POLYNOMIAL EFFECTIVE EQUIDISTRIBUTION

E. LINDENSTRAUSS, A. MOHAMMADI, AND Z. WANG

Abstract. We prove effective equidistribution theorems, with polynomial error rate, for orbits of the unipotent subgroups of $\text{SL}_2(\mathbb{R})$ in arithmetic quotients of $\text{SL}_2(\mathbb{C})$ and $\text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R})$.

The proof is based on the use of a Margulis function, tools from incidence geometry, and the spectral gap of the ambient space.

Résumé. Nous prouvons des théorèmes d’équidistribution effectifs, avec un taux d’erreur polynomial pour les orbites des sous-groupes unipotents de $\text{SL}_2(\mathbb{R})$ en quotients arithmétiques de $\text{SL}_2(\mathbb{C})$ et $\text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R})$.

La preuve est basée sur l’utilisation d’une fonction de Margulis, des outils de la géométrie d’incidence, et le trou spectral de l’espace ambiant.

1. Introduction

A landmark result of Ratner [Rat91b] states that if $G$ is a Lie group, $\Gamma$ a lattice in $G$ and if $u_t$ is a one-parameter Ad-unipotent subgroup of $G$, then for any $x \in G/\Gamma$ the orbit $u_t.x$ is equidistributed in a periodic orbit of some subgroup $L < G$ that contains both the one parameter group $u_t$ and the initial point $x$. We say an orbit $L.x$ of a group $L$ in some space $X$ is periodic if the stabilizer of $x$ in $L$ is a lattice in $L$, equivalently that the stabilizer of $x$ in $L$ is discrete and $L.x$ supports a unique $L$-invariant probability measure $m_{L,x}$; and $u_t.x$ is equidistributed in $L.x$ in the sense that

\begin{equation}
\frac{1}{T} \int_0^T f(u_t.x) dt \to \int f dm_{L,x} \quad \text{for any } f \in C_0(G/\Gamma).
\end{equation}

In order to prove this equidistribution result, Ratner first classified the $u_t$-invariant probability measures on $G/\Gamma$ [Rat90, Rat91a]; the proof also uses the non-divergence properties of unipotent flows established by Dani and Margulis [Mar71, Dan84, Dan86].

A key motivation behind Ratner’s equidistribution theorem for one parameter unipotent flows has been to establish Raghunathan’s conjecture regarding the possible orbit closures of groups generated by one-parameter unipotent groups; using the equidistribution theorem Ratner proved that if $G$ and $\Gamma$ are as above, and if $H < G$ is generated by one parameter unipotent groups, then for any $x \in G/\Gamma$ one has that $\overline{H.x} = L.x$ where $H \leq L \leq G$.

E.L. acknowledges support by ERC 2020 grant HomDyn (grant no. 833423).
A.M. acknowledges support by the NSF, grants DMS-1764246 and 2055122.
Z.W. acknowledges support by the NSF, grant DMS-1753042.
and \( L \cdot x \) is periodic. Important special cases of Raghunathan’s conjecture were proven earlier by Margulis and by Dani and Margulis using a different more direct approach, which in particular gave a proof of a rather strong form of the longstanding Oppenheim conjecture [Mar89, DM89, DM90].

These results have had a surprisingly rich trove of applications in number theory and beyond. One drawback of this method, compared to more traditional number theoretic techniques, is that these equidistribution results were neither effective nor quantitative. Indeed, Ratner’s proof relies heavily on the pointwise ergodic theorem and Lusin’s theorem, both rather fundamental theorems but that do not have good effective analogues. The Dani-Margulis method is somewhat easier to make effective and a result in this direction was given by Margulis and the first named author in [LM14]; moreover a general result in this direction was announced by Margulis, Shah and two of us (E.L. and A.M.) with the first installment of this work appearing in [LMMS19]. However, this only gives density properties of these flows, not equidistribution, and the rate of density obtained in this way is far from optimal (polylog at best, though in most cases one only has an iterative-log type bound).

In this paper we announce a quantitative equidistribution result for orbits of a one parameter unipotent group on quotients \( G/\Gamma \) where \( G \) is either \( \text{SL}_2(\mathbb{C}) \) or \( \text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R}) \) with a polynomial error rate, which is the first quantitative equidistribution statement for individual orbits of unipotent flows on quotients of semi-simple groups beyond the horospherical case. Our approach builds on the paper [LM21] by the first two authors, where an effective density result with a polynomial rate for orbits of a Borel subgroup of a subgroup \( H \cong \text{SL}_2(\mathbb{R}) \) of \( G \) was proved.

Recall that a group \( N < G \) is horospherical if there is some \( g \in G \) so that
\[
N = \{ h \in G : g^{-n}hg^n \to 1 \text{ as } n \to \infty \}.
\]
For instance, the one parameter unipotent group
\[
\left\{ \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} : r \in \mathbb{R} \right\}
\]
is horospherical in \( \text{SL}_2(\mathbb{R}) \) as are the groups
\[
\left\{ \begin{pmatrix} 1 & r + is \\ 0 & 1 \end{pmatrix} : r, s \in \mathbb{R} \right\} \quad \text{and} \quad \left\{ \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} , \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : r, s \in \mathbb{R} \right\}
\]
in \( \text{SL}_2(\mathbb{C}) \) and \( \text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R}) \), respectively. The classification of invariant measures and orbit closures for horospherical flows was established prior to Ratner’s work by Hedlund, Furstneberg, Dani, Veech and others, and this has been understood for some time also quantitatively since one can relate the distribution properties of individual \( N \) orbits to the ergodic theoretic properties of the action of \( g \) on \( G/\Gamma \) (cf. [7] for more details).

