ALMOST-PRIME TIMES IN HOROSPHERICAL FLOWS
ON THE SPACE OF LATTICES

TAYLOR MCADAM
(Communicated by Dmitry Kleinbock)

ABSTRACT. An integer is called almost-prime if it has fewer than a fixed number of prime factors. In this paper, we study the asymptotic distribution of almost-prime entries in horospherical flows on \( \Gamma \backslash \mathrm{SL}_n(\mathbb{R}) \), where \( \Gamma \) is either \( \mathrm{SL}_n(\mathbb{Z}) \) or a cocompact lattice. In the cocompact case, we obtain a result that implies density for almost-primes in horospherical flows where the number of prime factors is independent of basepoint, and in the space of lattices we show the density of almost-primes in abelian horospherical orbits of points satisfying a certain Diophantine condition. Along the way we give an effective equidistribution result for arbitrary horospherical flows on the space of lattices, as well as an effective rate for the equidistribution of arithmetic progressions in abelian horospherical flows.

1. INTRODUCTION

There is an intimate connection between number theory and dynamics on homogeneous spaces, from applications in Diophantine approximation [38, 40] to Margulis’s famous proof of the Oppenheim conjecture [48, 49]. One topic that has long interested both number theorists and dynamicists is the distribution of orbit points at sparse sequences of times, especially prime times. In [6] Bourgain proved the remarkable result that ergodic averages over primes converge almost-everywhere, however this is often not sufficient for applications in number theory which require information about a specific orbit or about every orbit for a given system. For example, Sarnak’s Möbius disjointness conjecture seeks to formalize the heuristic that primes are essentially randomly distributed by positing that the Möbius function

\[
\mu(n) = \begin{cases} 
0 & \text{if } n \text{ is not squarefree} \\
(-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes}
\end{cases}
\]

is asymptotically uncorrelated with any sequence of the form \( f(T^nx) \) for \( x \) in a compact metric space \( X \), \( f \in C(X) \), and \( T : X \to X \) a continuous map of zero topological entropy. The conjecture has been established in a variety of settings,
including affine transformations on compact abelian groups \([21, 44]\), translations on compact nilmanifolds \([31]\), and horocyclic flows on \(\Gamma \backslash \text{SL}_2(\mathbb{R})\) for \(\Gamma\) a lattice \([7]\) (generalized in \([54]\)), among others.

In \([31]\), Green and Tao are able to go further, proving an equidistribution result for prime times in nilflows. Similarly, it is conjectured that prime times in horocyclic (and more generally, horospherical) flows are equidistributed in the orbit closures containing them (see the conjecture of Margulis in \([29]\)). In this direction, Sarnak and Ubis have shown that primes are dense in a set of positive measure for equidistributing horocycle orbits on \(\text{SL}_2(\mathbb{R})/\text{SL}_2(\mathbb{Z})\) and that almost-primes (i.e., integers having fewer than a fixed number of prime factors) are dense in the whole space \([61]\). The purpose of this work is to provide a partial generalization to \([61]\) in the higher dimensional, horospherical setting.

Horospherical flows are a type of dynamical system arising naturally in the study of homogeneous spaces. A subgroup of a Lie group \(G\) is said to be horospherical if it is contracted (conversely, expanded) under iteration of the adjoint action of some element of \(G\) (see Section 2.2 for a more precise definition). It can be shown that any horospherical subgroup is unipotent, although not every unipotent subgroup can be realized as the horospherical subgroup corresponding to an element of \(G\). In general, horospherical flows are easier to study than arbitrary unipotent flows, as the mixing action of the expanding element can be used to provide dynamical information about a given horospherical flow.

Actions by horospherical and unipotent subgroups have been studied extensively. Hedlund proved in \([34]\) that the horocycle flow on \(\Gamma \backslash \text{SL}_2(\mathbb{R})\) for \(\Gamma\) cocompact is minimal, and Furstenberg later showed it to be uniquely ergodic \([28]\). These results were extended in \([70]\) and \([26]\) to more general horospherical flows on compact quotients of suitable Lie groups. For \(\Gamma\) non-uniform, Margulis proved that orbits of unipotent (hence horospherical) flows cannot diverge to infinity \([46]\), which was later refined in Dani’s nondivergence theorem \([17, 18]\). Dani also showed in \([15]\) (for the case of \(\Gamma \backslash \text{SL}_2(\mathbb{R})\) noncompact) and \([16, 19]\) (for more general noncompact homogeneous spaces) that horocyclic/horospherical flows have nice (finite volume, homogeneous submanifold) orbit closures and that every ergodic probability measure invariant under such a flow is the natural Lebesgue measure on some such orbit closure. In a series of breakthrough papers culminating in \([57]\) and summarized in \([58]\), Ratner resolved conjectures of Raghunathan and Dani by giving a complete description of unipotent orbit closures and unipotent-invariant measures on homogeneous spaces. Although similar in form to Dani’s theorems for horospherical flows, Ratner’s theorems require a very different method of proof that cannot be easily modified to provide a rate of convergence.

Recall that a subset of an orbit is said to equidistribute with respect to a given probability measure if it spends the expected amount of time in measurable subsets of the space. Often in applications to number theory it is important that equidistribution results be effective—that is, that the rate of convergence...
is known. Many of the above results for horospherical flows have since been effec-
tivized, in particular with polynomial rates (see [20, 38, 61, 11, 27, 64]). Many
authors have also considered the effective equidistribution of periods (that it,
closed horospherical orbits) in a variety of settings (e.g., [60, 63, 37, 42, 14],
although this list is by no means complete). In studying both periods and long
pieces of generic horospherical orbits, one can make use of the “thickening”
argument developed by Margulis in his thesis [45], which uses a known rate of
mixing for the semisimple flow with respect to which the given subgroup is horo-
spherical along with the expansion property to get a rate for the horospherical
flow. Some interesting effective results which do not pertain to horospherical
flows include [24, 31, 65, 23].

Many applications in number theory need the error in relevant approxima-
tions to be quantitatively controlled, requiring dynamical ingredients in such
proofs to be effective. For example, Green and Tao are able to prove their re-
sult about the equidistribution of primes in nilflows because they have strong
rates of convergence for the M"obius disjointness phenomenon, whereas in [7]
Bourgain–Sarnak–Ziegler do not get such a result for the horocycle flow because
their proof of M"obius disjointness relies on Ratner’s joining classification theo-
rem, which is not effective. Another example in the area of sparse equidistribu-
tion is Venkatesh’s use of effective equidistribution of the horocycle flow to pro-
vide a partial solution to a conjecture of Shah [62] by showing that orbits along
sequences of times of the form \( \{n^{1+\gamma}\}_{n\in\mathbb{N}} \) equidistribute in compact quotients
of \( SL_2(\mathbb{R}) \) for small \( \gamma \) [71]. In their work on prime times in the horocycle flow,
Sarnak and Ubis also rely on strong effective results that they develop, includ-
ing a quantitative version of Dani’s equidistribution theorem for the discrete
horocycle flow [61].

In this paper, we are interested in the asymptotic distribution of almost-
primes in horospherical flows on the space of lattices and on compact quotients
of \( SL_n(\mathbb{R}) \). Our main results are summarized in the following two theorems.

**Theorem 1.** Let \( \Gamma < SL_n(\mathbb{R}) \) be a cocompact lattice and \( u(t) \) be a horospherical
flow on \( \Gamma \backslash SL_n(\mathbb{R}) \) of dimension \( d \). Then there exists a constant \( M \) (depending
only on \( n \), \( d \), and \( \Gamma \)) such that for any \( x \in \Gamma \backslash SL_n(\mathbb{R}) \), the set
\[
\{ x u(k_1, k_2, \ldots, k_d) \mid k_i \in \mathbb{Z} \text{ has fewer than } M \text{ prime factors} \}
\]
is dense in \( \Gamma \backslash SL_n(\mathbb{R}) \).

In fact, the method of proof yields an effective density statement which de-
scribes how long it takes for almost-prime times in an orbit to hit any ball of
fixed radius (see the remark at the end of Section 5.1). The constant \( M \) depends
polynomially on \( n \) and \( d \) (see equation (56) for an explicit formula). The de-
pendence of \( M \) on \( \Gamma \) arises from the spectral gap of the action of \( SL_n(\mathbb{R}) \) on
\( \Gamma \backslash SL_n(\mathbb{R}) \). Since \( SL_n(\mathbb{R}) \) has Kazhdan’s property \((\Gamma)\) for \( n \geq 3 \), the dependence
on \( \Gamma \) can be removed in that setting, or in any setting where \( \Gamma \) varies over a
family of lattices with bounded spectral gap, such as when \( \Gamma \) is a congruence
lattice in \( n = 2 \).
In the noncompact case, the number of prime factors allowed in our almost-primes will depend on the rate of equidistribution for the continuous flow, which in turn depends on the basepoint due to the existence of proper orbit closures. The following definition provides a condition under which the basepoint has a persistently good rate of equidistribution.

**Definition 2.** For a horospherical flow $u(t)$ on $\text{SL}_n(\mathbb{Z}) \backslash \text{SL}_n(\mathbb{R})$, we say that $x = \text{SL}_n(\mathbb{Z})g$ is strongly polynomially $\delta$-Diophantine if there exists some sequence $T_i \to \infty$ as $i \to \infty$ such that

$$\inf_{w \in \Lambda^j(\mathbb{Z}^n) \backslash \{0\}} \sup_{t \in [0,T_i]^d} \left| w g u(t) \right| > T_i^\delta$$

for all $i \in \mathbb{N}$.

The definition of $\Lambda^j(\mathbb{Z}^n)$ and the norm are given in Section 2.8, and further motivation for this definition is provided in Section 5.2. See also [43, Definition 3.1] and the related discussion of its algebraic and dynamical implications.

With this definition, we have the following theorem.

**Theorem 3.** Let $u(t)$ be an abelian horospherical flow on $\text{SL}_n(\mathbb{Z}) \backslash \text{SL}_n(\mathbb{R})$ of dimension $d$, and let $x \in \text{SL}_n(\mathbb{Z}) \backslash \text{SL}_n(\mathbb{R})$ be strongly polynomially $\delta$-Diophantine for some $\delta > 0$. Then there exists a constant $M_\delta$ (depending on $\delta$, $n$, and $d$) such that

$$\left\{ xu(k_1, k_2, \ldots, k_d) \mid k_i \in \mathbb{Z} \text{ has fewer than } M_\delta \text{ prime factors} \right\}$$

is dense in $\text{SL}_n(\mathbb{Z}) \backslash \text{SL}_n(\mathbb{R})$.

The constant $M_\delta$ depends polynomially on $n$ and $d$ and inversely on $\delta$ (see equation (62) for an explicit formula).

A brief outline of the paper is as follows:

In Section 2, we establish the basic notation that will be used throughout the paper and introduce the key facts and theorems that we use in our analysis. We also prove a small corollary of the nondivergence theorem in [38] that applies to the particular setting of this paper.

In Section 3, we prove an effective equidistribution result for long orbits of arbitrary horospherical flows on the space of lattices. The proof makes use of the “thickening” argument of Margulis, leveraging the exponential mixing properties of the subgroup with respect to which the flow of interest is horospherical, which is itself a consequence of a spectral gap. The main result in this section is probably not surprising to experts, but the author was unable to locate a result in the literature that is stated in the way presented here.

In Section 4, we use the theorem from the previous section to derive an effective bound for equidistribution along multivariate arithmetic sequences of numbers.

\[1\] For Theorem 1, we first prove the result for abelian horospherical flows and then extend the result to arbitrary horospherical flows. However, the method we use to do this cannot be applied in the noncompact setting (see the remark at the end of Section 5.2). However, it is possible to adapt the methods in this paper in order to prove the result for arbitrary horospherical flows, which is the topic of a forthcoming paper.
entries in abelian horospherical flows on the space of lattices. In this result, we allow the arithmetic sequences in different coordinates to have different spacing, although we will not need it for Section 5. The techniques used in this section are heavily inspired by [71, Section 3] and also make use of the spectral gap, as well as some Fourier analysis and other analytic techniques.

In Section 5, we use the bound along arithmetic sequences as well as a combinatorial sieve theorem to obtain an upper and lower bound on averages over almost-prime entries in horospherical flows. We start with the case of $\Gamma$ cocompact, for which we obtain a result that implies Theorem 1 above. We then move to the case $\Gamma = \text{SL}_n(\mathbb{Z})$, where we prove a similar result for almost-primes in the orbits of points satisfying a strongly polynomially Diophantine condition, giving us Theorem 3.

Finally, in Section 6, we make some closing remarks and indicate possible extensions and areas for future research.

2. Notation and preliminaries

2.1. Some basic notation. Let $G = \text{SL}_n(\mathbb{R})$ for $n \geq 2$. Throughout most of this document, $\Gamma$ will denote $\text{SL}_n(\mathbb{Z})$, but we will also discuss the case where $\Gamma \leq G$ is a cocompact lattice. We are interested in the right actions of certain subgroups of $G$ on the right coset space $X = \Gamma \backslash G$.

The Haar measure $m_G$ on $G$ projects to a $G$-invariant measure $m_X$ on $X$. In this document, we will always take $m_X$ and $m_G$ to be normalized so that $m_X$ is a probability measure and so that the measure of a small set in $G$ equals the measure of its projection in $X$. We will use $|\cdot|$ to denote the standard Lebesgue measure on $\mathbb{R}^d$ and $dt$ to denote the differential with respect to Lebesgue measure for $t \in \mathbb{R}^d$.

We will use gothic letters to represent the Lie algebra of a Lie group (e.g., $\mathfrak{g}$ is the Lie algebra of $G$). Fix an inner product on $\mathfrak{g}$. This extends to a Riemannian metric on $G$ via left translation, which defines a left-invariant metric $d_G$ and a left-invariant volume form, which (by uniqueness) coincides with the Haar measure on $G$ up to scaling. This then induces a metric $d_X$ on $X$ of the form

$$d_X(\Gamma g_1, \Gamma g_2) = \inf_{\gamma_1, \gamma_2 \in \Gamma} d_G(\gamma_1 g_1, \gamma_2 g_2) = \inf_{\gamma \in \Gamma} d_G(g_1, \gamma g_2).$$

The same construction can be used to define a left-invariant metric $d_H$ for any subgroup $H \leq G$ by restricting the inner product to $h \subseteq \mathfrak{g}$. Note, however, that in general $d_H \neq d_G|_H$. Instead, we have that $d_G(h_1, h_2) \leq d_H(h_1, h_2)$ for $h_1, h_2 \in H$, since the infimum used to define the distance $d_G$ is taken over a larger set than in $d_H$. We will use the notation $B^G_r(h)$ to denote a ball of radius $r$ with respect to the metric $d_H$ around a point $h \in H$ (this is to distinguish these balls from the sets $B_T$ that we will define in Section 2.2). Also observe that every point has a neighborhood in which the left-invariant metric is Lipschitz equivalent to the metric derived from any matrix norm on $\text{Mat}_{n \times n}(\mathbb{R})$ (see [25, Lemma 9.12] for details).
Define the adjoint representation of $g \in G$ as the map $\text{Ad}_g : g \to g$ given by $Y \mapsto g Y g^{-1}$ for $Y \in g$.

In considering equidistribution questions, our space of test functions will be $C_c^\infty(X)$, the set of smooth, compactly supported (real- or complex-valued) functions on $X$. Define the action of $G$ on this space by $[g \cdot f](x) = f(xg^{-1})$ for $g \in G$ and $f \in C_c^\infty(X)$.

Finally, we will use the notation $a \ll b$ to indicate that $a$ is less than a fixed constant times $b$ and $a \asymp b$ to indicate that $a \ll b$ and $b \ll a$. In general, the implied constants may depend on $n$ and on the data of the dynamical system (more specifically, on $d$, the dimension of the horospherical subgroup). Any additional dependence of the constants will be indicated by a subscript (e.g., $\ll_f$ indicates that the implicit constant may depend on $n$, $d$, and $f$). In principle, the constants may also depend on the lattice $\Gamma$, although since we are primarily considering $\Gamma = \text{SL}_n(\mathbb{Z})$, we will not indicate this dependence with a subscript when $\Gamma$ is understood to be fixed in this way. We will also use the standard notation $O(f(x))$ to indicate a function whose absolute value is bounded by a constant times $|f(x)|$ as $x \to \infty$, where as before the constant may depend on $n$ and $d$, and any additional dependence will be indicated with a subscript.

2.2. Horospherical subgroups. A subgroup $U$ of $G$ is (expanding) horospherical with respect to an element $g \in G$ if $U = \{ u \in G | g^{-j}ug^j \to e \text{ as } j \to \infty \}$, where $e$ is the identity. In other words, elements of $U$ are contracted under conjugation by $g^{-1}$ and expanded under conjugation by $g$.

Define the one-parameter subgroup \{a_t\} for $t \in \mathbb{R} \in G$ by

$$a_t = \exp \left( t \, \text{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_N, \ldots, \lambda_N) \right)$$

where $m_1 + \cdots + m_N = n$ and $m_1 \lambda_1 + \cdots + m_N \lambda_N = 0$. Without loss of generality we may also assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$.

Let $U$ denote the block-upper-triangular unipotent subgroup given by

$$U = \left\{ \begin{pmatrix} I_{m_1} & & & * \\ 0 & I_{m_2} & & \\ & & \ddots & \\ & & & I_{m_N} \end{pmatrix} \right\}$$

where $I_m$ is the $m \times m$ identity matrix. Notice that $U$ is the horospherical subgroup corresponding to $a_t$ for $t > 0$. Similarly, define the contracting subgroup $U^-$ by

$$U^- = \left\{ \begin{pmatrix} I_{m_1} & & 0 \\ 0 & I_{m_2} & \ddots \\ & & \ddots & I_{m_N} & * \\ & & & 0 & I_{m_N} \end{pmatrix} \right\}$$
which is horospherical with respect to \( a_t \) for \( t < 0 \), and define \( U^0 \) to be the centralizer of \( a_t \) (\( t \neq 0 \)), given by

\[
U^0 = \left\{ \begin{pmatrix} B_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B_{m_N} \end{pmatrix} : B_i \in \text{GL}_{m_i}(\mathbb{R}), \det B_1 \cdots \det B_N = 1 \right\}.
\]

Let \( d_0 = \sum_{i=1}^{N} m_i^2 \) and note that \( d := \dim U = \dim U^- = \frac{1}{2}(n^2 - d_0) \) and \( \dim U^0 = d_0 - 1 \). All horospherical subgroups of \( G = \text{SL}_n(\mathbb{R}) \) are conjugate to a subgroup of the form given in (2), so we restrict our attention to \( U \) of this form.

Observe that \( U \) is diffeomorphic to \( \mathbb{R}^d \) through identification \( t \mapsto u(t) \) of the coordinates of \( \mathbb{R}^d \) with the matrix entries in the upper-right corner of (2).\(^2\) Note, however, that \( U \) and \( \mathbb{R}^d \) are only isomorphic as groups in the case that \( U \) is abelian, which occurs when \( a_t \) has precisely two eigenvalues.

The bi-invariant Haar measure \( m_{U} \) on \( U \) is the pushforward of Lebesgue measure on \( \mathbb{R}^d \) under this identification, and we may normalize it so that \( u([0,1]^d) \) has unit measure. Define an expanding family of Følner sets in \( U \) by

\[
B_T = a_{\text{log } T} u([0,1]^d) a_{-\text{log } T}
\]

for \( T \in \mathbb{R} \). One may verify that the preimage of \( B_T \) in \( \mathbb{R}^d \) is given by a box where \( x_k \in [0, T^{\lambda_i - \lambda_j}] \) for \( i < j \) if the coordinate \( x_k \) is mapped to the \((i,j)\)-block of (2) under our identification. Hence, \( m_{U}(B_T) = T^p \), where \( p = \sum_{i \neq j} m_i m_j (\lambda_i - \lambda_j) \).

