EFFECTIVE EQUIDISTRIBUTION OF TRANSLATES OF MAXIMAL HOROSPHERICAL MEASURES IN THE SPACE OF LATTICES

KATHRYN DABBS, MICHAEL KELLY, AND HAN LI

Abstract. Recently Mohammadi and Salehi-Golsefidy gave necessary and sufficient conditions under which certain translates of homogeneous measures converge, and they determined the limiting measures in the cases of convergence. The class of measures they considered includes the maximal horospherical measures. In this paper we prove the corresponding effective equidistribution results in the space of unimodular lattices. We also prove the corresponding results for probability measures with absolutely continuous densities in rank two and three. Then we address the problem of determining the error terms in two counting problems also considered by Mohammadi and Salehi-Golsefidy. In the first problem, we determine an error term for counting the number of lifts of a closed horosphere from an irreducible, finite-volume quotient of the space of positive definite $n \times n$ matrices of determinant one that intersect a ball with large radius. In the second problem, we determine a logarithmic error term for the Manin conjecture of a flag variety over $\mathbb{Q}$.

1. Introduction

Several important and recurring problems in homogeneous dynamics concern the equidistribution properties of closed unipotent orbits. These problems have been studied for many years by many authors, and they are important because of their connections to geometry and number theory. It often happens that one is interested in proving not only an equidistribution result, but also a quantitative bound on the discrepancy of the equidistribution. There are two reasons for this: (1) knowledge of the rate of equidistribution sheds light on the regularity and rigidity of the dynamics; and (2) in applications, particularly in counting problems, effective rates of equidistribution play a fundamental role in determining an error term for any relevant estimates.

A fundamental example of the equidistribution of closed unipotent orbits is the equidistribution of long closed horocycles in $M = \mathrm{SO}_2(\mathbb{R})\backslash\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$, the modular surface with the Poincaré metric. For any $y > 0$ there exists a unique closed horocycle h_y in M of length $\frac{1}{y}$, and h_y equidistributes in M as y tends to zero. See for example [3] or [20]. That is, if ν_y is the probability measure on M that puts uniform mass on h_y, then ν_y converges weakly to the uniform probability measure on M. See Sarnak's paper [15] for a generalization to general non-compact, finite-volume Riemann surfaces. From a dynamical point of view, using the fact that the horocycles h_y are geodesic translates of any fixed closed horocycle, one can prove the equidistribution using the mixing properties of the geodesic flow. This idea originates in the thesis of G. Margulis [12].

The discrepancy estimates for this equidistribution problem are well studied. There are currently two main approaches to obtain such estimates. In the method of Sarnak and Zagier [15, 20] one associates an Eisenstein series to each ν_y and uses the analytic continuation of the Eisenstein series (due to Selberg [16]) to produce an effective rate. Alternatively, using the ideas originating in Margulis's thesis, one can use the spectral gap for M to achieve an effective rate for the equidistribution of the long closed horocycles. It is a well known result of Zagier [20] that the rate of equidistribution is $O(y^{3/4-\epsilon})$ for each $\epsilon > 0$ if and only if the Riemann hypothesis is true. He also showed that the rate of equidistribution is at least $o(y^{1/2})$, which is of the same strength as the prime number theorem.

The story for long closed horocycles in general rank one spaces is similar: the full horocycle is expanded or contracted depending on the direction it is translated. In higher rank, however, the closed horospheres (the closed maximal unipotent orbits) can be simultaneously contracted and expanded as they are translated along a given geodesic. This complication has caught the attention of many mathematicians over the years, and there are many special circumstances for which we know how to control it. For instance, when investigating certain problems concerning Diophantine

Date: April 5, 2016.

approximation with weights, Kleinbock and Weiss [11] proved an equidistribution theorem for the translates of *minimal* horospherical measures[1] in the presence of simultaneous expansion and contraction. An effective form of their result was obtained in [9] and was generalized in the recent paper [10].

Another situation in which we know how to control simultaneous expansion and contraction was the subject of the recent work of Mohammadi and Salehi-Golsefidy [13]. They provide necessary and sufficient conditions under which translates of certain homogeneous measures (including the maximal horospherical measures) converge, and they determine the limiting measures in the cases of convergence. Similar results can be found in an earlier work of Shah and Weiss [17], in which a similar collection of translates is considered. To reiterate: the difficulty one encounters in the higher rank setting is that a closed horospherical orbits can both expand and contract while being translated in a particular direction. This phenomenon makes it difficult to determine the convergence of translates. Once more, it makes it difficult to achieve effective rates of convergence.

It is our objective in this paper to establish the rates of convergence for the main results in [13] in the space of unimodular lattices. Our first theorem establishes effective equidistribution for translates of maximal horospherical measures. This result is an analogue of a similar result for minimal horospherical measures originally obtained in an ineffective form in [11] and effective form in [9]. Our method of proof closely mirrors that of Kleinbock-Margulis [9] for the minimal horospherical case.

1.1. **Statement of Results.** Let $n > 2$ be an integer, $G = \mathrm{SL}_n(\mathbb{R})$, $\Gamma = \mathrm{SL}_n(\mathbb{Z})$, and A be the subgroup of G consisting of positive diagonal matrices. For an element $a \in A$ we will use the notation

$$a = \mathrm{diag}(a_1, ..., a_n).$$

Let $\Delta = \{\alpha_1, ..., \alpha_{n-1}\}$ be simple roots of G with respect to A given by

$$\alpha_i(a) = \frac{a_i}{a_{i+1}},$$

and let $\lambda_{\alpha_1}, ..., \lambda_{\alpha_{n-1}}$ be the corresponding fundamental weights

$$\lambda_{\alpha_i}(a) = a_1 \cdots a_i.$$

For each $E \subset \Delta$, let P_E be the associated standard parabolic subgroup (see [7]). For example, $P_\Delta = G$ and $P_\emptyset$ is the group of upper triangular matrices in G. Let Q_E be the group generated by the one parameter unipotent subgroups of P_E. The group $Q_\emptyset$ is a maximal unipotent subgroup of G, and for it we reserve the special notation $U = Q_\emptyset$. Finally, we let μ_E be the unique invariant probability measure supported on $Q_E\Gamma$ in G/Γ. We use the notation $\mu = \mu_\emptyset$, and set $m = \mu_\Delta$ for the G-invariant probability measure on G/Γ. We will now state a special case[2] of the main result of [13].

Theorem 1 ([13, Theorem 1])**.** *Let $\{a_k\} \subset A$ and $E \subset \Delta$. Then*

(1) If $\lambda_\alpha(a_k) \to 0$ as $k \to \infty$ for some $\alpha \notin E$, then $a_k\mu_E$ diverges in the space of Borel probability measures on G/Γ.

(2) Let $E \subset F \subset \Delta$. If $\lambda_\alpha(a_k) = 1$ for any $\alpha \notin F$ and $\lambda_\alpha(a_k) \to \infty$ as $k \to \infty$ for any $\alpha \in F \backslash E$, then $a_k\mu_E$ converges to μ_F as $k \to \infty$.

Theorem 1 can be thought of as identifying "cones" in A that govern the convergence of the translates of the measures μ_E. For each $E \subset \Delta$, let

$$\mathcal{C}_E = \{a \in A : \lambda_\alpha(a) > 1 \text{ for each } \alpha \in \Delta \setminus E\}.$$

[1] A horospherical subgroup is the unipotent radical of a proper parabolic subgroup. A horospherical subgroup is *minimal* if it is the unipotent radical of a *maximal* parabolic subgroup, and it is *maximal* if it is the unipotent radical of a *minimal* parabolic subgroup.

[2] In [13] this theorem is proved in much greater generality, e.g. G does not necessarily have to be $\mathbb{Q}$-split.

If $\{a_k\}$ tends to infinity *away from the boundary* in $\mathcal{C}_E$ (a notion made precise by the above theorem), then $a_k.\mu$ tends to μ_E. If $E = \emptyset$, then we call the set $\mathcal{C} = \mathcal{C}_\emptyset$ the convergence cone. Each of these cones contains the cone

$$\mathcal{A} = \{a \in A : \alpha(a) > 1 \text{ for each } \alpha \in \Delta\}$$

which we call the *positive* or *fundamental* Weyl chamber.

Our main result is an effective version of Theorem 1 for the translates of the maximal horospherical measure μ.

Theorem 2. *There exists a constant $\delta = \delta(n) > 0$ such that for any $\varphi \in C^\infty_{comp}(G/\Gamma)$ there exists a constant $C = C(\varphi, n) > 0$ such that for any $a \in A$*

$$\Big|\int_{G/\Gamma} \varphi(a.z)d\mu(z) - \int_{G/\Gamma} \varphi(z)dm(z)\Big| < C\left(\min_{\alpha\in\Delta} \lambda_\alpha(a)\right)^{-\delta}. \tag{1}$$

We remark the above theorem is trivial when $a \notin \mathcal{C}$. To see this suppose that $a \notin \mathcal{C}$ and observe (by the definition of $\mathcal{C}$) for any $\delta > 0$, $(\min_{\alpha\in\Delta} \lambda_\alpha(a))^{-\delta} \geq 1$ and by taking $C = 2\sup|\varphi|$ we find the inequality is always satisfied. Our next result is a generalization of Theorem 2, and is an effective version of Theorem 1. After a suitable decomposition of measures, its proof proceeds by repeatedly applying Theorem 2 to certain marginals of μ_E.

Theorem 3. *Let $E \subset F \subset \Delta$. There exists a constant $\delta = \delta(n) > 0$ such that for any $\varphi \in C^\infty_{comp}(G/\Gamma)$ there exists a constant $C = C(\varphi, n) > 0$ such that for any $a \in A$ we have*

$$\Big|\int_{G/\Gamma} \varphi(a.h)d\mu_E(h) - \int_{G/\Gamma} \varphi(h)d\mu_F(h)\Big| < C\left(\min_{\alpha\in F\setminus E} \lambda_\alpha(a)\right)^{-\delta}.$$

Our next result is an effective version of Theorem 1 for absolutely continuous measures. We are able to obtain effective results in the full convergence cone $\mathcal{C}$ when $n = 3$ and $n = 4$. For $n > 4$ we are able to prove an effective result for flows in a cone that is strictly larger than $\mathcal{A}$. However, in general, we are unable to handle the absolutely continuous case for the full convergence cone.

For each $j = 1, ..., n-1$ we define

$$\mathcal{C}_j = \{\mathrm{diag}(e^{r_1}, ..., e^{r_n}) \in A \ : \ \min\{r_i : i = 1, ..., j\} \geq \max\{r_s : s = j+1, ..., n\}\},$$

and $\tilde{\mathcal{C}} = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_{n-1}$. Let ν_U be the Haar measure on U that is equal to μ on a fundamental domain of Γ in G/Γ. Now we can state our result regarding absolutely continuous densities.