We also mention that while our result is the first quantitative equidistribution statement for individual orbits of unipotent flows on quotients
of semi-simple groups beyond the horospherical case there have been some other interesting quantitative equidistribution results. When $G$ is unipotent, an effective equidistribution result for unipotent flows on $G/\Gamma$ was given by Green and Tao in [GT12], and was a key ingredient in a series of works by Green, Tao and Ziegler about linear equations in primes. In the case of quotients of the skew product $G = \text{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$, Strombergsson [Str15] has an effective equidistribution result for one parameter unipotent orbits (which are not horospheric in $G$, but project to a horospheric group on $\text{SL}_2(\mathbb{R})$), and this has been generalized by several authors, in particular by Wooyeon Kim [Kim21] (using a completely different argument) to $\text{SL}_n(\mathbb{R}) \ltimes \mathbb{R}^n$. Moreover there is an important work of Einsiedler, Margulis and Venkatesh [EMV09] showing that periodic orbits of semisimple subgroups $H$ of a semisimple group $G$ are quantitatively equidistributed in an appropriate homogeneous subspace of $G/\Gamma$ if $\Gamma$ is a congruence lattice and $H$ has finite centralizer in $G$. Subsequently Einsiedler, Margulis, Venkatesh and the second named author by using Prasad’s volume formula and a more adelic viewpoint were able to prove such an equidistribution result for periodic orbits of maximal semisimple subgroups of $G$ when the subgroup is allowed to vary [EMMV20] with arithmetic applications (the equidistribution of periodic orbits of semisimple groups is also closely connected to the equidistribution of Hecke points; a quantitative treatment of such equidistribution was given by Clozel, Oh and Ullmo in [COU01]).

In a slightly different direction, Bourgain, Furman, Mozes and the first named author [BFLM11] gave a quantitative equidistribution result with exponential rates (this is analogous to polynomial rates in our problem) for random walks by automorphisms of the torus. In this case this equidistribution result was new even without rates. There have been several extensions of this result, in particular [HdS19] where a proximality assumption was removed. Kim in [Kim21] used the techniques of [BFLM11] to study $\text{SL}_n(\mathbb{R}) \ltimes \mathbb{R}^n$. Our work is also heavily influenced by [BFLM11].

We now proceed to describe our results and some of the ingredients of the proofs. Let $G = \text{SL}_2(\mathbb{C})$ or $G = \text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R})$. Let $\Gamma \subset G$ be a lattice, and put $X = G/\Gamma$. We let $m_X$ denote the $G$-invariant probability measure on $X$. Throughout the paper, we will denote by $H$ a subgroup of $G$ isomorphic to $\text{SL}_2(\mathbb{R})$, namely

$$\text{SL}_2(\mathbb{R}) \subset \text{SL}_2(\mathbb{C}) \quad \text{or} \quad \{(g, g) : g \in \text{SL}_2(\mathbb{R})\} \subset \text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R}).$$

For all $t, r \in \mathbb{R}$, let $a_t$ and $u_r$ denote the image of

$$
\begin{pmatrix}
  e^{t/2} & o \\
  0 & e^{-t/2}
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
  1 & r \\
  0 & 1
\end{pmatrix}.
$$

\footnote{As pointed out by Venkatesh in [Ven10], the distinction between horospheric and non-horospheric is not completely clear cut, and indeed the results of that paper can also be recast as a nonhorospheric equidistribution problem; cf. [FFT10] as well as [Kat19].}
in $H$, respectively.

We fix maximal compact subgroups $SU(2) \subset SL_2(\mathbb{C})$ and $SO(2) \times SO(2) \subset SL_2(\mathbb{R}) \times SL_2(\mathbb{R})$. Let $d$ be the right invariant metric on $G$ which is defined using the Killing form and the aforementioned maximal compact subgroups. This metric induces a metric $d_X$ on $X$, and natural volume forms on $X$ and its submanifolds. We define the injectivity radius of a point $x \in X$ using this metric. In the sequel, $\| \|$ denotes the maximum norm on $\text{Mat}_2(\mathbb{C})$ or $\text{Mat}_2(\mathbb{R}) \times \text{Mat}_2(\mathbb{R})$ with respect to the standard basis.

The following is the main results in this paper.

1.1. Theorem. Assume $\Gamma$ is an arithmetic lattice. For every $0 < \delta < 1$, every $x_0 \in X$ and large enough $T$ (depending explicitly on $\delta$ and the injectivity radius of $x_0$) at least one of the following holds.

(1) For every $\varphi \in C_0^\infty(X)$, we have
\[
\left| \int_0^1 \varphi(a_{\log T} u_r x_0) \, dr - \int \varphi \, d\nu_X \right| \leq C_1 S(\varphi) T^{-\delta_1},
\]
where $S(\varphi)$ is a certain Sobolev norm.

(2) There exists $x \in X$ such that $Hx$ is periodic with $\text{vol}(Hx) \leq T^\delta$, and
\[
d_X(x, x_0) \leq C_1 T^{-1/A}.\]

The constants $A, \kappa_1, \text{and } C_1$ are positive, and depend on $X$ but not on $\delta$ and $x_0$.

Theorem 1.1 can be viewed as an effective version of [Sha96, Thm. 1.4]. Combining Theorem 1.1 and the Dani-Margulis linearization method [DM91] (cf. also Shah [Sha91]), that allows to control the amount of time a unipotent trajectory spends near invariant subvarieties of a homogeneous space, we also obtain an effective equidistribution theorem for long pieces of unipotent orbits (more precisely, we use a sharp form of the linearization method taken from [LMMS19]).

1.2. Theorem. Assume $\Gamma$ is an arithmetic lattice. For every $0 < \delta < 1$, every $x_0 \in X$ and large enough $T$ (depending explicitly on $\delta$ and the injectivity radius of $x_0$) at least one of the following holds.

(1) For every $\varphi \in C_0^\infty(X)$, we have
\[
\left| \frac{1}{T} \int_0^T \varphi(u_r x_0) \, dr - \int \varphi \, d\nu_X \right| \leq C_2 S(\varphi) T^{-\delta_2},
\]
where $S(\varphi)$ is a certain Sobolev norm.