2.3. Measure decomposition. The product map \( U \times U^0 \times U^- \to G \) given by \((u,u^0,u^-) \mapsto uu^0u^-\) is a biregular map onto a Zariski open dense subset of \( G \) (see [50, Proposition 2.7]). In particular, if we let \( H = U^0 U^- \), this means that \( m_G(G \sim UH) = 0 \) and that the product map \((u,h) \mapsto uh\) is open and continuous. Additionally, it is not difficult to see that \( U \cap H = \{e\} \). Then by virtue of the fact that \( G \) is unimodular, we have that \( m_G \) restricted to \( UH \) is proportional to the pushforward of \( m_U \times m_H^t \) by the product map, where \( m_H^t \) is the right Haar measure on \( H \) (see, e.g., [25, Lemma 11.31] or [41, Theorem 8.32]). Note that we could equivalently use the left Haar measure on \( H \) and multiply by the modular function \( \Delta_H \), but for convenience of notation we will use the right Haar measure.

2.4. Sobolev norms. Fix a basis \( \mathcal{B} \) for the Lie algebra \( \mathfrak{g} \) of \( G \). Define the (right) differentiation action of \( \mathfrak{g} \) on \( C_c^\infty(X) \) by

\[
Yf(x) = \frac{d}{dt} f(x \exp(tY)) \bigg|_{t=0}
\]

\(^2\) One could also use the more standard map \( u(t) = \exp(it) \), where \( t : \mathbb{R}^d \to \mathbb{u} \) is any identification of \( \mathbb{R}^d \) with the Lie algebra \( \mathbb{u} \) of \( U \). We have chosen to use the former embedding for the ease of certain computations and because we will later restrict our attention to abelian horosphericals, for which the two maps coincide (up to scaling and permutations of the coordinates). However, whichever map is used does not substantively change the results presented here.
for $Y \in \mathcal{B}$ and $f \in C^\infty_c(X)$. Higher order derivatives of $f$ can then be expressed as monomials in the basis $\mathcal{B}$.

For $p \in [1, \infty)$ and $\ell \in \mathbb{N}$, the $(p, \ell)$-Sobolev norm of $f \in C^\infty_c(X)$ simultaneously controls the $L^p$-norm of all derivatives of $f$ up to order $\ell$. More precisely, let

$$\mathcal{S}_{p,\ell}(f) = \sum_{\deg(D) \leq \ell} \|Df\|_{L^p(X)}$$

where $D$ ranges over all monomials in $\mathcal{B}$ of degree $\leq \ell$. Observe that the Sobolev norm can be defined similarly for $C^\infty_c(G)$ and $C^\infty_c(H)$ where $H \subseteq G$, given a choice of basis for $\mathfrak{h} \subseteq \mathfrak{g}$.

We will only require the $(2, \ell)$- and $(\infty, \ell)$-Sobolev norms. When $p = 2$, we will drop the notation, letting $\mathcal{S}_\ell(f) = \mathcal{S}_{2,\ell}(f)$. When needed, we will use a superscript $\mathcal{S}_X$ to indicate a Sobolev norm for functions defined on $X$.

Some useful properties of these norms are as follows (see [71] or [37]):

(i) For $X$ a probability space, $f \in C^\infty_c(X)$, $p \in [1, \infty]$, and $\ell \leq \ell'$, $\mathcal{S}_{p,\ell}(f) \leq \mathcal{S}_{p,\ell'}(f)$.

(ii) For $f_1, f_2 \in C^\infty_c(X)$, $\mathcal{S}_{\infty,\ell}(f_1 + f_2) \leq \mathcal{S}_{\infty,\ell}(f_1) + \mathcal{S}_{\infty,\ell}(f_2)$.

(iii) For $f \in C^\infty_c(X)$ and $g \in G$, $\mathcal{S}_{\infty,\ell}(g \cdot f) \leq \|\text{Ad}_{g^{-1}}\| \mathcal{S}_{\infty,\ell}(f)$, where $\|\cdot\|$ is the operator norm on linear functions $g \to g$.

(iv) Let $L \subset G$ be compact. For $f \in C^\infty_c(X)$, $x \in X$,

$$|f(xg) - f(x)| \leq \mathcal{S}_{\infty,1}(f)\mathcal{D}_G(g, e)$$

for all $g \in L$.

(v) Let $X$ and $Y$ be Riemannian manifolds. For $f_1 \in C^\infty_c(X)$ and $f_2 \in C^\infty_c(Y)$,

$$\mathcal{S}_\ell^{X,Y}(f_1 \cdot f_2) \leq \mathcal{S}_{X,Y}(f_1) \mathcal{S}_\ell^{Y}(f_2).$$

2.5. **Approximation to the identity.** At times we will want to use smooth bump functions with small support as approximations to the identity, but we will need to know that the Sobolev norm of such functions can be controlled. For this we have the following lemma, which can be found in [37].

**Lemma 4** ([37, Lemma 2.4.7 (b)]). Let $Y$ be a Riemannian manifold of dimension $k$. Then for any $0 < r < 1$ and $y \in Y$, there exists a function $\theta \in C^\infty_c(Y)$ such that

(i) $\theta \geq 0$

(ii) $\text{supp} (\theta) \subseteq B_r^Y(y)$

(iii) $\int_Y \theta = 1$

(iv) $\mathcal{S}_\ell^{Y}(\theta) \leq r^{-\ell(k/2)}$.

---

3 The choice of the basis $\mathcal{B}$ is unimportant in the sense that choosing a different basis will lead to an equivalent norm. Likewise, we could use any norm on the components $\|Df\|_{L^p(X)}$ (here we have used the $l^1$-norm), but as all such norms are equivalent, the choice is unimportant.
2.6. The space of unimodular lattices. For $\Gamma = \text{SL}_n(\mathbb{Z})$, $X$ is noncompact and can be understood as the space of unimodular lattices (that is, lattices of covolume one) in $\mathbb{R}^n$ under the identification $\Gamma g \rightarrow \mathbb{Z}^n g$.

For $0 < \epsilon \leq 1$, define $L_\epsilon$ to be the set of lattices in $X = \text{SL}_n(\mathbb{Z}) \backslash \text{SL}_n(\mathbb{R})$ with no nonzero vectors shorter than $\epsilon$. That is, let,

$$L_\epsilon = \{ \Gamma g \in X | \|vg\| \geq \epsilon \text{ for all } v \in \mathbb{Z}^n \sim \{0\} \}$$

where the norm above can be taken to be any norm on $\mathbb{R}^n$, but for convenience we will use the max norm. By Mahler’s Compactness Criterion, $L_\epsilon$ is a compact set (for details and a proof, see [56, Corollary 10.9], [2, Theorem 5.3.2], or [25, Theorem 11.33]).

2.7. Injectivity radius. Given small $\epsilon > 0$, we want to find a radius $r > 0$ (depending on $\epsilon$) such that projection at $x$, given by

$$\pi_x : B_r^G(e) \rightarrow B_r^X(x)$$

$$g \rightarrow xg$$

is injective for all $x \in L_\epsilon$ (in fact, it is not difficult to see from the definition of the metric on $X$ that this will be an isometry). For this, we have the following lemma, which can be found in [40] or in a more general setting in [3] (see the proof of Lemma 11.2).

**Lemma 5** ([40, Proposition 3.5]). There exist constants $c_1, c_2 > 0$ (depending only on $n$) such that for any $0 < \epsilon < c_1$, the projection map $\pi_x : B_r^G(e) \rightarrow B_r^X(x)$ is injective for all $x \in L_\epsilon$, where $r = c_2\epsilon^n$.

2.8. Quantitative nondivergence. Let $\{e_1, \ldots, e_n\}$ be the standard basis on $\mathbb{R}^n$. Let $e_I = e_{i_1} \wedge \cdots \wedge e_{i_j}$ for a multi-index $I = (i_1, \ldots, i_j)$, where $1 \leq i_1 < \cdots < i_j \leq n$. Then $\{e_I\}$ is a basis for $\Lambda^j(\mathbb{R}^n)$, the $j$th exterior power of $\mathbb{R}^n$. Define the norm of $v = \sum_i w_i e_i \in \Lambda^j(\mathbb{R}^n)$ to be $\|v\| = \max_I |w_I|$. Denote by $\Lambda^j(\mathbb{Z}^n)$ the discrete subset of $\Lambda^j(\mathbb{R}^n)$ composed of linear combinations of basis vectors with integer coefficients. Notice that $g \in \text{GL}_n(\mathbb{R})$ acts on $\Lambda^j(\mathbb{R}^n)$ on the right by

$$(e_{i_1} \wedge \cdots \wedge e_{i_j}) g = (e_{i_1} g) \wedge \cdots \wedge (e_{i_j} g)$$

where the action extends to all of $\Lambda^j(\mathbb{Z}^n)$ via linearity. A vector in $\Lambda^j(\mathbb{Z}^n)$ is called primitive if it is not a multiple of any other vector in $\Lambda^j(\mathbb{Z}^n)$.

The following theorem quantitatively describes how often certain polynomial maps from $\mathbb{R}^d$ to $X$ land inside a compact set $L_\epsilon$. This is a special case of [38, Theorem 5.2], which itself extends results of [18] and [47]. The original theorem is stated for much more general $(C, \alpha)$-good functions, but we will only need the version below, which can be found as [40, Theorem 3.1].

In fact, this can be improved due to an old result called the Remez inequality (see [59, 10]), so that the bound becomes $(\epsilon/\rho)^{1/(n-1)}|B|$. We have decided to use the following less sharp bound in our treatment because we feel it may be more familiar to the reader and because it would take some time reformulate the Remez inequality for application to the space of lattices, which is not a primary concern of this paper.
Then the right action of $a_t$ on either the left or the right only changes matrix entries by scaling, each entry in the upper-right corner of $a_{\log}T u(t) a_{-\log}T a_{\log}R$ only depends linearly on a single coordinate of $t$. This means that for any matrix $g_0$, all entries of $\xi(t) = g_0 a_{\log}T u(t) a_{-\log}T a_{\log}R$ will be affine, satisfying condition (i).

Notice that $e_k a_{\log}R = R^{\lambda_k} e_k$ if $\lambda_k$ is the $k^{th}$ eigenvalue in the definition of $a_t$ in (1).

Then the right action of $a_{\log}R$ scales $e_1 \wedge \cdots \wedge e_j \in \Lambda^j(\mathbb{R}^d)$ by the product of all such corresponding factors. Since $R > 1$, the most $a_{\log}R$ can therefore contract any basis element is by the product of all scaling factors corresponding to negative eigenvalues of (1), that is, by $R^{-q}$, where $q = \sum_{i < 0} -m_i \lambda_i$. It then follows from the definition of the norm that

$$\|w_{\log}R\| \geq R^{-q} \|w\|$$

for any $w \in \Lambda^j(\mathbb{R}^d) \sim \{0\}$ and $j \in \{1, \ldots, n - 1\}$. Then for $\rho = \min(1, R_0/R^d)$, we have $0 < \rho \leq 1$ and also

$$\sup_{t \in [0,1]^d} \|w_{\xi(t)}\| = \sup_{t \in [0,1]^d} \|w_{g_0 a_{\log}T u(t) a_{-\log}T a_{\log}R}\| \geq R^{-q} \sup_{t \in [0,1]^d} \|w_{g_0 a_{\log}T u(t) a_{-\log}T}\| \geq R_0/R^d \geq \rho$$
for $j \in \{1, \ldots, n-1\}$ and primitive $u \in \Lambda^j(\mathbb{Z}^n) \sim \{0\}$, satisfying condition (ii).

Hence, by Theorem 6, we have

$$\left| \{ t \in [0, 1]^d : |\Gamma \xi(t) \notin L_c \} \right| \leq (c/\rho)^{1/d(n-1)}. \quad \square$$

### 2.9. Decay of matrix coefficients.

In order to obtain effective rates of equidistribution in Sections 3 and 4, we will need to use results on the effective decay of matrix coefficients.

Estimates of this type have a long and rich history, including Selberg’s celebrated 3/16 theorem for congruence quotients of $SL_2(\mathbb{Z})$, Kazhdan’s property (T), and works of Harish-Chandra, Cowling, Haagerup, Howe, and Oh. Far reaching extensions of Selberg’s work are also in place thanks to works of Jacques–Langlands, Burger–Sarnak, and Clozel. Our formulation here is taken from [37] (see [37, 36, 30, 23] for a more comprehensive history and discussion). In [53], Oh gives optimal bounds for $SL_n(\mathbb{R})$, $n \geq 3$.

**Theorem 8** ([37, Corollary 2.4.4]). Let $G = SL_n(\mathbb{R})$ and $\mathcal{X} = \Gamma \backslash G$ for a lattice $\Gamma$. There exists a constant $0 < \tilde{\beta} < 1$ such that for $f_1, f_2 \in C_{c}^{\infty}(X)$ and $g \in G$,

$$\left| \langle g \cdot f_1, f_2 \rangle_{L^2(X)} - \int_X f_1 dm_X \int_X f_2 dm_X \right| \ll e^{-\tilde{\beta} d_{c}(v,g)} \mathcal{J}_\ell(f_1) \mathcal{J}_\ell(f_2)$$

where $\ell$ is the dimension of maximal compact subgroup of $G$. When $n \geq 3$ the constant $\tilde{\beta}$ is independent of the lattice $\Gamma$, and when $n = 2$ it is independent of the lattice if $\Gamma$ is a congruence lattice.

For our specific applications, we have the following corollaries.

**Corollary 9.** Let the setting be as above. There exist $\beta_1, \beta_2 > 0$ such that

(i) For $f_1, f_2 \in C_{c}^{\infty}(X)$ and $t \geq 0$, we have

$$\left| \int_X f_1(xa_t)f_2(x)dm_X(x) - \int_X f_1 dm_X \int_X f_2 dm_X \right| \ll e^{-\beta_1 t} \mathcal{J}_\ell(f_1) \mathcal{J}_\ell(f_2).$$

(ii) For $f \in C_{c}^{\infty}(X)$ and $t \in \mathbb{R}^d$,

$$\left| \langle u(t)f, f \rangle_{L^2(X)} - \int_X f dm_X \right| \ll \max(1, |t|)^{-\beta_2} \mathcal{J}_\ell(f)^2.$$

### 3. Effective equidistribution of horospherical flows

Throughout this section, let $G = SL_n(\mathbb{R})$, $\Gamma = SL_n(\mathbb{Z})$, $X = \Gamma \backslash G$, and $U < G$ a horospherical subgroup of dimension $d$.

The following qualitative equidistribution statement is well known and follows from Ratner’s Theorems, although in this case simpler proofs may be used. As mentioned in the introduction, Dani proved a density result in this setting, and equidistribution results were proved in some special cases prior to Ratner’s work [15, 27, 64]. The method of proof for this general qualitative result is the same as for these special cases, however we were unable to locate an explicit reference for a result of this form which avoids the use of Ratner’s theorem.
Theorem 10. For every $x_0 = \Gamma g_0 \in X$, either

\begin{equation}
\frac{1}{m_U(B_T)} \int_{B_T} f(x_0u)dm_U(u) \xrightarrow{T \to \infty} \int_X f dm_X \quad \forall f \in C_c^\infty(X)
\end{equation}

or

\begin{equation}
\exists j \in \{1, \ldots, n-1\} \text{ and primitive } w \in \Lambda^j(\mathbb{Z}^n) \sim \{0\}
\end{equation}

such that $wg_0u = wg_0 \forall u \in U$.

The subgroup $L = \text{Stab}(wg_0)$ is the well-known intermediate subgroup $U \leq L \leq G$ such that $x_0U \subset x_0L$ with $x_0L$ supporting an $L$-invariant probability measure. We will call a point $x_0 \in X$ generic if it satisfies (4a). By ergodicity of the horospherical flow, the set of generic points has full measure in $X$.

Our main objective in this section is to prove the following quantitative refinement of this theorem.

Theorem 11. There exist constants $\gamma, C > 0$ (depending\footnote{Specifically, $\gamma = 2\beta_1/(n(3n+1)(n-1)^2 + 2)$, where $\beta_1$ is the constant from Corollary 9. If the Remez inequality is used instead of Theorem 6, then $\gamma$ depends only on $n$, with $\gamma = 2\beta_1/(n(3n+1)(n-1)^2 + 2)$.}) only on $n, d, \text{ and } \beta_1$) such that for every $x_0 = \Gamma g_0 \in X$ and $T > R > C$, either

\begin{equation}
\frac{1}{m_U(B_T)} \int_{B_T} f(x_0u)dm_U(u) - \int_X f dm_X \ll R^{-\gamma} \mathcal{S}_\infty(f) \quad \forall f \in C_c^\infty(X)
\end{equation}

or

\begin{equation}
\exists j \in \{1, \ldots, n-1\} \text{ and primitive } w \in \Lambda^j(\mathbb{Z}^n) \sim \{0\}
\end{equation}

such that $\|wg_0u\| < R^q \forall u \in B_T$

where $q = \sum_{\lambda_i < 0} -m_i\lambda_i$ and $\ell = n(n-1)/2$ is the dimension of maximal compact subgroup of $G$.

Intuitively, this theorem says that either the $U$-orbit of $x_0$ equidistributes in $X$ with a fast rate, or $x_0$ is close to a proper subspace of $X$ that is fixed by the action of $U$, where our notions of “fast” and “close” are quantitatively related.

Remark. Observe that for fixed $R$, any generic point $x_0$ will fail (5b) for large enough $T$. This is because there are only finitely many $w \in \Lambda^j(\mathbb{Z}^n) \sim \{0\}$ with $j \in \{1, \ldots, n-1\}$ such that $\|wg_0\| < R^q$ to begin with, so for (5b) to hold for all large times $T$ it must be the case that there is a single $w$ satisfying $\|wg_0u\| < R^q$ for all $u \in U$. Moreover, $\|wg_0u\|$ is the maximum of several polynomials in the coordinates of $U$. Hence, in order to be bounded for all time, it must in fact be constant. This implies $wg_0$ is fixed by $U$, which recovers the qualitative result of Theorem 10.

Remark. It is worth noting that this theorem says something even in the case that the basepoint is not generic. Consider the following finitary version of genericity: $x_0 \in X$ is said to be $R$-generic if it satisfies (5a) for all sufficiently large $T$. If $x_0$ is $R$-generic for all $R > 0$, then it is generic. On the other hand,
suppose that \( x_0 = \Gamma g_0 \) is not generic, so that there exists \( w \in \Lambda^j(\mathbb{Z}^n) \sim \{0\} \) with \( j \in \{1, \ldots, n-1\} \) such that \( wg_0 \) is fixed by \( U \). If additionally \( \|wg_0\| > R^q \) for large enough \( R \), then the theorem tells us that \((5a)\) holds for all \( T > R \) and thus \( x_0 \) is \( R \)-generic. In fact, if \( R \) is quite large, then the orbit of \( x_0 \) will be very nearly equidistributed, even though it is not generic. This relates to the equidistribution of proper orbit closures as their volume goes toward infinity (see, e.g., [60, 63, 37]).

**Remark.** The condition that \( w \) in \((5b)\) be primitive is conceptually useful but technically unnecessary, in that if there exists any \( w \in \Lambda^j(\mathbb{Z}^n) \sim \{0\} \) satisfying \((5b)\), then there will also exist a primitive vector that does so. Moreover, although the theorem is stated for Folner sets of the form \( B_T = a_{\log T}u(\{0, 1\}^d)_{a-\log T} \), it should hold equally well for sets of the form \( B_T = a_{\log T}u(B)_{a-\log T} \) for any ball \( B \subset \mathbb{R}^d \).

