \prec \alpha$. More precisely, we claim that there exists a constant $C''_\Gamma > 0$ such that for any given $X_0 > 0$ there exist some $X \geqq X_0$, $j \in \{1, ..., \kappa\}$ and $\gamma \in \Gamma$ such that

$$\mu_j(\gamma) \leqq X; \quad \gamma(\eta_j) \prec \alpha; \quad \text{and} \quad \rho(\gamma(\eta_j), \alpha) \leqq \frac{C''_\Gamma}{\sqrt{\mu_j(\gamma) X}}. \tag{4.9}$$

To prove this, let us assume from start that $X_0 > 10^4(1+C_\Gamma^2)\cdot\max(1, 1/C'_\Gamma)$; this ensures that all points on $\partial\mathbb{H}$ considered in the argument below lie close to each other in the ρ-metric, so that the relation $\prec$ is well defined. To start with, we choose $j \in \{1, ..., \kappa\}$ and $\gamma \in \Gamma$ so that (4.7) holds with X_0 in place of X, and so that $\mu_j(\gamma)$ is minimal with respect to this property. (This is possible since the set $\{\gamma \in \Gamma/[\mathsf{T}_j] \mid \mu_j(\gamma) \leqq X_0\}$ is finite for each j, by a compactness argument.) After making a proper choice of $X \geqq X_0$, we now have:

$$\mu_j(\gamma) \leqq X; \quad \rho(\gamma(\eta_j), \alpha) = \frac{C_\Gamma}{\sqrt{\mu_j(\gamma) X}} \quad \text{(equality!)} \tag{4.10}$$

$$\textit{and } \forall j', \forall \gamma' \in \Gamma : \ \mu_{j'}(\gamma') < \mu_j(\gamma) \Longrightarrow \rho(\gamma'(\eta_{j'}), \alpha) > \frac{C_\Gamma}{\sqrt{\mu_{j'}(\gamma') X}}.$$

If $\gamma(\eta_j) \prec \alpha$ then we are done; (4.9) holds with $C''_\Gamma = C_\Gamma$. Now assume $\alpha \prec \gamma(\eta_j)$. We then let $\alpha' \in \partial\mathbb{H}$ be the unique point satisfying $\rho(\alpha', \alpha) = \frac{C_\Gamma^2+1}{C'_\Gamma X}$ and $\alpha' \prec \alpha$. By (4.7), there exist $j' \in \{1, ..., \kappa\}$ and $\gamma' \in \Gamma$ such that $\mu_{j'}(\gamma') \leqq X$ and

$$\rho(\gamma'(\eta_{j'}), \alpha') \leqq \frac{C_\Gamma}{\sqrt{\mu_{j'}(\gamma') X}}. \tag{4.11}$$

Now $\gamma'(\eta_{j'}) \prec \gamma(\eta_j)$, for otherwise $\alpha' \prec \alpha \prec \gamma(\eta_j) \preceq \gamma'(\eta_{j'})$ and thus $\rho(\gamma(\eta_j), \alpha) < \rho(\gamma'(\eta_{j'}), \alpha')$, which by (4.10), (4.11) leads to $\mu_{j'}(\gamma') < \mu_j(\gamma)$, and in view of the second line of (4.10) and $\rho(\gamma'(\eta_{j'}), \alpha') > \rho(\gamma'(\eta_{j'}), \alpha)$ this leads to a contradiction against (4.11).

Furthermore, if $\alpha \prec \gamma'(\eta_{j'})$ then $\alpha \prec \gamma'(\eta_{j'}) \prec \gamma(\eta_j)$ so that $\rho(\gamma'(\eta_{j'}), \gamma(\eta_j)) < \rho(\alpha, \gamma(\eta_j))$, and using (4.8) and (4.10) we get $\mu_{j'}(\gamma') > (C'_\Gamma/C_\Gamma)^2 \cdot X$, and hence, via (4.11),

$$\frac{C_\Gamma^2+1}{C'_\Gamma X} = \rho(\alpha', \alpha) < \rho(\alpha', \gamma'(\eta_{j'})) \leqq \frac{C_\Gamma}{\sqrt{\mu_{j'}(\gamma')X}} < \frac{C_\Gamma^2}{C'_\Gamma \cdot X},$$

which is a contradiction. Hence $\gamma'(\eta_{j'}) \prec \alpha$ must hold. We have

$$\rho(\gamma'(\eta_{j'}), \alpha) \leqq \rho(\gamma'(\eta_{j'}), \alpha') + \rho(\alpha', \alpha) \leqq \Big(C_\Gamma + \frac{C_\Gamma^2+1}{C'_\Gamma}\Big)\frac{1}{\sqrt{\mu_{j'}(\gamma')X}}.$$

Hence (4.9) holds for j', γ', with $C''_\Gamma = \big(C_\Gamma + \frac{C_\Gamma^2+1}{C'_\Gamma}\big)$.

To reformulate (4.9), note that (using [Le, p. 105 (Ex. 2)]) we may assume that the representative g_0 for the fixed point $p \in \Gamma \setminus G$ has been chosen in such a way that $\alpha \in \mathbb{R}$, $|\alpha| \leqq 1$ and $|\alpha - \eta_j| > (2\kappa)^{-1}$ for all $j \in \{1, ..., \kappa\}$. Then $\mathsf{N}_j(\alpha) \neq \infty$, and whenever $\rho(\gamma(\eta_j), \alpha)$ is sufficiently small we have $\mathsf{N}_j(\gamma(\eta_j)) \neq \infty$ and $\big|\mathsf{N}_j(\gamma(\eta_j)) - \mathsf{N}_j(\alpha)\big| \ll \rho(\gamma(\eta_j), \alpha)$, where the implied constant depends on Γ and N_j, but not on α; furthermore $\gamma(\eta_j) \prec \alpha$ implies $\mathsf{N}_j(\gamma(\eta_j)) < \mathsf{N}_j(\alpha)$ on $\mathbb{R}$. Writing $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \mathsf{N}_j \gamma \mathsf{N}_j^{-1}$ we have $c \neq 0$ since $\mathsf{N}_j(\gamma(\eta_j)) \neq \infty$. Hence $\mathsf{N}_j(\gamma(\eta_j)) = \mathsf{N}_j \gamma \mathsf{N}_j^{-1}(\infty) = a/c$, and $|c| \geqq 1$ by Shimizu's lemma. Using $\mu(\mathsf{N}_j g \mathsf{N}_j^{-1}) \ll \mu(g)$, $\forall g \in G$, (where the implied constant depends on N_j), we also have

$$\mu_j(\gamma) = \inf_{n\in\mathbb{Z}} \mu\big(\gamma \mathsf{N}_j^{-1} \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) \mathsf{N}_j\big) \gg \inf_{n\in\mathbb{Z}} \mu\big(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)\big) \geqq c^2.$$

In view of these observations, it follows from (4.9) that there exists a constant $C_3 > 0$ (which depends on $\Gamma, \mathsf{N}_1, ..., \mathsf{N}_\kappa$ but not on α) such that for any $X_0 > 0$ there exist some $X > X_0$, $j \in \{1, ..., \kappa\}$ and $\gamma \in \Gamma$ such that if $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \mathsf{N}_j \gamma \mathsf{N}_j^{-1}$ then

$$1 \leqq |c| \leqq X \quad \text{and} \quad 0 < \mathsf{N}_j(\alpha) - \frac{a}{c} \leqq \frac{C_3}{X|c|}. \tag{4.12}$$

We are now able to conclude the proof fairly quickly. Given j, γ as in (4.12) we write $\left(\begin{smallmatrix} a_j & b_j \\ c_j & d_j \end{smallmatrix}\right) = \mathsf{N}_j g_0$, so that $\mathsf{N}_j(\alpha) = \mathsf{N}_j g_0(\infty) = \frac{a_j}{c_j}$ and $c_j \neq 0$. We also write

$$\begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix} = \mathsf{N}_j \gamma^{-1} g_0 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}$$

so that

$$\mathbf{c} = -ca_j + ac_j = cc_j\big(\frac{a}{c} - \frac{a_j}{c_j}\big) = cc_j\Big(\frac{a}{c} - \mathsf{N}_j(\alpha)\Big), \tag{4.13}$$

and, if X is sufficiently large (cf. (4.12)),

$$\mathbf{d} = c\big(-b_j + \frac{a}{c}d_j\big) = (1+\nu)\frac{c}{c_j}, \qquad (\text{for some } |\nu| < 0.1), \tag{4.14}$$

since $-b_j + a_j d_j/c_j = 1/c_j$.