(2) There exists $L \subset G$ which is a conjugate of $H$ and satisfies $\text{vol}(L \Gamma / \Gamma) \leq T^\delta$, and for every $r \in [0, T]$ there exists $g \in G$ with $\|g\| \leq T^\delta$ so that
\[
d_X(u_s x_0, g L \Gamma / \Gamma) \leq C_2 T^{-1/A'} \text{ for all } |s - r| \leq T^{1/A'}.
\]

(3) For every $r \in [0, T]$,
the injectivity radius at $a_{-(\log T)/A'} u_r x_0$ is $< T^{-(1-\delta)/A'}$.\]
The constants $A', \kappa_2,$ and $C_2$ are positive, and depend on $X$ but not on $\delta$ and $x_0$.

The assumption in Theorem 1.1 that $\Gamma$ is arithmetic may be relaxed. Let us say $\Gamma$ has algebraic entries if the following is satisfied: there is a number field $F$, a semisimple $F$-group $G$ of adjoint type, and a place $v$ of $F$ so that $F_v = \mathbb{R}$ and $G(F_v)$ and $G$ are locally isomorphic — in which case there is a surjective homomorphism from $G$ onto the connected component of the identity in $G(F_v)$ — and the image of $\Gamma$ in $G(F_v)$ (possibly after conjugation) is contained in $G(F)$. Every arithmetic lattice has algebraic entries, but there are lattices with algebraic entries that are not arithmetic.

Note that the condition that $\Gamma$ has algebraic entries is automatically satisfied if $\Gamma$ is an irreducible lattice in $\text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R})$ or if $G = \text{SL}(2, \mathbb{C})$. Indeed, by arithmeticity theorems of Selberg and Margulis, irreducible lattices in $\text{SL}_2(\mathbb{R}) \times \text{SL}_2(\mathbb{R})$ are arithmetic [Mar91, Ch. IX]. Moreover, by local rigidity, lattices in $\text{SL}_2(\mathbb{C})$ always have algebraic entries [GR70, Thm. 0.11] (see also [Sel60, Wei60, Wei64]).

1.3. Theorem. Assume $\Gamma$ is a lattice which has algebraic entries. For every $0 < \delta < 1$, every $x_0 \in X$ and large enough $T$ (depending explicitly on $\delta$ and the injectivity radius of $x_0$) at least one of the following holds.

1. For every $\varphi \in C^\infty_c(X)$, we have
   $$\left| \int_0^1 \varphi(a \log T u, x_0) \, dr - \int \varphi \, dm_X \right| \leq C_1 S(\varphi) T^{-\delta/2},$$
   where $S(\varphi)$ is a certain Sobolev norm.

2. There exists $x \in X$ with
   $$d_X(x, x_0) \leq C_1 T^{-1/A},$$
   satisfying the following: there are elements $\gamma_1$ and $\gamma_2$ in $\text{Stab}_H(x)$ with $\|\gamma_i\| \leq T^\delta$ for $i = 1, 2$ so that the group generated by $\{\gamma_1, \gamma_2\}$ is Zariski dense in $H$.

The constants $A, C_1$, and $C_2$ are positive, and depend on $X$ but not on $\delta$ and $x_0$.

The obstacle to effective equidistribution in Theorem 1.1 is much cleaner and simpler than in Theorem 1.2. This is not an artifact of the proof but a reflection of reality; a unipotent orbit may fail to equidistribute at the expected rate without it staying near a single period orbit of some subgroup $\{u_t\} < L < G$: one must allow a slow drift of the periodic orbit in the direction of the centralizer of $u_t$. Unlike the work of Shah in [Sha96], where (in particular) a non-effective version of Theorem 1.1 is proved relying on Ratner’s measure classification theorem for unipotent flows, our proof goes the other way, first establishing Theorem 1.1 and then reducing Theorem 1.2 from it using a linearization and non-divergence argument.

An extremely interesting analogue to unipotent flows on homogeneous spaces is given by the action of $\text{SL}_2(\mathbb{R})$ and its subgroups on strata of abelian
differentials. Let \( g \geq 1 \), and let \( \alpha = (\alpha_1, \ldots, \alpha_n) \) be a partition of \( 2g - 2 \). Let \( \mathcal{H}(\alpha) \) be the corresponding stratum of abelian differentials, i.e., the space of pairs \((M, \omega)\) where \( M \) is a compact Riemann surface with genus \( g \) and \( \omega \) is a holomorphic 1-form on \( M \) whose zeroes have multiplicities \( \alpha_1, \ldots, \alpha_n \). The form \( \omega \) defines a canonical flat metric on \( M \) with conical singularities and a natural area from. Let \( \mathcal{H}_1(\alpha) \) be the space of unit area surfaces in \( \mathcal{H}(\alpha) \). The space \( \mathcal{H}(\alpha) \) admits a natural action of \( \text{SL}_2(\mathbb{R}) \); this action preserves the unit area hyperboloid \( \mathcal{H}_1(\alpha) \).

A celebrated theorem of Eskin and Mirzakhani [EM18] shows that any \( P \)-invariant ergodic measure is \( \text{SL}_2(\mathbb{R}) \)-invariant and is supported on an affine invariant manifolds, where \( P \) denotes the group of upper triangular matrices in \( \text{SL}_2(\mathbb{R}) \). We shall refer to these measures as affine invariant measures. Moreover if we define, for any interval \( I \subset \mathbb{R} \) and \( x \in \mathcal{H}_1(\alpha) \), the probability measure \( \mu_I^x \) on \( \mathcal{H}_1(\alpha) \) by

\[
\mu_I^x = |I|^{-1} \int_I \delta_{u,x} \, ds,
\]

then Eskin, Mirzakhani and the second named author [EMM15] showed that for any \( x \in \mathcal{H}_1(\alpha) \) the limit

\[
\lim_{T \to \infty} \frac{1}{T} \int_{t=0}^T a_t \mu_{[0,1]}^x \, dt
\]

exists in weak* sense and is equal to an (\( \text{SL}(2, \mathbb{R}) \)-invariant) affine invariant probability measure with \( x \) in its support. On the other hand, there are several results, in particular by Chaika, Smillie and B. Weiss in [CSW20], that show that an analogue of Ratner’s equidistribution theorem (or our Theorem 1.2) fails to hold in this setting, for instance for some \( x \) the sequence of measure \( \mu_{[0,T]}^x \) may fail to converge as \( T \to \infty \), or may converge to a non-ergodic measure. We believe the following strengthening of (1.2), which we have learned has already been conjectured by Forni in [For21, Conj. 1.4], should however hold:

1.4. Conjecture. Let \( \mathcal{H}_1(\alpha) \) be the space of unit area surfaces in stratum of abelian differentials on a genus \( g \) surface whose zeros have multiplicities given by \( \alpha = (\alpha_1, \ldots, \alpha_n) \), and let \( x \in \mathcal{H}_1(\alpha) \). Then \( \lim_{t \to \infty} a_t \mu_{[0,1]}^x \) exists in the weak* sense and is equal to an affine invariant measure with \( x \) in its support.

Of course, once one establishes that \( \lim_{t \to \infty} a_t \mu_{[0,1]}^x \) exists, the rest follows from [EMM15].

2. Some preliminaries

We discuss the proofs of Theorems 1.1 and 1.3. As mentioned above, Theorem 1.2 is proved by combining Theorem 1.1 and the linearization techniques; in this announcement we focus on the proof of the former results. We note that the idea of using equidistribution of expanding translates of a
fixed piece of a $U$ orbit of the type $\{a_s u_s x : 0 \leq s \leq 1\}$ to deduce equidis-
tribution of a large segment of a non-translated $U$ orbit $\{u_s x : 0 \leq s \leq T\}$
is quite classical.

Let $U \subset N$ denote the group of upper triangular unipotent matrices in $H \subset G$, respectively, and let $A = \{a_t : t \in \mathbb{R}\} \subset H$. More explicitly, if $G = SL_2(\mathbb{C})$, then

$$N = \left\{ n(r, s) = \begin{pmatrix} 1 & r + is \\ 0 & 1 \end{pmatrix} : (r, s) \in \mathbb{R}^2 \right\}$$

and $U = \{n(r, 0) : r \in \mathbb{R}\}$; note that $n(r, 0) = u_r$ for $r \in \mathbb{R}$. Let

$$V = \{n(0, s) = v_s : s \in \mathbb{R}\};$$

if $G = SL_2(\mathbb{R}) \times SL_2(\mathbb{R})$, then

$$N = \left\{ n(r, s) = \begin{pmatrix} 1 & r + s \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix} : (r, s) \in \mathbb{R}^2 \right\}$$

and $U = \{n(r, 0) : r \in \mathbb{R}\}$. As before, $n(r, 0) = u_r$ for $r \in \mathbb{R}$. Let

$$V = \{n(0, s) = v_s : s \in \mathbb{R}\}.$$ 

In both cases, we have $N = UV$. Let us denote the transpose of $U$ by $U^\top$ and its elements by $u_r^\top$.

Let $g = \text{Lie}(G)$, that is, $g = sl_2(\mathbb{C})$ or $g = sl_2(\mathbb{R}) \oplus sl_2(\mathbb{R})$. Let $r = isl_2(\mathbb{R})$ if $g = sl_2(\mathbb{C})$ and $r = sl_2(\mathbb{R}) \oplus \{0\}$ if $g = sl_2(\mathbb{R}) \oplus sl_2(\mathbb{R})$. In either case $g = h \oplus r$ where $h = \text{Lie}(H) \cong sl_2(\mathbb{R})$, and both $h$ and $r$ are $\text{Ad}(H)$-invariant.

We fix a norm on $h$ by taking the maximum norm where the coordinates are given by $\text{Lie}(U)$, $\text{Lie}(U^\top)$ and $\text{Lie}(A)$; similarly fix a norm on $r$. By taking maximum of these two norms, we obtain a norm on $g$. All these norms will be denoted by $\|\|$. For all $\beta > 0$, we define $B^H_{\beta} := \exp(B_h(0, \beta)) \cdot \exp(B_r(0, \beta))$ where $B_{\bullet}(0, \beta)$ denotes the ball of radius $\beta$ in $\bullet$ with respect to $\|\|$. We also define

$$B^H_{\beta} := \{u_s^- : |s| \leq \beta\} \cdot \{a_t : |t| \leq \beta\} \cdot \{u_r : |r| \leq \beta\}$$

for all $0 < \beta < 1$.

For the sake of simplicity of the exposition here, let us assume $X$ is compact. Let $0 < \eta_0 < 1$ be so that the map $g \mapsto gx$ is injective on $B^G_{100\eta_0}$ for all $x \in X$.

3. From large dimension to equidistribution

We begin with an equidistribution theorem which is of independent interest. In the proof of Theorem 1.1 this proposition will be applied to each of the sets obtained in the bootstrap phase, see Proposition 6.1.
Let us recall the following quantitative decay of correlations for the ambient space $X$: There exists $0 < \kappa_0 \leq 1$ so that

$$\left| \int \varphi(gx)\psi(x) \, dm_X - \int \varphi \, dm_X \int \psi \, dm_X \right| \ll S(\varphi)S(\psi)e^{-\kappa_0d(e,g)}$$

for all $\varphi, \psi \in C^\infty_c(X) + \mathbb{C}$, where $m_X$ is the $G$-invariant probability measure on $X$ and $d$ is our fixed right $G$-invariant metric on $G$. See, e.g., [KM96, §2.4] and references there for (3.1); we note that $\kappa_0$ is absolute if $\Gamma$ is a congruence subgroup. This is known in much greater generality, but the cases relevant to our paper are due to Selberg and Jacquet-Langlands [Sel65, JL70].

The quantitative decay of correlation can be used to establish quantitative results regarding the equidistribution of translates of pieces of an $N$-orbit. Specifically we employ the results in [KM96], but there is rich literature around the subject; a more complete list of such papers can be found in [LM21, §1].