**Remark.** The “either/or” in the theorem statement is not meant to imply an exclusive or. The structure of the proof will be to show that for \( x_0, T, \) and \( R \) as in the theorem, not-(5b) implies (5a). This leads us to define the following \((T,R)\)-Diophantine basepoint condition for \( x_0 = \Gamma g_0 \in X \), which is simply the negation of condition \((5b)\):

\[(5c) \quad \forall j \in \{1, \ldots, n-1\} \text{ and } w \in \Lambda^j(\mathbb{Z}^n) \sim \{0\}, \exists u \in B_T \text{ s.t. } \|wg_0u\| \geq R^q.\]

**Proof of Theorem 11.** From now on let \( U \) be a horospherical subgroup as in \((2)\) corresponding to the one-parameter diagonal flow \( a_t \) as in \((1)\). Let \( x_0 = \Gamma g_0 \in X \) satisfy the basepoint condition in \((5c)\) for some \( T > R \). Then consider \( f \in C^\infty_c(X) \) and write, via a change of variables,

\[
I_0 := \frac{1}{m_U(B_T)} \int_{B_T} f(x_0u)dm_U(u)
\]

\[
= \frac{1}{m_U(B_{T/R})} \int_{B_{T/R}} f(x_0a_{\log R}ua_{-\log R})dm_U(u)
\]

\[
= \frac{1}{m_U(B_{T/R})} \int_{U} 1_{B_{T/R}}(u) f(x_0a_{\log R}ua_{-\log R})dm_U(u).
\]

We want to show that this quantity is close to \( \int f dm_X \), and from \((6)\) it almost looks as if we could apply the exponential mixing result of Corollary 9(i) to achieve this, however there are several significant barriers to doing so. Most obviously, the integral in \((6)\) is over \( U \) instead of \( X \). Furthermore, the “basepoint” \( x_0a_{\log R}u \) varies with \( u \), and will eventually spend time outside of any fixed compact subset of \( X \) for \( u \) coming from a large enough set. Finally, the function \( 1_{B_{T/R}} \) is not smooth.

We will first address the issue of smoothness by convolving the indicator function with a smooth approximation to the identity (Step 1). We will then apply the “thickening” argument of Margulis to obtain an integral over \( X \) from our integral over \( U \) (Step 2). Finally, we will deal with the moving basepoint by demonstrating that for most \( u \in B_{T/R} \) we have a uniformly good rate of equidistribution and that the size of the set on which this does not occur can be quantitatively
controlled (Step 3). This last step is where we will use the nondivergence result of Section 2.8.

**Step 1.** Let \( r \) be a small, positive number (to be determined) and let \( \theta \in C_c^\infty(U) \) be a nonnegative bump function supported on \( B_r^U(e) \) satisfying the approximate identity properties of Lemma 4. Then the convolution \( \int_U \theta(u') \mathbf{1}_{B_r^U}(u(u')^{-1})d\mu_U(u') \) is a smooth function approximating our original indicator function. If we substitute this function for \( \mathbf{1}_{B_r^U} \) in (6) and use the invariance property of the Haar measure, we get the integral

\[
I_{\text{smth}}(7) := \frac{1}{\mu_U(B_{T/R})} \int_U \int_U \theta(u') \mathbf{1}_{B_{T/R}}(u(u')^{-1})d\mu_U(u')f(x_0a_{\log R}ua_{-\log R})dm_U(u) \]

\[
= \frac{1}{\mu_U(B_{T/R})} \int_U \int_U \theta(u') \mathbf{1}_{B_{T/R}}(u)f(x_0a_{\log R}uu'a_{-\log R})dm_U(u)d\mu_U(u').
\]

Now observe that since \( \int \theta = 1 \), we may again use the invariance of the Haar measure to rewrite (6) as

\[
I_0 = \frac{1}{\mu_U(B_{T/R})} \int_U \mathbf{1}_{B_{T/R}}(uu')f(x_0a_{\log R}uu'a_{-\log R})dm_U(u) \int_U \theta(u')d\mu_U(u')
\]

\[
= \frac{1}{\mu_U(B_{T/R})} \int_U \int_U \theta(u') \mathbf{1}_{B_{T/R}}(uu')f(x_0a_{\log R}uu'a_{-\log R})dm_U(u)d\mu_U(u').
\]

From (7) and (8), we can see that

\[
|I_0 - I_{\text{smth}}| \leq \frac{1}{\mu_U(B_{T/R})} \int_U \theta(u')\mathcal{S}_\infty,0(f)\left(\int_U \mathbf{1}_{B_{T/R}}(uu') - \mathbf{1}_{B_{T/R}}(u)\right)dm_U(u')dm_U(u')
\]

\[
= \frac{\mathcal{S}_\infty,0(f)}{\mu_U(B_{T/R})} \int_U \theta(u')\mu_U(B_{T/R}\Delta B_{T/R}(u')^{-1})d\mu_U(u').
\]

But notice that since supp \( \theta \subseteq B_r^U(e) \), we know that \( u' \) is close to the identity, so \( u \) in this region can only shift \( B_{T/R} \) by a small amount. In fact, by pulling the measure back to \( \mathbb{R}^n \), one may compute directly that the size of the symmetric difference is bounded by

\[
m_U(B_{T/R}\Delta B_{T/R}(u')^{-1}) \ll (T/R)^{p-p_0}r
\]

for any \( u' \in B_r^U(e) \), where \( p_0 = \min_{i<j}(\lambda_i - \lambda_j) \). Combining this with (9) above and again using the fact that \( \theta \) integrates to 1, we see that

\[
|I_0 - I_{\text{smth}}| \ll \frac{(T/R)^{p-p_0}}{m_U(B_{T/R})}r\mathcal{S}_\infty,0(f) = (R/T)^{p_0}r\mathcal{S}_\infty,0(f) \leq r\mathcal{S}_\infty,0(f)
\]

since \( m_U(B_{T/R}) = (T/R)^p \) and \( T \geq R \).

Now that we know \( I_0 \) and \( I_{\text{smth}} \) can be made close, we want to know that \( I_{\text{smth}} \) is not too far from \( \int f dm_X \). Using Fubini’s Theorem, we can say

\[
I_{\text{smth}} = \frac{1}{\mu_U(B_{T/R})} \int_U \int_U \theta(u')f(x_0a_{\log R}uu'a_{-\log R})dm_U(u')d\mu_U(u)
\]
and we may also write
\[ \int_X f \, dm_X = \frac{1}{m_U(B_{T/R})} \int_{B_{T/R}} \left( \int_X f \, dm_X \right) dm_U(u). \]

Hence,
\[ (12) \quad \left| I_{\text{smth}} - \int_X f \, dm_X \right| \leq \frac{1}{m_U(B_{T/R})} \int_{B_{T/R}} \left| \int_U \theta(u') f(x_0 a_{\log R} uu' a_{-\log R}) dm_U(u') - \int_X f \, dm_X \right| dm_U(u). \]

**Step 2.** Now the expression inside the absolute value looks more similar to the mixing result of Corollary 9(i), but we are still integrating with respect to the wrong measure. We want an integral with respect to $m_X$, and although functions on $X$ integrate locally like their pullback by projection over $G$, the integral with which we are concerned is over the lower-dimensional ("thin") subspace $U$.

Define
\[ I_U(u) := \int_U \theta(u') f(x_0 a_{\log R} uu' a_{-\log R}) dm_U(u'). \]

to be the integral from inside (12) above. In order to apply Corollary 9(i), we will need to “thicken” this integral over $U$ to an integral over a neighborhood of the orbit in $G$ and then project to $X$.

Recall from Section 2.3 that $m_G = m_U \times m_{H}^{r}$, where $m_{H}^{r}$ is the right Haar measure on $H = U^{0}U^{-}$. Then let $\psi \in C_{c}^{\infty}(H)$ be an approximate identity supported on $B_{H}^{r}(e)$ as described in Lemma 4. Since $\int \psi = 1$, we may rewrite (13) as
\[ (14) \quad I_U(u) = \int_{H} \int_{U} \theta(u') \psi(h) f(x_0 a_{\log R} uu' a_{-\log R}) dm_U(u') dm_{H}^{r}(h). \]

Now define
\[ (15) \quad I_X(u) := \int_{H} \int_{U} \theta(u') \psi(h) f(x_0 a_{\log R} uu' h a_{-\log R}) dm_U(u') dm_{H}^{r}(h), \]

which differs from $I_U(u)$ only by the presence of the variable $h$ inside $f$. To see that $I_U(u)$ and $I_X(u)$ are close, observe that
\[ (16) \quad |I_U(u) - I_X(u)| \leq \int_{H} \int_{U} \theta(u') \psi(h) \left| f(\tilde{x}) - f(\tilde{x} a_{\log R} h a_{-\log R}) \right| dm_U(u') dm_{H}^{r}(h), \]

where $\tilde{x} = x_0 a_{\log R} uu' a_{-\log R}$. But since $f$ has bounded derivative,
\[ (17) \quad \left| f(\tilde{x}) - f(\tilde{x} a_{\log R} h a_{-\log R}) \right| \ll \mathcal{J}_{\infty,1}(f) d_{G}(e, a_{\log R} h a_{-\log R}) \]

by Sobolev property (iv). Furthermore, since conjugation by $a_{t}$ is non-expanding on the subgroup $H$ (recall that it fixes $U^{0}$ and contracts $U^{-}$), we
may see that
\begin{equation}
\tag{18}
\frac{d_G(e, a_{\log R} h a_{-\log R})}{d_G(e, h)} \leq d_H(e, h) \ll \epsilon \leq r
\end{equation}
for \( h \in \text{supp } \psi \subseteq B_r^H(e) \).

Then from (16), (17), and (18) and the fact that both \( \theta \) and \( \psi \) integrate to 1, we have
\begin{equation}
\tag{19}
|I_U(u) - I_X(u)| \ll \mathcal{L}_{\infty,1}(f)r.
\end{equation}
Now we want to verify that \( I_X(u) \) is not far from \( \int f \, dm_X \). By our measure decomposition, we can see (15) as an integral over \( G \):
\begin{equation}
\tag{20}
I_X(u) = \int_G \phi(g) f(x_0 a_{\log R} u g a_{-\log R}) \, dm_G(g),
\end{equation}
where the function \( \phi(uh) = \theta(u) \psi(h) \) is defined for all \( g \in UH \), hence it is defined almost-everywhere. In order to apply mixing, we want to further interpret \( I_X(u) \) as an integral over \( X \). To do this, let \( y = x_0 a_{\log R} u \), keeping in mind that \( y \) depends on \( u \). Then define \( \phi_y \in C_c^{\infty}(X) \) by \( \phi_y = \phi \circ \pi_y^{-1} \) where \( \pi_y : G \rightarrow X \) is natural projection at \( y \). Note, however, that \( \phi_y \) is only well-defined if \( \pi_y \) is injective on \( \text{supp } \phi = \text{supp } \theta \text{ supp } \psi \subseteq B_r^U(e) B_r^H(e) \). In a neighborhood of the identity, \( B_r^U(e) B_r^H(e) \subseteq B_{cr}(e) \) for a positive constant \( c \), since
\[
d_C(\alpha, e) \leq d_C(\alpha, u, e) + d_C(u, e) = d_C(h, e) + d_C(u, e) \ll d_H(h, e) + d_U(u, e) \leq 2r.
\]
Therefore, if \( \pi_y \) is injective on \( B_{cr}(e) \) for \( \alpha = x_0 a_{\log R} u \) (an assumption we will return to later) we can say from (20) that
\[
I_X(u) = \int_G \int_X \phi(g, \pi_y(x)) f(x a_{-\log R}) \, dm_G(g) \, dm_X(x).
\]
Since \( \int \phi_y \, dm_X = \int \phi \, dm_G = \int \theta \, dm_U \int \psi \, dm_H = 1 \), we can now apply the effective mixing result from Section 2.9 to obtain
\[
\left| I_X(u) - \int f \, dm_X \right| = \left| \int \phi_y(x) \, dm_G(g) \, dm_X(x) - \int f \, dm_X \right| \ll R^{-\beta_1} \mathcal{S}_\ell(\phi_y) \mathcal{S}_\ell(f).
\]
Then from property (v) in Section 2.4 and our bound on the Sobolev norm of an approximate identity (property (iv) in Section 2.5), we can say
\[
\mathcal{S}_\ell^X(\phi_y) = \mathcal{S}_\ell^G(\phi) \ll \mathcal{S}_\ell^U(\theta) \mathcal{S}_\ell^H(\psi) \ll r^{-\ell + d/2} r^{-\ell + d/2} = r^{-2\ell - (m^2 - 1)/2},
\]
\footnote{There is a slight subtlety here because we used the right Haar measure on \( H \), so the corresponding metric \( d_H \) is right-invariant, while \( d_G \) is left-invariant. In general, \( d_G \) restricted to \( H \) will be less than or equal to the corresponding left-invariant metric on \( H \). However, any left-invariant metric is Lipschitz equivalent to any right-invariant metric in a suitable neighborhood of the identity, so the above series of inequalities goes through for \( r \) small enough.}
where \( \tilde{d} = \dim H \). Thus if \( \pi_y \) is injective on \( B_{cr}^{\tilde{G}}(e) \), then

\[
I_X(u) - \int_X f \, dm_X \ll R^{-\beta_1} r^{-p_1} \mathcal{S}(f),
\]

where \( p_1 = 2\ell + (n^2 - 1)/2 \).

**Step 3.** However, as we have noted, \( y = x_0 a_{\log R} u \) depends on \( u \), which varies over \( B_{T/R} \) in (12). While we cannot ensure that \( \pi_y \) is injective on \( B_{cr}^{\tilde{G}}(e) \) for all \( u \in B_{T/R} \), we can say that the set on which this does not occur has small measure.

Recall from Lemma 5 that \( \pi_y : B_{cr}^{\tilde{G}}(e) \to B_{cr}^{\tilde{X}}(y) \) is injective for \( y \in L_\epsilon \) for \( r \) proportional to \( \epsilon^n \) and for \( \epsilon \) small enough. Furthermore, observe that condition (5c) is equivalent to the statement that for all \( f \in [1,\ldots,n-1] \) and primitive \( w \in \Lambda^1(\mathbb{Z}^d) \sim \{0\} \), there exists \( t \in [0,1]^d \) such that \( \|u g_0 a_{\log T} u(t) a_{-\log T}\| \geq R^q \).

Then by Corollary 7 in Section 2.8, we have that

\[
\left| \{ t \in [0,1]^d \mid x_0 a_{\log T} u(t) a_{-\log T} a_{\log R} \notin L_\epsilon \} \right| \ll \epsilon^{1/d(n-1)}.
\]

From this we find that

\[
\left| \{ t \in [0,1]^d \mid x_0 a_{\log T} u(t) a_{-\log T} a_{\log R} \notin L_\epsilon \} \right|
\]

\[
= \left| \{ t \in [0,1]^d \mid x_0 a_{\log R} a_{\log T/R} u(t) a_{-\log T/R} \notin L_\epsilon \} \right|
\]

\[
= m_U(\{ u \in B_1 \mid x_0 a_{\log R} a_{\log T/R} u a_{-\log T/R} \notin L_\epsilon \})
\]

\[
= m_U(\{ u \in B_{T/R} \mid x_0 a_{\log R} u \notin L_\epsilon \}) / m_U(B_{T/R})
\]

where the last equality can be verified using a change of variables. That is, for \( x_0 \) satisfying condition (5c), we have

\[
m_U(\{ u \in B_{T/R} \mid x_0 a_{\log R} u \notin L_\epsilon \}) \ll \epsilon^{1/d(n-1)} m_U(B_{T/R}).
\]

In other words, if we let \( E := \{ u \in B_{T/R} \mid x_0 a_{\log R} u \in L_\epsilon \} \), then (21) holds for all \( u \in E \) and \( m_U(B_{T/R} \sim E) \ll \epsilon^{1/d(n-1)} m_U(B_{T/R}) \). Thus, from (12), (19), and (21), we find

\[
\left| I_{\text{smith}} - \int_X f \, dm_X \right| \leq \frac{1}{m_U(B_{T/R})} \int_{B_{T/R}} \left| I_U(u) - \int_X f \, dm_X \right| \, dm_U(u)
\]

\[
\leq \frac{1}{m_U(B_{T/R})} \int_{B_{T/R}} \left| I_U(u) - I_X(u) \right| \, dm_U(u)
\]

\[
+ \frac{1}{m_U(B_{T/R})} \int_{B_{T/R}} \left| I_X(u) - \int_X f \, dm_X \right| \, dm_U(u)
\]

\[
\ll \mathcal{S}_\infty(f) r + \frac{1}{m_U(B_{T/R})} \int_E \left| I_U(u) - \int_X f \, dm_X \right| \, dm_U(u)
\]

\[
+ \frac{1}{m_U(B_{T/R})} \int_{B_{T/R} \sim E} \left| I_X(u) - \int_X f \, dm_X \right| \, dm_U(u)
\]

\[
= \mathcal{S}(f) r + \mathcal{S}_\infty(f) r + \frac{1}{m_U(B_{T/R})} \int_E f \, dm_U
\]

\[
+ \frac{1}{m_U(B_{T/R})} \int_{B_{T/R} \sim E} f \, dm_U
\]

\[
= \mathcal{S}(f) r + \mathcal{S}_\infty(f) r + \frac{1}{m_U(B_{T/R})} \int_E F \, dm_U
\]

\[
+ \frac{1}{m_U(B_{T/R})} \int_{B_{T/R} \sim E} f \, dm_U
\]
Finally, from this and (11), we have

\[ |I_0 - \int_X f \, dm_X| \leq |I_0 - I_{\text{smth}}| + |I_{\text{smth}} - \int_X f \, dm_X| \]

\[ \ll \left( \varepsilon^n + R^{-\beta_1} \varepsilon^{-p_1 n} + \varepsilon^{1/d(n-1)} \right) S_{\infty,\ell} (f) \]

where we have used that \( r \) is proportional to \( \varepsilon^n \), as well as Sobolev property (i).

Let \( p_2 := 1/d(n-1) \). Since \( n > p_2 \), the \( \varepsilon^n \) term above decays more quickly than other terms and can be ignored. To optimize the rate of decay, we set

\[ R^{-\beta_1} \varepsilon^{-p_1 n} = \varepsilon^{p_2}, \]

which implies

\[ \varepsilon = R^{-\beta_1 / (p_1 n + p_2)}. \]

Then so long as \( R \) is chosen sufficiently large so that \( \varepsilon \) (and subsequently \( r \)) are small enough to make Corollary 7 and Lemma 5 true (along with several other statements we made regarding neighborhoods of the identity), then we have demonstrated (5a) in Theorem 11 with the rate

\[ |I_0 - \int_X f \, dm_X| \ll R^{-\gamma} S_{\infty,\ell} (f) \]

where

\[ \gamma = \frac{\beta_1 p_2}{p_1 n + p_2} = \frac{2\beta_1}{nd(3n+1)(n-1)^2 + 2} \]

where we have written \( p_1, p_2, \) and \( \ell \) in terms of \( n \) and \( d \).

**Remark.** In the case of \( \Gamma \) cocompact, it follows from the above proof that we may remove dependence on the basepoint from our effective equidistribution statement. That is, for \( X = \Gamma \backslash G, \Gamma \leq G \) a cocompact lattice, and \( U \leq G \) a horospherical subgroup, we have that there exists \( \gamma > 0 \) (depending\(^6\) only on \( n \) and \( \beta_1 \)) such that for \( T \) large enough,

\[ \left| I_0 - \int_X f \, dm_X \right| \ll T^{-\gamma} S_{\infty,\ell} (f) \]

for any \( f \in C^\infty(X) \) and \( x_0 \in X \). This is because we only make use of the basepoint condition in Step 3, where we need it to deal with the moving basepoint and the fact that the injectivity radius depends on where we are in \( X \). However, in the compact setting, we have a uniform injectivity radius, so we may avoid this step altogether. Morally speaking, uniformity in the basepoint is due to the fact that in the compact setting the dynamics are minimal, so there are no

---

\(^6\) In this case we have \( \gamma = 2\beta_1 / ((3n+1)(n-1) + 2) \).
proper invariant subspaces near which an orbit can become trapped for long periods of time.

