

On equidistribution of polynomial sequences in quotients of $PSL_2(\mathbb{R})$

LAURITZ STRECK 

*Centre for Mathematical Sciences, University of Cambridge,
Wilberforce Road, Cambridge, CB3 0WA, UK
(e-mail: ls909@cantab.ac.uk)*

(Received 3 August 2023 and accepted in revised form 28 July 2025)

Abstract. In this paper, it is shown that for every lattice $\Gamma \subset PSL_2(\mathbb{R})$, there exists a $c > 0$ such that for any $0 \leq \gamma < c$, the sequence $ph(n^{1+\gamma})$ equidistributes for any $p \in \Gamma \backslash PSL_2(\mathbb{R})$, where h is the horocycle flow. This makes modest progress towards a conjecture of Shah and generalizes a result of Venkatesh [Sparse equidistribution problems, period bounds, and subconvexity. *Ann. of Math. (2)* **172**(2) (2010), 989–1094], who established the same equidistribution for co-compact lattices. The proof uses a dichotomy between good equidistribution estimates and approximability of $\{ph(t), t \leq T\}$ by closed horocycles of small period.

Key words: homogeneous dynamics, sparse equidistribution, horocycle flow, modular surface

2020 Mathematics Subject Classification: 37 (Primary); 11 (Secondary)

1. Introduction

Consider the (multiplicative) group $G := PSL_2(\mathbb{R})$ with a Haar measure μ_G . A lattice $\Gamma \subset G$ is a discrete subgroup such that the quotient $X := \Gamma \backslash G$ has a fundamental domain in G of finite Haar measure. The Haar measure then descends to a finite measure μ_X . We define the matrices

$$h(x) := \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad a(y) := \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}.$$

The *geodesic flow* at time t of $p \in X$ is defined by $g_t(p) := pa(e^t)$ and the *horocycle flow* at time t is defined by $h_t(p) := ph(t)$.

While the orbit $g_t(p)$ for $t \rightarrow \infty$ can behave quite irregularly depending on the initial point, the horocycle orbit $h_t(p)$ is known to behave much more rigidly. Before we detail the known results, we pin down some notation. We say that the orbit $h_t(p)$ equidistributes with respect to μ_X if for any compactly supported, continuous function f on X ,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(ph(t)) dt \rightarrow \int f d\mu_X.$$

Similarly, we say that the orbit equidistributes along a sequence $a_n \in \mathbb{R}$ with respect to μ_X if

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(ph(a_n)) \, dt \rightarrow \int f \, d\mu_X.$$

Lastly, a point $p \in X$ is called *periodic* if there is a $t_0 \in \mathbb{R}$ such that $p = ph(t_0)$. In this case, the horocycle orbit will be trapped in the periodic orbit and will never equidistribute with respect to μ_X ; the system $t \mapsto ph(t)$ is then isomorphic to the circle-rotation $x \mapsto x + t_0^{-1}$ on the torus $\mathbb{R}/\mathbb{Z}$. Below, we use ‘ $ph(a_n)$ equidistributes’ as a shorthand for ‘for all non-periodic $p \in X$, $ph(a_n)$ equidistributes with respect to μ_X ’. It was shown by Dani and Smillie that both $ph(t)$ for $t \in \mathbb{R}$ and $ph(n)$ for $n \in \mathbb{N}$ equidistribute.

It was subsequently asked what happens for sequences other than $\mathbb{N}$. Margulis conjectured that $ph(p_n)$, where p_n is the n th prime number, should also equidistribute. Shah conjectured that for any $\gamma \geq 0$, $ph(n^{1+\gamma})$ would equidistribute. We remark that these results follow for μ_X -almost every $p \in X$ from the work of Bourgain in a much more general context [1]. The challenge is really to establish equidistribution for *all* non-periodic $p \in X$.

Venkatesh made progress on Shah’s conjecture by showing that for co-compact Γ , there is a small $c = c(\Gamma) > 0$ such that for all $0 \leq \gamma < c$ and all $p \in X$, $ph(n^{1+\gamma})$ equidistributes [9]. Venkatesh’s proof operates by controlling arithmetic sequences of the type $ph(sn)$ for $n \in \{0, \dots, N-1\}$ with s small compared with n . Controlling these sparse sequences also means that the almost-primes equidistribute for co-compact Γ ; that is, for sufficiently big R , $ph(q)$ equidistributes, where q runs over all numbers having at most R many prime factors. That controlling sparse sequences is enough to control the almost-primes can be seen either using sieve methods or using the pseudo-random measure ν , introduced by Goldston and Yilmaz, and subsequently used by Green and Tao to show that the primes contain infinitely long arithmetic progressions [3, 4] (see [6] for a proof of the equidistribution of almost-primes using sieve methods and [7] for a proof using the pseudo-random measure ν).