Now let $\xi : [0,1] \to r$ be a smooth curve. Then using the quantitative results regarding equidistribution of translates of pieces of an $N$-orbit such as [KM96], one can show that for every $x \in X$,

$$a_{\tau}\{u_r \exp(\xi(s)) \cdot x : r, s \in [0,1]\}$$

is equidistributed in $X$ as $\tau \to \infty$ (with a rate which is polynomial in $e^{-\tau}$). The key point in the deduction of this equidistribution result from the equidistribution of shifted $N$ orbits is that conjugation by $a_{\tau}$ moves $u_r \exp(\xi(s))$ to the direction of $N$, hence the above average essentially reduces to an average on a $N$ orbit.

Roughly speaking, the following proposition states that one may replace the curve $\{\xi(s) : s \in [0,1]\}$ with a measure on $r$ so long as the measure has dimension $\geq 1 - \theta$, for an appropriate choice of $\theta$ depending on $\kappa_0$.

The precise formulation is the following.

**3.1. Proposition.** For any $\theta > 0$ and $c > 0$ there is a $\kappa_3$ so that the following holds: Let $0 < q < 10^{-6}$, and let $F \subset B_{r}(0, q)$ be a finite set satisfying

$$\frac{\#(F \cap B_{r}(0, b))}{\#F} \leq \rho^{-c}(b/q)^{1-\theta} \quad \text{for all } b \geq \rho$$

where $\rho < q^{10}$.

Then for all $x \in X$ and all $\log(1/q) \leq \tau \leq \frac{1}{10} \log(1/\rho)$, we have

$$\left| \int_{0}^{1} \frac{1}{\#F} \sum_{w \in F} \varphi(a_{\tau}u_r \exp(w)x) \, dr - \int \varphi \, dm_X \right| \ll_X$$

$$S(\varphi) \max \left( (q/\rho)^{2\kappa_3} \rho^{-2c} e^{2\tau \theta} q^{2\kappa_3/M} \right),$$

where $S(\varphi)$ is a certain Sobolev norm and $M$ an absolute constant.
The proof of this proposition is significantly more delicate than that of the "toy version" of a shifted curve, and relies on an adaptation of a projection theorem due to Käenmäki, Orponen, and Venieri \cite{kov17}, based on the works of Wolff \cite{vol00}, Schlag \cite{sch03}, and \cite{zah12}, in conjunction with a sparse equidistribution argument due to Venkatesh \cite{ven10}. These elements also played a crucial role in previous work by E.L. and A.M. \cite{lm21} regarding quantitative density for the action of $AU$ on the spaces we consider here.

The goal in the remaining steps is to show that unless Theorem 1.1(2) holds, we can find a subset $J \subset [0,1], \text{ where } [0,1] \setminus J \text{ has an exponentially small measure, and up to an exponentially small error the uniform measure on } \{a_{A \log T u_r x_0} : r \in J\} \text{ can be decomposed into counting measures on sets each of which satisfies the conditions in Proposition 3.1.}$

4. Inheritance of the Diophantine property

If part (2) in Theorem 1.1 holds, we are done. Let us, therefore, assume the alternative, which gives some Diophantine condition on the point $x_0$ in terms of its distance to nearby periodic $H$-orbits. The first step in the proof is to improve this Diophantine condition, perhaps not at $x_0$, but at some (indeed all except a set of exponentially small measure) point on the translation of a the $U$-orbit segment $\{u_r x_0 : r \in [0,1]\}$ by a big diagonal element $a_t$.

4.1. Proposition. There exist $D_0$ (absolute) and $C_3$ (depending on $X$) so that the following holds. Let $\delta > 0$. Let $x_0 \in X$ be so that $d_X(x,x_0) > T^{-1}$ for all $x$ with $\text{vol}(Hx) \leq T^\delta$ for a large parameter $T$ (depending on $\delta$). Then for all $t \geq D_0 \log T$, we have

$$\left| \left\{ r \in [0,1] : \text{There is } x \text{ with } \text{vol}(Hx) \leq T^\delta \text{ so that } d_X(x,a_t u_r x_0) \leq C_3 T^{-D_0 \delta} \right\} \right| \leq C_3 T^{-\delta}.$$ 

In the proof of Proposition 4.1 we consider each periodic orbit individually, and then use the fact that the number of periodic $H$-orbits with volume $\leq R$ in $X$ is $\ll R^6$, see e.g. \cite{mo20} §10 to conclude. The desired result for an individual orbit can be proved using Margulis functions for periodic $H$-orbits similar to those which were used in \cite{lm21} §9, see also \cite{emm15} Prop. 2.13.

It is worth mentioning that even though \cite{lmms19} Thm. 1.4] concerns long pieces of $U$-orbits and Proposition 4.1 concerns translates of pieces of $U$-orbits, similar tools are applicable here as well. In particular, a version of Proposition 4.1 can also be proved using the methods of \cite{lmms19}. 
5. Closing lemma

Let \( t > 0 \) be a large parameter and fix some \( \varepsilon^{0.01t} < \beta = e^{-\kappa t} < \eta_0 \); in our application, \( \kappa \) will be chosen to be \( \ll 1/D_0 \) where the implied constant depends on \( X \) and \( D_0 \) is as in Proposition 4.1.

For every \( \tau \geq 0 \), put
\[
E_\tau = B_{\beta}^H \cdot a_\tau \cdot \{ u_r : r \in [0,1] \} \subset H.
\]

If \( y \in X \) is so that the map \( h \mapsto hy \) is injective on \( E_\tau \), then \( \mu_{E_\tau,y} \) denotes the pushforward of the normalized Haar measure on \( E_\tau \) to \( E_\tau \cdot y \subset X \).

Let \( \tau \geq 0 \) and \( y \in X \). For every \( z \in E_\tau \cdot y \), put
\[
I_\tau(z) := \{ w \in \tau : 0 < \| w \| < \eta_0 \text{ and } \exp(w)z \in E_\tau \cdot y \};
\]
this is a finite subset of \( \tau \) since \( E_\tau \) is bounded — we will define \( I_\tau(h,z) \) for all \( h \in H \) and more general sets \( E \) in the bootstrap phase below.