4. EQUIDISTRIBUTION FOR ARITHMETIC SEQUENCES IN ABELIAN HOROSPHERICAL FLOWS

Let \( G = \SL_n(\mathbb{R}) \), \( \Gamma = \SL_n(\mathbb{Z}) \), and \( X = \Gamma \backslash G \). Let \( U \) be an upper triangular unipotent subgroup of the form
\[
U = \left\{ \begin{pmatrix} I_m & * \\ 0 & I_{n-m} \end{pmatrix} \right\}
\]
for \( m < n \). Note that \( U \cong \mathbb{R}^d \) as groups for \( d = m(n-m) \) under the identification \( u(t) \) which maps the coordinates of \( t \in \mathbb{R}^d \) to the matrix entries in the upper-right block of \( (24) \). Recall that the Haar measure on \( U \) is the Lebesgue measure on \( \mathbb{R}^d \) under this identification, which we normalize so that \( u([0,1]^d) \) has unit measure. Observe that \( U \) is horospherical with respect to the element
\[
a_t = \text{diag}(e^{(n-m)/m}, \ldots, e^{(n-m)/m}, e^{-(n-m)/m}, \ldots, e^{-(n-m)/m})
\]
for any \( t > 0 \) and that conjugation by \( a_t \) scales all entries in the upper-right block of \( U \) by \( e^{(n-m)/n} e^{tm/n} = e^t \). Hence, for this choice of \( a_t \), we have \( B_T = a_{\log T} u([0,1]^d) a_{-\log T} = u([0, T]^d) \). For this reason we will conflate the notation and write \( B_T \) for both \( [0, T]^d \subseteq \mathbb{R}^d \) and \( u([0, T]^d) \subseteq U \).

Let \( \psi \) be an additive character of \( U \) (so \( \psi(t) = e^{iat} \) for some \( a \in \mathbb{R}^d \)). Define measure \( \nu_T \) and (complex) measure \( \mu_{T,\psi} \) on \( X \) via duality: for \( f \in C_c^\infty(X) \) let
\[
\int_X f \, d\nu_T = \nu_T(f) := \frac{1}{|B_T|} \int_{B_T} f(x_0 u(t)) \, dt
\]
and
\[
\int_X f \, d\mu_{T,\psi} = \mu_{T,\psi}(f) := \frac{1}{|B_T|} \int_{B_T} \psi(t) \left( f(x_0 u(t)) - \int_X f \, d\mu_X \right) \, dt.
\]
Our main goal in this section is to obtain an effective rate of equidistribution along (multivariate) arithmetic sequences of inputs for the right action of \( U \) on \( X \). To do this, we first present the following lemma, the proof of which closely follows the proof of [71, Lemma 3.1] for the case of \( G = \SL_2(\mathbb{R}) \) and \( \Gamma \) cocompact.

**Lemma 12.** Let \( x_0 = \Gamma g_0 \in X \) satisfy (5C) for \( T > R > C \). Then there exists \( b > 0 \) such that for any \( f \in C_c^\infty(X) \) and additive character \( \psi \),
\[
|\mu_{T,\psi}(f)| \ll R^{-b} S_{\infty,\ell}(f),
\]
where \( \ell \) is as in Theorem 11.

**Remark.** As noted in [71], the significance of this lemma is that the implicit constant is independent of choice of \( \psi \). This can be shown for highly oscillatory \( \psi \) using integration by parts and for almost constant \( \psi \) using equidistribution of the horospherical flow directly, thus this lemma is most significant for \( \psi \) of moderate oscillation. The proof will use our effective equidistribution result...
as well as a variety of technical integral manipulations that nonetheless do not require any heavy machinery.

Proof. Let $1 \leq H \leq T$ and define a complex measure $\sigma_H$ on $U$ by

$$
\int_U g d\sigma_H = \sigma_H(g) := \frac{1}{|B_H|} \int_{B_H} \overline{\psi}(t) g(u(t)) dt
$$

for $g \in C_c^\infty(U)$.

Let $f * \sigma_H$ be the right convolution of $f$ by $\sigma_H$, i.e., for $x \in X$

$$
(f * \sigma_H)(x) = \int f(xu(t)^{-1}) d\sigma_H(t)
= \frac{1}{|B_H|} \int_{B_H} \overline{\psi}(t) f(xu(t)^{-1}) dt.
$$

Notice that by switching the order of integration (one may verify that the conditions of Fubini’s theorem are satisfied) and using invariance of the measure $m_X$, we have

$$
\int_X f * \sigma_H dm_X = \int_X \frac{1}{|B_H|} \int_{B_H} \overline{\psi}(t) f(xu(t)^{-1}) dt dm_X(x)
$$

$$
= \frac{1}{|B_H|} \int_{B_H} \overline{\psi}(t) \left( \int_X f(xu(t)^{-1}) dm_X(x) \right) dt
$$

$$
= \frac{1}{|B_H|} \int_{B_H} \overline{\psi}(t) \left( \int_X f dm_X \right) dt.
$$

Hence,

$$
\mu_{T,\psi}(f * \sigma_H) = \frac{1}{|B_T|} \int_{B_T} \psi(t) \left( f * \sigma_H(x_0u(t)) - \int_X f * \sigma_H dm_X \right) dt
$$

$$
= \frac{1}{|B_T|} \int_{B_T} \psi(t) \frac{1}{|B_H|} \int_{B_H} \overline{\psi}(s) \left( f(x_0u(t)u(s)^{-1}) - \int_X f dm_X \right) ds dt
$$

$$
= \frac{1}{|B_T||B_H|} \int_{B_T} \int_{B_H} \psi(t-s) \left( f(x_0u(t-s)) - \int_X f dm_X \right) ds dt
$$

since $\overline{\psi}(s) = \psi(-s)$ and $U \cong R^d$. Now by switching the order of integration and applying a change of variables, we get

$$
\mu_{T,\psi}(f * \sigma_H) = \frac{1}{|B_T||B_H|} \int_{B_H} \int_{B_T+s} \psi(t) \left( f(x_0u(t)) - \int_X f dm_X \right) dt ds.
$$

But we may also write

$$
\mu_{T,\psi}(f) = \frac{1}{|B_T|} \int_{B_T} \psi(t) \left( f(x_0u(t)) - \int_X f dm_X \right) dt
$$

$$
= \frac{1}{|B_T||B_H|} \int_{B_H} \int_{B_T} \psi(t) \left( f(x_0u(t)) - \int_X f dm_X \right) dt ds.
$$
Thus

\[
|\mu_{T,\psi}(f) - \mu_{T,\psi}(f * \sigma_H)| \leq \frac{1}{|B_T||B_H|} \int_{B_T} \int_{B_T \triangle (B_T - s)} |f(x_0 u(t)) - \int_X f dm_X| \, dt ds
\]

\[
\ll \frac{1}{|B_T||B_H|} \int_{B_T} |B_T \triangle (B_T - s)| \mathcal{F}_{\infty,0}(f) ds.
\]

But notice that $B_T \triangle (B_T - s)$ is simply the symmetric difference of two shifted cubes, the measure of which will be maximized when $s = (H, \ldots, H)$ (see Figure 1). Hence,

\[
|B_T \triangle (B_T - s)| \leq 2(T^d - (T - H)^d)
\]

\[
= 2(dT^{d-1}H - \cdots \pm dTH^{d-1} + H^d)
\]

\[
\ll T^{d-1}H.
\]

since $H \leq T$ implies that the leading term dominates. It follows that $\int_{B_T} |B_T \triangle (B_T - s)| ds \ll T^{d-1}H^{d+1}$. Thus,

\[
|\mu_{T,\psi}(f) - \mu_{T,\psi}(f * \sigma_H)| \ll \frac{T^{d-1}H^{d+1}}{|B_T||B_H|} \mathcal{F}_{\infty,0}(f) = \frac{H}{T} \mathcal{F}_{\infty,0}(f).
\]

Now consider

\[
|\mu_{T,\psi}(f * \sigma_H)|^2 = \left| \frac{1}{|B_T|} \int_{B_T} \psi(t) \left(f * \sigma_H(x_0 u(t)) - \int_X f * \sigma_H dm_X \right) dt \right|^2
\]

\[
\leq \frac{1}{|B_T|^2} \left( \int_{B_T} \left| f * \sigma_H(x_0 u(t)) - \int_X f * \sigma_H dm_X \right| dt \right)^2
\]

\[
= \frac{1}{|B_T|^2} \left( \left| f * \sigma_H(x_0 u(\cdot)) - \int_X f * \sigma_H dm_X \right|_{L^2(B_T)}^2 \right).
\]
By Cauchy–Schwarz, we know that

\[ |\mu_{T,\psi}(f * \sigma_H)|^2 \leq \frac{1}{|B_T|^2} \|1\|_{L^2(B_T)}^2 \left\| f * \sigma_H(x_0 u(\cdot)) - \int_X f * \sigma_H d\mu_X \right\|_{L^2(B_T)}^2. \]

Now, \( \|1\|_{L^2(B_T)}^2 = \int_{B_T} 1^2 dt = |B_T| \), and

\[
\left\| f * \sigma_H(x_0 u(\cdot)) - \int_X f * \sigma_H d\mu_X \right\|_{L^2(B_T)}^2
= \int_{B_T} \left| f * \sigma_H(x_0 u(t)) - \int_X f * \sigma_H d\mu_X \right|^2 dt
= |B_T| \nu_T \left( \left| f * \sigma_H - \int_X f * \sigma_H d\mu_X \right|^2 \right)
\]

which shows that

\[ |\mu_{T,\psi}(f * \sigma_H)|^2 \leq \nu_T \left( \left| f * \sigma_H - \int_X f * \sigma_H d\mu_X \right|^2 \right). \]

Hence, by (26) and (27), we have

\[ |\mu_{T,\psi}(f)| \leq |\mu_{T,\psi}(f) - \mu_{T,\psi}(f * \sigma_H)| + |\mu_{T,\psi}(f * \sigma_H)| \]
\[ \leq \frac{H}{T} J_{\infty,0}(f) + \nu_T \left( \left| f * \sigma_H - \int_X f * \sigma_H d\mu_X \right|^2 \right)^{1/2}. \]

To estimate \( \nu_T \left( \left| f * \sigma_H - \int_X f * \sigma_H d\mu_X \right|^2 \right) \), observe that

\[
\left| f * \sigma_H(x) - \int_X f * \sigma_H d\mu_X \right|^2
= \frac{1}{|B_H|^2} \int_{B_H} \int_{B_H} \psi(s_1)(\left| u(s_1)f(x) - \int_X f d\mu_X \right| ds_1)
\cdot \left( \int_{B_H} \psi(s_2)(\left| u(s_2)f(x) - \int_X f d\mu_X \right| ds_2) \right)
\]
\[ = \frac{1}{|B_H|^2} \int_{B_H} \int_{B_H} \psi(s_2 - s_1)(\left| u(s_1)f(x) - \int_X f d\mu_X \right| ds_1) \left( \left| u(s_2)f(x) - \int_X f d\mu_X \right| ds_2 \right). \]

When we apply \( \nu_T \) to this, we can change the order of integration so that the innermost integral is over \( B_T \), with the character \( \psi(s_2 - s_1) \) outside this integral. We may then integrate separately over the four terms we get by expanding the
where

\[ F(s_1, s_2) = \nu_T(u(s_1)f \cdot u(s_2)f) - \nu_T(u(s_1)f) \int_X f dm_X \]

(30)

Likewise,

\[ \nu_T(u(s_1)f) \int_X f dm_X = \int_X f dm_X + \nu_T(u(s_1)f) \left| f \int_X f dm_X \right|^2. \]

(29)

Thus for arbitrary \( f \in C_c^\infty(X) \) and \( x_0 \) satisfying the Diophantine basepoint condition (5c) with \( T > R > C \), we have

\[ \left| \nu_T(\tilde{f}) - \int_X \tilde{f} dm_X \right| = \frac{1}{|B_T|} \int_{B_T} \tilde{f}(x_0 u(t)) dt - \int_X \tilde{f} dm_X \ll R^{-\gamma} \mathcal{I}_{\infty, \ell}(|\tilde{f}|), \]

that is,

(31)

\[ \nu_T(\tilde{f}) = \int_X \tilde{f} dm_X + \mathcal{O}(R^{-\gamma} \mathcal{I}_{\infty, \ell}(|\tilde{f}|)). \]

Applying this to the function \( \tilde{f} = u(s_1)f \), we find that

\[ \nu_T(u(s_1)f) = \int_X u(s_1)f dm_X + \mathcal{O}(R^{-\gamma} \mathcal{I}_{\infty, \ell}(u(s_1)f)). \]

But by invariance of \( m_X \), we have

\[ \int_X u(s_1)f dm_X = \int_X (xu(s_1)^{-1}) dm_X(x) = \int_X f dm_X. \]

Thus

\[ \nu_T(u(s_1)f) \int_X f dm_X = \left| \int_X f dm_X \right|^2 + \mathcal{O} \left( R^{-\gamma} \mathcal{I}_{\infty, \ell}(u(s_1)f) \left| \int_X f dm_X \right| \right). \]

Furthermore, from Sobolev norm property (iii), we know that for \( f \in C_c^\infty(X) \) and \( h \in G \), we have \( \mathcal{I}_{\infty, \ell}(hf) \ll \|h\|^\ell \mathcal{I}_{\infty, \ell}(f) \), where \( \|h\| \) is the operator norm of \( \text{Ad}_{h^{-1}} \). Since the entries of \( u(s)^{-1} \) are bounded by \( \max(1, |s|) \), we have \( \|u(s)\| \ll \max(1, |s|)^2 \). Thus for \( s_1 \in [0, H] \) with \( H \geq 1 \), \( \mathcal{I}_{\infty, \ell}(u(s_1)f) \ll H^{2\ell} \mathcal{I}_{\infty, \ell}(f) \). Combining this with the bound \( \left| \int f dm_X \right| \ll \mathcal{I}_{\infty, \ell}(f) \ll \mathcal{I}_{\infty, \ell}(f) \), we find that

\[ \nu_T(u(s_1)f) \int_X f dm_X = \left| \int_X f dm_X \right|^2 + \mathcal{O}(R^{-\gamma} H^{2\ell} \mathcal{I}_{\infty, \ell}(f)^2). \]

Likewise,

\[ \nu_T(u(s_2)f) \int_X f dm_X = \left| \int_X f dm_X \right|^2 + \mathcal{O}(R^{-\gamma} H^{2\ell} \mathcal{I}_{\infty, \ell}(f)^2). \]
Therefore, (30) becomes simply

$$F(s_1, s_2) = \nu_T(u(s_1) f \cdot u(s_2) f) \left| \int_X f \, dm_X \right|^2 + \Theta(T^{-a} H^{2\ell} \mathcal{L}_{\infty, \ell}(f)^2).$$

Substituting this back into (29), we conclude that

$$\nu_T \left( \left| f * \sigma_H(x) - \int_X f \, dm_X \right|^2 \right) \ll \frac{1}{|BH|^2} \iint_{B_H \times B_H} \left| \nu_T(u(s_1) f \cdot u(s_2) f) - \int_X f \, dm_X \right|^2 \, ds_1 \, ds_2 + R^{-\gamma} H^{2\ell} \mathcal{L}_{\infty, \ell}(f)^2.$$

But now notice that

$$\int_X u(s_1) f \cdot u(s_2) f \, dm_X = \langle u(s_1) f, u(s_1) f \rangle_{L^2(X)} = \langle u(s_1 - s_2) f, f \rangle_{L^2(X)}$$

so by the triangle inequality, we can estimate

$$\left| \nu_T(u(s_1) f \cdot u(s_2) f) - \int_X f \, dm_X \right|^2 \leq \left| \nu_T(u(s_1) f \cdot u(s_2) f) - \int_X u(s_1) f \cdot u(s_2) f \, dm_X \right|$$

$$+ \left| \langle u(s_1 - s_2) f, f \rangle_{L^2(X)} - \int_X f \, dm_X \right|^2.$$ (33)

Again, by our distribution result in (31), we know that

$$\nu_T(u(s_1) f \cdot u(s_2) f) - \int_X u(s_1) f \cdot u(s_2) f \, dm_X \ll R^{-\gamma} \mathcal{L}_{\infty, \ell}(u(s_1) f \cdot u(s_2) f)$$

and by properties (ii) and (iii) of Sobolev norms, we have

$$\mathcal{L}_{\infty, \ell}(u(s_1) f \cdot u(s_2) f) \ll \mathcal{L}_{\infty, \ell}(u(s_1) f) \mathcal{L}_{\infty, \ell}(u(s_2) f) \ll H^{4\ell} \mathcal{L}_{\infty, \ell}(f)^2$$

for $s_1, s_2 \in [0, H]$. Thus, from (35), (34), and (33), equation (32) becomes

$$\nu_T \left( \left| f * \sigma_H(x) - \int_X f * \sigma_H \, dm_X \right|^2 \right) \ll \frac{1}{|BH|^2} \iint_{B_H \times B_H} \left| \langle u(s_1 - s_2) f, f \rangle_{L^2(X)} - \int_X f \, dm_X \right|^2 \, ds_1 \, ds_2 + R^{-\gamma} H^{4\ell} \mathcal{L}_{\infty, \ell}(f)^2.$$ (36)

Now from Corollary 9(ii), we know there exists $\beta_2 > 0$ such that for any $s \in \mathbb{R}^d$,

$$\left| \langle u(s) f, f \rangle_{L^2(X)} - \int_X f \, dm_X \right|^2 \ll \max(1, |s|)^{-\beta_2} \mathcal{L}_{\infty, \ell}(f)^2.$$ (37)

Then for $s = s_1 - s_2$, we have the following problem: We want to bound the integral in (36) by a power of $H$, but for $(s_1, s_2)$ close to the diagonal in $B_H \times B_H$ we cannot do better than a constant times the Sobolev norm of $f$ in (37). We will address this by integrating separately over a neighborhood of the diagonal.
that has small measure (depending on \(H\)) and away from the diagonal where \(\max(1,|s_1 - s_2|)\) is dominated by \(H\).

To make this precise, let \(D := \{(s_1, s_2) \in B_H \times B_H | s_1 = s_2\}\) be the diagonal of \(B_H \times B_H\) and define \(D_\epsilon := \{(s_1, s_2) \in B_H \times B_H | |s_1 - s_2| < \epsilon\}\). Notice that \(D\) is a \(d\)-dimensional subset of \(\mathbb{R}^d\) with diameter \(\sqrt{2dH}\). Furthermore, any point satisfying \(|s_1 - s_2| = \epsilon\) is distance \(\epsilon/\sqrt{2}\) from the diagonal, so \(D_\epsilon\) is an \((\epsilon/\sqrt{2})\)-neighborhood of \(D\) sitting inside \([0, H]^{2d}\). Thus \(D_\epsilon\) is contained within a box in \(\mathbb{R}^{2d}\) with \(d\) side-lengths of \(\sqrt{2dH}\) and \(d\) side-lengths of \(2\epsilon/\sqrt{2}\), so

\[
|D_\epsilon| \ll H^d \epsilon^d
\]

(see Figure 2). In particular, if \(\epsilon = H^\zeta\) (for \(0 < \zeta < 1\) to be determined), then

\[
|\{(s_1, s_2) \in B_H \times B_H | |s_1 - s_2| < H^\zeta\}| \ll H^{d(1+\zeta)}.
\]

In this region, the integrand is dominated by 1, so when we integrate over this region and divide by \(|B_H|^2 = H^{2d}\) (as we are doing in equation (36)), we get a term of order \(H^d(\zeta - 1)\mathcal{S}_{\infty, \ell}(f)^2\). On the other hand, for \(|s_1 - s_2| \geq H^\zeta\), we can say that

\[
\left| \langle u(s_1 - s_2)f, f \rangle_{L^2(X)} - \int_X f \, dm_X \right|^2 \ll \max(1, |s_1 - s_2|)^{-\beta_2} \mathcal{S}_{\infty, \ell}(f)^2 \\
\leq H^{-\epsilon \beta_2} \mathcal{S}_{\infty, \ell}(f)^2.
\]
Hence,

\[
\frac{1}{|B_H|^2} \iint_{B_H \times B_H} \left| \left\langle u(s_1 - s_2) f, f \right\rangle_{L^2(X)} - \left\langle \int_X f \, dm_X \right\rangle \right|^2 \, ds_1 \, ds_2
\]

(38)

\[\ll (H^{-\zeta \beta_2} + H^{d(\zeta - 1)}) \mathcal{S}_{\infty, \ell}(f)^2\]

\[= H^{-d \beta_2 + 2/(d + \beta_2)} \mathcal{S}_{\infty, \ell}(f)^2,\]

where we have chosen \(\zeta = d/(d + \beta_2)\) to optimize the error.