Using Venkatesh’s method in the case of a non-compact lattice, one can show that $ph(n^{1+\gamma})$ and $ph(q)$, q almost prime, equidistribute under the assumption of a Diophantine condition on p [5, 10, 11]. This Diophantine condition assures that the horocycle orbit $ph(t)$ equidistributes with rate $T^{-\varepsilon}$ for all T , which is needed for Venkatesh’s argument. Using the fact that for any point p , there are some times $T_i \rightarrow \infty$ such that $ph(t)$, $t \leq T_i$ equidistributes with error $T_i^{-\varepsilon}$, one can also deduce with the same method that the orbits $ph(n^{1+\gamma})$ and $ph(q)$, q almost prime, are dense.

However, showing equidistribution for all p is significantly harder, as there are p such that there are times T for which the equidistribution of $ph(t)$, $t \leq T$ is far worse than polynomial. In this case, Venkatesh’s method cannot be applied.

Sarnak and Ubis were the first to show such a sparse equidistribution result for all initial p . They showed that the almost-primes equidistribute for $\Gamma = PSL_2(\mathbb{Z})$, which is not co-compact [6]. It was subsequently proved by the author that the almost-primes equidistribute for all lattices Γ in $PSL_2(\mathbb{R})$ [7].

In this paper, the equidistribution of $ph(n^{1+\gamma})$ is established for small γ in the setting of a general lattice. This generalises Venkatesh's result from co-compact Γ to all lattices Γ in $PSL_2(\mathbb{R})$ and makes (modest) progress on the conjecture of Shah.

We make this precise in the result below, which is the main result of this paper. For this, we need some more notation and start by defining the metric d_X . The group $G = PSL_2(\mathbb{R})$ comes with a natural left-invariant metric d_G (see for example [2, Ch. 9]). This metric descends to X via $d_X(\Gamma g, \Gamma h) := \inf_{\gamma \in \Gamma} d_G(g, \gamma h)$. We also fix a point $p_0 \in X$ and define $\text{dist}(p) := d_X(p, p_0)$.

For two functions $f, g: U \rightarrow \mathbb{R}$, we write $f \ll g$ or $f = O(g)$ if there is a constant C such that $|f(x)| \leq C|g(x)|$ for all $x \in U$, where U is some domain. In this paper, this constant C implicit in the definition is always allowed to depend on the lattice Γ and the choice of γ , but nothing else. We write $f \sim g$ if both $f \ll g$ and $g \ll f$.

For a function $f \in C^4(X)$, let $\|f\|_{W^4}$ be its Sobolev norm in the Hilbert space $W^{4,2}$ involving the fourth derivative and let $\|f\|_{\infty,j}$ be the supremum norm of the j th derivatives. Define

$$\|f\| := \|f\|_{W^4} + \|f\|_{\infty,1} + \|f\|_{\infty,0};$$

this norm is the same one Strömbergsson used to show his equidistribution result [8]. In his result, the equidistribution properties of a horocycle piece $\{ph(t), 0 \leq t \leq T\}$ are measured in terms of the parameter

$$r(p, T) := T \exp(-\text{dist}(g_{\log T}(p))),$$

which will be an important quantity to measure the equidistribution properties throughout this paper; its significance and role in the proof will be discussed below in more detail. It is well known that $r(p, T) \rightarrow \infty$ as $T \rightarrow \infty$ for any non-periodic p .

We let β be the constant in Theorem 1.2; it ultimately comes from the rate of effective mixing. The constant in Theorem 1.1 can be taken to be $c = \beta/600$.

THEOREM 1.1. *For any lattice $\Gamma \subset PSL_2(\mathbb{R})$, there is a constant $c = c(\Gamma) > 0$ such that for any $0 \leq \gamma \leq c$, any non-periodic $p \in X$ and any function $f \in C^4(X)$ with $\|f\| = 1$,*

$$\left| \frac{1}{T} \sum_{n \leq T} f(ph(n^{1+\gamma})) - \int f d\mu_X \right| \ll r^{-\beta/4},$$

where $r = r(p, T^{1+\gamma})$.

To prove Theorem 1.1, we will split the range into different intervals and use Taylor expansion on each one. On an interval $[T_0, T_1]$, the function $t^{1+\gamma}$ will be approximately equal to $T_0^{1+\gamma} + (1+\gamma)T_0^\gamma(t - T_0)$, provided that T_0 is not too small and that the range is not too long. The question thus becomes how well $ph(ns)$ for $s \sim T^\gamma$ equidistributes. To control these sparse arithmetic sequences, we need two results.

The first one is the following theorem, which is a straightforward consequence of combining Strömbergsson's equidistribution result [8] with Venkatesh's method [9], as performed, for example, by Zheng [10].

THEOREM 1.2. [10, Theorem 1.2] *Let Γ be a non-compact lattice in G . Let $f \in C^4(X)$ with $\|f\| < \infty$ and $1 \leq s < T$. Then,*

$$\left| \frac{s}{T} \sum_{1 \leq j \leq T/s} f(ph(sj)) - \int f d\mu_X \right| \ll s^{1/2} r^{-\beta/2} \|f\|$$

for any initial point $p \in X$, where $r = r(p, T)$. The parameter $\frac{1}{6} > \beta > 0$ and the implied constant depend only on Γ .