Now take $C_2 > 1$ so large that $f(g) = 0$ whenever $\mathcal{Y}_\Gamma(g) \geqq C_2$. If X is sufficiently large we have $\mathbf{c}^2 < 1/2C_2$, and hence $\operatorname{Im} \mathsf{N}_j\gamma^{-1}g_0\mathbf{n}(t)(i) = (\mathbf{c}^2 + (\mathbf{c}t + \mathbf{d})^2)^{-1} > C_2$ for all $t \in \mathbb{R}$ satisfying $\left|t + \mathbf{d}/\mathbf{c}\right| < (\sqrt{2C_2}\mathbf{c})^{-1}$. Hence, for these t we have $\mathcal{Y}_\Gamma(g_0\,\mathbf{n}(t)) > C_2$ and $f(g_0\,\mathbf{n}(t)) = 0$. Notice that $\mathbf{d}/\mathbf{c} < 0$, by (4.12), (4.13), (4.14)! We now let

$$T_1 = \left|\frac{\mathbf{d}}{\mathbf{c}}\right| \quad \text{and} \quad T_2 = \left|\frac{\mathbf{d}}{\mathbf{c}}\right| + \frac{1}{\sqrt{2C_2}|\mathbf{c}|}.$$

One has $0 < T_1 < T_2 < 2T_1$, provided that X is sufficiently large. (To see this one uses (4.14), (4.12), and the fact that for each $j \in \{1, ..., \kappa\}$ and each $X > 0$ there exist only a finite number of double cosets $[\mathsf{T}_j]\gamma[\mathsf{T}_j] \subset \Gamma$ for which $|c| \leqq X$ in $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \mathsf{N}_j\gamma\mathsf{N}_j^{-1}$, cf., e.g., [I, Prop. 2.8].) As in the proof of Proposition 4.2 we now obtain

$$\left|\frac{1}{T_\ell}\int_0^{T_\ell} f(g\,\mathbf{n}(t))\,dt - \langle f\rangle\right| \geqq \frac{|\langle f\rangle|}{2\sqrt{2C_2}} \cdot \frac{1}{|\mathbf{c}|T_\ell} \geqq \frac{|\langle f\rangle|}{4\sqrt{2C_2}} \cdot \frac{1}{|\mathbf{d}|} \tag{4.15}$$

for at least one $\ell \in \{1, 2\}$. But (4.13) and (4.12) imply that $|\mathbf{c}| \leqq C_3|c_j|/X \leqq C_3|c_j/c|$. Using this and (4.14) we obtain $|\mathbf{d}| \leqq 2|c/c_j| \leqq 2C_3/|\mathbf{c}|$. Hence

$$\frac{1}{|\mathbf{d}|} \geqq \frac{1}{\sqrt{2C_3}} \cdot \sqrt{\frac{|\mathbf{c}|}{|\mathbf{d}|}} \geqq \frac{1}{\sqrt{2C_3}} \cdot \frac{1}{\sqrt{T_\ell}}. \tag{4.16}$$

The desired conclusion follows from (4.15) and (4.16), by repeating the above argument for a sequence of X-values tending to ∞. □

5. Ω-results from Fourier expansions

In this section we obtain further Ω-results, using more difficult methods than those in §4. Our proofs here exploit the fact that the horocycle segment $\{p\,\mathbf{n}(t) \mid 0 \leqq t \leqq T\}$ for given T can be shown to lie close to a subsegment of a *closed* horocycle, and then use explicit computations together with known facts about the Fourier coefficients of the individual eigenfunctions on $\Gamma\backslash\mathbb{H}$ in a way reminiscent of what was done in [St]. We conclude the section by giving the proof of Theorem 2, using our results from this section and the preceding one.

Proposition 5.1. *There exists a function $f \in \mathcal{H}^\infty$ with $\langle f\rangle = 0$ and positive constants C_1, C_2, C_3 which only depend on Γ, such that the following holds. For any $p \in \Gamma\backslash G$ and any $u_0 \geqq C_1$ such that $C_1(\log u_0)^{5/2} \leqq \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \leqq u_0/C_1$ and such that the function $u \mapsto \mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum at $u = u_0$, we have for $T = u_0/C_2$:*

$$\left|\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt\right| \geqq \frac{C_3}{\sqrt{r}\cdot\log(r+2)}, \qquad \textit{where } r = \frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}.$$

We let $E_k(z, \frac{1}{2}+iR)$ denote the Eisenstein series on $\Gamma\setminus\mathbb{H}$ associated to the cusp η_k (cf., e.g., [He, p. 280]). Recall that for each $j \in \{1,...,\kappa\}$ we have a Fourier expansion

$$\begin{aligned} E_k(z,\tfrac{1}{2}+iR) =& \delta_{kj} y_j^{\frac{1}{2}+iR} + \varphi_{kj}(\tfrac{1}{2}+iR) y_j^{\frac{1}{2}-iR} \\ &+ \sum_{n\neq 0} c_n \sqrt{y_j} K_{iR}(2\pi|n|y_j)\, e(nx_j), \end{aligned} \tag{5.1}$$

where $x_j + iy_j := \mathsf{N}_j(z)$, the coefficients c_n depend on R, j, k, and $\varphi_{kj}(s)$ is an element of the scattering matrix $\Phi(s) = (\varphi_{kj}(s))_{k,j=1,...,\kappa}$ (cf. [He, pp. 280–281]).

The following lemma gives information on the size of the contribution from the constant terms in (5.1) to the horocycle integral.

Lemma 5.2. *Given Γ, there exists a bounded piecewise continuous function $h(R)$ on $[1,10]$, complex constants $\beta_1,...,\beta_\kappa$, and a constant $0<C_4<1$ such that for each $j\in\{1,...,\kappa\}$, $T>0$, and for each positive function $y(t)$ satisfying $0<y(0)<C_4$ and $|y(t)-y(0)|<C_4\cdot y(0)$ for all $t\in[0,T]$, we have*

$$\begin{aligned} &\Big|\frac{1}{T}\int_0^T \int_1^{10} h(R)\Big(\beta_j y(t)^{\frac{1}{2}+iR} + \sum_{k=1}^{\kappa} \beta_k \varphi_{kj}(\tfrac{1}{2}+iR) y(t)^{\frac{1}{2}-iR}\Big)\, dR\, dt\Big| \\ &\qquad\qquad \geqq C_4 \cdot \frac{\sqrt{y(0)}}{|\log y(0)|}. \end{aligned} \tag{5.2}$$

Proof. Let $h_0(R)$ be a fixed C^∞-function on $[0,1]$ satisfying $0 \leqq h_0(R) \leqq 1$ and $h_0(0)=1$, $h_0(1)=0$. Take constants $R_0\in[2,4]$ and $\alpha\in\mathbb{C}$, and define

$$h(R) = \begin{cases} h_0(R-R_0) & \text{if } R\in[R_0,R_0+1] \\ \alpha\cdot h_0(R-2R_0) & \text{if } R\in[2R_0,2R_0+1] \\ 0 & \text{otherwise.} \end{cases}$$