Together, the bounds in (36) and (38) imply that

\[
\nu_T \left( \left| f * \sigma_H(x) - \int_X f * \sigma_H \, dm_X \right|^2 \right) \ll \left( R^{-\gamma} H^{4\ell} + H^{-d \beta_2/(d + \beta_2)} \right) \mathcal{S}_{\infty, \ell}(f)^2.
\]

Finally, from (28) and (39), we have

\[
|\mu_{T, \psi}(f)| \ll \left( T^{-1} H + R^{-\gamma/2} H^{2\ell} + H^{-d \beta_2/(2d + 2 \beta_2)} \right) \mathcal{S}_{\infty, \ell}(f).
\]

Since \(\gamma < 1\) and \(R < T\), the first term decays more quickly than the second, and

\[
H^{-d \beta_2/(2d + 2 \beta_2)} = R^{-\gamma/2} H^{2\ell}
\]

\[H = R^{(d + \beta_2)/(4d + 4\ell \beta_2 + d \beta_2)}.
\]

This demonstrates the claim that

\[
|\mu_{T, \psi}(f)| \ll R^{-b} \mathcal{S}_{\infty, \ell}(f)
\]

where

\[
b = \frac{d \beta_2 \gamma}{8d \ell + 8 \ell \beta_2 + 2d \beta_2}
\]

(40)

\[= \frac{d \beta_1 \beta_2}{(nd(3n + 1)(n - 1)^2 + 2)(2n(n - 1)(d + \beta_2) + d \beta_2)}
\]

where we have used the formula for \(\gamma\) in (22).

We will now use this lemma to establish an effective equidistribution bound along multivariate arithmetic sequences.

Let \(K_1, \ldots, K_d \geq 1\) and define \(K\) to be the diagonal matrix

\[K := \text{diag}(K_1, \ldots, K_d) = \begin{pmatrix} K_1 & & \\ & \ddots & \\ & & K_d \end{pmatrix}
\]

and \(|K| = \det(K) = K_1 K_2 \cdots K_d\).

We want to understand the behavior of

\[
S := \sum_{K \in \mathbb{Z}^d \atop K \in B_T} f(x_0 u(Kk)).
\]

(41)
For equidistribution, we want this to be close to $\#(k \in \mathbb{Z}^d \mid Kk \in B_T) \int_X f dm_X \approx \frac{T^d}{|K|} \int_X f dm_X$. For $x_0$ satisfying a basepoint property, we have the following result.

**Theorem 13.** Let $K = \text{diag}(K_1, \ldots, K_d)$ with $T \geq K_1, \ldots, K_d \geq 1$ and determinant $|K|$. Then for all $x_0 \in X$ satisfying (5c) with $T > R > C_K$, we have

$$\left| \sum_{k \in \mathbb{Z}^d \mid Kk \in B_T} f(x_0 u(Kk)) - \frac{T^d}{|K|} \int_X f dm_X \right| \ll \left( T^d R^{-b/(d+1)}|K|^{-d/(d+1)} + \frac{T^{d-1} \max_i K_i}{|K|} \right) \mathcal{R}_\infty, \ell(f)$$

where $C_K := \max(C, (2/\min_i K_i)^{(d+1)/b}|K|^{1/b})$ with $C$ and $\ell$ as in Theorem 11.

**Proof.** Let $\delta > 0$ be small (to be determined) and define the single-variable hat function

$$g_\delta(t) := \max(\delta^{-2}(\delta - |t|), 0)$$

for $t \in \mathbb{R}$ and (through slight abuse of notation) the multivariable function

$$g_\delta(t) := g_\delta(t_1) \cdots g_\delta(t_d)$$

for $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. Notice that $\int_{\mathbb{R}^d} g_\delta(t) dt = 1$ and supp $(g_\delta) \subseteq [-\delta, \delta]^d$.

Define an approximation to the sum $S$ by

$$S_{\text{approx}} := \int_{B_T} \left( \sum_{k \in \mathbb{Z}^d \mid Kk \in B_T} g_\delta(t - Kk) f(x_0 u(t)) \right) dt.$$  

That is, instead of averaging $f$ over the lattice points of $K\mathbb{Z}^d$, we average over small neighborhoods around the lattice points using the bump function $g_\delta$, since $\sum_k g_\delta(t - Kk)$ is supported on a disjoint union of $\delta$-cubes centered around the points of $K\mathbb{Z}^d$ (that is, so long as $\delta < \min_i K_i/2$).

We want to show that $S_{\text{approx}}$ can be written

$$S_{\text{approx}} = \sum_{k \in \mathbb{Z}^d \mid Kk \in B_T} \int_{[-\delta, \delta]^d + Kk} g_\delta(t - Kk) f(x_0 u(t)) dt + r(T, K, f, d)$$

where $r(T, K, f, d)$ is an error term depending on $T, K_1, \ldots, K_d, f$, and dimension $d$. To see observe, that in both (42) and (43) we are integrating $f$ against a sum of bump functions supported on a disjoint union of $\delta$-cubes centered at the lattice points of $K\mathbb{Z}^d$. However, in (42) we are integrating over the region shaded in solid gray in Figure 3, whereas in (43) we are integrating over the region shaded with diagonal lines (that is, we are only integrating against the hat functions whose centers intersect $B_T$). Thus all of the possible error comes from integrating over those $\delta$-cubes that intersect the boundary of $B_T$. Consider a face of $B_T$ that is orthogonal to the $i$th standard basis vector. It will
Figure 3. The area shaded in solid gray indicates the region over which we are integrating in the definition of $S_{\text{approx}}$, whereas the area shaded with diagonal lines represents the region over which we are integrating in our estimate of $S_{\text{approx}}$ given in (43). The difference between the two integrals can be bounded by the number of $\delta$-cubes intersecting the boundary of $B_T$ multiplied by the supremum of $f$. 

intersect at most $T/K_j + O(1)$ of these cubes along an edge in the $j^{th}$ direction for $j \neq i$. Hence, the total number of cubes that face intersects can be bounded by

$$\frac{T}{K_1} \cdot \frac{T}{K_{i-1}} \cdot \frac{T}{K_{i+1}} \cdots \frac{T}{K_d} = T^{d-1} \frac{K_i}{|K|}.$$ 

Since $g_\delta$ integrates to one, the error that results from integrating over one of these $\delta$-cubes is bounded by $\mathcal{J}_{\infty,0}(f)$. Then considering all the faces of $B_T$, we see that the error satisfies

$$|r(T,K,f,d)| \ll \mathcal{J}_{\infty,0}(f) \sum_{i=1}^{d} T^{d-1} \frac{K_i}{|K|} \ll T^{d-1} \frac{\max_i K_i}{|K|} \mathcal{J}_{\infty,0}(f).$$
Then by a change of variables in (43), we have

\[(44) \quad S_{\text{approx}} = \left( \sum_{K \in \mathbb{Z}^d} \int_{[-\delta, \delta]^d} g_\delta(s) f(x_0 u(Kk + s)) ds \right) + r(T, K, f, d).\]

Also, since \(\int_{[-\delta, \delta]^d} g_\delta(s) ds = 1\), we may rewrite the definition of \(S\) in (41) as

\[
S = \sum_{K \in \mathbb{Z}^d} \int_{[-\delta, \delta]^d} g_\delta(s) f(x_0 u(Kk)) ds
\]

and combining this with (44), we obtain

\[
|S_{\text{approx}} - S| \leq \left( \sum_{K \in \mathbb{Z}^d} \int_{[-\delta, \delta]^d} g_\delta(s) ds \right) |f(x_0 u(Kk + s)) - f(x_0 u(Kk))| ds
\]

\[+ |r(T, K, f, d)|.
\]

But note that from property (iv) of Sobolev norms, we have

\[
|f(x_0 u(Kk + s)) - f(x_0 u(Kk))| \ll \mathcal{S}_{\infty, 1}(f) |s| \ll \mathcal{S}_{\infty, 1}(f) \delta
\]

for \(s \in [-\delta, \delta]^d\). Together with our error bound, this implies that

\[
|S_{\text{approx}} - S| \ll \left( \sum_{K \in \mathbb{Z}^d} \int_{[-\delta, \delta]^d} g_\delta(s) ds \right) \delta \mathcal{S}_{\infty, 1}(f) + T^{d-1} \frac{\max_i K_i}{|K|} \mathcal{S}_{\infty, 0}(f)
\]

\[= \left( \# \{K \in \mathbb{Z}^d | Kk \in B_T \} \delta + T^{d-1} \frac{\max_i K_i}{|K|} \right) \mathcal{S}_{\infty, 1}(f)
\]

once again, because \(\int_{[-\delta, \delta]^d} g_\delta(s) ds = 1\). But \(\# \{K \in \mathbb{Z}^d | Kk \in B_T \} \approx T^d / |K|\), also with an error of magnitude \(\ll T^{d-1} \max_i K_i / |K|\) (for reasons analogous to those illustrated in Figure 3). Therefore

\[(45) \quad |S_{\text{approx}} - S| \ll \left( \frac{T^d}{|K|} \delta + \frac{T^{d-1} \max_i K_i}{|K|} \right) \mathcal{S}_{\infty, 1}(f).
\]

To show that \(S_{\text{approx}}\) and \(\frac{T^d}{|K|} \int_X f dm_X\) are close, we observe that by Poisson summation,

\[
\sum_{K \in \mathbb{Z}^d} g_\delta(t - Kk) = \sum_{K \in \mathbb{Z}^d} g_\delta(t + Kk)
\]

\[= \sum_{K \in \mathbb{Z}^d} \tilde{g}_\delta(K^{-1}t + k)
\]

\[(46) \quad = \sum_{K \in \mathbb{Z}^d} \psi_{K^{-1}k}(t) \tilde{g}_\delta(k).
\]
where \( \psi_{K^{-1}k}(t) = e^{2\pi i k(K^{-1}t)} \) and \( \widehat{g_\delta} \) is the multivariate Fourier transform of \( g_\delta(x) = g_\delta(Kx) \). When we substitute (46) into the definition of \( S_{\text{approx}} \) given in (42), we get

\[
S_{\text{approx}} = \int_{B_T} \left( \sum_{k \in \mathbb{Z}^d} \psi_{K^{-1}k}(t) \widehat{g_\delta}(k) \right) f(x_0 u(t)) \, dt
\]

\[
= \sum_{k \in \mathbb{Z}^d} \widehat{g_\delta}(k) \left( \int_{B_T} \psi_{K^{-1}k}(t) f(x_0 u(t)) \, dt \right)
\]

where Fubini’s Theorem allows us to switch the order of the sum and the integral. Similarly,

\[
\frac{T^d}{|K|} \int_X f \, dm_X = \left( \int_{B_T} g_\delta(t - Kk) \, dt + \Theta \left( \frac{T^{d-1} \max_i K_i}{|K|} \right) \right) \int_X f \, dm_X
\]

\[
= \sum_{k \in \mathbb{Z}^d} \widehat{g_\delta}(k) \left( \int_{B_T} \psi_{K^{-1}k}(t) \int_X f \, dm_X \, dt \right) + \Theta \left( \frac{T^{d-1} \max_i K_i}{|K|} \mathcal{S}_{\infty,0}(f) \right)
\]

where we have used that \( |\int_X f \, dm_X| \leq \mathcal{S}_{\infty,0}(f) \). Thus

\[
\left| S_{\text{approx}} - \frac{T^d}{|K|} \int_X f \, dm_X \right| = \left| \sum_{k \in \mathbb{Z}^d} \widehat{g_\delta}(k) \int_{B_T} e^{2\pi i k(K^{-1}t)} \left( f(x_0 u(t)) - \int_X f \, dm_X \right) \, dt \right|
\]

\[
+ \Theta \left( \frac{T^{d-1} \max_i K_i}{|K|} \mathcal{S}_{\infty,0}(f) \right)
\]

(47)

\[
(48)
\]

Then since \( R > C \), we can apply Lemma 12 to obtain

\[
\left| S_{\text{approx}} - \frac{T^d}{|K|} \int_X f \, dm_X \right| \ll f \cdot T^d R^{-b} \mathcal{S}_{\infty,\ell}(f) \sum_{k \in \mathbb{Z}^d} \widehat{g_\delta}(k) + \frac{T^{d-1} \max_i K_i}{|K|} \mathcal{S}_{\infty,0}(f)
\]

(by direct computation we can see that \( \widehat{g_\delta} \) is positive). Observe how it was crucial here that the result in Lemma 12 was uniform over characters.

Finally, again by Poisson summation, we have

\[
\sum_{k \in \mathbb{Z}^d} \widehat{g_\delta}(k) = \sum_{k \in \mathbb{Z}^d} g_\delta(k)
\]

\[
= \sum_{k \in \mathbb{Z}^d} g_\delta(Kk)
\]

\[
= g_\delta(0, \ldots, 0) = \delta^{-d}
\]

since \( \text{supp } (g_\delta) \subseteq [-\delta, \delta]^d \) and \( \delta < \min_i K_i/2 \) implies that \( g_\delta(Kk) = 0 \) for \( k \neq (0, \ldots, 0) \). Substituting this into equation (48), combining it with (45), and using
property (i) of Sobolev norms, we get
\[ S - \frac{T^d}{|K|} \int_X f \, dm_X \ll \left( T^d R^{-b} \delta^{-d} + \frac{T^d}{|K|} \delta + \frac{T^{d-1} \max_i K_i}{|K|} \right) \mathcal{I}_{\infty, \ell}(f). \]

We can optimize the first two terms by choosing \( \delta = (|K|/R^b)^{1/(d+1)}. \) Observe that our only restriction on \( \delta \) was that \( \delta < \min_i K_i/2. \) This will be achieved with our choice of \( \delta \) so long as \( R > (2/\min K_i)^{(d+1)/b}|K|^{1/b}. \) Thus, under these conditions,
\[ S - \frac{T^d}{|K|} \int_X f \, dm_X \ll \left( T^{d-b/(d+1)} |K|^{-d/(d+1)} + \frac{T^{d-1} \max_i K_i}{|K|} \right) \mathcal{I}_{\infty, \ell}(f). \]

If \( K \) has all diagonal entries of equal weight (in abuse of notation, say all of weight \( K \)) then we get the following corollary which will be of use to us in the next section.

**Corollary 14.** Let \( T \geq K \geq 1. \) There exists a constant \( \tilde{C} > 0 \) (depending only on \( n \) and \( d \)) such that for all \( x_0 \in X \) satisfying (5c) with \( T > R > \tilde{C}, \) we have
\[ \left| \sum_{k \in \mathbb{Z}^d \atop K k \in B_T} f(x_0 u(Kk)) - \frac{T^d}{K^d} \int_X f \, dm_X \right| \ll T^{d-b/(d+1)} K^{-d/(d+1)} \mathcal{I}_{\infty, \ell}(f). \]

**Proof.** This is a straightforward application of the previous theorem, observing that in this case \( (2/\min K_i)^{(d+1)/b}|K|^{1/b} = (2^{d+1}/K)^{1/b} \leq 2^{(d+1)/b} \) since \( K \geq 1. \) Thus the theorem holds with \( \tilde{C} = \max(C, 2^{(d+1)/b}). \) Moreover, the second error term in Theorem 13 in this case is simply \( T^{d-1} K^{1-d}, \) and since \( K, R < T \) and \( b < 1, \) this term decays more quickly than the first and can be ignored. \( \square \)

**Remark.** For \( X = \Gamma \backslash G \) where \( \Gamma \) is a cocompact lattice, we have the following basepoint-independent versions of Lemma 12, Theorem 13, and Corollary 14.

**Lemma 15.** There exists \( b > 0 \) (depending\(^7\) on \( n, d, \) and \( \Gamma \)) such that for all \( T \) large enough, we have
\[ |\mu_{T, \psi}(f)| \ll T^{-b} \mathcal{I}_{\infty, \ell}(f) \]
for any \( f \in C^\infty(X), \) \( x_0 \in X, \) and additive character \( \psi. \)

\(^7\) In this case, \( b = d \beta_1 \beta_2/(3n+1)(n-1) + 2\beta_1(2n+1)(2n-1)/(d+\beta_2) + d \beta_2. \) Since \( \beta_1 \) and \( \beta_2 \) depend on the spectral gap for the action of \( \text{SL}_n(\mathbb{R}) \) on \( X \) we may remove dependence on \( \Gamma \) for \( n \geq 3 \) or for \( n = 2 \) if \( \Gamma \) is a congruence lattice.
**Theorem 16.** Let $K = \text{diag}(K_1, \ldots, K_d)$ with $T \geq K_1, \ldots, K_d \geq 1$ and determinant $|K|$. Then for all $T$ large enough, we have

$$\left| \sum_{k \in \mathbb{Z}^d \atop Kk \in \mathcal{B}_T} f(x_0 u(Kk)) - \frac{T^d}{|K|} \int_X f \, dm_X \right| \ll T^{d-b/(d+1)} |K|^{-d/(d+1)} \mathcal{S}_{\infty, \ell}(f)$$

for all $f \in C^\infty(X)$ and $x_0 \in X$.

**Corollary 17.** Let $T \geq K \geq 1$. Then for all $T$ large enough, we have

$$\left| \sum_{k \in \mathbb{Z}^d \atop Kk \in \mathcal{B}_T} f(x_0 u(Kk)) - \frac{T^d}{K^d} \int_X f \, dm_X \right| \ll T^{d-b/(d+1)} K^{-d^2/(d+1)} \mathcal{S}_{\infty, \ell}(f)$$

for all $f \in C^\infty(X)$ and $x_0 \in X$.

The proofs of these results are completely analogous to the corresponding proofs for $\text{SL}_n(\mathbb{Z}) \backslash \text{SL}_n(\mathbb{R})$, but use the basepoint-independent equidistribution result stated in (23) instead of Theorem 11.

5. Sieving and orbits along almost-primes

5.1. $\Gamma$ cocompact. Let $\Gamma$ be a cocompact lattice in $G = \text{SL}_n(\mathbb{R})$ and let $u(t)$ be an abelian horospherical flow on $X = \Gamma \backslash G$, as in Section 4. We know that the orbit of $u(t)$ equidistributes with a uniform rate for all $x_0 \in X$, and that as a consequence we have a uniform rate of equidistribution along multivariable arithmetic sequences of the form given in Corollary 17. Here and throughout this section, assume $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$.

We want to understand the behavior of orbits at almost-prime entries of $u(t)$. More precisely, we want to understand averages of positive $f \in C_c^\infty(X)$ over points in $B_T$ that have entries with fewer than a certain fixed number of primes in their prime factorization.

To investigate this question, we will use the following combinatorial sieve theorem (see [32, Theorem 7.4], [35, Sections 6.1-6.4] for a form more similar to that stated here).