In the cases where r is big compared with T (say $r \geq T^\epsilon$ for some absolute ϵ), this theorem in itself is enough to show equidistribution of the sequence $ph(n^{1+\nu})$.

The result below will be used to deal with the case in which the equidistribution is bad. It was proved by the author in [7] to show equidistribution of almost-primes. Its proof uses ideas of Sarnak and Ubis [6] and has parallels to [8], whose proof in turn uses ideas going back to Marina Ratner. This result encompasses the dichotomy mentioned in the abstract.

LEMMA 1.3. [7, Lemma 1.3] *Let Γ be a lattice in $G = PSL_2(\mathbb{R})$ and let $X = \Gamma \backslash G$. Let $p \in X$ and $T \geq 0$. Let $\delta > 0$ and $K \leq T$.*

There is an interval $I_0 \subset [0, T]$ of size $|I_0| \leq \delta^{-1} K^2$ such that: for all $t_0 \in [0, T] \setminus I_0$, there is a segment $\{\xi h(t), t \leq K\}$ of a closed horocycle approximating $\{ph(t_0 + t), 0 \leq t \leq K\}$ of order δ , in the sense that

$$\text{for all } 0 \leq t \leq K, \quad d_X(ph(t_0 + t), \xi h(t)) \leq \delta.$$

The period $P = P(t_0, p)$ of this closed horocycle is at most $P \ll r(p, T)$.

Moreover, one can assure $P \gg \eta^2 r$ for some $\eta > 0$ by weakening the bound on I_0 to $|I_0| \leq \max(\delta^{-1} K^2, \eta T)$.

2. On the behaviour of the equidistribution parameter in Theorem 1.2

Except for Lemma 1.3 itself, we will also need some of the other material in [7, Ch. 4] to prove Theorem 1.1. We recall some of the material, going slightly beyond what is presented in [7].

It is well known that $G \cong T_1 \mathbb{H}$, where $\mathbb{H}$ is the upper half-plane with the hyperbolic metric. Then, $X = \Gamma \backslash G$ has as fundamental domain a set $T_1 F$, where F is a geodesic polygon in $\mathbb{H}$, that is, a polygon with finitely many vertices with the edges being pieces of geodesics [2]. This fundamental polygon F has finitely many vertices touching the boundary of the upper half-plane, either at the axis with imaginary part equal to zero or at infinity. After identifying vertices that are in the same orbit under the action of Γ , one gets the cusps of X , which we will denote by $r_1, \dots, r_n$. Any such cusp r_i is in 1–1 correspondence to an element $\gamma_i \in \Gamma$ with the property that γ_i fixes r_i and that γ_i is conjugated to $h(1)$ (see [7, Lemma 3.1]). For each cusp, there are elements $\sigma_i \in G$ such that $\sigma_i r_i = \infty$ and $\sigma_i \gamma_i \sigma_i^{-1} = h(1)$.

For $g \in G$, we define $Y_i^0(g) := \text{Im}(\sigma_i g)$, where

$$\text{Im} \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) := \frac{1}{c^2 + d^2}$$

is the imaginary part of the the matrix projected to $\mathbb{H}$. We also set for $p = \Gamma g_p \in X$, $y_i^0(p) := \max_{\gamma \in \Gamma} Y_i^0(\gamma g_p)$.

It was shown in [7, Lemma 4.1] that there exist disjoint neighbourhoods $C_i \subset X$ of each cusp r_i with $K = X \setminus \cup C_i$ being compact such that for any $p \in C_i$, $\exp(\text{dist}(p)) \sim y_i^0(p)$ (while of course $\exp(\text{dist}(p)) \sim 1$ for $p \in K$). Arguing as in the [7, proof of 1 in Lemma 4.1], one also sees that if $p = \Gamma g_p \in C_i$ and g_p is such that $Y_i^0(g_p) = y_i^0(p)$, then for any $\gamma \in \Gamma$, either $Y_i^0(\gamma g_p) \ll 1$ or $Y_i^0(\gamma g_p) = Y_i^0(g_p)$ (which is the case in which $\sigma_i \gamma g_p = h(n) \sigma_i g_p$ and $\gamma = (\gamma_i)^n$ for some n). This implies in particular that there is an absolute constant $C = C(\Gamma)$ such that if g_p is such that $Y_i^0(g_p) \geq C$, then

$$Y_i^0(g_p) \sim y_i^0(p) \sim \exp(\text{dist}(p)),$$

where the second equivalence holds because $y_i^0(p) \geq C$ implies that $p \in C_i$ for C sufficiently big.

From the expression above, the reader sees the relation to the equidistribution parameters

$$r(q, K) := K \exp(-\text{dist}(g_{\log K}(q)))$$

appearing in Theorems 1.1 and 1.2.