By repeated integration by parts one then finds that, as $y\to 0$,

$$\begin{aligned} &\int_1^{10} h(R)\Big(\beta_j y^{\frac{1}{2}+iR} + \sum_{k=1}^{\kappa} \beta_k \varphi_{kj}(\tfrac{1}{2}+iR) y^{\frac{1}{2}-iR}\Big)\, dR \\ &\qquad\qquad = \frac{i\sqrt{y}}{\log y}\big(A_j(y)+o(1)\big), \end{aligned} \tag{5.3}$$

where

$$\begin{aligned} A_j(y) =& \beta_j y^{iR_0} - \sum_{k=1}^{\kappa} \beta_k \varphi_{kj}(\tfrac{1}{2}+iR_0) y^{-iR_0} \\ &+ \alpha\Big(\beta_j y^{2iR_0} - \sum_{k=1}^{\kappa} \beta_k \varphi_{kj}(\tfrac{1}{2}+2iR_0) y^{-2iR_0}\Big). \end{aligned}$$

Note that $A_j(y)$ is periodic in the sense that $A_j(ye^{2\pi/R_0}) = A_j(y)$ for all $y > 0$; notice also that $|A_j'(y)| \ll y^{-1}$ as $y \to 0$. Hence we see that to prove the lemma it suffices to show that there is a choice of $R_0, \beta_1, ..., \beta_\kappa, \alpha$ such that

$$\inf_{y>0} \big|A_j(y)\big| > 0, \qquad \text{for all } \ j \in \{1, ..., \kappa\}. \tag{5.4}$$

Let us write $\delta_j := |\beta_j| - \big|\sum_k \beta_k \varphi_{kj}(\frac{1}{2} + iR_0)\big|$ and let M be the set of those j for which there exists some $k \neq j$ such that $\varphi_{kj}(s) \not\equiv 0$. We have $\varphi_{jj}(s) \not\equiv 0$ for each j, since $\varphi_{jj}(s)$ has a simple pole at $s = 1$ (cf., e.g., [I, §6.4] or [He, pp. 286–287]). Hence, for any *generic* choice of $R_0, \beta_1, ..., \beta_\kappa$ with $R_0 \in [2, 4]$, $\beta_k \in \mathbb{C}$, $0 < |\beta_k| \leqq 1$, we have $\delta_j \neq 0$ for all $j \in \mathsf{M}$. Then, for each $j \in \mathsf{M}$ and all $y > 0$ we have

$$\Big|\beta_j y^{iR_0} - \sum_k \beta_k \varphi_{kj}(\tfrac{1}{2} + iR_0) y^{-iR_0}\Big| \geqq \big|\delta_j\big| > 0.$$

This implies that (5.4) holds for each $j \in \mathsf{M}$, provided that we keep $|\alpha| < (\kappa+1)^{-1} \inf_{j\in\mathsf{M}} |\delta_j|$. (Recall that $\big|\varphi_{kj}(\frac{1}{2} + 2iR_0)\big| \leqq 1$, since $\Phi(s) = (\varphi_{kj}(s))$ is unitary for $\mathrm{Re}\ s = \frac{1}{2}$.)

It remains to treat the case $j \notin \mathsf{M}$; then $\big|\varphi_{jj}(\frac{1}{2} + iR)\big| = 1$ for all $R \in \mathbb{R}$ since $\Phi(\frac{1}{2} + iR)$ is unitary, and hence $\delta_j = 0$. We take $\alpha_1, \alpha_2 \in \mathbb{C}$ such that $\alpha_\ell^2 = -\varphi_{jj}(\frac{1}{2} + i\ell R_0)$, $\ell = 1, 2$. Then

$$A_j(y) = 2\beta_j\big(\alpha_1 \cdot \mathrm{Re}(y^{iR_0}/\alpha_1) + \alpha\alpha_2 \cdot \mathrm{Re}(y^{2iR_0}/\alpha_2)\big),$$

and hence for (5.4) to hold it suffices that $\alpha\alpha_2/\alpha_1 \notin \mathbb{R}$ and $\alpha_1^4 \neq -\alpha_2^2$, viz., $\varphi_{jj}(\frac{1}{2} + iR_0)^2 \neq \varphi_{jj}(\frac{1}{2} + 2iR_0)$. Notice that since $\varphi_{jj}(s)$ has a simple pole at $s = 1$ but is analytic for $\mathrm{Re}\ s > 1$ we certainly have $\varphi_{jj}(2s - \frac{1}{2})\varphi_{jj}(s)^{-2} \not\equiv 1$. Hence, by choosing $R_0, \beta_1, ..., \beta_\kappa$ generic as above, and *then* taking $\alpha \in \mathbb{C}$ generic subject to $|\alpha| < (\kappa + 1)^{-1} \inf_{j\in\mathsf{M}} |\delta_j|$, we make (5.4) hold for all j, and we are done. □

Proof of Proposition 5.1. We take $h, \beta_1, ..., \beta_\kappa, C_4$ as in Lemma 5.2. We may assume $|h(R)| \leqq 1$ and $|\beta_k| \leqq 1$ for all R, k. Now define

$$f(z) = \int_1^{10} h(R) \sum_{k=1}^{\kappa} \beta_k \cdot E_k\big(z, \tfrac{1}{2} + iR\big)\, dR. \tag{5.5}$$

As usual, f is viewed as a function on $\Gamma \setminus G$ via the projection $\Gamma \setminus G \ni g \mapsto g(i) \in \Gamma \setminus \mathbb{H}$.

We will choose C_1 and C_2 at the end of the proof, but we will assume $C_1 \geqq C_2 \geqq 1000$ from start. Arguing as in the proof of Proposition 4.2, we find that whenever $p \in \Gamma \setminus G$ and $u_0 \geqq C_1$ satisfy all our assumptions, there

exist a representative $g \in G$ for p and some $j \in \{1, ..., \kappa\}$ such that

$$\mathsf{N}_j g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad 0 < c < C_1^{-1}, \quad |d| \geqq \sqrt{C_1/2}, \tag{5.6}$$
$$u_0 = \frac{|d|}{c}, \quad \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) = \mathrm{Im}\ \mathsf{N}_j g\,\mathbf{a}(u_0)(i) = \frac{1}{2c|d|}.$$

In particular, $u_0 \geqq |d| > 10$, and thus the condition $C_1(\log u_0)^{5/2} \leqq \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0))$ implies

$$c|d|\big(\log|d|\big)^{5/2} \leqq \frac{1}{2C_1}. \tag{5.7}$$

Given j, g as above, we define $x(t), y(t) \in \mathbb{R}$ by

$$x(t) + i\,y(t) := \mathsf{N}_j g\,\mathbf{n}(t)(i). \tag{5.8}$$

Then, for $f_0(z) = E_k(z, \frac{1}{2} + iR)$ with Fourier expansion as in (5.1), and any $T > 0$, we have

$$\begin{aligned} \frac{1}{T}\int_0^T f_0(g\,\mathbf{n}(t)(i))\,dt =& \frac{1}{T}\int_0^T \Big(\delta_{kj} y(t)^{\frac{1}{2}+iR} + \varphi_{kj}(\tfrac{1}{2}+iR) y(t)^{\frac{1}{2}-iR}\Big)\,dt \\ &+ \sum_{n\neq 0} c_n \frac{1}{T}\int_0^T \sqrt{y(t)}\,K_{iR}(2\pi|n|y(t))\,e(nx(t))\,dt. \end{aligned} \tag{5.9}$$

By a direct computation one finds that

$$\begin{aligned} &y(t) = ((ct+d)^2 + c^2)^{-1}, \qquad x(t) = \frac{b}{d} + \frac{1}{d}(ct^2 + dt + c)y(t); \\ &y'(t) = -2c(ct+d)y(t)^2, \qquad x'(t) = (d^2 + 2ctd + c^2t^2 - c^2)y(t)^2; \\ &x''(t) = -2c(d+ct)(d^2 + 2ctd + c^2t^2 - 3c^2)y(t)^3. \end{aligned} \tag{5.10}$$