Let \( 0 < \alpha < 1 \). Define the function \( f_\tau : E_\tau \cdot y \to [1,\infty) \) as follows
\[
f_\tau(z) = \begin{cases} 
\sum_{w \in I_\tau(z)} \| w \|^{-\alpha} & \text{if } I_\tau(z) \neq \emptyset \\
\eta_0^{-\alpha} & \text{otherwise}
\end{cases}.
\]

The following proposition supplies an initial dimension which we will bootstrap in the next phase. Roughly speaking, it asserts that points in \( E_{8t} \cdot x_0 \) (possibly after removing an exponentially small set of exceptions) are separated transversal to \( H \), unless \( x_0 \) is extremely close to a periodic \( H \) orbit.

5.1. Proposition. Assume \( \Gamma \) is arithmetic. There exists \( D_1 \) (which depends on \( \Gamma \) explicitly) satisfying the following. Let \( D \geq D_1 \) and \( x_0 \in X \). Then for all large enough \( t \) at least one of the following holds.

(1) There is a subset \( E' \subset E_{8t} \) with \( m_H(E_{8t} \setminus E') \ll_X \beta m_H(E_{8t}) \) such that for all \( h \in E' \) we have the following
   (a) \( h \mapsto h \cdot x_0 \) is injective on \( E_t \).
   (b) For all \( z \in E_t \cdot h \cdot x_0 \), we have
   \[
f_t(z) \leq e^{Dt}.
   \]

(2) There is \( x \in X \) such that \( Hx \) is periodic with
   \[
   \text{vol}(Hx) \leq e^{Dt} \quad \text{and} \quad d_X(x,x_0) \leq e^{(-D+D_1)t}.
   \]

This proposition is where the arithmeticity assumption on \( \Gamma \) is used. If we replace the assumption that \( \Gamma \) is arithmetic with the weaker requirement that \( \Gamma \) has algebraic entries, we get a version of this proposition where part (2) is replaced with the following.

(2') There is \( x \in X \) with
\[
d_X(x,x_0) \leq e^{(-D+D_1)t},
\]
satisfying the following: there are elements \( \gamma_1 \) and \( \gamma_2 \) in \( \text{Stab}_H(x) \) with \( \| \gamma_i \| \leq e^{D_1t} \) for \( i = 1,2 \) so that the group generated by \( \{ \gamma_1, \gamma_2 \} \) is Zariski dense in \( H \).
6. Improving the dimension

Proposition 5.1 shows that up to an exponentially small error, the set \( \{a_r u_r x_0 : r \in [0,1] \} \) has a small positive dimension transversal to \( H \) at controlled scales. The objective in this step is to show that by applying elements of the form \( a_r u_r \), for a fixed \( \ell \) and a random \( r \in [0,1] \), we can inductively improve this dimension transversal to \( H \) at controlled scales to \( \alpha \) (which will be chosen to be smaller than but quite close to 1). This is achieved by investigating the evolution of a certain Margulis function (cf. the survey \cite{EM22}, though the Margulis function we use here is somewhat intricate).

To state the main result which is Proposition 6.1, we need some notation. Let \( t \) be large enough so that the conclusion of Proposition 5.1 holds. This \( t \) is related to the \( T \) of Theorems 1.1 and 1.3 in the following way: we take \( t \) so that \( e^{D_1 t} = T^6 \).

Let
\[
E = B^H_\beta \cdot \{ u_r : |r| \leq \eta_0 \},
\]
where \( \beta = e^{-\kappa t} \) for some small \( \kappa > 0 \). (More explicitly, we will fix some \( 0 < \varepsilon \leq 10^{-8} \) and let \( \kappa = 10^{-6} \varepsilon / D \) where \( D \geq D_1(D_0 + 1) \).

It will be more convenient to approximate translations \( \{a_r u_r x_0 : r \in [0,1] \} \) with sets which are a disjoint union of local \( E \)-orbits as we now define. Let \( F \subset B_\epsilon(0, \beta) \) be a finite set with \( \#F \geq e^{t/2} \), and let \( y \in X \). For every \( w \in F \), let \( E_w \subset E \) be a Borel set so that \( m_H(E_w) \geq \beta^4 \) and \( m_H(E_w \triangle (B^H_\beta \cdot E_w)) \leq \beta m_H(E_w) \), where \( m_H \) denotes a fixed Haar measure on \( H \). Put
\[
E = \bigcup E_w \cdot \{ \exp(w)y : w \in F \}.
\]
We equip \( E \) with the probability measure \( \mu_E = \sum_w \frac{1}{m_H(E_w)} \sum_w \mu_{w,y} \) where \( \mu_{w,y} \) denotes the pushforward of \( m_H|_{E_w} \) to \( E_w \cdot \exp(w)y \) for every \( w \in F \).

Let \( \theta \) be a small constant depending on the decay of matrix coefficients in \( G/T \) (the exact value we shall use is \( \theta = \kappa_0^2 / 500 M \), where \( \kappa_0 \) is as \eqref{eq:kappa0} and \( M \) as in Proposition 3.1). Let \( \alpha = 1 - \theta \).

As was mentioned before, we fix \( 0 < \varepsilon \leq 10^{-8} \). Let \( \ell = 0.01 \varepsilon t \), and let \( \nu_\ell \) be the probability measure on \( H \) defined by
\[
\nu_\ell(\varphi) = \int_0^1 \varphi(a_r u_r \varphi) \, dr \quad \text{for all } \varphi \in C_c(H);
\]
let \( \nu_\ell^{(n)} = \nu_\ell \ast \cdots \ast \nu_\ell \) denote the \( n \)-fold convolution of \( \nu_\ell \) for all \( n \in \mathbb{N} \).

The following is the main statement.