Theorem 4. *There exists a constant $\delta = \delta_n > 0$ such that for any compact subset L of G/Γ, for any $f \in C^\infty_{comp}(U)$, $\varphi \in C^\infty_{comp}(G/\Gamma)$, there exists a constant $C = C(f, \varphi, L, n) > 0$ such that for any $z \in L$ and $a \in \tilde{\mathcal{C}}$ we have*

$$\Big|\int_U f(u)\varphi(auz)d\nu_U(u) - \int_U f \cdot \int_{G/\Gamma} \varphi\Big| < C \cdot \left(\min_{\alpha\in\Delta} \lambda_\alpha(a)\right)^{-\delta}. \tag{2}$$

We remark that Theorem 4 can be obtained from Theorem 1.3 of Kleinbock-Margulis [9] in conjunction with Fubini's theorem. This was pointed out to us by Kleinbock. The key point is that if $a \in \mathcal{C}$ completely expands a minimal horospherical marginal of ν_U, then the equidistribution of that particular marginal will force the equidistribution of $f d\nu_U$. Theorem 1.3 of Kleinbock-Margulis [9] exactly describes the equidistribution of minimal horospherical measures.

Notice that $\tilde{\mathcal{C}} \not\subset \mathcal{C}$. For instance, $a = \mathrm{diag}(e^1, e^{-2}, e^1) \in \mathcal{C}_1 \subset \tilde{\mathcal{C}}$ but $a \notin \mathcal{C}$. So Theorem 4 is only non-trivial for $a \in \tilde{\mathcal{C}} \cap \mathcal{C}$. See the remark after the statement of Theorem 2. In [13] it was pointed out by Mohammadi and Salehi-Golsefidy that when $n = 5$

$$a_0 = \mathrm{diag}(e^6, e^7, e^{-12}, e^9, e^{10})$$

is an element of $\mathcal{C}$, but it does not fully expand a minimal horospherical subgroup. It follows that $a_0 \notin \tilde{\mathcal{C}}$ (this can be shown directly) and, consequently, that $\tilde{\mathcal{C}}$ is not even convex when $n = 5$.

From here it is an easy exercise to show that $\mathcal{C} \subset \tilde{\mathcal{C}}$ if and only if $n = 3$ or 4. Thus we have the following corollary of Theorem 4.

Corollary 1. *Suppose $n = 3$ or 4. There exists a constant $\delta = \delta_n > 0$ such that for any compact subset L of G/Γ, for any $f \in C^\infty_{comp}(U)$, $\varphi \in C^\infty_{comp}(G/\Gamma)$, there exists a constant $C = C(f, \varphi, L, n) > 0$ such that for any $z \in L$ and $a \in A$ we have*

$$\left| \int_U f(u)\varphi(a.uz)d\nu_U(u) - \int_U f \cdot \int_{G/\Gamma} \varphi \right| < C \cdot \left(\min_{\alpha\in\Delta} \lambda_\alpha(a) \right)^{-\delta}. \tag{3}$$

1.2. **Applications.** In our first application we consider a geometric counting problem first considered in [5]. Let $K = \mathrm{SO}_n(\mathbb{R}) \leq G$ and $X = K\backslash G$ be the corresponding Riemannian symmetric space arising from G. If U is a maximal unipotent subgroup of G, then $\mathcal{U} = K\backslash KgU$ is a horosphere in X and all horospheres in X can be realized in this way. We let Ξ be the space of horospheres in X. Let $\mathcal{M} = X/\Gamma$ and let $\pi : X \to \mathcal{M}$ be the covering map. Suppose that $\mathcal{U}$ is a horosphere in X such that $\overline{\mathcal{U}} = \pi(\mathcal{U})$ is closed in $\mathcal{M}$. We are interested in estimating how many lifts of $\overline{\mathcal{U}}$ intersect a given ball $B(x, R)$ in X. That is, we wish to analyze the asymptotic behavior of the quantity

$$\#\left\{\mathcal{U} \in \Xi \ : \ \pi(\mathcal{U}) = \overline{\mathcal{U}} \text{ and } \mathcal{U} \cap B(x,R) \neq \emptyset\right\}. \tag{4}$$

In the rank one case ($n = 2$) it was shown by Eskin and McMullen [5] that the quantity in Equation 4 is asymptotic to the volume of $B(x, R)$ (times a suitable constant). The analogous result for higher rank ($n > 2$) was established by Mohammadi and Salehi-Golsefidy [13]. Our first theorem is an effective form of this result for $G = \mathrm{SL}_n(\mathbb{R})$. In principle, the Eskin-McMullen example can be made effective using Sarnak's effective equidistriubtion of low-lying horocycles [15]. We prove here, as far as we know, the first effective result for this counting problem in higher rank.

Theorem 5. *Let $\overline{\mathcal{U}}$ be a closed horosphere in $\mathcal{M}$ and $x_0 \in X$ be the identity coset. Then there is a constant $C > 0$, depending only on the dimension, and $\delta > 0$ such that*

$$\begin{aligned}\#\left\{\mathcal{U} \in \Xi : \pi(\mathcal{U}) = \overline{\mathcal{U}} \textit{ and } \mathcal{U} \cap B(x_0,R) \neq \emptyset\right\} = C\frac{\mathrm{vol}(\overline{\mathcal{U}})}{\mathrm{vol}(\mathcal{M})}&\mathrm{vol}(B(x_0,R)) \\ &+ O\left(\mathrm{vol}(B(x_0,R))^{1-\delta}\right).\end{aligned}$$

To prove the above theorem we only need to use the effective equidistribution for directions coming from the interior of $\mathcal{A}$. Consequently, our proof of Theorem 5 can be adapted to prove [13, Theorem 3] using only the wavefront lemma of Eskin-McMullen [5].

For our second application we consider the Manin conjecture for flag varieties over $\mathbb{Q}$. This problem was solved for generalized flag varieties by Franke, Manin, and Tschinkel in [6]. Their proof uses Langland's analytic continuation of higher rank Eisenstein series, and the method of that paper produces what is essentially the best possible error term. Here we will prove an effective form of their theorem that produces an inferior error term, but by using our effective equidistribution results in place of Eisenstein series. A dynamical proof of a more general result[3] is provided in [13], and it is this proof that we effectivize. Consider the standard representation of G on $\mathbb{R}^n$. It is well known that the stabilizer of any flag in $\mathbb{R}^n$ is a parabolic subgroup of G. Conversely, any parabolic subgroup of G stabilizes a flag in $\mathbb{R}^n$. It then follows that any flag variety over $\mathbb{Q}$ can be realized as $X = G/P$ for some parabolic subgroup P of G. The anticanonical line bundle of X is induced by a character ρ_P of P by $\mathcal{L} = G \times \mathbb{R}/\sim$ where $(g, x) \sim (gp, \rho_P(p)x)$. It follows from [2, Section 12] that ρ_P is the highest weight of a unique irreducible representation $\eta : G \to GL(V)$ which is strongly rational over $\mathbb{Q}$, there is a $v_0 \in V(\mathbb{Q})$ such that

$$P = \{g \in G \ : \ \eta(g)[v_0] = [v_0]\}$$

[3]In [13] Mohammadi and Salehi-Golsefidy are also able to handle the counting for heights with respect to arbitrary metrized line bundles.

where $[v_0]$ is the point corresponding to v_0 in $\mathbb{P}(V)$, and X is homeomorphic to $\eta(G)[v_0] \subset \mathbb{P}(V)$. Our counting will take place in this orbit and we henceforth identify X with $\eta(G)[v_0]$. We define a function $H : \mathbb{P}(V)(\mathbb{Q}) \to \mathbb{R}^+$ by $H([v]) = \|v\|$ where v is a primitive integral point corresponding to the point $[v]$ and $\|\cdot\|$ is the Euclidean norm on V. Using $H(\cdot)$ we define the (anticanonical) height $h : X \to \mathbb{R}^+$ on X by

$$h(\eta(g)[v_0]) = H(\eta(g)[v_0]). \tag{5}$$

We are interested in the asymptotic behavior of

$$N(T) = \#\{x \in X(\mathbb{Q}) \ : \ h(x) \leq T\}.$$

In [6, Theorem 5] it was proven that there exists a polynomial p of degree $rk(\mathrm{Pic}(X))$, such that

$$N(T) = Tp(\log(T)) + o(T)$$

as $T \to \infty$. It is not difficult to show that their method shows that the error term $o(T)$ can be replaced with $O(T^{1-\epsilon})$ for some $\epsilon > 0$. We are able to prove the following.

Theorem 6. *Let X and h be as above. Then there exists a constant $\delta > 0$, and a polynomial $p(t)$ of degree $k = rk(\mathrm{Pic}(X))$ such that*

$$\#\{x \in X(\mathbb{Q}) \ : \ h(x) \leq T\} = Tp(\log(T))\left(1 + o(\log(T)^{-\delta})\right) \tag{6}$$

as $T \to \infty$.

In the proofs of the previous two theorems we use a well developed counting technique which is due to Duke-Rudnick-Sarnak [4] and that has been employed by a number of authors. We recommend the survey [14] of Hee Oh for an overview of the method as well as its various applications.

1.3. **Further Remarks and References.** After the initial submission of this paper we learned of a recent preprint of Shi [19] that generalizes the main results of [10] and includes a generalization [19, Theorem 1.5] of our Theorem 4. In both [19, Theorem 1.5] and our Theorem 4 above it is required that the translates of the maximal horospherical measure contain a minimal horospherical marginal which is completely expanded. In rank four and greater, this is not always possible (see the example in [13, §2] which is mentioned in the remarks following Theorem 4 above). The proofs of both Theorem 4 and [19, Theorem 1.5] follow the approach of Kleinbock-Margulis [9] which is summarized in §2.1. It will be apparent from the remarks in §2.1 that Shi's proofs can be modified to prove a generalization of our Theorem 2 when G is a higher rank semisimple Lie group without compact factors.

Currently it seems that new ideas are needed to prove a full generalization of [13, Theorem 2] or even Corollary 1 in rank greater than three. See also the remarks below the statement of Theorem 1.4 in [19]. In [13] more general ineffective versions of the above theorems were proved, and it would be desirable to treat their effectivization for applications. In particular, it would be interesting to prove an effective version of [13, Theorem 1] (i.e. a generalization of our Theorem 3), and to treat the Manin problem for a generalized flag variety with respect to an arbitrary metrized line bundle. We plan to revisit these questions in a follow-up paper. The purpose of this paper is to report this progress in a concrete setting: the space of unimodular lattices.