We now let $T = u_0/C_2 = |d|/C_2 c$. We then have $ct \leqq cT = |d|/C_2$ for all $t \in [0, T]$. Notice also that $T \geqq 1$, i.e. $c \leqq |d|/C_2$. Recall $C_2 \geqq 1000$. It now follows that for all $t \in [0, T]$,

$$\big|y(t)^{-1} - d^2\big| = \big|c^2t^2 + 2ctd + c^2\big| < \frac{3}{C_2}d^2,$$

and similarly,

$$\big|y(t) - 1/d^2\big| \leqq \frac{4}{C_2 d^2}; \qquad \big|x'(t) - 1/d^2\big| \leqq \frac{20}{C_2 d^2}. \tag{5.11}$$

Similarly, using $C_2 \geqq 1000$ and (5.10), we find that for all $t \in [0, T]$,

$$\big|x'(t)^{-1}\big| \ll d^2, \qquad \big|y'(t)\big| \ll c/|d|^3, \qquad \big|x''(t)\big| \ll c/|d|^3, \tag{5.12}$$

where the implied constants are absolute. (These inequalities express in a precise way the fact that the horocycle segment $\{x(t) + iy(t) \mid t \in [0, T]\}$ is "almost horizontal".)

We now consider the last integral in (5.9). Since $x'(t) > 0$ for all $t \in [0,T]$ (by (5.11)) we may integrate by parts as follows:

$$\begin{aligned}
&\frac{1}{T}\int_0^T \sqrt{y(t)}\, K_{iR}(2\pi|n|y(t))\, e(nx(t))\, dt \\
&= \frac{1}{T}\Big[\sqrt{y(t)}\, K_{iR}(2\pi|n|y(t))\, x'(t)^{-1} \cdot \frac{e(nx(t))}{2\pi i n}\Big]_0^T \\
&\quad - \frac{1}{T}\int_0^T \frac{d}{dt}\Big(\sqrt{y(t)}\, K_{iR}(2\pi|n|y(t))\, x'(t)^{-1}\Big)\, \frac{e(nx(t))}{2\pi i n}\, dt
\end{aligned} \tag{5.13}$$

We have the following convenient bounds, which hold uniformly for all $1 \leqq R \leqq 10$ and $v > 0$ (cf. [Wa, pp. 77(2), 78(6), 202(1)]),

$$\big|K_{iR}(v)\big| \ll e^{-v}, \qquad \big|K'_{iR}(v)\big| \ll v^{-1}e^{-v/2}.$$

Notice also $2\pi|n|y(t) > 2|n|/d^2$ for all $t \in [0,T]$, by (5.11). Using these facts and (5.12) we obtain

$$\begin{aligned}
&\Big|\sqrt{y(t)}\, K_{iR}(2\pi|n|y(t))\, x'(t)^{-1}\Big| \ll |d| e^{-|n|/d^2}, \\
&\Big|\frac{d}{dt}\Big(\sqrt{y(t)}\, K_{iR}(2\pi|n|y(t))\, x'(t)^{-1}\Big)\Big| \ll c\, e^{-|n|/d^2}.
\end{aligned}$$

These bounds hold for all $t \in [0,T]$, and the implied constants are absolute. Hence, by (5.13) (and using $c \ll |d|/T$),

$$\Big|\frac{1}{T}\int_0^T \sqrt{y(t)}\, K_{iR}(2\pi|n|y(t))\, e(nx(t))\, dt\Big| \ll \frac{|d|}{T} \cdot \frac{e^{-|n|/d^2}}{|n|}. \tag{5.14}$$

The coefficients c_n in (5.1) are known to satisfy the bound $\sum_{1\leqq|n|\leqq X} |c_n| \ll_\Gamma X\sqrt{\log(1+X)}$ for all $X > 0$, uniformly for $j,k \in \{1,...,\kappa\}$ and $R \in [1,10]$, cf. [St, Prop. 4.1]. Using this fact, (5.14), $|d| > 10$ (cf. (5.6)), and summation by parts, we obtain

$$\sum_{n\neq 0}\Big|c_n \frac{1}{T}\int_0^T \sqrt{y(t)}\, K_{iR}(2\pi|n|y(t))\, e(nx(t))\, dt\Big| \ll_\Gamma \frac{|d|}{T}\big(\log|d|\big)^{3/2}.$$

Combining this with (5.5), (5.9) and $|h(R)| \leqq 1$, $|\beta_k| \leqq 1$, we get

$$\begin{aligned}
\Big|\frac{1}{T}\int_0^T \int_1^{10} h(R)\Big(\beta_j y(t)^{\frac12+iR} + \sum_{k=1}^{\kappa}\beta_k\varphi_{kj}(\tfrac12+iR)y(t)^{\frac12-iR}\Big)\, dR\, dt \\
- \frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\, dt\Big| \leqq C_5 \frac{|d|}{T}\big(\log|d|\big)^{3/2},
\end{aligned}$$

where C_5 is a positive constant which only depends on Γ.

On the other hand, it is clear from (5.6) and (5.11) that if both C_1 and C_2 have been chosen sufficiently large (depending on C_4), then $0 < y(0) < C_4$ and $|y(t) - y(0)| < C_4 \cdot y(0)$ for all $t \in [0,T]$, and hence (5.2) holds. By (5.6) and (5.11), the right side in (5.2) is $> (C_4/10)|d|^{-1}\big(\log|d|\big)^{-1}$. Furthermore,

if C_1, C_2 have also been chosen so that $C_1 > 10C_2C_5/C_4$, then it follows using (5.7) and $T = |d|/C_2 c$ that

$$C_5 \frac{|d|}{T}\big(\log|d|\big)^{3/2} < \frac{C_4}{20}|d|^{-1}\big(\log|d|\big)^{-1}.$$

Hence we obtain

$$\Big|\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt\Big| > \frac{C_4}{20}|d|^{-1}\big(\log|d|\big)^{-1}. \tag{5.15}$$

Regarding the right side in this inequality, note that $\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) \leqq C_2/(2c|d|)$ (by $T = u_0/C_2$ and (2.5), (5.6)), and thus $r = T/\mathcal{Y}_\Gamma(p\,\mathbf{a}(T)) \geqq 2d^2/C_2^2$. Hence (5.15) implies the desired inequality. □

The next two propositions give relevant Ω-results in the presence of small eigenvalues on $\Gamma\backslash\mathbb{H}$.

Proposition 5.3. *Assume that ϕ ($\not\equiv 0$) is a Maass waveform on $\Gamma\backslash\mathbb{H}$ of eigenvalue $\lambda \in (0, \frac{1}{4})$. Write $\lambda = s(1-s)$, $s \in (\frac{1}{2}, 1)$. Then there exists a positive constant C such that for any $p \in \Gamma\backslash G$ for which $\{p\,\mathbf{n}(t)\}$ is a non-closed horocycle, there is a sequence $1 < T_1 < T_2 < \ldots$ with $\lim_{k\to\infty} T_k = \infty$ such that for each $T = T_k$,*

$$\Big|\frac{1}{T}\int_0^T \phi(p\,\mathbf{n}(t))\,dt\Big| \geqq C\cdot T^{s-1}. \tag{5.16}$$

As usual, ϕ is viewed as a function on $\Gamma\backslash G$ via the standard projection $\Gamma\backslash G \to \Gamma\backslash\mathbb{H}$.

Proof. Let $\mathsf{m}_\Gamma = \inf_{g\in G}\mathcal{Y}_\Gamma(g) > 0$. We consider any $u_0 \geqq 1000$ such that $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \leqq u_0$ and $u \mapsto \mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum at $u = u_0$.