Observation 2.1. There is an absolute $c_0 = c_0(\Gamma) > 0$ such that for any T and any p , if there is a representative g_p of p and an i such that for $\sigma_i g_p =: \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\max(T^2 c^2, d^2) \leq c_0 T$, then $r(p, T) \sim \max(T^2 c^2, d^2)$.

Proof. We have that

$$2 \max(T^2 c^2, d^2) \geq T(c^2 T + d^2 T^{-1}) = Y_i^0(g_{\log T}(g_p))^{-1} T.$$

Thus, $Y_i^0(g_{\log T}(g_p)) \geq \frac{1}{2} c_0^{-1}$, which shows that

$$\exp(\text{dist}(g_{\log T}(p))) \sim Y_i^0(g_{\log T}(g_p))$$

by the argument above, provided that c_0 is sufficiently small. $\square$

3. Proof of Theorem 1.1

We start by approximating $t^{1+\gamma}$ with sparse arithmetic sequences. More precisely, we write

$$t^{1+\gamma} = T_0^{1+\gamma} + (1+\gamma)T_0^\gamma(t - T_0) + O(T^{-1/6})$$

on $[T_0, T_0 + T^{1/3}]$ for $T_0 \geq T^{5/6}$ using Taylor expansion.

We will split into several cases. To govern in which case we are, we fix some $\varepsilon > 0$ and impose that $\gamma < (\varepsilon\beta/6)$. We will see at the end which value of ε makes everything work (which will turn out to be $\varepsilon = 1/100$).

To apply the results about sparse equidistribution, we are thus tasked with evaluating expressions of the form

$$\left| \frac{1}{K} \sum_{n \leq K} f(qh((1+\gamma)T_0^\gamma n)) - \int f d\mu_X \right|$$

for $q = ph(T_0^{1+\gamma})$ and $T^{1/6} \leq K \leq T^{1/3}$, given some $T_0 \leq T$. In the case where $r(q, K) \geq T^\varepsilon$, Theorem 1.2 is enough to deduce good equidistribution.

If $r(q, K) \leq T^\varepsilon$, then $g_{\log K}(q)$ must lie in the neighbourhood C_i of some cusp r_i , as explained in the previous section. In this case, there is a (essentially unique) representative g_q of q such that $r(q, K) \sim \max(K^2 c^2, d^2)$, where we set

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} := \sigma_i g_q,$$

now and for the next couple of pages.

One then has to split into two more cases. The distinction between these cases is governed by

$$W_q := \left| \frac{d}{c} \right|.$$

The relevance of this W_q is that it measures the time it takes until one gets from bad to good equidistribution again. More precisely, by Observation 2.1,

$$r(q, K) \sim \begin{cases} d^2, & K \leq W_q, \\ d^2 \frac{K^2}{W_q^2}, & K \geq W_q, \end{cases} \quad (3.1)$$

as long as $r(q, K) \leq c_0 K$.

This means that even if q and K are such that $r(q, K) \leq T^\varepsilon$, one has that $r(q, T^\varepsilon W_q) \geq T^{2\varepsilon}$. Together with Theorem 1.2, this will be good enough to show effective equidistribution under all assumptions except for those of Proposition 3.1. Under those assumptions, which encompass the most interesting case, almost the entire horocycle orbit $\{ph(t), t \leq T^{1+\gamma}\}$ is close to periodic horocycle orbits of small period. In this case, one will need Lemma 1.3 to conclude.

PROPOSITION 3.1. *Let Γ and $\gamma < c$ be as in Theorem 1.1, and let $\varepsilon = 1/100$. Let $p \in X$ and T be such that $r(p, T^{1+\gamma}) \leq T^{4\varepsilon}$ and $W_p \geq T^{1-\varepsilon}$. Then, for f as in Theorem 1.1,*

$$\left| \frac{1}{T} \sum_{n \leq T} f(ph(n^{1+\gamma})) - \int f d\mu_X \right| \ll r^{-\beta/4}.$$

To prove Theorem 1.1, we will first show how one can reduce its proof to Proposition 3.1 using Observation 2.1 and Theorem 1.2. We will then prove Proposition 3.1.

Proof of Theorem 1.1 assuming Proposition 3.1. Say we are given some t_0 and set $q = ph(t_0^{1+\gamma})$. If $r := r(q, T^{1/6}) \geq T^\varepsilon$, then we know by Theorem 1.2 that for any f with $\|f\| \leq 1$,

$$\left| \frac{1}{T^{1/6}} \sum_{n \leq T^{1/6}} f(qh((1+\gamma)t_0^\gamma n)) - \int f d\mu_X \right| \ll T^{\gamma/2} r^{-\beta/2} \leq r^{-\beta/4},$$

where we recall $\gamma \leq (\varepsilon\beta/6)$. We are thus done unless there is a q such that $r = r(q, T^{1/6}) \leq T^\varepsilon$. As we saw in §2, then with c and d as defined previously,

$$r \sim \max(T^{2/6}c^2, d^2). \quad (3.2)$$

If $c^2T^{2/6}$ attains the maximum in (3.2), or equivalently, if $W_q \leq T^{1/6}$, then $r(q, T^{1/4}) \sim T^{1/6}r \geq T^{1/6}$ by (3.1) and we are done by Theorem 1.2. We can thus assume $W_q \geq T^{1/6}$. The following claim shows how one can improve the lower bound on W_q further.