Organization of the paper. We begin by proving our effective equidistribution theorems in Section 2. In Section 2.2 and Section 2.3 we recall some results we will need from [9, 11] regarding Margulis's thickening technique and establishing a quantitative recurrence result for translates of maximal unipotent orbits (see Corollary 2 and Corollary 3). Then we finish Section 2 with the proofs of Theorem 2, Theorem 3, and Theorem 4 in Section 2.4, Section 2.5, and Section 2.6 respectively.

In Section 3 we prove Theorem 5 and then prove Theorem 6 in Section 4.

2. Translates of horospherical measures

While we have stated our main results in terms of the multiplicative form of Δ, we will find it convenient to prove our results in additive form. That is, we take logs, and instead of considering elements in A we consider elements in its Lie algebra $\mathfrak{a}$, the vector space of traceless diagonal matrices. More specifically, for any $a \in A$, we may write $a = \exp(\mathrm{diag}(t_1, \dots, t_n))$, where $t_1, \dots, t_n \in \mathbb{R}$. Then, abusing the notation, we let $\Delta = \{\alpha_1, \dots, \alpha_{n-1}\}$, where

$$\alpha_i(\mathrm{diag}(t_1, \dots, t_n)) = t_i - t_{i+1}.$$

The set Δ is a standard choice of simple roots of $\mathfrak{g} = \mathfrak{sl}_n(\mathbb{R})$. The corresponding fundamental weights are given by

$$\lambda_{\alpha_i}(\mathrm{diag}(t_1, \dots, t_n)) = t_1 + \cdots + t_i.$$

Then the cones $\mathcal{A}$ and $\mathcal{C}_E$ may be identified with their logarithms as follows:

$$\mathcal{A} = \{X \in \mathfrak{a} : \alpha(X) > 0 \quad \text{for each} \quad \alpha \in \Delta\},$$

and for each $E \subset \Delta$

$$\mathcal{C}_E = \{X \in \mathfrak{a} : \lambda_\alpha(X) > 0 \quad \text{for each} \quad \alpha \in \Delta \setminus E\}.$$

2.1. **An overview of the method.** Our goal in Section 2 is to prove the effective equidistribution results in Theorem 2 and Theorem 4. The proofs of the two theorems use similar ideas. We will provide an overview of these ideas for Theorem 2 and then we will comment on the additional complications that must be dealt with in the proof of Theorem 4.

Let $z_0 = e\Gamma \in G/\Gamma$ be the identity coset, $a = g_t = \exp(t\boldsymbol{\theta})$, where $t > 0$, and $\boldsymbol{\theta} \in \mathcal{C}$ is on the unit sphere of $\mathfrak{a}$. We assume, as we may, that the test function φ in Theorem 2 satisfies $\int_{G/\Gamma} \varphi dm = 0$. Let ξ be a smooth function supported in $B_U(r)$ with $\int_U \xi = 1$. Then plainly

$$\int_{U.z_0} \varphi(g_t z) d\mu(z) = \int_U \int_{U.z_0} \xi(u) \varphi(g_t z) d\mu(z) d\nu_U(u).$$

As g_t lies in the interior of the convergence cone $\mathcal{C}$, we can write $g_t = a_t b_t$ where a_t is a perturbation lying in the interior of the positive Weyl chamber and b_t still lies in the interior of the convergence cone $\mathcal{C}$. Since $b_{-t} u b_t \in U$ and the measure μ on $U.z_0$ is left invariant ($z \mapsto b_{-t} u b_t z$), we have

$$\int_{U.z_0} \varphi(g_t z) d\mu(z) = \int_{U.z_0} \varphi(a_t b_t z) d\mu(z) = \int_{U.z_0} \int_U \xi(u) \varphi(a_t u b_t z) d\nu_U(u) d\mu(z).$$

Now we are in a position to estimate the above integral. To accomplish this we write $U.z_0$ as $U.z_0 = B_1 \cup B_2$, where $B_1 := \{z \in U.z_0 \ : \ b_t \cdot z \notin K\}$ consists of those z not returning to a properly chosen large compact subset K of G/Γ; and write

$$\int_{U.z_0} \int_U \xi(u) \varphi(a_t u b_t z) d\nu_U(u) d\mu(z) = I + II \tag{7}$$

where

$$I := \int_{B_1} \int_U \xi(u) \varphi(a_t u b_t z) d\nu_U(u) d\mu(z)$$

and

$$II := \int_{B_2} \int_U \xi(u) \varphi(a_t u b_t z) d\nu_U(u) d\mu(z).$$

To prove Theorem 2 it suffices to show that the integrals I and II in (7) are both small. To show that integral I is small, we will prove in Section 2.3 that the measure of B_1 is small. In other words, most of the points of $U.z_0$, translated by b_t, will return to the compact set K. As we will see, the return is guaranteed by the fact that b_t lies in the interior of the cone $\mathcal{C}$. To show that integral II is small, we will use a result of Kleinbock-Margulis [9] on the effective equidistribution of the full expanding horospherical orbits. Their result will be recalled in Section 2.2.

The proof of Theorem 4 is quite similar to the proof just outlined but there is a crucial difference.

Following the discussion above, but replacing $d\mu(z)$ by $f(z)d\mu(z)$, we come to a situation where we choose sets B_1 and B_2 (which now depend on the choice of f) and estimate the integrals I and II. It turns out that estimating I is manageable. However the estimate of II is based on effective equidistribution of the full expanding horospherical orbits of [9] (which is Proposition 1 below). After applying this result the Sobolev norm of $h \in H \mapsto f(b_{-t}hb_t z)$ makes an appearance where H is the horospherical subgroup appearing in Proposition 1 below. We control this Sobolev norm by choosing b_t so that H is completely expanded by conjugation with b_t. But it is not always possible to choose $b_t \in \mathcal{C}$ in this way while also choosing a_t to lie in $\mathcal{A}$. (§2 of [13] provides an example of such a flow g_t. See the remarks following Theorem 4 above.) This is why we are not presently able to prove Theorem 4 for all $a \in \mathcal{C}$. So the crucial difference in the proofs of Theorem 2 and Theorem 4: when f is a constant function (as in Theorem 2) there is no need to control its Sobolev norm! Without the need to control the Sobolev norm, the proof of Theorem 2 goes through without restricting the factorization $g_t = a_t b_t$.

2.2. **Effective equidistribution of expanding horospheres.** Fix a right-invariant metric 'dist' on G which gives rise to the corresponding metric on $SO(n)\backslash G$. The following result is essentially [9, Theorem 2.3].

Proposition 1. *Let $\{a_t : t > 0\}$ be a diagonal flow in G and H the full expanding horospherical subgroup of $\{a_t : t > 0\}$. Let $z \in G/\Gamma$, $f \in C^\infty_{comp}(H)$, and $0 < r < 1$ be such that the map $g \mapsto g.z$ is injective on $B_G(2r)\mathrm{supp}(f) \subset G$. Then for any $t > 0$ and any smooth function φ on G/Γ with $\int_{G/\Gamma} \varphi = 0$ one has that*

$$\left| \int_H f(h)\varphi(a_t hz)d\nu_H(h) \right| \ll r \cdot \|\varphi\|_{\mathrm{Lip}} \cdot \int_H |f| + r^{-k} \cdot \|f\|_\ell \cdot \|\varphi\|_\ell \cdot e^{-\gamma \mathrm{dist}(a_t, e)} \tag{8}$$

where $\gamma > 0$ is an absolute constant and $k, \ell \in \mathbb{Z}^+$, where $k > 2\ell$ depends on ℓ and $\dim(H)$, and the implied constant is absolute.

We shall not reproduce the proof of the above proposition since it is nearly identical to the proof of [9, Theorem 2.3]. They prove the above proposition for the special case $a_t = \mathrm{diag}(e^{t/m}, ..., e^{t/m}, e^{-t/n}, ..., e^{-t/n})$, but the general case follows easily.

2.3. **Quantitative non-divergence of unipotent flows.** For $\varepsilon > 0$ define

$$K_\varepsilon := \pi(\{g \in G : ||g\boldsymbol{v}|| \geq \varepsilon \quad \text{for all} \quad \boldsymbol{v} \in \mathbb{Z}^n \setminus \{0\}\}).$$

In other words, K_ε consists of unimodular lattices in $\mathbb{R}^n$ whose first minimum is at least ε. By Mahler's compactness criterion, K_ε is a compact subset of G/Γ. Kleinbock and Margulis proved in [8] that certain polynomial maps cannot escape K_ε except on a set of small measure. See Theorem 5.2 from [8]. This result was generalized in [1]. The following Theorem from [9] is a special case of Theorem 6.2 from [1].

Theorem 7 ([9, Theorem 3.1])**.** *Let $\phi : \mathbb{R}^d \to GL_n(\mathbb{R})$ be a map such that all coordinates are polynomial of degree not greater than l, and let B be a ball in $\mathbb{R}^d$ such that for any $k = 1, \ldots, n-1$ and any $\boldsymbol{v} \in \bigwedge^k(\mathbb{Z}^n)\backslash\{0\}$, $||\phi(x)\boldsymbol{v}|| \geq 1$ for some $x \in B$. Then for any positive $\varepsilon \leq 1$,*

$$\lambda(\{x \in B : \pi(\phi(x)) \notin K_\varepsilon\}) \ll \varepsilon^{\frac{1}{dl}} \lambda(B),$$

where λ is the Lebesgue measure on $\mathbb{R}^d$ and $|| \cdot ||$ is the Euclidean norm.

Let $d := \frac{n^2-n}{2}$, and let $\{X_1, \ldots, X_d\}$ be a basis for the Lie algebra $\mathfrak{u}$ of U. Define

$$\Theta : \mathbb{R}^d \to G \quad \text{by} \quad \Theta(s_1, \ldots, s_d) = \exp(s_1 X_1) \ldots \exp(s_d X_d).$$

Let $g_{\mathbf{t}} := \mathrm{diag}(e^{t_1}, \cdots, e^{t_n})$, and define $T_{\min} := \min_{1 \leq j < n} t_1 + \cdots + t_j$. We will apply Theorem 7 with $\phi : \mathbb{R}^d \to G$ defined by

$$\phi : s \mapsto g_{\mathbf{t}} \Theta(s) g$$

for a fixed $g_{\mathbf{t}} \in \mathcal{C}$ and $g \in G$. It is easy to see that this choice of ϕ satisfies the first condition of Theorem 7. We will use the next proposition to show that ϕ satisfies the second condition.