**Theorem 18 ([32, Theorem 7.4]).** Let $A = \{a_n\}$ be a sequence of nonnegative numbers and let $P = P(z) = \prod_{p < z} p$ be the product of primes less than $z$. Define

$$S(A, P) = \sum_{(n, P) = 1} a_n \quad \text{and} \quad S_K(A) = \sum_{n \equiv 0 \mod K} a_n.$$

Then suppose
(i) There exists a multiplicative function $g(K)$ on $K$ squarefree such that

$$S_K(A) = g(K)X + r_K(A)$$

and for some $c_1 > 1$, we have $0 \leq g(p) < 1 - \frac{1}{c_1}$ for all primes $p$.

(ii) $A$ has level distribution $D(X)$, i.e., there is $\epsilon > 0$ such that

$$\sum_{K < D \text{ squarefree}} |r_K(A)| \leq C_e X^{1-\epsilon}.$$ 

(iii) $A$ has sieve dimension $r$, i.e., there exists a constant $c_2 > 0$ such that for all $2 \leq w \leq z$, we have

$$-c_2 \leq \sum_{w \leq p \leq z} g(p) \log p - r \log \frac{z}{w} \leq c_2.$$ 

Then for $s > 9r$, $z = D^{1/s}$, and $X$ large enough (depending on $\epsilon$, $C_e$, and $r$), we have

$$S(A, P) \approx_{c_1, c_2, r} X \left( \frac{X}{\log X} \right)^r.$$ 

In the context of our problem, we want to define

$$S(A, P) := \sum_{\substack{k \in B_T \mod K \mid k_1 \cdots k_d = n \gcd(k_1, \ldots, k_d, P) = 1}} f(x_0 u(k))$$

where $f \in C_c^\infty(X)$, $f \geq 0$, and $P$ is the product of primes less than $z$ (to be determined). That is, we are summing over integer points in $B_T$ with entries containing no primes smaller than $z$ in their prime factorizations. Then let

$$A = \{a_n\} := \left\{ \sum_{\substack{k \in B_T \mid k_1 \cdots k_d = n}} f(x_0 u(k)) \right\}$$

and observe that

$$S_K(A) := \sum_{n=0}^{\tilde{B}_T} \sum_{\substack{k \in B_T \mid k_1 \cdots k_d = n}} f(x_0 u(k)) = \sum_{k \in \tilde{B}_T} f(x_0 u(k))$$

where $\tilde{B}_T = (0, T]^d$ (since the index $n$ starts at 1 we want to avoid counting terms of the form $K$ divides 0).

Notice that $K|k_1 \cdots k_d$ if and only if $K|k_1 \cdots (k_i + K) \cdots k_d$, that is, the collection of points that we are summing over is periodic with period $K$ in each coordinate. Thus we can rewrite $S_K(A)$ as a sum over cubic grids of side length $K$ based at each point in the first box $B_K$:

$$\sum_{k \in \tilde{B}_T} f(x_0 u(k)) = \sum_{k \in \tilde{B}_K} \left( \sum_{k \in \tilde{B}_T} f(x_0 u(k)u(Kk)) + \mathcal{O}(T^{d-1} K^{1-d} \mathcal{S}_{\infty, 0}(f)) \right)$$

(49)
Figure 4. In $S_K(A)$ we are summing over the integer points in $\tilde{B}_T$ such that $K|k_1 \cdots k_2$ (filled in black). We may do this by summing over shifted grids based at each of the points in the first box $\tilde{B}_K$ (filled in gray). However, this introduces an error determined by $\mathcal{S}_{\infty,0}(f)$ and the number of points in each of these shifted grids falling outside $B_T$ (filled in white). The number of such points can be bounded by $T^{d-1}K^{1-d}$, as we have seen before.

where the error arises from the fact that a point $\hat{k} + Kk$ for $Kk \in B_T$ may, in fact, fall outside of $B_T$ (see Figure 4). From Corollary 17, we know that at each basepoint $x_k = x_0u(k)$, we have

$$
\sum_{Kk \in B_T} f(x_k u(Kk)) = \frac{T^d}{K^d} \int f dm_X + \Theta\left(T^{d-b/(d+1)}K^{-d^2/(d+1)}\mathcal{S}_{\infty,\ell}(f)\right). 
$$

If we let

$$
G_d(K) := \#\{k \in \tilde{B}_K | k_1 \cdots k_d \equiv 0 \mod K\}
$$

then (49) together with (50) says that

$$
S_K(A) = \sum_{Kk \in B_T \atop K|k_1 k_2 \cdots k_d} f(x_k u(k)) = \frac{G_d(K)}{K^d} \mathcal{X} + r(f, K, T)
$$
where \( \mathcal{X} = T^d \int f \, dm_X \) and

\[
|r(f, K, T)| \ll T^d \int \left( \frac{T}{\log T} \right)^d f dm_X.
\]

This suggests that our function \( g(K) \) in Theorem 18 should be \( g_d(K)/K^d \). It remains to show that this function satisfies the sieve axioms (i) and (iii) and that the corresponding remainders satisfy axiom (ii) for appropriately chosen \( D(\mathcal{X}) = D(T) \). Verifying these conditions gives us the following theorem.

**Theorem 19.** Let \( u \) be a \( d \)-dimensional abelian horospherical flow on \( X = \Gamma \backslash \text{SL}_n(\mathbb{R}) \) for \( \Gamma \) cocompact, and let \( P \) be the product of primes less than \( T^a \) for \( \alpha = b/9d^2 \), where \( b \) is the constant from Lemma 15. Then for any \( x_0 \in X \), positive \( f \in C_c(X) \), and \( T \) large enough (depending on \( \alpha \), \( n \), \( d \), \( \Gamma \), and \( f \)), we have

\[
\sum_{\substack{k \in B_T \\gcd(k_1 \cdots k_d, P) = 1}} f(x_0 u(k)) \approx \left( \frac{T}{\log T} \right)^d \int f \, dm_X.
\]

**Remark.** The implicit constants in the conclusion of this theorem depend only on \( d \), however \( b \) depends on \( n \), \( d \), and \( \Gamma \), where dependence on \( \Gamma \) may be removed if \( n \geq 3 \) or if \( \Gamma \) is a congruence lattice.

**Remark.** Let \( \phi(x, y) \) be the number of positive integers \( \leq x \) not divisible by any prime \( \leq y \). It is known that

\[
\phi(x, y) = \frac{x \omega(\log x/\log y) - y}{\log y} + \mathcal{O} \left( \frac{x}{(\log y)^2} \right)
\]

where \( \omega : [1, \infty) \to [1/2, 1] \) is the Buchstab function. Thus, the number of integers in \([0, T]\) not divisible by any prime less than \( T^a \) for \( \alpha < 1 \) is given by

\[
\phi(T, T^a) = \frac{\omega(1/\alpha) T}{\alpha \log T} - \frac{T^a}{\alpha \log T} + \mathcal{O} \left( \frac{T}{(\alpha \log T)^2} \right).
\]

Thus the number of points \( k \in B_T \) such that \( \gcd(k_1 \cdots k_d, P) = 1 \) where \( P \) is the product of primes less than \( T^a \) is \( \phi(T, T^a) T^d \), which grows asymptotically like \( (T/\log T)^d \) as \( T \to \infty \). Although our result above only states that there is an upper and lower bound with respect to this quantity, it hints that there may be underlying equidistribution behavior.

**Remark.** Notice that \( G_2(K) = \sum_{f=1}^K \gcd(K, f) \) is Pillai’s arithmetical function,\(^8\) a multiplicative function first considered by Cesàro and rediscovered by Pillai in [55] which counts the number of non-congruent solutions to the equation \( k_1 k_2 \equiv 0 \mod K \). From the definition of \( g_d \) in (51), we can see that \( g_d(K) \)

---

\(^8\)This is a classical function that has been well studied. In terms of Dirichlet convolution, we have the useful identities \( G_2 = \text{Id} \ast \phi \) and \( G_2 = \mu \ast (\text{Id} \ast \tau) \), where \( \phi \) is Euler’s totient function, \( \mu \) is the Möbius function, and \( \tau \) is the divisor function. In [8], Broughan used this to derive a closed form for the Dirichlet series in terms of the Riemann zeta function, as well as an asymptotic formula for the partial sums of the Dirichlet series. The asymptotics for partial sums of the Dirichlet series were later refined by [5], [9], and [66]. The values of \( G_2(K) \) for \( K = 1, 2, 3, \ldots \) are given as sequence A018804 in the OEIS [1].
counts the number of non-congruent solutions to $k_1 k_2 \cdots k_d \equiv 0 \pmod{K}$, so it can be considered a generalization of Pillai’s arithmetical function.\footnote{Other generalizations of Pillai’s arithmetical function have been studied. Examples include [13], [67], [4], [33], and [68], however none of these include the generalization given here. In [69], Tóth considers a generalization that is very similar to ours, and in the notation of that paper, $G_d(K) = A_{d-1}(K) K^{d-1}$. Sieve axioms (i) and (iii) can thus be considered corollaries of results proved in [69], but we prove them independently in order to keep the paper self-contained. Tóth also gives a formula for the Dirichlet series of this generalization in terms of the Dirichlet series of a related arithmetic function, however we will need an explicit estimate for the partial sums of the Dirichlet series where it does not converge in order to verify sieve axiom (ii).} The generalized Pillai’s functions $G_d$ have several interesting properties and interpretations that are not necessary for the proof of Theorem 19. We have included these properties in Appendix A for those interested.

Proof. We need to show that sieve axioms (i), (ii), and (iii) are satisfied for

$$S_K(A) := \sum_{k \in \mathcal{B}_T} f(x_0 u(k)) = g(K) X^d + r(f, K, T)$$

where $g(K) = G_d(K) / K^d$, $X = T^d \int f \, dm_X$, and

$$|r(f, K, T)| \ll_T G_d(K) T^{d-b/(d+1)} K^{-d^2/(d+1)} J_{\infty, \ell}(f).$$

For $K_1$ and $K_2$ coprime, the Chinese Remainder Theorem implies that there is a bijection between $(k_1, \ldots, k_d) \in \mathcal{B}_{K_1} K_2$ such that $k_1 \cdots k_d \equiv 0 \pmod{K_1 K_2}$ and $(\ell_1, \ldots, \ell_d, \ell_1', \ldots, \ell_d') \in \mathcal{B}_{K_1} K_2$ such that $\ell_1 \cdots \ell_d \equiv 0 \pmod{K_1}$ and $\ell_1' \cdots \ell_d' \equiv 0 \pmod{K_2}$, where $k_i$ is the unique integer in $\{1, \ldots, d\}$ such that $k_i \equiv \ell_i \pmod{K_1}$ and $k_i \equiv \ell_i' \pmod{K_2}$. By counting the number of solutions in both settings, we have that $G_d(K_1 K_2) = G_d(K_1) G_d(K_2)$, which shows that $G_d$ (and hence $g$) is multiplicative.

For $p$ prime and $k \in \mathcal{B}_p$, $k_1 \cdots k_d \equiv 0 \pmod{p}$ implies that $k_i = p$ for some $i$. Then the number of such solutions $G_d(p)$ will be the total number of $k \in \mathcal{B}_p$ except for those with $k_i \in \{1, \ldots, p-1\}$ for all $i \in \{1, \ldots, d\}$. Thus for $p$ prime, we have

$$G_d(p) = p^d - (p-1)^d.$$  

Therefore

$$0 < g(p) = \frac{p^d - (p-1)^d}{p^d} = 1 - \left(\frac{p-1}{p}\right)^d \leq 1 - \left(\frac{2-1}{2}\right)^d = 1 - \frac{1}{2^d}$$

since $p \geq 2$. So sieve axiom (i) is satisfied with, e.g., $c_1 = 2^{d+1}$.

For prime $p$, we have the bound $G_d(p) = p^d - (p-1)^d < d p^{d-1}$. By the multiplicity of $G_d$, this implies that for arbitrary squarefree $K$, we have
Almost-primes in horospherical flows

\[ G_d(K) < d^{\omega(K)} K^{d-1}, \] where \( \omega(K) = \Omega(K) \) is the number of (distinct) prime factors of \( K \). Then

\[
\sum_{K < D} \text{squarefree} |r(f, K, T)| \ll \mathcal{S}_{\infty, \ell}(f) T^{d-b/(d+1)} \sum_{K < D} \text{squarefree} G_d(K) K^{-d^2/(d+1)}
\]

\[
\ll \mathcal{S}_{\infty, \ell}(f) T^{d-b/(d+1)} \sum_{K < D} d^{\omega(K)} K^{-1/(d+1)}.
\]

Now observe that for any \( \epsilon_1 > 0 \), \( \omega(K) < \epsilon_1 \log(K) \) for all but a finite number of squarefree \( K \), so by appropriately adjusting the implicit constant (depending on \( \epsilon_1 \)) we may write

\[
\sum_{K < D} |r(f, K, T)| \ll \mathcal{S}_{\infty, \ell}(f) T^{d-b/(d+1)} \sum_{K < D} K^{-1/(d+1)+\epsilon_1}
\]

\[
\ll \mathcal{S}_{\infty, \ell}(f) T^{d-b/(d+1)} \sum_{K < D} K^{-1/(d+1)+\epsilon_1}.
\]

Now if we let \( D = T^\eta \) for any \( \eta < b/d \), say \( \eta = b/d - 2(d+1) \epsilon \) for \( \epsilon > 0 \), and set \( \epsilon_1 = d^2 \epsilon / b \), we get that

\[
\sum_{K < D} |r(f, K, T)| \ll \mathcal{S}_{\infty, \ell}(f) T^{d(1-\epsilon)} \ll \mathcal{S}_{\infty, \ell}(f) T^{d(1-\epsilon)}
\]

which demonstrates sieve axiom (ii).

Finally, to verify sieve axiom (iii), notice that (by the binomial theorem)

\[
g(p) = p^d - (p-1)^d = d \frac{d}{p} - \sum_{i=2}^d a_i p^i
\]

where \( a_i = (-1)^i \binom{d}{i} \). Since \( \sum_{j=1}^\infty \log(j) / j^i \) converges for any \( i > 1 \), we have that

\[
\left| \sum_{w \leq p \leq z} \sum_{i=2}^d \frac{a_i \log p}{p^i} \right| \leq \sum_{i=2}^d |a_i| \sum_{j=1}^\infty \frac{\log j}{j^i} = C_2
\]

and by a corollary of the Prime Number Theorem, we know that

\[
\sum_{p \leq x} \frac{\log p}{p} = \log(x) + \mathcal{O}(1).
\]

Hence,

\[
\sum_{p \leq z} \frac{\log p}{p} - \sum_{p < w} \frac{\log p}{p} = \log(z) - \log(w) + \mathcal{O}(1)
\]

\[
\sum_{w \leq p \leq z} \frac{\log p}{p} = \log \frac{z}{w} + \mathcal{O}(1),
\]

i.e., there exists \( C'_2 \) such that

\[
\left| \sum_{w \leq p \leq z} \frac{\log p}{p} - \log \frac{z}{w} \right| \leq C'_2
\]
for all $2 \leq w \leq z$. Putting (53), (54), and (55) together, we see that
\[
\left| \sum_{w \leq p < z} g(p) \log p - d \log \frac{z}{w} \right| \leq d \left| \sum_{w \leq p < z} \log p - \log \frac{z}{w} \right| + \sum_{w \leq p < z} \sum_{i=2}^{d} \frac{a_i \log p}{p^i} \leq dC_2' + C_2,
\]
which shows that axiom (iii) is satisfied with sieve dimension $r = d$ and $c_2 = dC_2' + C_2$.

Since we have demonstrated that sieve axioms (i), (ii), and (iii) hold, we have the conclusion of Theorem 18. This, along with the various dependencies of the constants in that theorem on $n, d, f, \text{ and } \Gamma$, implies our result. \qed

From this theorem we may easily deduce the density of almost-prime times in arbitrary horospherical flows.

**Theorem 1.** Let $u(t)$ be a horospherical flow of dimension $d$ on $X = \Gamma \backslash \text{SL}_n(\mathbb{R})$ for $\Gamma$ cocompact. Then there exists a constant $M$ (depending only on $n, d, \text{ and } \Gamma$) such that for any $x_0 \in X$, the set
\[
\{x_0u(k_1, k_2, \ldots, k_d) | k_i \in \mathbb{Z} \text{ has fewer than } M \text{ prime factors} \}
\]
is dense in $X$.

**Remark.** Explicitly, we may take $M$ to be any integer satisfying
\[
M > \frac{9d((3n+1)(n-1)+2)(2n(n-1)(d+\beta_2)+d\beta_2)}{\beta_1 \beta_2}
\]
where $\beta_1, \beta_2$ are as in Corollary 9 for the element $a_t$ is as in (25). For such $a_t$, $\beta_1$ may be taken to be $\frac{1}{2}(1 - \epsilon) \min(m, n - m)$ for $\epsilon > 0$ (see [52, Theorem A] and following discussion regarding $\text{SL}_n(\mathbb{R})$). Furthermore, if we write $u(t_1, \ldots, t_d)$ as $kak'$ with $k, k' \in K$ and $a = \text{diag}(a_1, \ldots, a_n) \in A$ and apply [52, Theorem A] along with the fact that $|a_i| < 1 + |t|$, we can deduce that $1 - \epsilon \leq \beta_2$ for $\epsilon > 0$.

For example, considering an abelian horospherical flow in $\text{SL}_3(\mathbb{R})$ ($n = 3$, $d = 2$, $\beta_1 = 1/2 + \epsilon$, $\beta_2 = 1 - \epsilon$) we may take $M = 30097$. In general, $M$ is $\Omega(n^4d^2\beta_1^{-1}\beta_2^{-1})$, so in the minimal case where $d = n - 1$, we have $M = \Omega(n^5)$, and in the maximal case where $d = n(n-1)/2$ we have $M = \Omega(n^7)$.

**Proof.** First consider $u(t)$ abelian. If an integer $k < T$ has no prime factors less than $T^\alpha$, then it must have fewer than $1/\alpha$ prime factors total. Hence, if we take $f$ to be a positive function supported on any small neighborhood, Theorem 19 tells us that we can take $T$ large enough so that averaging $f$ over integer points in $B_T$ with no prime factors less than $T^\alpha$ has a positive lower bound. This means that the set of almost-prime times with fewer than $M$ prime factors hitting any neighborhood is nonempty, where $M = 1/\alpha$. From Theorem 19, we have $\alpha < b/9d^2$, then substituting for $b$ gives us the formula in (56).

To move from abelian to arbitrary horospherical flows, observe that any horospherical flow can be written as the product of a unipotent flow (not necessarily
horospherical) and an abelian horospherical flow. Explicitly, an element of $U$ given by

$$
\begin{pmatrix}
I_{m_1} & & \\
0 & I_{m_2} & \\
& & \ddots \\
& & & I_{m_N}
\end{pmatrix}
$$

$$
\begin{pmatrix}
a_{ij} \\
b_{ij} \\
& \ddots \\
& & \ddots \\
0 & & & I_{m_N}
\end{pmatrix}
$$

can be expressed in the form

$$
\begin{pmatrix}
I_{m_1} & & \\
0 & I_{m_2} & \\
& & \ddots \\
& & & I_{m_N}
\end{pmatrix}
$$

$$
\begin{pmatrix}
a_{ij} \\
b_{ij} \\
& \ddots \\
& & \ddots \\
0 & & & I_{m_N}
\end{pmatrix}
$$

$$
\begin{pmatrix}
I_{m_1} & & \\
0 & I_{m_2} & \\
& & \ddots \\
& & & I_{m_N}
\end{pmatrix}
$$

Let $d = m_1(n - m_1)$, and let $\bar{u}(t_1, \ldots, t_d)$ represent the abelian horospherical flow on the right and $v(t_{d+1}, \ldots, t_d)$ represent the unipotent flow on the left, so that

$$u(t_1, \ldots, t_d) = v(t_{d+1}, \ldots, t_d) \bar{u}(t_1, \ldots, t_d)$$

for any $t_1, \ldots, t_d \in \mathbb{R}^d$. Then from the density result in the abelian case, we know that the set

$$\{ \bar{x} \bar{u}(k_1, k_2, \ldots, k_d) | k_i \in \mathbb{Z} \text{ has fewer than } M \text{ prime factors} \}$$

is dense for appropriately chosen $M$ and $\bar{x} = x_0 v(k_{d+1}, \ldots, k_d)$, where $\{k_{d+1}, \ldots, k_d\}$ are any fixed $M$-almost-primes. This is a subset of the almost-prime times in the larger horospherical flow, so the result follows.