CLAIM 3.2. *Let $q = ph(t_0^{1+\gamma})$ such that $r \leq T^\varepsilon$. Set $W := W_q$. If $W \leq T^{1-\varepsilon}$, then for $K = W^{1+\varepsilon}$ and for f with $\|f\| \leq 1$,*

$$\left| \frac{1}{K} \sum_{0 \leq n \leq K} f(ph((t_0 + n)^{1+\gamma})) - \int f d\mu_X \right| \ll r^{-\beta/4}.$$

Proof of Claim 3.2. Fix some $W^{1+\varepsilon} \geq s \geq W^{1+\varepsilon/2}$ and note that then $c^2s^2 \sim W^{-2}s^2d^2 \gg d^2$. Thus,

$$\begin{aligned} r(qh(s), T^{1/3}) &\sim \max(T^{2/6}c^2, (d + cs)^2) \sim \max(T^{2/6}c^2, c^2s^2) \\ &= c^2s^2 \sim \left(\frac{s}{W}\right)^2 r \geq rW^\varepsilon \geq rT^{\varepsilon/6}, \end{aligned}$$

where the first equivalence is due to Observation 2.1, which is applicable because $(s/W)^2r \ll T^{3\varepsilon}$. Applying Theorem 1.2 shows that

$$\begin{aligned} &\left| \frac{1}{T^{1/3}} \sum_{n \leq T^{1/3}} f(ph(t_0^{1+\gamma} + s)h((1+\gamma)(t_0 + s)^\gamma n)) - \int f d\mu_X \right| \\ &\ll T^{\gamma/2} T^{-\varepsilon\beta/12} r^{-\beta/2} \leq r^{-\beta/2}. \end{aligned}$$

Now, we use Taylor approximation as above to split the orbit of $(t_0 + n)^\gamma$ with $n \leq K$ into different ranges $[s, s + T^{1/3}]$ and note that for all but a $W^{-\varepsilon/2}T^\gamma$ proportion of s , one has $(t_0 + s)^{1+\gamma} - t_0^{1+\gamma} \geq W^{1+\varepsilon/2}$. As $W^{-\varepsilon/2}T^\gamma \leq T^{-\varepsilon/12} \leq r^{-\beta/4}$, the claim is shown. $\square$

We have thus shown the conclusion of Theorem 1.1 unless there is a $q = ph(t_0)$ such that $r(q, T^{1/6}) \leq T^\varepsilon$ and $W_q \geq T^{1-\varepsilon}$. We let c and d be as defined above and note that in the case considered, $r(q, T^{1/6}) \sim \max(c^2T^{1/3}, d^2) = d^2$ by definition of W_q . By (3.1), this implies that

$$r(q, T^{1+\gamma}) \ll d^2 \frac{T^{2(1+\gamma)}}{W_q^2} \ll T^{4\varepsilon}.$$

Lastly, to get an error term in $r(p, T^{1+\gamma})$ instead of $r(q, T^{1+\gamma})$, we note that $g_q h(-t_0^{1+\gamma})$ is a representative of p and that because $(d - ct_0^{1+\gamma})^2 \ll d^2 T^{2(1+\gamma)} W^{-2} \ll T^{4\varepsilon}$,

$$r(p, T^{1+\gamma}) \sim \max(c^2T^{2(1+\gamma)}, (d - ct_0)^2) \ll T^{4\varepsilon},$$

where the first equivalence is due to Observation 2.1. We have thus reduced the proof of Theorem 1.1 to the assumptions of Proposition 3.1. $\square$

It now only remains to show Proposition 3.1, which is the main part of the proof of Theorem 1.1.

Proof of Proposition 3.1. Let p and T be given such that $r := r(p, T^{1+\gamma}) \leq T^{4\varepsilon}$ and $W := W_p \geq T^{1-\varepsilon}$. Here, $W = |d/c|$, with c and d as defined in Observation 2.1. We also let $g := g_p$ and σ_i be as in Observation 2.1. We invoke Lemma 1.3 to split the orbit $[0, T^{1+\gamma}]$ into pieces of length $K = T^{1/3}$. As in [7, proof of Lemma 1.3 in Ch. 4], we now parametrize the orbit using the equation

$$\sigma_i gh(W + s) = lh(s) = h\left(\alpha - \frac{Rs}{s^2 + 1}\right) a\left(\frac{R}{s^2 + 1}\right) k(-\operatorname{arccot} s),$$

where $l := \sigma_i gh(W) =: (\alpha + iR, -i)$ is the highest point of the horocycle orbit and

$$k(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

is the (subsequently unimportant) rotation component.