Then by arguing as before there is a representative $g \in G$ for p and some $j \in \{1, \ldots, \kappa\}$ such that

$$\mathsf{N}_j g = \begin{pmatrix} a & b \\ c & d\end{pmatrix}, \quad c > 0, \quad |d| \geqq 2^{-\frac{1}{2}}, \quad u_0 = \frac{|d|}{c}, \tag{5.17}$$

$$\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) = \mathrm{Im}\ \mathsf{N}_j g\,\mathbf{a}(u_0)(i) = \frac{1}{2c|d|}, \quad c|d| \leqq (2\mathsf{m}_\Gamma)^{-1}.$$

For each $j \in \{1, \ldots, \kappa\}$ we have a Fourier expansion

$$\phi(z) = c_0^{(j)} y_j^{1-s} + \sum_{n\neq 0} c_n^{(j)} \sqrt{y_j} K_{s-\frac{1}{2}}(2\pi|n|y_j)\, e(nx_j) \tag{5.18}$$

(cf. (2.11); if ϕ is a cusp form then $c_0^{(j)} = 0$), where $x_j + iy_j := \mathsf{N}_j(z)$. As in the proof of Proposition 5.1 we now have, for any $T > 0$,

$$\begin{aligned}\int_0^T \phi(g\,\mathbf{n}(t)(i))\,dt \\ = \int_0^T \Big(c_0^{(j)} y(t)^{1-s} + \sum_{n\neq 0} c_n^{(j)} \sqrt{y(t)}\, K_{s-\frac{1}{2}}(2\pi|n|y(t))\, e(nx(t))\Big)\,dt,\end{aligned} \tag{5.19}$$

where $x(t), y(t)$ are as in (5.8), (5.10). We introduce the following notation

$$F_1(X) = \sqrt{y(T)}\, K_{s-\frac{1}{2}}(2\pi X y(T)) \cdot x'(T)^{-1} \cdot X^{-1}.$$
$$F_2(X) = \sqrt{y(0)}\, K_{s-\frac{1}{2}}(2\pi X y(0)) \cdot x'(0)^{-1} \cdot X^{-1}.$$
$$F_3(X,t) = \frac{d}{dt}\Big(\sqrt{y(t)}\, K_{s-\frac{1}{2}}(2\pi X y(t)) \cdot x'(t)^{-1}\Big) \cdot X^{-1},$$
$$A_t(X) = \sum_{1 \leqq |n| \leqq X} c_n^{(j)} e(nx(t)) \cdot \mathrm{sgn}(n).$$

Now assume $0 < T \leqq 10^{-3} u_0$. Then (5.11), (5.12) hold with "$C_2 = 1000$", and in particular $x'(t) > 0$ for all $t \in [0, T]$. Integrating by parts first as in (5.13) and then with respect to X, we obtain for each $M \in \mathbb{Z}^+$:

$$\begin{aligned}
(5.20)\qquad & \sum_{|n|>M} c_n^{(j)} \int_0^T \sqrt{y(t)}\, K_{s-\frac{1}{2}}(2\pi |n| y(t))\, e(nx(t))\, dt \\
&= \frac{1}{2\pi i}\Big[\int_{M+0}^{\infty} F_1(X)\, dA_T(X) - \int_{M+0}^{\infty} F_2(X)\, dA_0(X) \\
&\qquad\qquad - \int_0^T \Big(\int_{M+0}^{\infty} F_3(X,t)\, dA_t(X)\Big)\, dt\Big] \\
&= \frac{1}{2\pi i}\Big[-F_1(M) A_T(M) - \int_M^{\infty} F_1'(X)\, A_T(X)\, dX \\
&\quad + F_2(M) A_0(M) + \int_M^{\infty} F_2'(X)\, A_0(X)\, dX \\
&\quad + \int_0^T \Big(F_3(M,t) A_t(M) + \int_M^{\infty} \Big(\frac{d}{dX} F_3(X,t)\Big) A_t(X)\, dX\Big)\, dt\Big].
\end{aligned}$$

The above manipulations are easily justified using the absolute bounds below. Since $\frac{1}{2} < s < 1$, we have for all $v > 0$:

$$(5.21)\qquad \big|K_{s-\frac{1}{2}}(v)\big| \ll v^{\frac{1}{2}-s} e^{-v/2}, \qquad \big|K'_{s-\frac{1}{2}}(v)\big| \ll v^{-\frac{1}{2}-s} e^{-v/2},$$
$$\big|K''_{s-\frac{1}{2}}(v)\big| \ll v^{-\frac{3}{2}-s} e^{-v/2}.$$

(Cf. [Wa, pp. 77(2), 78(6), 202(1)]. The implied constants depend on s.) Using these bounds and (5.11), (5.12) (with "$C_2 = 1000$"), we find by a direct computation:

$$(5.22)\quad \big|F_1(X)\big| \ll |d|^{2s} X^{-\frac{1}{2}-s}, \qquad \big|F_1'(X)\big| \ll |d|^{2s} X^{-\frac{3}{2}-s};$$
$$\big|F_2(X)\big| \ll |d|^{2s} X^{-\frac{1}{2}-s}, \qquad \big|F_2'(X)\big| \ll |d|^{2s} X^{-\frac{3}{2}-s};$$
$$\big|F_3(X,t)\big| \ll c|d|^{2s-1} X^{-\frac{1}{2}-s}, \qquad \Big|\frac{d}{dX} F_3(X,t)\Big| \ll c|d|^{2s-1} X^{-\frac{3}{2}-s}.$$

Furthermore, we have $\big|A_t(X)\big| \ll X^{3/2-s}$ for all $X > 0$ and all t (cf. [St, Prop. 5.1]; if ϕ is a cusp form the exponent $3/2-s$ can be replaced by $1/2+\varepsilon$,

cf. [Ha]). Using this bound together with (5.22) and (5.20) we obtain

$$\Big|\sum_{|n|>M} c_n^{(j)} \int_0^T \sqrt{y(t)}\, K_{s-\frac12}(2\pi|n|y(t))\, e(nx(t))\, dt\Big| \leqq K|d|^{2s} M^{1-2s}, \tag{5.23}$$

where K depends only on Γ and ϕ.

On the other hand, we have the following lemma:

Lemma 5.4. *There exist positive constants $C_1 = C_1(\Gamma,\phi)$ and $M_0 = M_0(\Gamma,\phi)$ such that for each integer $M \geqq M_0$ there is some $C_2 > 1$ such that for any $j \in \{1,...,\kappa\}$ and any numbers $b,c,d \in \mathbb{R}$ with $c>0$, $|d| \geqq C_2$ and $c|d| \leqq (2\mathsf{m}_\Gamma)^{-1}$, there is some positive number $T \leqq 10^{-3}\mathsf{m}_\Gamma d^2$ such that*

$$\Big|\int_0^T \Big(c_0^{(j)} y(t)^{1-s} + \sum_{1\leqq|n|\leqq M} c_n^{(j)} \sqrt{y(t)} K_{s-\frac12}(2\pi|n|y(t))\, e(nx(t))\Big)\, dt\Big| \geqq C_1|d|^{2s}. \tag{5.24}$$

(Here $x(t), y(t)$ are defined as in (5.10) for the given b,c,d.)

We first complete the proof of Proposition 5.3 using Lemma 5.4. Let K be as in (5.23) and C_1, M_0 as in Lemma 5.4. We fix some integer $M \geqq M_0$ so large that $KM^{1-2s} \leqq \frac12 C_1$, and then let C_2 be as in Lemma 5.4.