6.1. Proposition. Let \( x_0 \in X \), and assume that Proposition 5.1(2) does not hold for \( D, x_0, \) and \( t \). Let \( h \in E' \), where \( E' \) is as in Proposition 5.1(1), and put \( x_1 = hx_0 \). Let
\[
J := [d_1 - 10^4 \varepsilon^{-1/2}, d_1] \cap \mathbb{N},
\]
where \( d_1 = 100\left[\frac{12\beta - 3}{2z}\right] \). For every \( d \in J \), there is a collection
\[
\Xi_d = \{E_{d,i} : 1 \leq i \leq N_d\}
\]
of sets defined as in (6.1) for some \( F_{d,i} \subset B_\epsilon(0, \beta) \), so that both of the following hold:

1. Put \( w = e^{-\sqrt{\epsilon}t} \). Let \( d \in J \), \( 1 \leq i \leq N_d \), and let \( w_0 \in B_\epsilon(0, \beta) \). Then for every \( w \in B_\epsilon(w_0, \rho) \) and all \( b \geq e^{-t/2} \), we have
\[
(6.2) \quad \#(B_t(w, b) \cap B_\epsilon(w_0, \rho) \cap F_{d,i}) \leq e^{\epsilon t}(b/\rho)^\alpha.
\]

2. For every \( \varphi \in C^\infty_c(X) \), we have
\[
(6.3) \quad \left| \int \varphi \, d\nu_d^{(d_1)} \ast \mu_{E_{d,i}} x_1 - \sum_{d,i} \int \varphi \, d\nu_d^{(d_1-d)} \ast \mu_{E_{d,i}} \right| \ll \text{Lip}(\varphi) \beta^{\kappa_4}
\]
where \( c_{d,i} \geq 0 \) and \( \sum_{d,i} c_{d,i} = 1 - O(\beta^{\kappa_4}) \), Lip(\varphi) is the Lipschitz norm of \( \varphi \), and \( \kappa_4 \) and the implied constants depend on \( X \).

Roughly speaking, the proposition states that up to an exponentially small error, \( \nu_d^{(d_1)} \ast \mu_{E_{d,i}} x_1 \) may be decomposed as \( \sum_{d,i} c_{d,i} \nu_d^{(d_1-d)} \ast \mu_{E_{d,i}} \) where
\[
\sum_{d,i} c_{d,i} = 1 - O(\beta^{\kappa_4}) \quad \text{(see (6.3))}
\]
and for all \( d \in J \) and \( 1 \leq i \leq N_d \) the dimension of \( E_{d,i} \) transversal to \( H \) at controlled scales is \( \geq \alpha \) (see (6.2)).

Combining Proposition 6.1 with the previous discussion, we may complete the proof of Theorem 1.1 (and Theorem 1.3). A brief outline of this deduction follows: For all \( k \in \mathbb{N} \) and all \( h \in \text{supp}(\nu^{(k)}_d) \), \( hE_{d,i} x_1 \) remains exponentially close to
\[
\{a_{k\ell+9t}u_r x_0 : r \in [0, 1]\}.
\]
Recall that \( x_1 = h x_0 \) where \( h \in E' \subset E_\delta \), hence the 9t term in the amount we flow by the group \( a_* \).

Let \( m = \left[\sqrt{\epsilon t}/\ell\right] = \left[100\epsilon^{-1/2}\right] \) and \( s = (d_1 + m)\ell + 9t \). The above observation and Proposition 6.1, thus, reduce the study of \( \int_0^1 \varphi(a_s u_r x_0) \, dr \) to the investigation of integrals of the form
\[
\int \varphi \, d\nu^{(n+m)}_d \ast \mu_{E}
\]
where \( E = E_{d,i} \) has dimension at least \( \alpha \) transversal to \( H \) at controlled scales, see (6.2), and \( 0 \leq n = d_1 - d \leq 10^4 \epsilon^{-1/2} \). Note that for any such \( n \), we have
\[
\sqrt{\epsilon t} \leq (n + m)\ell \leq \frac{10^4\ell}{\sqrt{\epsilon}} + 2\sqrt{\epsilon t} = 102\sqrt{\epsilon t} \leq \frac{\kappa_0^2}{4M\theta} \sqrt{\epsilon t}
\]
Thus, Proposition 3.1 applied with \( \theta = 1 - \alpha, \ c = 2\epsilon, \ \rho = e^{-\sqrt{\epsilon}t}, \ \rho = e^{-t/2} \), and \( \tau = (n + m)\ell \), implies that
\[
(6.4) \quad \left| \int \int \varphi(h y) \, d\mu_{E}(y) \, d\nu^{(n+m)}_d(h) - \int \varphi \, dm x \right| \ll S(\varphi) e^{-\epsilon t}
\]
where the implied constants depend on $X$. Theorem 1.1 follows.

The proof of Propositions 6.1 relies on the evolution of the Margulis function $f_{E,\varrho,R}$ defined below. For every $(h,z) \in H \times \mathcal{E}$, put

$$I_{E,\varrho}(h,z) := \{ w \in \tau : 0 < \|w\| < \varrho \text{ and } \exp(w)hz \in h\mathcal{E} \}.$$ 

Since $\mathcal{E}$ is bounded, $I_{E,\varrho}(h,z)$ is a finite set for all $(h,z) \in H \times \mathcal{E}$.

For every $R \geq 0$, define the modified and localized Margulis function $\hat{f}_{E,\varrho,R} : H \times \mathcal{E} \to [1, \infty)$ as follows:

$\hat{f}_{E,\varrho,R}(h,z) = \varrho^{-\alpha},$ and if $\#I_{E,\varrho}(h,z) > R$, put

$$\hat{f}_{E,\varrho,R}(h,z) = \min \left\{ \sum_{w \in I} \|w\|^{-\alpha} : I \subset I_{E,\varrho}(h,z) \text{ and } \#(I_{E,\varrho}(h,z) \setminus I) = R \right\}.$$

We begin the outline of the proof with the following observation: the set $\mathcal{E}_t \cap I_1$ gives rise to sets $\mathcal{E}$ which are defined as in (6.1) and satisfy $\hat{f}_{E,\varrho,0}(e,z) \leq e^{Dt}$ for all $z \in \mathcal{E}$, see Proposition 5.1.