Given an $M \leq T$, we then have that $ph(M^{1+\gamma} + t), t \leq T^{1/3}$ is at distance at most $O(T^{-1/6})$ from the orbit on a periodic horocycle $\xi h(t), t \leq T^{1/3}$ with its period being equal to y^{-1} , where

$$y := \frac{R}{(M^{1+\gamma} - W)^2 + 1}.$$

By the second clause in Lemma 1.3, we can assume $r \gg y^{-1} \gg \delta^2 r$ except on an interval of proportion δ , where δ is to be chosen later. Using Taylor approximation on $t^{1+\gamma}$, we thus want to bound

$$\left| \frac{(1+\gamma)M^\gamma}{T^{1/3}} \sum_{(1+\gamma)M^\gamma n \leq T^{1/3}} f(\xi h((1+\gamma)M^\gamma n)) - \int f d\mu_X \right|.$$

However, we may run into problems here: if for example $y^{-1} = (1+\gamma)M^\gamma$, the points do not equidistribute at all in the periodic horocycle. To deal with this and related obstructions, we proceed similarly to in [7, the proof of Claim 5.2]. For notational convenience, we set $s := (1+\gamma)M^\gamma$. Let $q \in \mathbb{N}$ with $y^{-1} \leq q \leq ys^{-1}T^{1/3}$ be such that

$$\left| sy - \frac{a}{q} \right| \leq \frac{y^{-1}s}{qT^{1/3}}$$

for some a coprime to q (such q exists by the pigeonhole principle). The problem case occurs if q is small compared with y^{-1} . If, however, q is sufficiently big, there are so many distinct points in the interval $[0, y^{-1}]$ that they cannot help being dense enough to approximate $\int_0^1 f(\xi h(ty^{-1})) dt$ by force, as we show now.

CLAIM 3.3. *If $q \geq y^{-3}$, then*

$$\left| \frac{s}{T^{1/3}} \sum_{sn \leq T^{1/3}} f(\xi h(sn)) - \int_0^1 f(\xi h(ty^{-1})) dt \right| \ll y \ll \delta^{-2} r^{-1},$$

where q, s, y and ξ all depend on M .

Proof of Claim 3.3. (The argument in the proof of this claim was suggested by Adrián Ubis.) We set $F(t) := f(\xi h(ty^{-1}))$, which is one periodic. Because the function f is 1-Lipschitz with respect to the hyperbolic metric, the function F is y^{-1} -Lipschitz. We wish to show

$$\left| \frac{s}{T^{1/3}} \sum_{sn \leq T^{1/3}} F(nsy) - \int_0^1 F(t) dt \right| \ll y.$$

For this, we note that as for any n ,

$$\left| sny - n \frac{a}{q} \right| \leq n \frac{y^{-1}s}{qT^{1/3}},$$

we have

$$\begin{aligned} \frac{s}{T^{1/3}} \sum_{sn \leq T^{1/3}} F(nsy) &= O\left(\frac{y^{-2}}{q}\right) + \frac{s}{T^{1/3}} \sum_{sn \leq T^{1/3}} F\left(n \frac{a}{q}\right) \\ &= O\left(\frac{y^{-2}}{q}\right) + O\left(\frac{qs}{T^{1/3}}\right) + \frac{1}{q} \sum_{j=0}^{q-1} F\left(\frac{ja}{q}\right) \end{aligned}$$

by the periodicity of F . As a is coprime to q , it does not play a role in the last average and can be dropped. Furthermore, for any $t \leq (1/q)$,

$$F\left(\frac{j}{q}\right) = O\left(\frac{y^{-1}}{q}\right) + F\left(\frac{j}{q} + t\right),$$

so

$$\begin{aligned} \frac{1}{q} \sum_{j=0}^{q-1} F\left(\frac{j}{q}\right) &= O\left(\frac{y^{-1}}{q}\right) + \frac{1}{q} \sum_{j=0}^{q-1} \int_0^1 F\left(\frac{j+t}{q}\right) dt \\ &= O\left(\frac{y^{-1}}{q}\right) + \int_0^1 F(t) dt. \end{aligned}$$

As both $y^{-2}q^{-1}$ and $qsT^{-1/3}$ are $O(y)$, this implies the claim. $\square$

By Strömbergsson's result [8],

$$\left| y \int_0^{y^{-1}} f(\xi h(t)) dt - \int f d\mu_X \right| \ll y^\beta \ll (\delta^{-2}r^{-1})^\beta,$$

so we see from Claim 3.3 that

$$\left| \frac{(1+\gamma)M^\gamma}{T^{1/3}} \sum_{n \leq T^{1/3}} f(\xi h((1+\gamma)M^\gamma n)) - \int f d\mu_X \right| \ll (\delta^2r)^{-\beta}$$

unless there is a $q \leq y^{-3} \leq r^3$ and a coprime to q such that

$$\left| (1+\gamma)M^\gamma y - \frac{a}{q} \right| \ll M^\gamma y^{-1} T^{-1/3} \leq r T^{-1/3+\gamma}.$$

To conclude the proof of Theorem 1.1, we just have to show that this is a very exceptional occurrence.