By assumption $\{p\,\mathbf{n}(t) \mid t\in\mathbb{R}\}$ is not a closed horocycle. Hence, for any given number $C_3 > 0$ we can find some $u_0 \geqq 1000$ such that $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \leqq C_3 u_0$ and such that $u \mapsto \mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum at $u = u_0$ (cf. Lemma 4.3). Assume $C_3 < \frac12 C_2^{-2}$. Defining j,g,c,d as in (5.17) we then obtain $|d| \geqq (2C_3)^{-\frac12} > C_2$. Hence by Lemma 5.4 there is some $T \in \big(0, 10^{-3}\mathsf{m}_\Gamma d^2\big]$ such that (5.24) holds. Notice that $T \leqq 10^{-3}\mathsf{m}_\Gamma d^2 < 10^{-3}u_0$, because of (5.17); hence (5.23) holds, and in view of (5.19) and $KM^{1-2s} \leqq \frac12 C_1$ we now obtain

$$\Big|\int_0^T \phi(g\,\mathbf{n}(t)(i))\, dt\Big| \geqq \frac{C_1}{2}|d|^{2s} \geqq \frac{C_1}{2}\Big(\frac{10^3}{\mathsf{m}_\Gamma}\Big)^s T^s. \tag{5.25}$$

In other words, (5.16) holds, with $C = \frac12 C_1 (10^3/\mathsf{m}_\Gamma)^s$. We may now repeat the above construction for a sequence of values of C_3 satisfying $C_3 \to 0$. We will then have $|d| \geqq (2C_3)^{-\frac12} \to \infty$, and hence because of the first inequality in (5.25), we must have $T \to \infty$ for the corresponding sequence of T-values. This completes the proof. □

It remains to prove Lemma 5.4. We first prove an auxiliary result.

Lemma 5.5. *Let Γ, ϕ, s be as above, fix $j \in \{1,...,\kappa\}$ and some $r \in \mathbb{C}$, and define*

$$f(x) = rx + \sum_{n\neq 0} c_n |n|^{-\frac12 - s} sgn(n) e(nx), \qquad (c_n = c_n^{(j)}).$$

Then $f(x)$ is non-constant on every nonempty open interval in $\mathbb{R}$.

Proof. Notice that the sum defining $f(x)$ is uniformly absolutely convergent, becuase of $\frac{1}{2} < s < 1$ and the Rankin-Selberg bound $\sum_{1 \leqq |n| \leqq N} |c_n|^2 \ll N$. Hence $f(x)$ is continuous. Notice also that $f(x) - rx$ is periodic with period 1.

Now assume that there are numbers $\alpha < \beta < \alpha + 1$ such that $f(x) = C$ for all $x \in (\alpha, \beta)$. We let $\mathsf{K}_M(x)$ be Fejer's kernel function, $\mathsf{K}_M(x) = \sum_{|n| \leqq M} \frac{M-|n|}{M} e(nx) = \frac{1}{M}\big(\frac{\sin \pi M x}{\sin \pi x}\big)^2$. We then have, for all $x_0 \in \mathbb{R}$, $M \in \mathbb{Z}^+$,

$$\begin{aligned} &\int_0^1 \Big(f(x) - rx\Big)\mathsf{K}'_M(x_0 - x)\, dx \\ &= 2\pi i \cdot \sum_{1 \leqq |n| \leqq M} \frac{M - |n|}{M} \cdot c_n |n|^{\frac{1}{2}-s} e(nx_0). \end{aligned} \tag{5.26}$$

Fix some $\eta < (\beta - \alpha)/2$, and keep $x_0 \in [\alpha + \eta, \beta - \eta]$. By periodicity, we may rewrite the integral as $\int_\alpha^\beta + \int_\beta^{\alpha+1}$. By our assumption, $f(x) - rx$ is differentiable in for $x \in (\alpha, \beta)$ with constant derivative $-r$. We integrate by parts once in $\int_\alpha^\beta$ and use $0 \leqq \mathsf{K}_M(x) \ll_\eta M^{-1}$, $\big|\mathsf{K}'_M(x)\big| \ll_\eta 1$ for all x with $||x|| \geqq \eta$ (where $||x||$ denotes distance to the nearest integer), and $\int_\alpha^\beta \mathsf{K}_M(x_0 - x)\, dx \leqq \int_0^1 \mathsf{K}_M(x)\, dx = 1$. We then find that the expression in (5.26) is uniformly bounded for all $M \in \mathbb{Z}^+$ and $x_0 \in [\alpha + \eta, \beta - \eta]$.

Now define, for $X > 0$,

$$\begin{aligned} S_{x_0}(X) &= \sum_{1 \leqq |n| \leqq X} c_n |n|^{\frac{1}{2}-s} e(nx_0); \\ A_{x_0}(X) &= \int_0^X S_{x_0}(Y)\, dY = \sum_{1 \leqq |n| \leqq X} (X - |n|) c_n |n|^{\frac{1}{2}-s} e(nx_0). \end{aligned} \tag{5.27}$$

Then by what we have proved, $\big|A_{x_0}(M)\big| \ll M$ for all $M \in \mathbb{Z}^+$ and all $x_0 \in [\alpha + \eta, \beta - \eta]$. Furthermore, $A_{x_0}(X) = 0$ for $0 < X \leqq 1$, and if $M \leqq X < M + 1$ for some $M \in \mathbb{Z}^+$ then

$$\big|A_{x_0}(X) - A_{x_0}(M)\big| = |X - M| \cdot \Big| \sum_{1 \leqq |n| \leqq M} c_n |n|^{\frac{1}{2}-s} e(nx_0) \Big| \ll M^{2-2s}$$

by [St, Prop. 5.1] (or [Ha], if ϕ is a cusp form) and integration by parts. Hence

$$\big|A_{x_0}(X)\big| \ll X, \tag{5.28}$$

uniformly for all $X > 0$ and $x_0 \in [\alpha + \eta, \beta - \eta]$.

After an auxiliary conjugation, we may assume that $\mathsf{N}_j = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. We now have (cf. (5.18) and (5.27)),

$$\phi(x_0+iy) = c_0 y^{1-s} + \sqrt{y}\int_{1/2}^{\infty} K_{s-\frac12}\big(2\pi y X\big)\cdot X^{s-\frac12}\, dS_{x_0}(X)$$
$$= c_0 y^{1-s} + \sqrt{y}\int_{1/2}^{\infty} \frac{d^2}{dX^2}\Big(K_{s-\frac12}\big(2\pi y X\big)\cdot X^{s-\frac12}\Big)\, A_{x_0}(X)\, dX.$$

But $\Big|\frac{d^2}{dX^2}\Big(K_{s-\frac12}\big(2\pi y X\big)\cdot X^{s-\frac12}\Big)\Big| \ll y^{\frac12-s}X^{-2}e^{-yX}$ (cf. (5.21)). Using this bound and (5.28), we obtain $\big|\phi(x_0+iy)\big| \ll y^{1-s}\log(1/y)$ for y small, and in particular $\phi(x_0+iy)\to 0$ as $y\to 0$, uniformly for $x_0\in[\alpha+\eta,\beta-\eta]$. This contradicts $\phi\not\equiv 0$ and the fact that the horocycle segment $[\alpha+\eta,\beta-\eta]+iy$ becomes asymptotically equidistributed in $\Gamma\backslash\mathbb{H}$ as $y\to 0$ (cf., e.g., [S] or [St], and recall our assumption $\mathsf{N}_j = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$). $\square$

Proof of Lemma 5.4. To start with, let $j\in\{1,...,\kappa\}$, $M\in\mathbb{Z}^+$ and $b,c,d\in\mathbb{R}$ be arbitrary numbers such that $c>0$, $|d|\geqq\sqrt{10^3/\mathsf{m}_\Gamma}$ and $c|d|\leqq(2\mathsf{m}_\Gamma)^{-1}$. We write $T_0 = 10^{-3}\mathsf{m}_\Gamma d^2$, and let $x(t), y(t)$ be defined by (5.10). Note that $1\leqq T_0 < 10^{-3}|d|/c$, and hence (5.11) and (5.12) hold for all $t\in[0,T_0]$ with "$C_2 = 1000$".