**Remark.** In fact, we can make Theorem 1 effective by examining the proof of Theorem 19. We will get a positive lower bound from the sieve in Theorem 18 so long as the main term surpasses the error term from sieve axiom (ii), i.e., so long as

$$C_{\epsilon} \mathcal{X}^{1-\epsilon} \ll \frac{\mathcal{X}}{(\log \mathcal{X})^{r}}.$$

From (52) we see that this means that for fixed $\epsilon$ (suppressing dependence on $\Gamma$) we require

$$\mathcal{G}_{\infty, \ell}(f) T^{d(1-\epsilon)} \ll \epsilon \frac{T^d \int f dm_X}{(d \log T + \log \int f dm_X)^d}.$$

\footnote{The dimension of the relevant abelian subgroup is $m_1(n - m_1)$, but we can always bound this in terms of $d$ if desired.}
For any small ball of radius $r > 0$, choose $f$ as in Lemma 4 supported on that ball. For such an $f$, we get a positive lower bound in Theorem 19 if

$$r^{-(f+(n^2-1)/2)} \ll_c c^{de/(\log T)^d}.$$ 

Recall that $\ell = n(n-1)/2$. It is sufficient to require that

$$r^{-((2n^2-n-1)/2)} \ll_c c^de/2$$

$$T^{-de/(2n^2-n-1)} \ll_c r.$$ 

This gives us the following effective density statement: For any $x_0 \in X$, the subset of its $U$-orbit consisting of integer times of norm less than $T$ with fewer than $M$ prime factors is $O_{c}(T^{-de/(2n^2-n-1)})$-dense in $X$, where $M = 9d^2/(b - 2d(d+1)e)$.

### 5.2. The space of lattices

Now consider $X = \Gamma \backslash G$ for the non-cocompact lattice $\Gamma = \text{SL}_n(\mathbb{Z})$. Since we no longer have a uniform rate of equidistribution for our abelian horospherical flow $u(s)$, we will consider a basepoint $x_0 = \Gamma g_0 \in X$ satisfying a Diophantine condition of the following form.

**Definition 2.** We say that $x = \Gamma g$ is strongly polynomially $\delta$-Diophantine if there exists a sequence $T_i \to \infty$ as $i \to \infty$ such that

$$\inf_{u \in \Lambda/\Lambda(\mathbb{Z}) \sim \{0\}} \sup_{j=1, ..., n-1} \| w g u(t) \| > T_i^\delta$$

for all $i \in \mathbb{N}$.

The motivation for this definition is that, as in the compact setting, we will want to apply sieving to learn about integer points having few prime factors. However, unlike in the compact case, we do not have a uniform rate of equidistribution, so we must consider the effect of the basepoint. For a given time-scale $T$, to obtain information about almost-primes of a certain order, we would want $R$ in the basepoint condition (5c) to look like a small power of $T$ (say $T^\delta$). However, a theorem like that of Theorem 19 will require $T$ be “large enough,” which depends on the function $f$, and so any fixed time-scale $T$ is insufficient. Moreover, the constant $\delta$ we are able to take at one time-scale may not work for a different time-scale, which affects the number of prime factors we allow for our almost-prime points. The condition given in Definition 2 ensures that for any function (hence any neighborhood in $X$) we will be able to find a time-scale large enough so that our sieving provides positive information about almost-primes of the same, fixed order.

Before moving on to the main theorem of this section, we briefly remark that this definition is a meaningful one. In view of results in [39], we see that not only do such points exist, but any generic point for the flow $u$ will satisfy this definition for some positive $\delta$. 

---

11 Recall that a subset is $\delta$-dense if any ball of radius $\delta$ contains a point in the subset.
Theorem 20. Let $u$ be a $d$-dimensional abelian horospherical flow on $X = \text{SL}_n(\mathbb{Z}) \backslash \text{SL}_d(\mathbb{R})$ and let $x_0 \in X$ be strongly polynomially $\delta$-Diophantine. Let $P$ be the product of primes less than $T^n$ for $\alpha < \delta b n / 9 d (d^2 + b n x)$, where $b$ is the constant from Lemma 12 and $\kappa = \min(m, n - m)$ for $m$ as in (24). Then for any positive $f \in C_c^\infty(X)$ there exists a sequence $T_i \to \infty$ as $i \to \infty$ where

$$\sum_{k \in B_{T_i}} f(x_0 u(k)) \approx \left(\frac{T_i}{\log T_i}\right)^d \int f \, dm_X.$$  

Proof. Let $f \in C_c^\infty(X), f \geq 0$, and let $u$ be an abelian horospherical flow as given in (24) of Section 4. As in the compact setting, we want to use our equidistribution theorem for arithmetic sequences to say that

$$S_K(A) := \sum_{k \in B_{T_i}} f(x_0 u(k))$$

(58)

$$= \sum_{k \in B_K} \left( \sum_{k \in B_{T_i}} f(x_0 u(k)) u(K k) \right) + O(T_i d^{-1} K^{1-d})$$

$$= g(K) A + r(f, K, T_i)$$

where $g(K) = G_d(K) / K^d$, $A = T_i^d \int f \, dm_X$, and the error terms can be suitably controlled. Unfortunately, we cannot apply the same equidistribution result to the shifted basepoints $x_0 u(k)$ since they will not necessarily satisfy the same Diophantine condition. However, since $K$ is understood to be small in comparison to the $T_i$, all of the points in $B_K$ lie comparatively close to $x_0$. Then since the Diophantine property varies continuously, we expect the points in this region to satisfy a Diophantine condition not much worse than that of $x_0$, and in fact we can make this quantitative.

Observe that if $x_0$ is strongly polynomially $\delta$-Diophantine, it means that condition (5c) holds for the sequence of parameters $T = T_i$ and $R = T_i^{\delta / q}$, where $q = \sum_{\Lambda_i < 0} -m_i \lambda_i = d / n$ for abelian $u$ of this form. That is, for all $j \in \{1, \ldots, n - 1\}$ and $w \in \Lambda^j(\mathbb{Z}^n) \sim \{0\}$, we have

(59) \[ \exists t \in [0, T_i]^d \text{ s.t. } \|w g_0 u(t)\| = \|w g_0 u(k) u(t) u(-k)\| \geq T_i^{d \delta}. \]

Recall that any $w \in \Lambda^j(\mathbb{R}^n)$ can be written as a sum $w = \sum_I w_I e_I$ over multi-indices $I = (i_1, \ldots, i_j)$ with $0 < i_j < \cdots < i_1 < n$, coefficients $w_I \in \mathbb{R}$, and basis elements $e_I = e_{i_1} \wedge \cdots \wedge e_{i_j}$ where $\{e_i\}_{1 \leq i \leq n}$ is the standard basis on $\mathbb{R}^n$. Recall also that the norm above is defined by

$$\|w\| = \max_I |w_I|$$

and that $G$ acts linearly on $\Lambda^j(\mathbb{R}^n)$ by sending a basis vector $e_{i_1} \wedge \cdots \wedge e_{i_j}$ to

$$(e_{i_1} \wedge \cdots \wedge e_{i_j}) g = (e_{i_1} g) \wedge \cdots \wedge (e_{i_j} g).$$
Since our abelian horospherical subgroup has the form given in (24), we can write an arbitrary \( u \in B_K^{-1} = u([-K,0]_d) \) as

\[
(60) \quad u = \begin{pmatrix}
I_m & \begin{pmatrix} a_1(m+1) & \cdots & a_1n \\
\vdots & & \vdots \\
a_{m(m+1)} & \cdots & a_{mn}
\end{pmatrix} \\
0 & I_{n-m}
\end{pmatrix}
\]

where \( a_{ij} \in [-K,0] \) for all \( 1 \leq i \leq m \) and \( m+1 \leq j \leq n \). One may verify that

\[
e_i u = e_i + a_i(m+1) \ e_{m+1} + \cdots + a_in \ e_n
\]

for \( 1 \leq i \leq m \), and

\[
e_i u = e_i
\]

for \( m+1 \leq i \leq n \). Hence, when we take wedge products \( (e_i, u) \wedge \cdots \wedge (e_j, u) \), we cannot get a coefficient of order greater than \( K^m \), since only the first \( m \) transformed basis vectors have nontrivial coefficients and none of these coefficients have magnitude greater than \( K \). On the other hand, we cannot get a coefficient of order larger than \( K^{n-m} \), since only the basis vectors \( e_{m+1} \) through \( e_n \) carry nontrivial coefficients. Thus if we let \( \kappa := \min(m,n-m) \), we find that

\[
\left\| (e_i, u) \wedge \cdots \wedge (e_j, u) \right\| \ll K^\kappa.
\]

Then for general \( w \in \Lambda^j(\mathbb{R}^n) \) and \( u \in B_K \), we have

\[
\left\| wu \right\| \ll K^\kappa \left\| w \right\|.
\]

Thus from (59), we can say that for any \( w \in \Lambda^j(\mathbb{Z}^n) - \{0\}, j \in \{1, \ldots, n-1\} \), there exists \( t \in [0, T_i]^d \) such that

\[
\left\| w^g_0 u(\tilde{k}) u(t) \right\| \gg \left\| w^g_0 u(\tilde{k}) u(-\tilde{k}) \right\| \geq T_i^\delta,
\]

so

\[
\left\| w^g_0 u(\tilde{k}) u(t) \right\| \gg T_i^\delta/K^\kappa.
\]

That is, for any \( u(\tilde{k}) \in B_K \), the shifted basepoint \( x_0 u(\tilde{k}) \) satisfies a Diophantine condition of the form (5c) with new parameter proportional to \( (T_i^\delta/K^\kappa)^{1/q} \). From Corollary 14, this implies that for \( T_i \) large enough (i.e., for \( i \) large enough), we have equidistribution with

\[
(61) \quad \sum_{k \in B_{T_i}} f(x_0 u(k)) = \frac{T_i^d}{K^d} \int_X f dm_X + O(T_i^d (T_i^\delta/K^\kappa)^{-nb/d(d+1)} K^{-d^2/(d+1)})
\]

for any \( k \in B_K \). Using this in (58), we find that

\[
S_k(A) = g(K) \mathcal{X} + r(f, K, T_i)
\]

where

\[
|r(f, K, T_i)| \ll G_d(K) T_i^{d-\delta nb/d(d+1)} K^{(Xnb-d^2)/d(d+1)} \mathcal{X}_\infty(f).
\]
Since we have already shown that the function \( g(K) = G_d(K)/K^d \) satisfies sieve axioms (i) and (iii) with sieve dimension \( d \), it remains to verify sieve axiom (ii).

As before, we know that for any \( \epsilon > 0 \),
\[
\sum_{K < D} |r(f, K, T_i)| \ll f T_i^{-d-\delta nb/d(d+1)} \sum_{K < D} G_d(K)/K^{(d-\kappa nb)/d(d+1)} \ll f \epsilon T_i^{-d-\delta nb/d(d+1)} D^{(d+\kappa nb)/d(d+1)+\epsilon}.
\]

Now let \( D = T_i^{\eta} \) for \( \eta < \delta bn/(d^2 + \kappa bn) \), say, e.g.,
\[
\eta = \delta bn/(d^2 + \kappa bn) - 2\epsilon d^2(d+1)/(d^2 + \kappa bn)
\]
for \( \epsilon > 0 \), and set \( \epsilon_1 = (d^2 + \kappa bn)\epsilon/\delta bn \). Then
\[
\sum_{K < D} |r(f, K, T_i)| \ll f \epsilon T_i^{d(1-\epsilon)} \ll f \mathcal{X}^{1-\epsilon}.
\]

Thus for given \( f \), the conclusion of Theorem 18 holds for all \( i \) large enough, which gives us Theorem 20.

As before, if we consider positive \( f \) supported on a neighborhood of \( X \), the above theorem tells us that we may take \( i \) large enough so that we have a positive lower bound on averages over almost-prime points with fewer than \( 1/\alpha > 9d(d^2 + \kappa nb)/\delta bn \) prime factors, hence such points are dense in \( X \). This gives us the theorem for the space of lattices from the introduction.

**Theorem 3.** Let \( u(t) \) be an abelian horospherical flow of dimension \( d \) on \( X = \text{SL}_n(\mathbb{Z}) \backslash \text{SL}_n(\mathbb{R}) \) and let \( x_0 \in X \) be strongly polynomially \( \delta \)-Diophantine for some \( \delta > 0 \). Then there exists a constant \( M_\delta \) (depending\(^{12} \) on \( \delta \), \( n \), and \( d \)) such that
\[
\{x_0 u(k_1, k_2, \ldots, k_d) \mid k_i \in \mathbb{Z} \text{ has fewer than } M_\delta \text{ prime factors}\}
\]
is dense in \( X \).

**Remark.** Explicitly, we have
\[
M_\delta > \frac{9d}{\delta n \beta_1 \beta_2} \cdot \frac{1}{d} \cdot \frac{n^2 (n+1)(n-1)^2 + 2}{d + 2} + 2 \cdot \frac{(2n(n-1)(d + \beta_2) + d \beta_2)}{d + \beta_2} + n \kappa \beta_1 \beta_2.
\]
For an abelian horospherical flow in \( \text{SL}_3(\mathbb{R}) \) (\( n = 3 \), \( d = 2 \), \( \beta_1 = 1/2 - \epsilon \), \( \beta_2 = 1 - \epsilon \), \( \kappa = 1 \)) we may take \( M_\delta = 220723 \delta^{-1} \). In general, we have \( M_\delta = \Theta(n^5 d^4 \delta^{-1} \beta_1^{-1} \beta_2^{-1}) \), so it is \( \Theta(n^9 \delta^{-1}) \) in the minimal case and \( \Theta(n^{12} \delta^{-1}) \) in the maximal case.\(^{13}\)

**Remark.** Unfortunately, we cannot easily generalize from abelian to arbitrary flows as we did in the compact setting. As before, we may write an arbitrary horospherical flow as
\[
u(t_1, \ldots, t_d) = \nu(t_{d+1}, \ldots, t_d) \tilde{u}(t_1, \ldots, t_d)
\]

\(^{12}\) Strictly speaking, \( M_\delta \) also depends on \( \kappa \), however \( \kappa \leq n/2 \). Additionally, the constants \( \beta_1 \) and \( \beta_2 \) depend solely on \( n \) (for the \( a_1 \) and \( u(t) \) we fixed at the beginning of Section 4) and are \( \Theta(1) \) in any case.

\(^{13}\) If we use the Remez inequality instead of Theorem 6, we have \( M_\delta = \Theta(n^5 d^4 \delta^{-1} \beta_1^{-1} \beta_2^{-1}) \), and for the example of an abelian flow in \( \text{SL}_3(\mathbb{R}) \), we may take \( M_\delta = 111283 \delta^{-1} \).
where \( v \) is unipotent, \( \tilde{u} \) is abelian horospherical, and \( \tilde{d} = m_1(n - m_1) \). However, it is not clear how the Diophantine condition for the flow \( u \) at the point \( x_0 \) relates to any sort of Diophantine condition for the flow \( \tilde{u} \) at the point \( x_0 v(k_{d+1}, \ldots, k_d) \) where \( \{k_{d+1}, \ldots, k_d\} \) are almost-prime. Note that we do not consider 0 to be almost-prime, so the set of points of the form \( x_0 \tilde{u}(k_1, \ldots, k_d) \) is not a subset of the larger flow. Moreover, the Diophantine condition depends on the flow under consideration, and it is possible to be Diophantine for a horospherical flow but not for a horospherical subset of that flow.

**Remark.** As in the compact setting, it is possible to extract an effective density statement from the proof of Theorem 20 of the form: For any strongly polynomially \( \delta \)-Diophantine \( x_0 \in X \), there exists a sequence of times \( T_i \to \infty \) as \( i \to \infty \) such that the subset of its orbit consisting of integer times of norm less than \( T_i \) with fewer than \( M_\delta \) prime factors is \( O(T_i^{-\delta}) \)-dense in \( L_{T_i} \) (recall from Section 2.6 that \( L_\epsilon \) is the compact subset of \( X \) consisting of lattices with no nonzero vectors of norm less than \( \epsilon \)).

6. Conclusion

In this paper we gave an effective equidistribution result for horospherical flows on the space of lattices and an effective rate of equidistribution for arithmetic sequences of entries in abelian horospherical flows on both the space of lattices and compact quotients of \( \text{SL}_n(\mathbb{R}) \). We then use sieve methods to derive an upper and lower bound for averages over almost-prime entries in abelian horospherical flows. In the compact setting, we have as a result the density of almost-prime times in arbitrary horospherical orbits, where the number of prime factors depends only on the dynamical system and not on the basepoint. In the space of lattices, we consider the orbits of points satisfying a Diophantine condition with parameter \( \delta \) and we prove the density of almost-prime times where the number of prime factors depends on the system and on \( \delta \).

There are several improvements and generalizations of this work that can be readily imagined, some of which are currently underway. It seems likely that the methods used here can be generalized to quotients of connected, semisimple Lie groups by lattices. It also seems possible that methods similar to those used in [61] could be adapted to remove dependence on the basepoint in the non-compact case, yielding a uniform result for the density of almost-primes in the orbits of any generic point. For the case of \( \Gamma \) non-uniform, one may generalize from abelian horospherical flows to arbitrary horospherical flows using different methods than in the compact setting, which is the topic of a forthcoming paper. Additionally, the sieve methods used Section 5 can be modified to learn about averages over points \( (k_1, \ldots, k_d) \in \mathbb{R}^d \) satisfying \( \gcd(P(k_1, \ldots, k_d), P) = 1 \), where \( P \) is a suitably nice irreducible polynomial (note that we considered the case where \( P(k_1, \ldots, k_d) = k_1 k_2 \cdots k_d \)). Finally, there are many places throughout the paper where given estimates may be improved upon, possibly leading to sharper effective results and a smaller number of prime factors.
Of course, the more natural question is not what happens at almost-prime times, but what happens at prime times. Unfortunately, it does not seem possible at present to use these methods to establish results about true primes, and additional ingredients or a wholly different approach may be required. However, this result is significant in that it continues to lend support to the conjecture, already suggested by [61], that prime times in horospherical orbits are dense and possibly equidistributed.

**Appendix A. Properties of the function $G_d$**

Recall that we defined the generalized Pillai’s function $G_d : \mathbb{N} \to \mathbb{N}$ by

$$G_d(K) := \# \left\{ k \in \tilde{B}_K \mid k_1 \cdots k_d \equiv 0 \mod K \right\}.$$  

We want to prove the following properties of this function.

**Lemma 21.** For $d \geq 1$, the following hold:

1. $G_d$ is multiplicative.
2. (Behavior at primes) Let $p$ be a prime. Then
   $$G_d(p) = p^d - (p - 1)^d.$$  
3. (Squarefree bound) For $K$ squarefree,
   $$G_d(K) < d^{\omega(K)}K^{d-1}.$$  
4. (Iterated sum formula)
   $$G_d(K) = \sum_{k_{d-1}=1}^{K} \cdots \sum_{k_1=1}^{K} \gcd(K, k_1 \cdots k_{d-1}).$$  
5. (Recursive formula) Let $\text{Id}^d(K) = K^d$. Then
   $$G_{d+1} = \text{Id}^d \cdot (\phi \cdot G_d).$$  
6. (Dirichlet series bound)\(^\dagger\) For real $x > 1$ and $s < d$,
   $$\sum_{K \leq x} \frac{G_d(K)}{K^s} \ll_{s,d} x^{d-s} (\log x)^{d-1}.$$  

To do this, let us first recall a few basic facts from number theory. For any function $f : \mathbb{N} \to \mathbb{R}$, we have

\[(63) \sum_{i=1}^{K} f(\gcd(K, i)) = \sum_{j | K} f(j) \phi(K/j)\]

where $\phi$ is Euler’s totient function, i.e., $\phi(n)$ is the number of positive integers less than $n$ that are relatively prime with $n$. This formula dates back to the work of Cesàro and is sometimes referred to as Cesàro’s formula (cf. [12] or [22]).