Fortunately, this is what one would expect: if we let

$$I_{q,a} := \left\{ v \in \mathbb{R} : \left| v - \frac{a}{q} \right| \leq rT^{-1/3+\gamma} \right\}$$

denote the problem intervals for $q \leq r^3$ and $(a, q) = 1$, we note that they are proportional to $rT^{-1/3+\gamma}$. Moreover, given distinct intervals $I_{q_1, a_1}, I_{q_2, a_2}$, the gap between them is at least of order r^{-6} , as

$$\left| \frac{a_1}{q_1} - \frac{a_2}{q_2} \right| \geq \frac{1}{q_1 q_2} \geq r^{-6}.$$

As $r \ll T^{4\varepsilon}$, this means that the set $E := \bigcup_{q \leq r^3, (a,q)=1} I_{q,a}$ makes up only a tiny proportion of the entire range. Unless the function

$$G(t) := \frac{t^\gamma R}{(t^{1+\gamma} - W)^2 + 1} = t^\gamma y$$

is highly concentrated on a small part of its range, our problem case $\{t \leq T : (1 + \gamma)G(t) \in E\}$ will thus only occur on a negligible proportion of $[0, T]$. The following claim shows that G does not behave in this unusual manner.

CLAIM 3.4. *For all but a $O(\delta + \delta^{-5}r^7T^{-1/3+\gamma})$ proportion of $t \leq T$, there does not exist $q \leq r^3$ such that*

$$\left| (1 + \gamma)G(t) - \frac{a}{q} \right| \leq rT^{-1/3+\gamma}.$$

Before we show the claim, we show how it implies Proposition 3.1. The claim implies that at most a small proportion of the intervals we split $[0, T]$ into when applying Taylor approximation will be bad; for the others, we know equidistribution from Claim 3.3. Collecting all the different error terms together,

$$\left| \frac{1}{T} \sum_{n \leq T} f(ph(n^{1+\gamma})) - \int f d\mu_X \right| \ll \delta + \delta^{-5}r^7T^{-1/3+\gamma} + (\delta^{-2}r^{-1})^\beta,$$

where the error terms come from, in that order, Lemma 1.3 and Claim 3.4, the contribution of the problem intervals $I_{q,a}$ on which the sequence $\xi(1 + \gamma)M^\gamma n$ does not equidistribute in the periodic horocycle, and the comparison with $\int f d\mu_X$ on the good intervals. Setting $\delta = r^{-1/10}$ takes care of the first and third terms, while, recalling that $r \ll T^{4\varepsilon}$, we can control the second term by setting $\varepsilon = 1/100$. This concludes the proof of Proposition 3.1 (and thus also the proof of Theorem 1.1) with only Claim 3.4 left to be shown. $\square$

Proof of Claim 3.4. To show this claim, we use the following simple lemma, whose proof is left to the reader as an exercise.

LEMMA 3.5. Let $I \subset \mathbb{R}$ be an open interval and let $G: I \rightarrow \mathbb{R}$ be continuously differentiable such that $0 < c \leq |G'(t)| \leq C$ for all $t \in I$. Let $\theta > 0$ and let $a_1 < b_1 < a_2 < \dots < a_{n-1} < b_{n-1} < a_n$ be real numbers with the property that $b_i - a_i \leq \theta(a_{i+1} - b_i)$ for all $1 \leq i < n$. Then, for $E := (a_1, b_1) \cup \dots \cup (a_{n-1}, b_{n-1})$,

$$|\{t \in I : G(t) \in E\}| \leq 2\theta Cc^{-1}|I|$$

provided that $|I| \geq \theta Cc^{-1}$.

To apply this to the function

$$G(t) = \frac{t^\gamma R}{(t^{1+\gamma} - W)^2 + 1} = t^\gamma y$$

in which we are interested, we need to calculate its derivative. We see that

$$\frac{dy}{dt}(t) = -\frac{2(1+\gamma)t^\gamma(t^{1+\gamma} - W)R}{((t^{1+\gamma} - W)^2 + 1)^2} = -\frac{2y(1+\gamma)t^\gamma(t^{1+\gamma} - W)}{(t^{1+\gamma} - W)^2 + 1}$$

and thus

$$G'(t) = yt^{\gamma-1} \left(\gamma - \frac{2(1+\gamma)t^{1+\gamma}(t^{1+\gamma} - W)}{(t^{1+\gamma} - W)^2 + 1} \right).$$

We recall that in Lemma 1.3, we exclude an interval J_0 of proportion δ to assure $r^{-1} \ll y \ll \delta^{-2}r^{-1}$. We also exclude a set J_1 comprising two intervals of proportion δ to assure $t \geq \delta T$ and $|W/t^{1+\gamma} - 1| \geq \delta$. This assures that $r^{-1}T^{\gamma-1} \ll yt^{\gamma-1} \ll \delta^{-3}r^{-1}T^{\gamma-1}$ on the range $[0, T] \setminus (J_0 \cup J_1)$. If we can bound the expression in the bracket in a similar manner up to factors of powers of δ^{-1} , the claim will follow from Lemma 3.5.