Proposition 2. *Let $\rho : G \to GL(V)$ be a representation on a finite-dimensional vector space V with no nonzero G-invariant vectors. Then there exist $\alpha > 0$ and $c_1 > 0$ such that for any $\boldsymbol{v} \in V$ and $g_{\mathbf{t}} \in \mathcal{C}$,*

$$\sup_{u \in B_U(r)} ||\rho(g_{\mathbf{t}} u)\boldsymbol{v}|| \geq c_1 e^{\alpha T_{\min}} ||\boldsymbol{v}||,$$

where c_1 depends on r, the representation, and choice of norm, and α depends on the representation.

Before we can prove Proposition 2, we need to prove the following representation-theoretic lemma. The proof is similar to [11], but here we need to consider more general diagonal elements: any $g_{\mathbf{t}} \in \mathcal{C}$.

Lemma 1. *Let (ρ, V) be a representation as in Proposition 2, and define*

$$V^U = \{\boldsymbol{v} \in V : u\boldsymbol{v} = \boldsymbol{v} \quad \text{for all} \quad u \in U\}.$$

Then there exist $\alpha > 0$ and $c_0 > 0$ such that for any $\boldsymbol{v} \in V^U$ and $g_{\mathbf{t}} \in \mathcal{C}$,

$$||\rho(g_{\mathbf{t}})\boldsymbol{v}|| \geq c_0 e^{\alpha T_{\min}} ||\boldsymbol{v}||,$$

where c_0 depends on the choice of norm and α depends on the representation.

Proof. Let A be the subgroup of positive diagonal matrices in G. Let $\mathfrak{a}$ and $\mathfrak{u}$ be the Lie algebras of A and U respectively. Note that A normalizes U, so V^U is a $\rho(A)$-invariant subspace. Then we can define the $\rho(A)$-equivariant projection $p : V \to V^U$, and we can write $V^U = \bigoplus_{\chi \in \Psi} V_\chi$, where Ψ is a finite set of weights and

$$V_\chi = \{\boldsymbol{v} \in V : \rho(\exp X)v = e^{\chi(X)}v \quad \text{for all} \quad X \in \mathfrak{a}\}.$$

Let $E_{i,j}$ be the $n \times n$ matrix with 1 in the ij^{th} entry and 0 otherwise, and define $F_{i,j} := E_{i,i} - E_{j,j}$. For $i = 1, \ldots, n-1$, define $G(i)$ to be the Lie subgroup of G whose Lie algebra is $\mathfrak{g}(i) := \langle E_{i,i+1}, E_{i+1,i}, F_{i,i+1}\rangle$. Note that each $G(i)$ is a copy of $\mathrm{SL}_2(\mathbb{R})$ in G. Also note that $\{\mathfrak{g}(i) : 1 \leq i < n\}$ generates $\mathfrak{g}$, the Lie algebra of G. Thus $\{G(i) : 1 \leq i < n\}$ generates G.

Every vector in V_χ is fixed by $\rho(u)$ for every $u \in U$, so in particular it is fixed by $\rho(\exp E_{i,i+1})$. Then by the representation theory of $\mathrm{SL}_2(\mathbb{R})$, $\chi(F_{i,i+1}) = m - 1$, where m is the dimension of the representation. Note that $\chi(F_{i,i+1}) = 0$ if and only if ρ is the trivial representation of $G(i)$ on V_χ. Since V contains no nonzero vectors fixed by G and G is generated by $\{G(i) : 1 \leq i < n\}$, there is some i such that $\chi(F_{i,i+1}) > 0$; call it i_0. Then

$$\begin{aligned} \chi(\mathrm{diag}(t_1, \ldots, t_n)) &= \chi(t_1 F_{1,2} + (t_1 + t_2)F_{2,3} + \cdots + (t_1 + \cdots + t_{n-1})F_{n,n-1}) \\ &\geq T_{\min}\chi(F_{i_0,i_0+1}). \end{aligned}$$

Thus for any $g_{\mathbf{t}} \in \mathcal{C}$ and any $\boldsymbol{v} \in V_\chi$, $||\rho(g_{\mathbf{t}})\boldsymbol{v}|| \geq e^{\chi(F_{i_0,i_0+1})T_{\min}} ||\boldsymbol{v}|| := e^{\alpha_0 T_{\min}} ||\boldsymbol{v}||$ where $\alpha_0 > 0$.

Without loss of generality, we may assume that $|| \cdot ||$ is the sup norm with respect to a basis of $\rho(A)$-eigenvectors. Then, for any $g_{\mathbf{t}} \in \mathcal{C}$ and any $\boldsymbol{v} \in V^U$,

$$||\rho(g_{\mathbf{t}})\boldsymbol{v}|| \geq c_0 e^{\alpha T_{\min}} ||\boldsymbol{v}||.$$

□

Now, combining Lemma 1 with [18, Lemma 5.1], we can prove the proposition.

Proof of Proposition 2. Let $p : V \to V^U$ be as in the proof of Lemma 1. Now

$$\begin{aligned}\sup_{u\in B_U(r)} ||\rho(g_{\mathbf{t}}u)\boldsymbol{v}|| &\geq \sup_{u\in B_U(r)} ||p(\rho(g_{\mathbf{t}}u)\boldsymbol{v})|| \\ &= \sup_{u\in B_U(r)} ||\rho(g_{\mathbf{t}})p(\rho(u)\boldsymbol{v}))|| \\ &\geq c_0 e^{\alpha T_{\min}} \sup_{u\in B_U(r)} ||p(\rho(u)\boldsymbol{v})|| && \text{by Lemma 1} \\ &\geq c_1 e^{\alpha T_{\min}} ||\boldsymbol{v}|| && \text{by [18, Lemma 5.1].}\end{aligned}$$

□

Let $D(\boldsymbol{\theta}) = \min_{\alpha\in\Delta} \lambda_\alpha(\boldsymbol{\theta})$.

Corollary 2. *Let $\boldsymbol{\theta}$ be on the unit sphere in $\mathcal{C}$ and $b_t = b_t^{\boldsymbol{\theta}} = \operatorname{diag}(e^{\theta_1 t}, \cdots, e^{\theta_n t})$. Then, for any compact subset L in G/Γ, there exists $\kappa = \kappa(n) > 0$ and a $T_1 = T_1(r, L, \boldsymbol{\theta}) \gg_{r,L} D(\boldsymbol{\theta})^{-1}$ such that for every $0 < \varepsilon < 1$, any $z \in L$, and any $t \geq T_1$,*

$$\nu_U(\{u \in B_U(r) : b_t uz \notin K_\varepsilon\}) \ll \varepsilon^\kappa \cdot \nu_U(B_U(r)).$$

Proof. By Proposition 2 applied to the irreducible representations of G on $\bigwedge^j(\mathbb{R}^n)$,

$$\sup_{u\in B_U(r)} ||b_t ug\boldsymbol{v}|| \geq c_1 e^{\alpha t} ||g\boldsymbol{v}||.$$

Then for any $g \in \pi^{-1}(L)$ and any $\boldsymbol{v} \in \bigwedge^j(\mathbb{Z}^n)\backslash\{0\}$,

$$\sup_{u\in B_U(r)} ||b_t ug\boldsymbol{v}|| \geq c_2 e^{\alpha t},$$

since L is compact and $\bigwedge^j(\mathbb{Z}^n)$ is discrete. Define T to be such that $c_2 e^{\alpha T} = 1$, and define $\phi(s) := b_t\Theta(s)g$, where $s \in \mathbb{R}^d$. Let O be a neighborhood of 0 in $\mathbb{R}^d$ such that $\Theta(O) = B_U(r)$. Then there is some $s \in O$ such that for any $t \geq T$, $||b_t\Theta(s)gv|| \geq 1$. Then by Theorem 7,

$$\lambda(\{s \in O : b_t\Theta(s)z \notin K_\varepsilon\}) \ll \varepsilon^{\frac{1}{d(n-1)}} \lambda(O).$$

Then, since λ and ν_U are absolutely continuous with respect to each other,

$$\nu_U(\{u \in B_U(r) : b_t uz \notin K_\varepsilon\}) \ll \varepsilon^{\frac{1}{d(n-1)}} \cdot \nu_U(B_U(r)).$$

It now remains to show the dependence of $T = T_1$ on r, L, and $\boldsymbol{\theta}$. Solving for T yields $T = -\alpha^{-1}\log(c_2)$. The constant c_2 depends on r and L and $\alpha = c\min_{\alpha\in\Delta} \lambda_\alpha(\boldsymbol{\theta}) = cD(\boldsymbol{\theta})$ for some $c > 0$ depending on the choice of representation coming from Proposition 2. □

Corollary 3. *Let $\boldsymbol{\theta}$ be on the unit sphere in $\mathcal{C}$ and $b_t = b_t^{\boldsymbol{\theta}} = \operatorname{diag}(e^{\theta_1 t}, \cdots, e^{\theta_n t})$. Then there exists $\kappa = \kappa(n) > 0$ and $T_2 = T_2(\boldsymbol{\theta}) \gg D(\boldsymbol{\theta})^{-1}$, such that for any $0 < \varepsilon < 1$ and any $t \geq T_2$,*

$$\mu(\{z \in U.z_0 : b_t z \notin K_\varepsilon\}) \ll \varepsilon^\kappa.$$

Proof of Corollary 3. Since $U.z_0$ is periodic, $U \cap \Gamma$ is a uniform lattice in U. Then there is a relatively compact fundamental domain, Ω, for $U/U\cap\Gamma$ in U. Cover each $u \in \Omega$ by $B_U(u, r)$ so that $\Omega \subseteq \bigcup_{u\in\Omega} B_U(u, r)$ and π is injective on $B_U(u, r)$. Let $O(\log u, r)$ be a ball in $\mathfrak{u}$ such that $\exp(O(\log u, r)) = B_U(u, r)$. By the Besicovitch covering theorem, there exists a constant c_d, depending only on the dimension d, such that

$$\Omega \subseteq \bigcup_{i=1}^{c_d} \bigcup_{B\in\mathcal{O}_i} B,$$

where each $\mathcal{O}_i$ is a collection of disjoint balls $B_U(u, r)$. By Corollary 2, for $t \gg_r D(\boldsymbol{\theta})^{-1}$

$$\nu_U(\{u \in B_U(u, r) : b_t uz \notin K_\varepsilon\}) \ll \varepsilon^{\frac{1}{dn}} \cdot \nu_U(B_U(u, r)).$$