Let $\mathcal{E}$ be one of these sets, then up to an exponentially small error, $\nu_t \ast \mu_{\mathcal{E}}$ may be decomposed into $\sum c_j' \mu_{\mathcal{E}_j}$ where $c_j' \geq 0$ and $\sum c_j' = 1 - O(\beta^*)$. This can be seen by first decomposing $\nu_t \ast \mu_{\mathcal{E}}$ into a combination of measures supported on subsets of $X$ which are exponentially thin in the direction of $U^-$ (note that $a_0 \mathcal{E}_\beta$ will be exponentially thin in the direction of $U^-$), and then smearing these measures with $B^H_\beta$. Continuing inductively, we obtain the decomposition (6.3).

The fact that the energy estimate (6.2) is also satisfied for the terms appearing in (6.3) is at the heart of the proposition. The proof of this fact is based on the following bootstrap step: for all but an exponentially small set of $r \in [0, 1]$ and all but an exponentially small set of $z \in \mathcal{E}$,

$$\hat{f}_{E,\varrho,\beta^{-1}}(a_\ell u_r, z) \leq e^{-\ell/2} e^{Dt} + 4e^{\ell} \varrho^{-\alpha} \psi_{E,\varrho}(a_\ell u_r, z),$$

where $\psi_{E,\varrho}(h,z) := \max\{ \#I_{E,\varrho}(h,z), 1 \}$.

The proof of (6.5) relies on the aforementioned works of Wolff, Schlag, and Zahl [Wol00, Sch03, Zah12], see also the discussion below.

The set $a_\ell u_r \mathcal{E}$ can now be used to construct sets $\mathcal{E}'$ (defined as in (6.1), and in view of (6.5), the estimate for $\hat{f}_{E,\varrho,\beta^{-1}}$ is improved. Continuing inductively and using $Dt - 0.5d_1 \ell \leq 3\ell/4$, after $d \leq d_1$ steps, we have

$$\hat{f}_{E',\varrho,\beta^{-1}}(e,z) \leq e^{\ell} \varrho^{-\alpha} \psi_{E',\varrho}(e,z),$$

which implies the dimension estimate (6.2) for the set $F' \subset B_\ell(0,\beta)$ which is used in the definition of $\mathcal{E}'$, see (6.1).

We emphasize that our inductive scheme produces sets $\mathcal{E}'$ at every step $0 < d \leq d_1$ with an improved bound on $\hat{f}_{E',\varrho,\beta^{-1}}$, however, it does not

\footnote{If $\Gamma$ is nonuniform, this Margulis function needs to be modified accordingly; but in this section we limit ourselves to the compact case.}
guarantee that (6.6) is only satisfied for \( d \in J \). On the other hand, and as it was discussed above, the fact that our stopping times \( d \) satisfy \( d_1 - d \leq 10^4 \varepsilon^{-1/2} \) is essential for us when we apply Proposition 3.1. We remedy this issue as follows: if (6.2) holds for some \( E' \) defined at step \( d < d_1 - 10^4 \varepsilon^{-1/2} \), then we use the above inductive scheme to show that starting with \( E' \), the process again terminates (i.e., (6.6) is satisfied) in at most \( 10^4 \varepsilon^{-1/2} \)-many further steps.

On a related note, it should also be mentioned that ideally one would like to replace the interval of possible choices of \( d \in J \) with the singleton \( \{d_1\} \), i.e., to show that after exactly \( d_1 \) steps, one obtains sets which satisfy (6.2). Indeed, such statement can be obtained if one is content with restricting to \( \alpha < 1/2 \) — this can be achieved using estimates analogues to [EMM98, Lemma 5.1] where one replaces the integral over \([0, 1]\) with integrals over much smaller intervals and by conditioning the random walk.

However, it is essential for us to work with \( \alpha = 1 - \theta \) where \( \theta > 0 \) is rather small. For such choices of \( \alpha \), there are vectors \( w \in \mathfrak{r} \) where the growth of \( \|a_t u_r w\| \) is too slow. Indeed, in general, one can only guarantee that

\[
\int_0^1 \|a_t u_r w\|^{-\alpha} \, dr \ll e^{-\theta t} \|w\|^{-\alpha}.
\]

Using this general fact (which is an exercise in linear algebra) as an input, one can prove a version of Proposition 6.1 where

\[
[d_1 - \frac{10^4}{\sqrt{\varepsilon}}, d_1]
\]

is replaced by

\[
[d_1 - \frac{36}{\theta \sqrt{\varepsilon}}, d_1],
\]
in particular, the length of the interval cannot be made smaller than \((\theta \sqrt{\varepsilon})^{-1}\).

As it was discussed above, this improvement is pivotal to our analysis, and it is made possible by bringing to bear [Wol00, Sch03, Zah12] in this step of the argument as well. Indeed, the poor rate in (6.7) is closely related to the existence of double zeroes for the map \( r \mapsto (a_t u_r w)_{12} \) (the (1,2)-entry of \( a_t u_r w \)); we thus use [Wol00, Sch03, Zah12], to control tangencies between the two parabolas \( \{(u_r w_1)_{12} : r \in [0, 1]\} \) and \( \{(u_r w_2)_{12} : r \in [0, 1]\} \) for most pairs \( w_1, w_2 \in \mathfrak{F} \) and most \( r \in [0, 1] \). This yields an improved version of (6.7) which we use crucially.

References


E.L.: The Einstein Institute of Mathematics, Edmond J. Safra Campus, Givat Ram, The Hebrew University of Jerusalem, Jerusalem, 91904, Israel
E-mail address: elon@math.huji.ac.il

A.M.: Department of Mathematics, University of California, San Diego, CA 92093
E-mail address: amohammadi@ucsd.edu

Z.W.: Pennsylvania State University, Department of Mathematics, University Park, PA 16802
E-mail address: zhirenw@psu.edu