Let us denote the integral in (5.24) by $J_M(T)$. Introduce the new variable $u(t) := x(t)-x(0)$ in $J_M(T)$. By (5.11), $\frac12 < d^2u'(t) < 2$ for all $t\in[0,T_0]$, and hence $u=u(t)$ gives a bijective C^1-correspondence between $t\in[0,T_0]$ and $u\in[0,U_0]$, where $U_0 = u(T_0)\in\big(\frac12\cdot 10^{-3}\mathsf{m}_\Gamma, 2\cdot 10^{-3}\mathsf{m}_\Gamma\big)$. Using this together with $y(t)\leqq 2/d^2$ (cf. (5.11)) and

$$K_{s-\frac12}(2\pi v) = k_s\cdot v^{\frac12-s} + O_s(v^{s-\frac12}) \qquad \text{as } v\to 0, \tag{5.29}$$

(cf. [Wa, pp. 77(2), 78(6)]; $k_s\in\mathbb{R}$ and $k_s\neq 0$), we obtain, for all $T\in[0,T_0]$:

$$\Big|J_M(T) - I_M\big(u(T)\big)\Big| \leqq C(M)\cdot|d|^{2-2s}, \tag{5.30}$$

where

$$I_M(v) := \int_0^v \Big[c_0^{(j)} + k_s\cdot\sum_{1\leqq|n|\leqq M} c_n^{(j)}|n|^{\frac12-s}\, e\Big(n\big(x(0)+u\big)\Big)\Big]\cdot\frac{y(t)^{1-s}}{x'(t)}\, du,$$

and $C(M)$ is a positive constant which only depends on Γ,ϕ,s,M.

Let us now define, for $u,U\in[0,U_0]$,

$$h(u) := \frac{y(t)^{1-s}}{x'(t)} \qquad\text{and}\qquad \mathcal{I}_M(U) := \int_0^U \frac{I_M'(u)}{h(u)}\, du.$$

It then follows directly from our definitions that

$$\mathcal{I}_M(U) = \frac{k_s}{2\pi i}\Big[f^{(M)}\big(x(0)+U\big) - f^{(M)}\big(x(0)\big)\Big], \tag{5.31}$$

where

$$f^{(M)}(x) = \frac{2\pi i c_0^{(j)}}{k_s} \cdot x + \sum_{1 \leqq |n| \leqq M} c_n^{(j)} |n|^{-\frac{1}{2}-s} \mathrm{sgn}(n) e(nx). \tag{5.32}$$

But we have $f^{(M)}(x) \to f(x)$ as $M \to \infty$, uniformly for all $x \in \mathbb{R}$, where $f(x)$ is as in Lemma 5.5 with $r = 2\pi i c_0^{(j)}/k_s$. It now follows from Lemma 5.5 that there exist constants $C_3 > 0$ and $M_0 > 0$, which only depend on Γ, ϕ, s, such that for all integers $M \geqq M_0$ we have

$$\inf_{x \in \mathbb{R}/\mathbb{Z}} \sup \Big\{ \big| f^{(M)}(x+U) - f^{(M)}(x) \big| \quad \Big| \quad 0 \leqq U \leqq \tfrac{1}{2} \cdot 10^{-3} \mathsf{m}_\Gamma \Big\} \geqq C_3. \tag{5.33}$$

On the other hand, integrating by parts in the definition of $\mathcal{I}_M(U)$, we see that

$$\mathcal{I}_M(U) = \frac{I_M(U)}{h(U)} - \frac{I_M(0)}{h(0)} + \int_0^U \frac{h'(u)}{h(u)^2} I_M(u)\, du, \tag{5.34}$$

and using (5.11), (5.12) one checks that $\big|h(u)\big|^{-1} \ll |d|^{-2s}$ and $\big|h'(u)\big| \ll c|d|^{2s+1}$ for all $u \in [0, U_0]$, and we also have $cU_0 \ll_\Gamma c \ll_\Gamma |d|^{-1}$. This implies

$$\big|\mathcal{I}_M(U)\big| \leqq C_4 |d|^{-2s} \sup_{u \in [0,U]} \big|I_M(u)\big|, \qquad \forall U \in [0, U_0], \tag{5.35}$$

where C_4 is a constant which only depends on Γ. But (5.31), (5.33) and (5.35) imply that for any integer $M \geqq M_0$, there exists some $u \in (0, U_0]$ such that $\big|I_M(u)\big| \geqq C_3 |k_s| |d|^{2s}/2\pi C_4$. Now recall (5.30). Given $M \geqq M_0$ we take $C_2 > \max\big(1, \sqrt{10^3/\mathsf{m}_\Gamma}\big)$ so large that $|d| \geqq C_2$ implies $C(M) \cdot |d|^{2-2s} \leqq C_3 |k_s| |d|^{2s}/4\pi C_4$. It then follows that for any $j \in \{1, ..., \kappa\}$ and any numbers $b, c, d \in \mathbb{R}$ satisfying $c > 0$, $|d| \geqq C_2$ and $c|d| \leqq (2\mathsf{m}_\Gamma)^{-1}$, there exists some $T \in \big(0, 10^{-3} \mathsf{m}_\Gamma d^2\big]$ such that $\big|J_M(T)\big| \geqq C_3 |k_s| |d|^{2s}/4\pi C_4$. $\square$

Proposition 5.6. *Assume that ϕ ($\not\equiv 0$) is a residual eigenfunction on $\Gamma \backslash \mathbb{H}$ of eigenvalue $\lambda \in (0, \frac{1}{4})$. Write $\lambda = s(1-s)$, $s \in (\frac{1}{2}, 1)$. Let $j \in \{1, ..., \kappa\}$ be such that $c_0^{(j)} \neq 0$ (cf. (5.18)). Then there exist positive constants C_1, C_2 such that the following holds. For any $p \in \Gamma \backslash G$ and any $u_0 \geqq C_1$ such that $j_\Gamma(p\,\mathbf{a}(u_0)) = j$, $C_1 \leqq \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \leqq u_0$ and such that the function $u \mapsto \mathcal{Y}_\Gamma(p\,\mathbf{a}(u))$ takes a local maximum at $u = u_0$, we have for $T = u_0/1000$:*

$$\Big| \frac{1}{T} \int_0^T \phi(p\,\mathbf{n}(t))\, dt \Big| > C_2 \cdot \Big(\frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))} \Big)^{s-1}.$$

Proof. This is similar to the proof of Proposition 5.3, but easier. Introducing g, j, a, b, c, d as usual, we repeat the argument from (5.20) to (5.23) to prove

$$\Big| \sum_{n \neq 0} c_n^{(j)} \frac{1}{T} \int_0^T \sqrt{y(t)}\, K_{s-\frac{1}{2}}(2\pi |n| y(t))\, e(nx(t))\, dt \Big| \ll T^{-1} |d|^{2s}. \tag{5.36}$$

We now have $c_0^{(j)} \neq 0$ in (5.18), and the term $|c_0^{(j)}| y_j^{1-s} \gg |c_0^{(j)}| \cdot |d|^{2s-2}$ will dominate over (5.36), provided that we have taken C_1 sufficiently large. □

Proof of Theorem 2. The first assertion in Theorem 2 follows from Proposition 3.1, using (1.1), (1.6), (2.7). If $\{p\,\mathbf{n}(t)\}$ is a closed horocycle then $\delta_p = 0$, and the second assertion in Theorem 2 is obvious.

From now on we assume that $\{p\,\mathbf{n}(t)\}$ is not closed. Notice that we always have $0 \leqq \delta_p \leqq \frac{1}{2}$, and if $\delta_p = \frac{1}{2}$ then the second assertion in Theorem 2 follows from Proposition 4.4. If $\delta_p = 1 - s_1 < \frac{1}{2}$ then the same assertion follows from Proposition 5.3.

It now remains to treat the case where $\delta_p = (1-\alpha_{p,j})(1-s_1^{(j)}) < 1-s_1 \leqq \frac{1}{2}$ for some j. Note that we then necessarily have $0 < \alpha_{p,j} \leqq 1$.