Then we have

$$\begin{aligned}
\nu_U(\{u\in\Omega : b_tuz\notin K_\varepsilon\}) &\le \nu_U(\{u\in\bigcup_{i=1}^{c_d}\bigcup_{B\in O_i} B : b_tuz\notin K_\varepsilon\})\\
&\le \sum_{i=1}^{c_d}\sum_{B\in O_i}\nu_U(\{u\in B : b_tuz\notin K_\varepsilon\})\\
&\ll \sum_{i=1}^{c_d}\sum_{B\in O_i}\varepsilon^{\frac{1}{d(n-1)}}\cdot\nu_U(B)\\
&= \sum_{i=1}^{c_d}\varepsilon^{\frac{1}{d(n-1)}}\cdot\nu_U(\bigcup_{B\in O_i} B)\\
&\le c_d\cdot\varepsilon^{\frac{1}{d(n-1)}}\cdot\nu_U(\Omega).
\end{aligned}$$

Since Ω is a fundamental domain and μ is a probability measure,

$$\mu(\{z\in U.z_0 : b_tz\notin K_\varepsilon\})\ll\varepsilon^{\frac{1}{d(n-1)}}.$$

□

2.4. Proof of Theorem 2.

Proof. It suffices to prove the result for $a\in\mathcal{C}$ since $\min_{\alpha\in\Delta}\lambda_\alpha(a)\le 1$ if $a\notin\mathcal{C}$ and the theorem (in the case $a\notin\mathcal{C}$) would follow by taking the implied constant to be a multiple of $\sup|\varphi|$. So we may suppose $a\in\mathcal{C}$. Write $a=g_t=\exp(t\boldsymbol{\theta})$ where $t>0$, $\boldsymbol{\theta}\in\mathcal{C}$ (reverting notation back to the Lie algebra) is on the unit sphere of $\mathfrak{a}$. Then the term appearing on the right hand side of (2) can be written as

$$\left(\min_{\alpha\in\Delta}\lambda_\alpha(a)\right)^{-\delta}=e^{-\delta tD}$$

where $D=D(\boldsymbol{\theta})=\min_{\alpha\in\Delta}\lambda_\alpha(\boldsymbol{\theta})=\operatorname{dist}(\boldsymbol{\theta},\partial\mathcal{C})$, and where the λ_α's appearing on the left hand side are understood to be multiplicative (as they are in (2)).

The proof follows the outline in Section 2.1. Notice that it suffices to prove that Equation 1 is valid whenever $t\gg 1/D$ because if $t\ll 1/D$, then $e^{-cDt}\gg e^{-c}$ and the left hand side of Equation 1 is trivially bounded by $2\sup|\varphi|$. Therefore we assume that $t\gg D(\boldsymbol{\theta})^{-1}$ and that φ has mean zero.

By [9, Lemma 2.2] there exists a smooth function ξ on U, whose support is contained in $B_U(r)$, satisfying $\xi\ge 0$, $\int_U\xi=1$, and $\|\xi\|_\ell\ll r^{-(k-\ell)}$. Note that a suitable r will be chosen later.

Write $g_t=a_tb_t$, where a_t lies in the interior of the positive Weyl chamber $\mathcal{A}$ and b_t lies in the interior of the convergence cone $\mathcal{C}$. Then

$$\int_{U.z_0}\varphi(g_tz)d\mu(z)=\int_{U.z_0}\varphi(a_tb_tz)d\mu(z)=\int_{U.z_0}\int_U\xi(u)\varphi(a_tub_tz)d\nu_U(u)d\mu(z)$$

where z_0 is the identity coset. To estimate the above integral, we partition $U.z_0$ as $U.z_0=B_1\cup B_2$ and write

$$\int_{U.z_0}\int_U\xi(u)\varphi(a_tub_tz)d\nu_U(u)d\mu(z)=I+II,$$

where

$$I:=\int_{B_1}\int_U\xi(u)\varphi(a_tub_tz)d\nu_U(u)d\mu(z)\quad\text{and}\quad II:=\int_{B_2}\int_U\xi(u)\varphi(a_tub_tz)d\nu_U(u)d\mu(z).$$

Let $\epsilon=e^{-\beta t}$, where β will be chosen later, and set

$$B_1:=\{z\in U.z_0\ :\ b_t\cdot z\notin K_\epsilon\}.$$

With this choice for B_1 we have, by Corollary 3,

$$|I| \le \sup|\varphi|\mu(B_1)\int_U \xi(u)d\mu(u) \le \sup|\varphi|\epsilon^{\kappa_1}.$$

To bound integral II, we define B_2 to be the complement of B_1 in $U.z_0$. Now we trivially have

$$|II| \le \int_{B_2}\left|\int_U \xi(u)\varphi(a_t u b_t z)d\nu_U(u)\right| d\mu(z).$$

Since a_t is in the interior of the positive Weyl chamber and U is the full expanding horospherical subgroup of a_t, we may apply Proposition 1, with $f = \xi$ and $H = U$, to estimate the innermost integral.

In order to satisfy the hypotheses of this proposition, we select r so that $z \mapsto g.z$ is injective on $B_G(2r)B_U(r) \subset G$ for each $z \in B_2$. The injectivity radius of a set $L \subset X_n$ is defined by

$$r(L) := \inf_{z\in L}\sup\{r > 0 \ : \ z \mapsto g.z \text{ is injective on } B_G(r)\}.$$

By [9, Proposition 3.5] the injectivity radius of K_ϵ satisfies $r(K_\epsilon) \ge c\epsilon^n$ for some $c > 0$. It follows from the definition of the injectivity radius that $r(B_2) \ge r(K_\epsilon)$ since $B_2 \subset K_\epsilon$. Therefore, if we take $r = c\epsilon^n/3 = (c/3)e^{-n\beta t}$, then $z \mapsto g.z$ is injective on $B_G(2r)B_U(r) \subset G$ for each $z \in B_2$. Therefore, by Proposition 1 and the assumptions on ξ, we have

$$\begin{aligned} |II| &\ll \mu(B)\left(r\cdot\|\varphi\|_{\text{Lip}}\cdot\int_U|\xi| + r^{-k}r^{-(k-\ell)}\cdot\|\varphi\|_\ell\cdot e^{-\gamma\text{dist}(a_t,e)}\right) \\ &\ll (c/3)e^{-n\beta t}\|\varphi\|_{\text{Lip}} + \cdot\|\varphi\|_\ell\cdot e^{-\gamma\text{dist}(a_t,e)}e^{\beta t(2k-\ell)n}. \end{aligned}$$

There is a number $\eta = \eta(a_1) > 0$ such that $e^{-\gamma\text{dist}(a_t,e)} \le e^{-\gamma\eta t}$, and so we may write

$$|I| + |II| \le C\left(e^{-\beta\kappa t} + e^{-\beta t n} + e^{(\beta(2k-\ell)n-\gamma\eta)t}\right)$$

where C depends only on φ and k. Note that $\kappa < 1$ and so we can choose β to equalize the exponents and we see that

$$\beta = \frac{\gamma\eta}{(2k-\ell)-\kappa}.$$

The only term above which depends on the flow is η. Recall that $D(\boldsymbol{\theta}) = \min_{\alpha\in\Delta}\lambda_\alpha((\theta)) = \lambda_\beta(\boldsymbol{\theta})$ is equal to $\text{dist}(\boldsymbol{\theta}, \partial\mathcal{C})$. Therefore we can choose $a_1 \in \mathcal{A}$ very close to a multiple of λ_β. That is, we can always choose the factorization $g_t = a_t b_t$ so that a_t is close to $\mathbb{R}\lambda_\beta$ with magnitude approximately $\text{dist}(\boldsymbol{\theta}, \partial\mathcal{C})$. Therefore the factorization can be chosen so that $\eta = \tilde{c}\text{dist}(\boldsymbol{\theta}, \partial\mathcal{C})$ for some $\tilde{c} > 0$ and the constant δ appearing in the statement of the theorem can be taken to be

$$\delta = \frac{\tilde{c}\gamma}{(2k-\ell)-\kappa} > 0.$$

Therefore

$$|I| + |II| \ll e^{-\delta t D}.$$

This proves the result. □

2.5. **Proof of Theorem 3.** *To simplify notation and to keep this section brief, we will prove the case $E = \emptyset$ as the general case is similar.* Our proof of Theorem 3 is basically an induction argument using Theorem 2 as the base case. To describe the basic idea behind the proof let us first give an explicit description of the groups Q_F. A subset $F \subset \Delta$ can be described as

$$F = \{\alpha_{i_1}, ..., \alpha_{i_\ell}\} \subset \Delta.$$

We will find it more convenient to work with the complement $\mathcal{F} = \Delta \setminus F$ of F in Δ rather than with F itself. Finally we can describe $Q_{\mathcal{F}}$ in matrix form as

$$Q_{\mathcal{F}} = \begin{pmatrix} \mathrm{SL}_{k_1}(\mathbb{R}) & * & \dots & * \\ & \mathrm{SL}_{k_2}(\mathbb{R}) & & \vdots \\ & & \ddots & * \\ & & & \mathrm{SL}_{k_{\ell+1}}(\mathbb{R}) \end{pmatrix}$$

where $k_1 = i_1$, $k_2 = i_2 - i_1$,..., $k_\ell = i_\ell - i_{\ell-1}$, $k_{\ell+1} = n - (k_1 + \cdots + k_\ell)$. Notice that if $F = \emptyset$, then $Q_{\mathcal{F}} = Q_\Delta = \mathrm{SL}_n(\mathbb{R})$.

Now we are in a position in which we can outline the basic idea behind the proof. By the same reasoning as in the beginning of the proof of Theorem 2, we may suppose that $a \in \mathcal{C}$. Let $\nu_{\mathcal{F}} = \nu_{Q_{\mathcal{F}}}$ denote the Haar measure which is equal to $\mu_{\mathcal{F}}$ in G/Γ when restricted to a fundamental domain of Γ. Then $\nu_{\mathcal{F}}$ decomposes into a product measure according to

$$\nu_{\mathcal{F}} = \nu_{\mathrm{SL}_{k_1}(\mathbb{R})} \otimes \cdots \otimes \nu_{\mathrm{SL}_{k_{\ell+1}}(\mathbb{R})} \otimes \nu_{W_{\mathcal{F}}}$$

where

$$W_{\mathcal{F}} = \begin{pmatrix} I_{k_1} & * & \dots & * \\ & I_{k_2} & & \vdots \\ & & \ddots & * \\ & & & I_{k_{\ell+1}} \end{pmatrix}.$$

The proof then proceeds by applying Theorem 2 to each SL block on the diagonal. Of course we must deal with the translates of the factor $W_{\mathcal{F}}$, but a Jacobian argument shows that the measure is invariant.