\(^\dagger\) This property can be used as an alternative way to verify sieve axiom (ii) in the proofs of Theorems 19 and 20.
We now have everything we need to complete the proof.

(v) We will proceed by induction on $d$. Hence,

$$G_d(K) = \sum_{k_{d-1}=1}^{K} \cdots \sum_{k_1=1}^{K} \gcd(K, k_1 \cdots k_{d-1}).$$

So, for example, (63) says that $\sum_{i=1}^{K} f(\gcd(K, i)) = (f * \phi)(K)$. Recall that the convolution of two multiplicative arithmetic functions is again multiplicative. We now have everything we need to complete the proof.

Proof. Properties (i)-(iii) were proved as part of the proof of Theorem 19.

(iv) Notice that to specify a point $k \in \mathbb{B}_K$ such that $k_1 \cdots k_d \equiv 0 \mod K$, we can choose $k_1$ through $k_{d-1}$ independently to be any integers between 1 and $K$, but then the remaining coordinate $k_d$ must be a multiple of $K/\gcd(K, k_1 \cdots k_{d-1})$, that is, the last coordinate must contain all primes in $K$ not contained in any of the previous coordinates. Since there are $\gcd(K, k_1 \cdots k_{d-1})$ multiples of $K/\gcd(K, k_1 \cdots k_{d-1})$ less than or equal to $K$, the total number of points counted in this way is given by

$$G_d(K) = \sum_{k_{d-1}=1}^{K} \cdots \sum_{k_1=1}^{K} \gcd(K, k_1 \cdots k_{d-1}).$$

(v) We will proceed by induction on $d$. For the base case, we have $G_2(K) = \text{Id} * \phi = \text{Id} * (\phi \cdot 1) = \text{Id} * (\phi \cdot G_1)$, which is a well-known formula for Pillai's arithmetical function. Then suppose $G_d = \text{Id}^{d-1} * (\phi \cdot G_{d-1})$ for $d \geq 2$ and consider $G_{d+1}$.

Notice that for any integers $k$, $n$, and $m$, we can write $\gcd(k, nm) = \gcd(k, n \gcd(k, m))$, that is, we can throw out all the primes in $m$ that are not in $k$. Furthermore, since $\gcd(k, m)|k$, we can write

$$\gcd(k, n \gcd(k, m)) = \gcd(k, m) \gcd(k/\gcd(k, m), n).$$

Hence, $G_{d+1}$ may be written

$$G_{d+1}(K) = \sum_{k_{d+1}=1}^{K} \cdots \sum_{k_1=1}^{K} \gcd(K, k_1 \cdots k_{d+1})$$

$$= \sum_{k_{d+1}=1}^{K} \cdots \sum_{k_1=1}^{K} \gcd(K, k_{d+1}) \gcd(K/\gcd(K, k_{d+1}), k_1 \cdots k_{d-1})$$

$$= \sum_{k_{d+1}=1}^{K} \gcd(K, k_{d+1}) \left( \sum_{k_{d-1}=1}^{K} \cdots \sum_{k_1=1}^{K} \gcd(K/\gcd(K, k_{d+1}), k_1 \cdots k_{d-1}) \right)$$

But now notice that the function $\gcd(K/\gcd(K, k_{d+1}), k_1 \cdots k_{d-1})$ is periodic with period $K/\gcd(K, k_{d+1})$ in each coordinate $k_i$ for $i = 1, \ldots, d - 1$. Thus

$$\sum_{k_1=1}^{K} \gcd(K/\gcd(K, k_{d+1}), k_1 \cdots k_{d-1})$$

$$= \gcd(K, k_{d+1}) \sum_{k_1=1}^{K/\gcd(K, k_{d+1})} \gcd(K/\gcd(K, k_{d+1}), k_1 \cdots k_{d-1})$$
for \( i = 1, \ldots, d - 1 \). Therefore,
\[
G_{d+1}(K) = \sum_{k_d=1}^{K} \gcd(K, k_d)^d \left( \sum_{k_{d-1}=1}^{K/\gcd(K, k_d)} \cdots \sum_{k_1=1}^{K/\gcd(K, k_d)} \gcd(K/\gcd(K, k_d), k_1 \cdots k_{d-1}) \right)
\]
\[
= \sum_{k_d=1}^{K} \gcd(K, k_d)^d G_d(K/\gcd(K, k_d)).
\]

But by Cesàro's formula, this is simply
\[
G_{d+1}(K) = \sum_{j|K} j^d \phi(K/j) G_d(K/j)
\]

Finally, we can express this in terms of Dirichlet convolution as
\[
G_{d+1}(K) = (\text{Id}^d * (\phi \cdot G_d))(K)
\]
which completes our proof by induction.

(vi) Once again, we proceed by induction on \( d \). Observe that for \( d = 1 \), we have
\[
\sum_{K \leq x} \frac{G_1(K)}{K^s} = \sum_{K \leq x} \frac{1}{K^s}.
\]
When \( 0 \leq s < 1 \), we have that \( 1/K^s \) is decreasing, and \( \sum_{K \leq x} 1/K^s \leq 1 + \int_1^x 1/t^s \, dt \). On the other hand, when \( s < 0 \), we have that \( 1/K^s \) is increasing, and \( \sum_{K \leq x} 1/K^s \leq \int_1^{x+1} 1/t^s \, dt \). In either case, we have
\[
\sum_{K \leq x} \frac{G_1(K)}{K^s} \ll_{s} x^{1-s},
\]
which is the desired bound for \( d = 1 \).

Now suppose that for \( d \geq 1 \) we have \( \sum_{K \leq x} G_d(K)/K^s \ll_{s,d} x^{d-s} (\log x)^{d-1} \) for all \( x > e \) and \( s < d \). By the complete multiplicativity of \( \text{Id}^{-s} \) and the recursive formula for \( G_d \), we can write
\[
G_{d+1}(K)/K^s = (\text{Id}^{d-s} * (\text{Id}^{-s} \cdot \phi \cdot G_d))(K).
\]
Also note that a Dirichlet product \((f * g)(K) = \sum_{j|K} f(j) g(K/j)\) can be seen as a sum over pairs of positive integers \((n, m)\) whose product is \( K \), i.e.,
\[
(f * g)(K) = \sum_{n,m} f(n) g(m).
\]
Hence, the sum
\[
\sum_{K \leq x} (f * g)(K) = \sum_{n,m} f(n) g(m) = \sum_{n \leq x} f(n) \sum_{m \leq x/n} g(m)
\]
is a sum over pairs of integers whose product is no greater than \( x \). Also notice that \( \phi(n) < n \) for any positive integer \( n \). Thus for any \( s < d + 1 \) and \( x > e \), we
may write

\[
\sum_{K \leq x} \frac{G_{d+1}(K)}{K^s} = \sum_{K \leq x} \frac{1}{n^{s-d}} \sum_{m \leq x/n} G_d(m) \frac{1}{m^s} \\
= \sum_{n \leq x} \frac{1}{n^{s-d}} \sum_{m \leq x/n} \phi(m) G_d(m) \frac{1}{m^s} \\
< \sum_{n \leq x} \frac{1}{n^{s-d}} \sum_{m \leq x/n} G_d(m) \frac{1}{m^{s-1}}.
\]

Then since \( s < d + 1 \), we have \( s - 1 < d \). Also, notice that for \( n < x/e \), we have \( x/n > e \), so the induction hypothesis applies to sums over \( m \leq x/n \) for \( n \) in this region. On the other hand, for \( n \geq x/e \), we have \( x/n \leq e \), so a sum over \( m \leq x/n \) is only a sum over the first two terms, \( m = 1 \) and \( m = 2 \), and can thus be bounded by a constant (depending on \( s \) and \( d \)). Hence, we may write

\[
\sum_{K \leq x} \frac{G_{d+1}(K)}{K^s} < \sum_{n \leq x} \frac{1}{n^{s-d}} \sum_{m \leq x/n} G_d(m) \frac{1}{m^{s-1}} + \sum_{x/e \leq n \leq x} \frac{1}{n^{s-d}} \sum_{m \leq x/n} G_d(m) \frac{1}{m^{s-1}} + \sum_{x/e \leq n \leq x} \frac{1}{n^{s-d}}.
\]

Observe that \( \sum_{x/e \leq n \leq x} \frac{1}{n^{s-d}} \ll x^{d-1} \) (this can be seen with a calculation similar to that of the base case). On the other hand, the function \( \log(x/t)^{d-1}/t \) is positive and decreasing in the region \((1, x/e)\), so we may bound the sum by the first term plus the corresponding integral:

\[
\sum_{n < x/e} \frac{\log(x/n)^{d-1}}{n} \leq (\log x)^{d-1} + \int_1^{x/e} \frac{\log(x/t)^{d-1}}{t} dt.
\]

With the substitution \( u = \log(x/t) \), we find that

\[
\int_1^{x/e} \frac{\log(x/t)^{d-1}}{t} dt = \int_1^{\log x} u^{d-1} du = \frac{(\log x)^d - 1}{d}.
\]

In total, we have that

\[
\sum_{K \leq x} \frac{G_{d+1}(K)}{K^s} \ll x^{d+1-s} \left( 1 + (\log x)^{d-1} + (\log x)^d \right)
\ll x^{d+1-s} (\log x)^d
\]

since \( x > e \), and this completes the proof.
Acknowledgments. I would like to thank my advisor, Amir Mohammadi, for giving me the idea to work on this problem and for providing important guidance throughout the process. I also want to thank Manfred Einsiedler, Hee Oh, and Dubi Kelmer for helpful discussions, as well as the anonymous referees whose valuable suggestions improved the paper significantly. This work was partially completed with the support of funds from the NSF RTG Awards DMS-1502651 and DMS-1148490.

REFERENCES

[1] A018804, On-line encyclopedia of integer sequences, https://oeis.org/A018804 (accessed Feb. 2, 2018).
[2] M. B. Bekka and M. Mayer, Ergodic Theory and Topological Dynamics of Group Actions on Homogeneous Spaces, Cambridge University Press, 2000.
[3] Y. Benoist and H. Oh, Effective equidistribution of \( S \)-integral points on symmetric varieties, Ann. Inst. Fourier (Grenoble), 62 (2012), 1889–1942.
[4] O. Bordellès, Mean values of generalized gcd-sum and lcm-sum functions, J. Integer Seq., 10 (2007), Article 07.9.2, 13pp.
[5] O. Bordellès, A note on the average order of the gcd-sum function, J. Integer Seq., 10 (2007), Article 07.3.3, 4pp.
[6] J. Bourgain, An approach to pointwise ergodic theorems, in Geometric Aspects of Functional Analysis (Israel Math. Seminar), Lecture Notes in Math., 1318, Springer Berlin, 1988, 204–223.
[7] J. Bourgain, P. Sarnak and T. Ziegler, Disjointness of Möbius from horocycle flows, Developments in Mathematics, 28 (2013), 67–83.
[8] K. A. Broughan, The gcd-sum function, J. Integer Seq., 4 (2001), Article 01.2.2, 19pp.
[9] K. A. Broughan, The average order of the Dirichlet series of the gcd-sum function, J. Integer Seq., 10 (2007), Article 07.4.2, 6pp.
[10] J. A. Brudnyi and M. I. Ganzburg, On an extremal problem for polynomials in \( n \) variables, Math. USSR Izv., 7 (1973), 344–355.
[11] M. Burger, Horocycle flow on geometrically finite surfaces, Duke Math. J., 61 (1990), 779–803.
[12] E. Cesáro, Étude moyenne di plus grand commun divisior de deux nombres, Annali di Matematica Pura ed Applicata (1867-1897), 13 (1885), 235–250.
[13] J. Chidambaramswamy and R. Sitaramachandra Rao, Asymptotic results for a class of arithmetical functions, Monatsh. Math., 99 (1985), 19–27.
[14] K. Dabbs, M. Kelly and H. Li, Effective equidistribution of translates of maximal horospherical measures in the space of lattices, J. Mod. Dyn., 10 (2016), 229–254.
[15] S. G. Dani, Invariant measures of horospherical flows on noncompact homogeneous spaces, Invent. Math., 47 (1978), 101–138.
[16] S. G. Dani, Invariant measures and minimal sets of horospherical flows, Invent. Math., 64 (1981), 357–385.
[17] S. G. Dani, On orbits of unipotent flows on homogeneous spaces, Ergodic Theory Dynam. Systems, 4 (1984), 25–34.
[18] S. G. Dani, On orbits of unipotent flows on homogeneous spaces II, Ergodic Theory Dynam. Systems, 6 (1986), 167–182.
[19] S. G. Dani, Orbits of horospherical flows, Duke Math. J., 53 (1986), 177–188.
[20] S. G. Dani and G. A. Margulis, Asymptotic behaviour of trajectories of unipotent flows on homogeneous spaces, Proc. Indian Acad. Sci. Math. Sci., 101 (1991), 1–17.
[21] H. Davenport, Multiplicative Number Theory, Graduate Texts in Mathematics, 74, Springer-Verlag, 1980.
[22] L. E. Dickson, History of the Theory of Numbers. Vol. I: Divisibility and Primality, Chelsea Publishing Co., 1966.
[23] M. Einsiedler, G. A. Margulis, A. Mohammadi and A. Venkatesh, Effective equidistribution and property tau, *Effective Equidistribution and Property (τ)*, (2019), 1–77.

[24] M. Einsiedler, G. A. Margulis and A. Venkatesh, Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces, *Invent. Math.*, 177 (2009), 137–212.

[25] M. Einsiedler and T. Ward, *Ergodic Theory with a View Towards Number Theory*, Springer-Verlag London, 2011.

[26] R. Ellis and W. Perrizo, Unique ergodicity of flows on homogeneous spaces, *Israel J. Math.*, 29 (1978), 276–284.

[27] L. Flaminio and G. Forni, Invariant distributions and time averages for horocycle flows, *Duke Math. J.*, 119 (2003), 465–540.

[28] H. Furstenberg, The unique ergodicity of the horocycle flow, in *Recent Advances in Topological Dynamics (Proc. Conf., Yale Univ., New Haven, CT, 1972)*, Lecture Notes in Math., 318, Springer, Berlin, 1973, 95–115.

[29] G. A. Hedlund, Fuchsian groups and transitive horocycles, *Duke Math. J.*, 2 (1936), 530–542.

[30] A. Katok and R. J. Spatzier, First cohomology of Anosov actions of higher rank abelian groups and applications to rigidity, *Publications mathématiques de l'IHÉS*, 79 (1994), 131–156.

[31] D. Y. Kleinbock and G. A. Margulis, Bounded orbits of nonquasiunipotent flows on homogeneous spaces, in *Sinaï’s Moscow Seminar on Dynamical Systems*, Amer. Math. Soc. Transl. Ser. 2, 171, Adv. Math. Soc., Providence, RI, 1996, 141–172.

[32] D. Y. Kleinbock and G. A. Margulis, Logarithm laws for flows on homogeneous spaces, *Invent. Math.*, 138 (1999), 451–494.

[33] G. A. Margulis, On effective equidistribution of expanding translates of certain orbits in the space of lattices, in *Number Theory, Analysis and Geometry*, Springer, Boston, 2012, 385–396.

[34] A. W. Knapp, *Lie Groups Beyond an Introduction*, Progress in Mathematics, 140, Birkhäuser Boston, Inc., 1996.

[35] M. Lee and H. Oh, Effective equidistribution of closed horocycles for geometrically finite surfaces, preprint, arXiv:1202.0848, (2012).

[36] G. A. Margulis, On the action of unipotent groups in the space of lattices, in *Lie Groups and Their Representations (Proc. Summer School, Bolyai, János Math. Soc., Budapest, 1971)*, Wiley, New York, 1975, 365–370.
[48] G. A. Margulis, Formes quadratiques indéfinies et flots unipotents sur les espaces homogènes, *C. R. Acad. Sci. Paris Sér. I Math.*, 304 (1987), 249–253.

[49] G. A. Margulis, Discrete subgroups and ergodic theory, in *Number Theory, Trace Formulas and Discrete Groups* (Symposium in honor of Atle Selberg, Oslo, 1987), Academic Press, Boston, 1989, 377–398.

[50] G. A. Margulis and G. M. Tomanov, Invariant measures for actions of unipotent groups over local fields on homogeneous spaces, *Invent. Math.*, 116 (1994), 347–392.

[51] A. Nevo and P. Sarnak, Prime and almost prime integral points on principal homogeneous spaces, *Acta Math.*, 205 (2010), 361–402.

[52] H. Oh, Tempered subgroups and representations with minimal decay of matrix coefficients, *Bull. Math. Soc. France*, 126 (1998), 355–380.

[53] H. Oh, Uniform pointwise bounds for decay of matrix coefficients of unitary representations and applications to Kazhdan constants, *Duke Math. J.*, 113 (2002), 133–192.

[54] R. Peckner, Möbius disjointness for homogeneous dynamics, *Duke Math. J.*, 167 (2018), 2745–2792.

[55] S. S. Pillai, On an arithmetic function, *J. of the Annamalai Univ.*, 2 (1933), 243–248.

[56] M. S. Raghunathan, *Discrete Subgroups of Lie Groups*, Springer-Verlag, New York-Heidelberg, 1972.

[57] M. Ratner, On Raghunathan’s measure conjecture, *Ann. of Math. (2)*, 134 (1991), 545–607.

[58] M. Ratner, Invariant measures and orbit closures for unipotent actions on homogeneous spaces, *Geom. Funct. Anal.*, 4 (1994), 236–257.

[59] E. J. Remez, Sur une propriété des polynômes de Tchebyscheff, *Comm. Inst. Sci. Kharkow*, 13 (1936), 93–95.

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

[61] P. Sarnak and A. Ubis, The horocycle flow at prime times, *J. Math. Pures Appl. (9)*, 103 (2015), 575–618.

[62] N. A. Shah, Limit distributions of polynomial trajectories on homogeneous spaces, *Duke Math. J.*, 75 (1994), 711–732.

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

[64] A. Strömbergsson, On the deviation of ergodic averages for horocycle flows, *J. Mod. Dyn.*, 7 (2013), 291–328.

[65] A. Strömbergsson, An effective Ratner equidistribution result for $SL(2,\mathbb{R}) \ltimes \mathbb{R}^2$, *Duke Math. J.*, 164 (2015), 843–902.

[66] Y. Tanigawa and W. Zhai, On the gcd-sum function, *J. Integer Seq.*, 11 (2008), Article 08.2.3, 11pp.

[67] L. Tóth, A generalization of Pillai’s arithmetical function involving regular convolutions, Proceedings of the 13th Czech and Slovak International Conference on Number Theory (Ostravice, 1997), *Acta Math. Inform. Univ. Ostraviensis*, 6 (1998), 203–217.

[68] L. Tóth, A survey of gcd-sum functions, *J. Integer Seq.*, 13 (2010), Article 10.8.1, 23pp.

[69] L. Tóth, Another generalization of the gcd-sum function, *Arab. J. Math. (Springer)*, 2 (2013), 313–320.

[70] W. A. Veech, Unique ergodicity of horospherical flows, *Amer. J. Math.*, 99 (1977), 827–859.

[71] A. Venkatesh, Sparse equidistribution problems, period bounds and subconvexity, *Ann. of Math. (2)*, 172 (2010), 989–1094.

Taylor Mcdam <taylor.mcadam@yale.edu>: Department of Mathematics, Yale University, 10 Hillhouse Ave., New Haven, CT 06520-8283, USA