To do this, we note that for $t \in [0, T] \setminus J_1$,

$$\left| \frac{1}{(t^{1+\gamma} - W)^2 + 1} - \frac{1}{(t^{1+\gamma} - W)^2} \right| = O(\delta^{-4}T^{-4(1+\gamma)}),$$

which implies

$$G'(t) = yt^{\gamma-1} \left(\gamma + \frac{2(1+\gamma)}{W/t^{1+\gamma} - 1} + O(\delta^{-4}T^{-2}) \right).$$

We set $J_2 := \{t : |W/t^{1+\gamma} - (1 - (2+\gamma)/\gamma)| \geq \delta\}$, which is the interval of proportion δ on which the second term roughly cancels out the first. We then have that

$$\delta \ll \gamma \left| \frac{W}{t^{1+\gamma}} - 1 \right|^{-1} \left| \frac{W}{t^{1+\gamma}} - 1 + \frac{(2+\gamma)}{\gamma} \right| = \left| \gamma + \frac{2(1+\gamma)}{W/t^{1+\gamma} - 1} \right| \ll \delta^{-1}$$

on $[0, T] \setminus (J_1 \cup J_2)$, which implies that

$$\delta r^{-1}T^{\gamma-1} \ll |G'(t)| \ll \delta^{-4}r^{-1}T^{\gamma-1}$$

on $[0, T] \setminus (J_0 \cup J_1 \cup J_2)$. We can now apply Lemma 3.5 to each of the intervals left. Recalling that each problem interval $I_{q,a}$ is of length $rT^{-1/3+\gamma}$ and the gap between any two successive intervals is of size at least $0.9r^{-6}$, we find that

$$\frac{1}{T} |\{t \in [0, T] \setminus (J_0 \cup J_1 \cup J_2) : (1+\gamma)G(t) \in E\}| \ll \delta^{-5}r^7T^{-1/3+\gamma},$$

where, as before, $E = \bigcup_{q \leq r^3, (a,q)=1} I_{q,a}$. This shows Claim 3.4, which was the last missing piece in the proof of Theorem 1.1. $\square$

Acknowledgements. This is a follow-up paper to [7], which is based on the master's thesis I did at the Hebrew University of Jerusalem in 2020. As such, I am thankful for the support by my thesis advisor Tamar Ziegler and by Elon Lindenstrauss, who also suggested that the result in the present paper should be achievable with the ideas in [7]. I thank my PhD supervisor Péter Varjú for giving me the freedom to finish the work on these two papers while doing my PhD with him. Above all, I am grateful to Adrián Ubis, who suggested the argument used in the proof of Claim 3.3 in his review of the previous paper, simplifying the proof in [7] considerably. Without getting this new perspective on the material two years later, I would not even have thought of revisiting the problem solved in this paper. The author received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 803711)

REFERENCES

- [1] J. Bourgain. Pointwise ergodic theorems for arithmetic sets. *Publ. Math. Inst. Hautes Études Sci.* **69** (1989), 5–45; with an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.
- [2] M. Einsiedler and T. Ward. *Ergodic Theory: With a View Towards Number Theory (Graduate Texts in Mathematics, 259)*. Springer, London, 2011.
- [3] D. A. Goldston and C. Y. Yildirim. Higher correlations of divisor sums related to primes I: triple correlations. *Integers* **3** (2003), A05.
- [4] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)* **167** (2004), 481–547.
- [5] T. McAdam. Almost-primes in horospherical flows on the space of lattices. *J. Mod. Dyn.* **15** (2018), 277–327.
- [6] P. Sarnak and A. Ubis. The horocycle flow at prime times. *J. Math. Pures Appl. (9)* **103** (2011), 575–618.
- [7] L. Streck. Non-concentration of primes in $\Gamma \backslash PSL_2(\mathbb{R})$. *Israel J. Math.* doi.org/10.1007/s11856-025-2744-z. Published online 27 March 2025.
- [8] A. Strömbergsson. On the deviation of ergodic averages for horocycle flows. *J. Mod. Dyn.* **7**(2) (2013), 291–328.
- [9] A. Venkatesh. Sparse equidistribution problems, period bounds and subconvexity. *Ann. of Math. (2)* **172**(2) (2010), 989–1094.
- [10] C. Zheng. Sparse equidistribution of unipotent orbits in finite-volume quotients. *J. Mod. Dyn.* **10** (2016), 1–21.
- [11] C. Zheng. On the density of some sparse horocycles. *Proc. Indian Acad. Sci. Math. Sci.* **134**(1) (2024), 0006.