To begin, we observe that the group U can be written as

$$U = \begin{pmatrix} U_{k_1} & * & \dots & * \\ & U_{k_2} & & \vdots \\ & & \ddots & * \\ & & & U_{k_{\ell+1}} \end{pmatrix}$$

where U_m is the group of $m \times m$ unipotent upper triangular matrices. The Haar measure ν_U evidently admits the factorization

$$\nu_U = \nu_{U_{k_1}} \otimes \cdots \otimes \nu_{U_{k_{\ell+1}}} \otimes \nu_{W_{\mathcal{F}}}.$$

The corresponding factorization for $a = g_t = \exp(t\boldsymbol{\theta})$ is given by

$$a = g_t = g_t^{(1)} \otimes \cdots \otimes g_t^{(\ell+1)} \tag{9}$$

where $g_t^{(j)}$ is the corresponding block of length k_j in g_t. In this way we see that

$$g_t \nu_U = g_t^{(1)} \nu_{U_{k_1}} \otimes \cdots \otimes g_t^{(\ell+1)} \nu_{U_{k_{\ell+1}}} \otimes g_t \nu_{W_{\mathcal{F}}}.$$

To prove Theorem 3 we will use the following two lemmas.

Lemma 2. *If $g_t \in \mathcal{C}_{\mathcal{F}}$, then $g_t \nu_{W_{\mathcal{F}}} = \nu_{W_{\mathcal{F}}}$.*

Proof. We can write $\nu_{W_{\mathcal{F}}} = \nu_{M(k_1, n-k_1)} \otimes \cdots \otimes \nu_{M(n-k_{\ell+1}, k_{\ell+1})}$ where $M(r,s) = M_{r\times s}(\mathbb{R})$ is the space of $r \times s$ matrices, hence

$$g_t \nu_{W_{\mathcal{F}}} = g_t^{(1)} \nu_{M(k_1, n-k_1)} \otimes \cdots \otimes g_t^{(\ell+1)} \nu_{M(n-k_{\ell+1}, k_{\ell+1})}$$

and $g_t^{(j)}\nu_{M(k_j,n-k_1-\cdots-k_{j-1})} = \mathrm{Jac}(g_t)\nu_{M(k_j,n-k_1-\cdots-k_{j-1})}$. But $g_t^{(j)} = (c_1(t),\dots,c_{k_j}(t))$ acts by dilating the i^{th} row by $c_i(t)$, and so $\mathrm{Jac}(g_t^{(j)}) = \prod_{i=1}^{k_j} c_i(t)^{n-k_1-\cdots-k_{j-1}} = 1$. So

$$g_t^{(j)}\nu_{M(k_j,n-k_1-\cdots-k_{j-1})} = \nu_{M(k_j,n-k_1-\cdots-k_{j-1})}$$

and the lemma follows. □

Lemma 3. *If $\nu_t^{(1)}, \nu_t^{(1)}$ are probability measures converging to $\nu^{(1)}, \nu^{(2)}$ effectively as*

$$|\nu_t^{(i)}(f_i) - \nu^{(i)}(f_i)| \ll e^{-\gamma_i t},$$

where the implied constant depends only on $\sup|f_i|$, $\|f_i\|_{C^k}$ and $\|f_i\|_{Lip}$, then the measure $\nu_t^{(1)} \otimes \nu_t^{(2)}$ converges to $\nu^{(1)} \otimes \nu^{(2)}$ effectively as

$$|\nu_t^{(1)} \otimes \nu_t^{(2)}(F) - \nu^{(1)} \otimes \nu^{(2)}(F)| \ll \max_{i=1,2}\left\{e^{-\gamma_i t}\right\},$$

where the implied constant may depend on $\sup|F|$, $\|F\|_{C^k}$, $\|F\|_{Lip}$, and the measure of the support of F.

Proof. This argument is a standard application of the triangle inequality. Observe

$$\begin{aligned}
|\nu_t^{(1)} \otimes \nu_t^{(2)}(F) - \nu^{(1)} \otimes \nu^{(2)}(F)| &\le |\nu_t^{(1)} \otimes \nu_t^{(2)}(F) - \nu_t^{(1)} \otimes \nu^{(2)}(F)| \\
&\quad + |\nu_t^{(1)} \otimes \nu^{(2)}(F) - \nu^{(1)} \otimes \nu^{(2)}(F)| \\
&\le \int_{X_1} \left|\nu_t^{(2)}(F(x_1,\cdot)) - \nu^{(2)}(F(x_1,\cdot))\right| d\nu_t^{(1)}(x_1) \\
&\quad + \int_{X_2} \left|\nu_t^{(1)}(F(\cdot,x_2)) - \nu^{(1)}(F(\cdot,x_2))\right| d\nu^{(2)}(x_2) \\
&\ll e^{-\gamma_1 t} + e^{-\gamma_2 t}.
\end{aligned}$$

□

Now we can finish off the proof of Theorem 3.

Proof of Theorem 3. Let $f \in C^\infty_{comp}(G/\Gamma)$. Then by Lemma 2

$$g_t\mu = g_t^{(1)}\nu_{U_{k_1}/\Gamma\cap U_{k_1}} \otimes \cdots \otimes g_t^{(\ell+1)}\nu_{U_{k_{\ell+1}}/\Gamma\cap U_{k_{\ell+1}}} \otimes \nu_{W_\mathcal{F}/\Gamma}.$$

But by Theorem 2 for each j there exists a constant $c = c_n, D_j = D_j(\boldsymbol{\theta}) > 0$ such that

$$\left|g_t^{(j)}\nu_{U_{k_j}/\Gamma\cap U_{k_j}}(f) - \nu_{\mathrm{SL}_{k_j}(\mathbb{R})/\mathrm{SL}_{k_j}(\mathbb{Z})}(f)\right| \ll E(f)e^{-cD_jt} \ll \tilde{E}(f)e^{-cD_jt}$$

where $E(f) = \max\{\|f\|_\ell, \|f\|_{Lip}\}$ and $\tilde{E}(f) = \max\{m(\mathrm{supp}(f))\|f\|_{C^\ell}, \|f\|_{Lip}\}$. By inductively applying Lemma 3 we obtain

$$\left|\nu_{W_\mathcal{F}/\Gamma} \otimes \prod_j g_t^{(j)}\nu_{U_{k_j}/\Gamma}(f) - \nu_{W_\mathcal{F}/\Gamma} \otimes \prod_j \nu_{\mathrm{SL}_{k_j}(\mathbb{R})/\mathrm{SL}_{k_j}(\mathbb{Z})}(f)\right| \ll \tilde{E}(f)e^{-\min_j cD_jt}.$$

Therefore there exists a constant $c = c_n > 0$ such that

$$|g_t\mu(f) - \nu_\mathcal{F}(f)| \ll e^{-c(\min_j D_j)t}.$$

It remains to show that $\min_j D_j = \min_{\alpha\in\mathcal{F}^c} \lambda_\alpha(\boldsymbol{\theta})$ (recall we are working with the complement of F in Δ). To see this, observe that with our choice of Δ the fundamental weights are given by

$$\lambda_{\alpha_j}(\boldsymbol{\theta}) = \theta_1 + \cdots + \theta_j \tag{10}$$

and since $\boldsymbol{\theta} \in \mathcal{C}_\mathcal{F}$

$$\lambda_{\alpha_j}(\boldsymbol{\theta}) = 0$$

if and only if $j = k_i$ for $i = 1, ..., \ell$. Recall the decomposition (9) and notice that if $i_s < r < i_{s+1}$, then by (10)

$$\begin{aligned} \lambda_{\alpha_r}(\log(g_t)) &= t(\lambda_{\alpha_{i_s}}(\boldsymbol{\theta}) + (\theta_{i_s+1} + \cdots + \theta_r)) \\ &= t(\theta_{i_s+1} + \cdots + \theta_r) \\ &= \tilde{\lambda}_{\beta_r}(\log(g_t^{(s+1)})) \end{aligned}$$

for some fundamental weight $\tilde{\lambda}_{\beta_r}$ of $\mathrm{SL}_{k_{s+1}}(\mathbb{R})$. In particular

$$D_s = \min_{i_s < r < i_{s+1}} \theta_{i_s+1} + \cdots + \theta_r$$

and, converting back to multiplicative notation (as at the end of the proof of Theorem 2) we have the desired result. □

2.6. **Proof of Theorem 4.** In the following proof we use the fact that for each $a \in \mathcal{C}_j$, the horospherical subgroup corresponding to a is

$$H_j = \left\{ \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \; : \; A \in M_{j \times (n-j)} \right\}.$$

Proof. Without loss of generality we may assume that φ has mean zero. By [9, Lemma 2.2] there is a function $\xi \in C^{\infty}_{comp}(H_j)$ that can be chosen so that $\|\xi\|_\ell \ll r^{-(k-\ell)}$, $\xi \geq 0$, $\int_{H_j} \xi = 1$, and the support contained in $B_{H_j}(r)$ where $r = e^{-\beta t}$ with β is to be chosen later. Arguing as in the beginning of the proof of Theorem 2 we may suppose that $a \in \mathcal{C}$. Write $a = g_t = \exp(t\boldsymbol{\theta}) = a_t b_t$ where $t > 0$, $\boldsymbol{\theta} \in \mathcal{C}_j$ is on the unit sphere of $\mathfrak{a}$, $t^{-1} \log a_t$ is a multiple of

$$(1/j, ..., 1/j, -1/(n-j), ..., -1/(n-j)),$$

and $b_t \in \mathcal{C}_j$. Note that the action of b_t on H_j (the horospherical subgroup of a_t) is non-contracting. Clearly

$$\int_U f(u)\varphi(g_t u z) d\nu_U(u) = \int_{H_j} \int_U \xi(h) f(u) \varphi(a_t b_t u z) d\nu_U(u) d\nu_{H_j}(h)$$

for each $z \in L$. Notice that the horospherical subgroup H_j is contained in U. Using the change of variables $u \mapsto b_{-t} h b_t u$ and the left invariance of the measure ν_U we get that

$$\int_U f(u)\varphi(g_t u z) d\nu_U(u) = \int_U \int_{H_j} f(b_{-t} h b_t u) \xi(h) \varphi(a_t h b_t u z) d\nu_{H_j}(h) d\nu_U(u).$$

Now we are in a position to estimate the above integral. To accomplish this break the integral into 2 pieces. We have

$$\mathrm{dist}(e, b_{-t} h b_t) \leq e^{-\rho t} \mathrm{dist}(e, h)$$

for any $h \in H_j$ where ρ depends only on g_t. Notice the supports of the functions

$$u \mapsto f_u(h) = f(b_{-t} h b_t u)$$

are contained in $B = \mathrm{supp}(f) B_U(e^{-(\rho+\beta)t})$. Suppose $t > T_1 > 0$ is taken large enough so that $r = e^{-\beta t} < r_0/2$ and $\mu(B) \leq 2\mu(\mathrm{supp}(f))$, where T_1 is from Corollary 3 and r_0 is the injectivity radius of L. Let $\epsilon = (2/c)^{1/n} e^{-\beta t/n}$ and define

$$\Omega = \{u \in U \; : \; b_t u z \notin K_\epsilon\} \quad \text{and} \quad \Phi = B - \Omega.$$