First assume $s_1^{(j)} = \frac{1}{2}$. Then take C_1, C_2, C_3 and f as in Proposition 5.1. Let some $\delta > \delta_p$ be given; we may then find a number $0 < \alpha' < \alpha_{p,j}$ such that $\delta_p < \frac{1}{2}(1-\alpha') < \delta$. It now follows from (1.6) and (2.7) that there exist arbitrarily large numbers u_1 for which $j_\Gamma(p\,\mathbf{a}(u_1)) = j$ and $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_1)) \geqq u_1^{\alpha'} > 1000$. As usual, given such a number u_1 there is a representative $g \in G$ for p such that

$$\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_1)) = \operatorname{Im} \mathsf{N}_j g \mathbf{a}(u_1)(i) = \frac{1}{c^2 u_1 + d^2/u_1}, \quad \mathsf{N}_j g = \begin{pmatrix} * & * \\ c & d \end{pmatrix}, \quad c > 0$$

($c = 0$ is impossible since $\{p\,\mathbf{n}(t)\}$ is non-closed). Letting $u_0 = |d|/c$ we have (by Lemma 3.2) $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) = (2c|d|)^{-1} \geqq \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_1))$, and, if $u_1 \leqq u_0$,

$$\frac{u_0}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0))} = 2d^2 \leqq 2(c^2 u_1^2 + d^2) = \frac{2u_1}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_1))} \leqq 2u_1^{1-\alpha'} \leqq 2u_0^{1-\alpha'},$$

whereas if $u_0 \leqq u_1$, exactly the same conclusion is reached as follows:

$$\frac{u_0}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0))} = \frac{2u_0^2}{u_1^2} c^2 u_1^2 \leqq \frac{2u_0^2}{u_1^2} \frac{u_1}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_1))} \leqq \frac{2u_0^2}{u_1^2} u_1^{1-\alpha'} \leqq 2u_0^{1-\alpha'}.$$

It also follows from $\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \geqq \mathcal{Y}_\Gamma(p\,\mathbf{a}(u_1)) \geqq u_1^{\alpha'}$ that $u_0 \to \infty$ as $u_1 \to \infty$. Hence $u_0/\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \to \infty$ as $u_1 \to \infty$ (cf. Lemma 4.3), and it is now clear that for each sufficiently large number u_1 as above, the corresponding u_0 satisfies all the assumptions in Proposition 5.1, and hence we have for $T = u_0/C_2$:

$$\Big|\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt\Big| \geqq \frac{C_3}{\sqrt{r}\cdot\log(r+2)}, \qquad \text{where } r = \frac{T}{\mathcal{Y}_\Gamma(p\,\mathbf{a}(T))}.$$

But by the above inequalities and (2.5), $r \leqq u_0/\mathcal{Y}_\Gamma(p\,\mathbf{a}(u_0)) \leqq 2u_0^{1-\alpha'} = 2C_2^{1-\alpha'} T^{1-\alpha'}$, and hence

$$\Big|\frac{1}{T}\int_0^T f(p\,\mathbf{n}(t))\,dt\Big| \gg \frac{T^{-\frac{1}{2}(1-\alpha')}}{\log T} \gg T^{-\delta}.$$

The argument shows that there exist arbitrary large values of T for which this holds.

The proof in the remaining case, i.e. $\frac{1}{2} < s_1^{(j)} < 1$, is entirely similar except that we use Proposition 5.6 instead of Proposition 5.1. □

References

[BR] J. Bernstein and A Reznikov, Analytic continuation of representations and estimates of automorphic forms, Ann. of Math., **150** (1999), 329–352.

[Bo] A. Borel, *Automorphic forms on $SL_2(\mathbb{R})$,* Cambridge University Press, 1997.

[Bu] D. Bump, *Automorphic Forms and Representations*, Cambridge University Press, 1997.

[Bur] M. Burger, Horocycle flow on geometrically finite surfaces, Duke Math. J. **61** (1990), 779–803.

[C] N. I. Chernov, On Sinai-Bowen-Ruelle Measures on Horocycles of 3-D Anosov Flows, Geom. Dedicata **68** (1997), 359–369.

[D] S. G. Dani, On uniformly distributed orbits of certain horocycle flows, Ergodic Theory Dynam. Systems **2** (1982), 139–158.

[DS] S. G. Dani, J. Smillie, Uniform distribution of horocycle orbits for Fuchsian groups, Duke Math. J. **51** (1984), 185–194.

[FF] L. Flaminio and G. Forni, Invariant distributions and time averages for horocycle flows, to appear in Duke Math. J., [Available electronically at: http://www-gat.univ-lille1.fr/~flaminio/]

[Ha] J. L. Hafner, Some Remarks on Odd Maass Wave Forms (and a correction to "Zeros of L-functions attached to Maass forms"), Math. Z. **190**, 113–128 (1985)), Math. Z. **196**, 129–132 (1987).

[He] D. A. Hejhal, *The Selberg Trace Formula for $PSL(2,\mathbb{R})$*, Vol. 2, Lecture Notes in Math. **1001**, Springer-Verlag, Berlin, 1983.

[I] H. Iwaniec, *Introduction to the Spectral Theory of Automorphic Forms,* Biblioteca de la Revista Matemática Iberoamericana, Madrid, 1995.

[La] S. Lang, $SL(2,\mathbb{R})$, Addison-Wesley, Reading, Mass., 1975.

[Le] J. Lehner, *Discontinuous Groups and Automorphic Functions*, AMS Math. Surveys No. 8, 1964.

[M] G. W. Mackey, *The Theory of Unitary Group Representations*, Univ. of Chicago Press, 1976.

[Ma] G. A. Margulis, Problems and conjectures in rigidity theory, in *Mathematics: frontiers and perspectives*, AMS, Providence, RI, 2000, pp. 161–174

[MP] M. V. Melián and D. Pestana, Geodesic excursions into cusps in finite volume hyperbolic manifolds, Michigan Math. J. **40** (1993) 77–93.

[Mi] T. Miyake, *Modular Forms,* Springer-Verlag, New York, 1989.

[N] E. Nelson, *Analytic vectors,* Ann. of Math. (2) **70** (1959), 572–615.

[P] S. J. Patterson, Diophantine approximation in Fuchsian groups, Phil. Trans. Soc. London, **282** (1976), 527–563.

[R1] M. Ratner, The rate of mixing for geodesic and horocycle flows, Ergodic Theory Dynam. Systems **7** (1987), 267–288.

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

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

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

[S] N. Shah, Limit distributions of expanding translates of certain orbits on homogeneous spaces, Proc. Indian Acad. Sci. (Math. Sci.) **106** (1996), 105–125. [Available electronically at: www.arXiv.org.]

[Sh] H. Shimizu, On discontinuous groups acting on the product of the upper half planes, Ann. Math., II. Ser. **77** (1963) 33–71.

[St] A. Strömbergsson, On the uniform equidistribution of long closed horocycles, preprint, 2003.

[Su] D. Sullivan, Disjoint spheres, approximation by imaginary quadratic numbers and the logarithm law for geodesics, Acta Math. **149** (1982) 215–237.

[V1] S. L. Velani, Diophantine approximation and Hausdorff dimension in Fuchsian groups, Math. Proc. Cambridge Philos. Soc. **113** (1993) 343–354.

[V2] S. L. Velani, Geometrically finite groups, Khintchine-type theorems and Hausdorff dimension, Math. Proc. Cambridge Philos. Soc. **120** (1996) 647–662.

[W] N. R. Wallach, Real Reductive Groups I, II, Academic Press, 1988 and 1992.

[Wa] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2d ed., Cambridge Univ. Press, Cambridge, 1944.

Andreas Strömbergsson, Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544, U.S.A; astrombe@princeton.edu