Then we write

$$\int_U \int_{H_j} f(b_{-t} h b_t u) \xi(h) \varphi(a_t h b_t u z) d\nu_{H_j}(h) d\nu_U(u) = I + II$$

where

$$I = \int_\Omega \int_{H_j} f(b_{-t} h b_t u) \xi(h) \varphi(a_t h b_t u z) d\nu_H(h) d\nu_U(u)$$

and

$$II=\int_{\Phi}\int_{H_j} f(b_{-t}hb_tu)\xi(h)\varphi(a_thb_tuz)d\nu_H(h)d\nu_U(u).$$

For I we have by Corollary 3 and the assumptions on t

$$|I|\le \mu(\Omega)\sup|f|\sup|\varphi|\int_{H_j}\xi(h)d\nu_{H_j}(h)\ll \epsilon^{\kappa_2}2\mu(\mathrm{supp}(f))\sup|f|\sup|\varphi|$$
$$\ll_{\varphi,f,n} e^{-\beta\kappa_2 t/n}.$$

For II we have by Proposition 1 to obtain

$$|II|\le \mu(\Phi)\left(r\cdot\|\varphi\|_{\mathrm{Lip}}\cdot\int_{H_j}|f_u|+r^{-k}\cdot\|f_u\|_\ell\cdot\|\varphi\|_\ell\cdot e^{-\gamma\mathrm{dist}(a_t,e)}\right).$$

As a compactly supported smooth function on H, the Sobelev norm of $f_u(h)$ is controlled by the norms of f and ξ because the conjugation by b_{-t} on H_j is non-expanding and hence the derivatives coming from f do not increase. In particular (by [9, Lemma 2.2])

$$\|f_u\|_\ell \ll_f r^{-(\ell+j(n-j)/2)}$$

so we find that for some $\eta=\eta(\boldsymbol{\theta})>0$ we have

$$|II|\ll_{f,\varphi} e^{-\beta t}+e^{-(\gamma\eta-(k+\ell+j(n-j)/2)\beta)t}.$$

Now we find that

$$\begin{aligned}|I|+|II| &\ll_{f,\varphi} e^{-\beta\kappa_2 t/n}+e^{-\beta t}+e^{-(\gamma\eta-(k+\ell+j(n-j)/2)\beta)t}\\ &\ll_{f,\varphi} e^{-\beta\min\{\kappa_2/n,1\}t}+e^{-(\gamma\eta-(k+\ell+j(n-j)/2)\beta)t}.\end{aligned}$$

We now choose

$$\beta=\frac{\gamma\eta n}{\min\{\kappa_2/n,1\}+k+\ell+j(n-j)/2}.$$

To finish the proof we argue as in the end of the proof of Theorem 2 to obtain the desired result. □

3. Counting Lifts of Horospeheres

Let $d(\cdot,\cdot)$ be a right G-invariant and $K-$bi-invariant Riemannian metric on X that is induced by the Killing form and satisfies

$$d(x_0,x_0a)=\|\log(a)\|$$

for each $a\in A$, where $\|\cdot\|$ is the Euclidean norm on $\mathfrak{a}$. Let $G=KAU$ be the usual Iwasawa decomposition of G and let $B(x,R)$ be the ball of radius R centered at $x\in X$ with respect to d.

3.1. **Proof of Theorem 5.** The proof of Theorem 5 is based on a counting argument that utilizes our main equidistribution theorem. To indicate the basic approach we must introduce a few technical facts. Firstly, it was proven in [13] that $\overline{B}_R=\{gU\ :\ d(x_0,x_0g)\le R\}$ admits the decomposition

$$\overline{B}_R=KA_RU/U \tag{11}$$

where $A_R=A\cap\tilde{B}(R)$ and $\tilde{B}_R=\{g\in G: d(x_0,x_0g)\le R\}$.

Let $\mathcal{U}=K\backslash Kg_0U$ and $\overline{\mathcal{U}}=\pi(\mathcal{U})$ where $\pi: X\to X/\Gamma$ is the natural projection. We observe that there exists an $a_0\in A$ such that $Kg_0U=Ka_0U$. This is guaranteed by the Iwasawa decomposition. In Theorem 5 we are interested in the asymptotic behavior of the function

$$N(R)=\#\{\gamma\in\Gamma\ :\ \mathcal{U}\gamma\cap B(x_0,R)\ne\emptyset\}.$$

The function

$$F_R(g\Gamma)=\sum_{\gamma\in\Gamma/\Gamma\cap U}1_R(g\gamma a_0U),$$

where $1_R(x)$ is the characteristic function of $\overline{B}_R$, satisfies $F_R(\Gamma) = N(R)$. Let

$$f(R) = \int_{A_R \cap \mathcal{C}} \rho'_\Delta(a) da,$$

where $\rho'_\Delta(a) = \exp(\langle \rho_\Delta, \log(a)\rangle)$ and ρ_Δ is the sum of the positive roots. We will also need the following proposition showing the existence of certain approximate identities, also known as mollifiers.

Proposition 3. *For each $\epsilon > 0$ with $\epsilon \ll 1$ there exists a function $\Psi_\epsilon \in C^\infty_{comp}(G/\Gamma)$ such that (1) Ψ_ϵ is supported in a ball of radius ϵ centered at Γ, (2) Ψ_ϵ is non-negative with integral equal to 1, (3) $\sup \Psi_\epsilon \ll \epsilon^{-dim(G)}$, (4) $\|\Psi_\epsilon\|_{Lip} \ll \epsilon^{1-\dim(G)}$, and (5) for each $\ell \geq 1$ we have $\|\Psi_\epsilon\|_\ell \ll \epsilon^{-(\ell+\dim(G))}$, and $\|\Psi_\epsilon\|_{C^\ell} \ll \epsilon^{-(\ell+\dim(G))}$.*

Proof. It suffices to prove the result in $\mathbb{R}^N$ for $\epsilon \ll 1$ where $N = \dim(G)$. Let $\Psi \in C^\infty_{comp}(\mathbb{R}^N)$ be non-negative with integral equal to 1 that is supported in the ball of radius 1 centered at the origin. Define

$$\Psi_\epsilon(\boldsymbol{x}) = \epsilon^{-N}\Psi(\epsilon^{-1}\boldsymbol{x}).$$

Then by construction Ψ_ϵ satisfies (1)-(3). Item (4) also follows from the definition:

$$\|\Psi_\epsilon\|_{Lip} = \sup_{\|\boldsymbol{x}\|,\|\boldsymbol{y}\|<\epsilon} \frac{|\Psi_\epsilon(\boldsymbol{x}) - \Psi_\epsilon(\boldsymbol{y})|}{\|\boldsymbol{x}-\boldsymbol{y}\|} = \sup_{\|\boldsymbol{x}\|,\|\boldsymbol{y}\|<1} \frac{\epsilon^{-N}|\Psi(\boldsymbol{x}) - \Psi(\boldsymbol{y})|}{\epsilon^{-1}\|\boldsymbol{x}-\boldsymbol{y}\|} = \epsilon^{1-N}\|\Psi\|_{Lip}.$$

Similarly item (5) follows from the observation that if $D = d/dx_{i_1}\cdots d/dx_{i_\ell}$, then $\|D\Psi_\epsilon\|_{L^p} = \epsilon^{-\ell}\|D\Psi\|_{L^p}$ when $0 < p < \infty$, which follows by a change of variables, and $\|D\Psi_\epsilon\|_{L^\infty} = \epsilon^{-(N+\ell)}\|D\Psi\|_{L^\infty}$. Here $\|\cdot\|_{L^p}$ is the usual $L^p(\mathbb{R}^N)$ norm. ☐

Our proof of Theorem 5 is based on the following two lemmas. The first lemma, Lemma 4, is a direct consequence of Theorem 2, and the second lemma, Lemma 5, is essentially a corollary of Lemma 4. We will prove the lemmas after we prove Theorem 5.

Lemma 4. *There exists $\delta > 0$ such that for every $\Psi \in C^\infty_{comp}(G/\Gamma)$*

$$|\langle F_R, \Psi\rangle - \kappa f(R)\langle 1, \Psi\rangle| \ll E(\Psi)e^{(\|\rho_\Delta\|-\delta)R}, \tag{12}$$

where

$$\kappa = c_n \mathrm{vol}(K)\frac{\mathrm{vol}(\overline{\mathcal{U}})}{\mathrm{vol}(\mathcal{M})}$$

for some constant $c_n > 0$ that only depends on n and $E(\Psi) = \max\{\|\Psi\|_{Lip}, \|\Psi\|_\ell, \sup|\Psi|\}$.

Lemma 5. *Suppose $\Psi_\epsilon \in C^\infty_c(G/\Gamma)$ is the approximate identity supported in $\tilde{B}_\epsilon$ given in Proposition 3. If $\epsilon = e^{-cR}$ for some $c > 0$, then*

$$|F_R(\Gamma) - \langle F_R, \Psi_\epsilon\rangle| \ll R^{n-2}e^{(\|\rho_\Delta\|-c)R} + E(\Psi_\epsilon)e^{(\|\rho_\Delta\|-\delta)R}. \tag{13}$$

where $E(\Psi_\epsilon) = \max\{\|\Psi_\epsilon\|_{Lip}, \|\Psi_\epsilon\|_\ell, \sup|\Psi_\epsilon|\}$.

Proof of Theorem 5. Assume Lemma 4 and Lemma 5. Take Ψ_ϵ to be the approximate identity given by Proposition 3 supported in the ball of radius $\epsilon = e^{-cR}$ for some $c > 0$ to be chosen later. Then it is immediate that

$$\max\{\|\Psi_\epsilon\|_{Lip}, \|\Psi_\epsilon\|_\ell, \sup|\Psi_\epsilon|\} \ll \epsilon^{-(N+\ell)} = e^{c(N+\ell)R}.$$

Then for some $u, \delta > 0$

$$\begin{aligned} |N(R) - \text{main term}| &= |F_R(\Gamma) - \text{main term}| \\ &\leq |F_R(\Gamma) - \langle F_R, \Psi_R\rangle| + |\langle F_R, \Psi_R\rangle - \text{main term}| \\ &\ll R^u e^{(\|\rho_\Delta\|-c)R} + e^{c(N+\ell)R}e^{(\|\rho_\Delta\|-\delta)R}. \end{aligned}$$

We now select $0 < c < \delta(N + \ell + 1)^{-1}$ to obtain

$$|N(R) - \text{main term}| \ll e^{(\|\rho_\Delta\| - c)R}.$$

To finish the proof we observe that for each $s > 0$ we have $e^{(\|\rho_\Delta\| - s)R} \ll_s \text{vol}(B(x_0, R))$. In particular there is a number $0 < q < 1$ such that

$$e^{(\|\rho_\Delta\| - c)R} \ll_c (\text{vol}(B(x_0, R)))^{1-q}$$

This can be seen by choosing q such that $(\|\rho_\Delta\| - c) < \|\rho_\Delta\|(1 - q)$. Plainly we may take any $0 < q < c\|\rho_\Delta\|^{-1}$ and conclude that

$$|N(R) - \text{main term}| \ll (\text{vol}(B(x_0, R)))^{1-q}.$$

$\square$

Proof of Lemma 4. Note that ρ_Δ is twice the sum of the positive root and so it is in the positive Weyl chamber. From the proof of Lemma 30 from [13] we find that

$$\langle F_R, \Psi \rangle = \nu(\pi(U)) \int_K \int_{A_R} \int_{G/\Gamma} \overline{\Psi(g'\Gamma)} d(ka\mu_U)(g')\rho'_\Delta(a)dadk$$

where $A_R = A \cap \tilde{B}_R$, and ρ'_Δ is given in [13]. For simplicity let us assume that A_R is centered at the identity. Choose $\delta_1 > 0$ such that

$$Y = \{w \in \mathfrak{a} : \langle w, \rho_\Delta \rangle > (1 - \delta_1)\|\rho_\Delta\|\|w\|\}$$

is contained in $\mathcal{A}$ and write

$$A_R = \Omega_1 \cup \Omega_2$$

where $\Omega_1 = A_R \cap Y$ and Ω_2 is the complement of Ω_1 in A_R. We will decompose the Haar measure da on A in "polar coordinates" as

$$da = d\exp(r\boldsymbol{\theta}) = r^{n-2}drd\sigma(\boldsymbol{\theta}) \tag{14}$$

where $\boldsymbol{\theta}$ is an element of the unit sphere of $\mathfrak{a}$, $r > 0$, dr is a multiple of the Lebesgue measure on $\mathbb{R}$, and $d\sigma$ is the measure on the unit sphere S^{n-2} inherited from the Lebesgue measure. Observe that for each $k \in K$

$$\begin{aligned}
&\left| \int_{\Omega_1} \int_{G/\Gamma} \overline{\Psi(g'\Gamma)} d(ka\mu_U)(g')\rho'_\Delta(a)da - \int_{\Omega_1} \int_{G/\Gamma} \overline{\Psi(g'\Gamma)} dm(g')\rho'_\Delta(a)da \right| \\
\leq \;& \int_{\Omega_1} \left| \int_{G/\Gamma} \overline{\Psi(g'\Gamma)} d(ka\mu_U)(g') - \int_{G/\Gamma} \overline{\Psi(g'\Gamma)} dm(g') \right| \rho'_\Delta(a)da \\
= \;& \int_{\mathscr{S}} \int_0^R \left| \int_{G/\Gamma} \overline{\Psi(g'\Gamma)} d(k\exp(r\boldsymbol{\theta}))\mu_U)(g') - \int_{G/\Gamma} \overline{\Psi(g'\Gamma)} dm(g') \right| \rho'_\Delta(\exp(r\boldsymbol{\theta}))d\sigma(\boldsymbol{\theta}) \\
\ll \;& C \int_{\mathscr{S}} \int_0^R e^{-cD(\boldsymbol{\theta})r} \rho'_\Delta(\exp(r\boldsymbol{\theta}))r^{n-1}drd\sigma(\boldsymbol{\theta}) \\
\ll \;& C \int_{\mathscr{S}} \int_0^R e^{-crD(\boldsymbol{\theta})} e^{r\langle \rho_\Delta, \boldsymbol{\theta} \rangle} r^{n-2}drd\sigma(\boldsymbol{\theta}) \\
\ll \;& CR^{n-1}e^{R(\|\rho_\Delta\| - \alpha_s)},
\end{aligned}$$

where we have applied Theorem 2 on the third to last line and $C = C(\Psi, n)$ is the constant appearing in the statement of Theorem 2. Here $\mathscr{S}$ is the intersection of the unit sphere and Ω_1, and $\alpha_s = \min D(\boldsymbol{\theta}) > 0$ where the minimum is over all $\boldsymbol{\theta} \in \mathscr{S}$ and $D(\boldsymbol{\theta}) = \text{dist}(\boldsymbol{\theta}, \partial\mathcal{C})$ as in the proof of Theorem 2.

Replacing Ω_1 with Ω_2 above we have upon a trivial estimation

$$\Bigg|\int_{\Omega_1}\int_{G/\Gamma} \overline{\Psi(g'\Gamma)}d(ka\mu_U)(g')\rho'_\Delta(a)da - \int_{\Omega_1}\int_{G/\Gamma} \overline{\Psi(g'\Gamma)}dm(g')\rho'_\Delta(a)da\Bigg| \ll (\sup|\Psi|)R^{n-1}e^{R\|\rho_\Delta\|(1-\delta_1)}.$$

Following the proof of Theorem 2 it is easily seen that $C \ll E(\Psi)$ where the implied constant depends only on n. Therefore we have

$$\left|\langle F_R,\Psi\rangle - \mathrm{vol}_H(K)\nu(\pi(U))\int_{A_Ra_0}\rho'_\Delta(a)da\,\langle 1,\Psi\rangle\right| \ll \mathrm{vol}(K)E(\Psi)R^{n-1}e^{R(\|\rho_\Delta\|-\delta_2)}$$

where $\delta_2 = \min\{\alpha_s, \|\rho_\Delta\|\delta_1\}>0$. □

Proof of Lemma 5. Let $d = n-2$ be the dimension of A_R. Observe that for each $g \in \tilde{B}_\epsilon$ we have

$$F_{R-\epsilon}(g\Gamma) \le F_R(\Gamma) \le F_{R+\epsilon}(g\Gamma), \tag{15}$$

which implies

$$\langle F_{R-\epsilon},\Psi_\epsilon\rangle \le F_R(\Gamma) \le \langle F_{R+\epsilon},\Psi_\epsilon\rangle\,. \tag{16}$$

Then by Equation 16, we find that

$$\begin{aligned} |F_R(\Gamma) - \langle F_R,\Psi_\epsilon\rangle| &\le \langle F_{R+\epsilon},\Psi_\epsilon\rangle - \langle F_{R-\epsilon},\Psi_\epsilon\rangle \\ &\ll \int_{A_{R+\epsilon}-A_{R-\epsilon}}\rho'_\Delta(a)da + E(\Psi_\epsilon)\mu_A(\psi_S(A_{R+\epsilon}))e^{(\|\rho_\Delta\|-\delta)(R+\epsilon)} \end{aligned}$$

where $E(\Psi_\epsilon) = \max\{\sup|\Psi_\epsilon|, \|\Psi_\epsilon\|_\ell, \|\Psi_\epsilon\|_{Lip}\}$. Observe for Z a small spherical shell about ρ_Δ

$$\int_{A_{R+\epsilon}-A_{R-\epsilon}}\rho'_\Delta(a)da = \int_Z\int_{R-\epsilon}^{R+\epsilon} e^{\langle\rho_\Delta,\boldsymbol{\theta}\rangle r}r^{n-1}drd\sigma(\boldsymbol{\theta}) + O(e^{(\|\rho_\Delta\|-\delta)R}).$$

By repeated applications of integration by parts we have for any $\omega \in \mathfrak{a}$

$$\int_Z\int_{R-\epsilon}^{R+\epsilon} e^{\langle\omega,\boldsymbol{\theta}\rangle r}r^{d-1}drd\sigma(\boldsymbol{\theta}) = \int_Z\sum_{p=0}^{d-1}(-1)^p\, c_{d,p}(\boldsymbol{\theta})e^{\langle\omega,\boldsymbol{\theta}\rangle r}r^{d-1-p}\Big|_{R-\epsilon}^{R+\epsilon} d\sigma(\boldsymbol{\theta})$$

where $c_{d,p}(\boldsymbol{\theta}) > 0$ for each $\boldsymbol{\theta} \in Z$. Observe for real numbers p, q we have

$$(R+\epsilon)^pe^{q(R+\epsilon)} - (R-\epsilon)^pe^{q(R-\epsilon)} \ll R^pe^{qR}\sinh(q\epsilon)$$

and $\sinh(q\epsilon) = q\epsilon + O(\epsilon^3)$ as $\epsilon \to 0$. To see this consider (after factoring out an R^p)

$$\epsilon = e^{-cR} \mapsto (1+\epsilon/R)^p = \left(1 - \frac{c\epsilon}{\log\epsilon}\right)^p = 1 - \frac{cp\epsilon}{\log(\epsilon)} + o(c\epsilon/\log(\epsilon)).$$

Therefore

$$(R+\epsilon)^pe^{q(R+\epsilon)} - (R-\epsilon)^pe^{q(R-\epsilon)} = R^pe^{qR}\left(e^{q\epsilon} - e^{-q\epsilon} + O_{c,p}(\epsilon/\log(\epsilon))\right).$$

It then follows (take $q = \|\rho_\Delta\|$) that

$$\begin{aligned} |F_R(\Gamma) - \langle F_R,\Psi_\epsilon\rangle| &\ll R^{d-1}\epsilon e^{\|\rho_\Delta\|R} + E(\Psi_\epsilon)e^{(\|\rho_\Delta\|-\delta)(R+\epsilon)} \\ &= R^{d-1}e^{(\|\rho_\Delta\|-c)R} + E(\Psi_\epsilon)e^{(\|\rho_\Delta\|-\delta)(R+\epsilon)}. \end{aligned}$$ □

4. Proof of Theorem 6

In this section we consider the problem of counting the number of rational points on a flag variety with respect to the anticanonical line bundle and prove Theorem 6.

Let $X = G/P_E$ where P_E is a standard parabolic subgroup of G determined by E. Since $\rho_E \in \mathcal{A}$, it follows that there is a unique finite dimensional irreducible representation $\eta : G \to GL(V)$ for which ρ_E is the highest weight. Moreover, there exists a $v_0 \in V(\mathbb{Q})$ such that

$$P_E = \{g \in G \ : \ \eta(g)[v_0] = [v_0]\}$$

and X is homeomorphic to the orbit $\eta(G)[v_0]$. We will now define the height on X with respect to L.

Let $H : \mathbb{P}(V)(\mathbb{Q}) \to \mathbb{R}^+$ be defined by $H([v]) = \|v\|$ where $[v]$ is the point in projective space corresponding to $v \in V$ corresponding to a primitive v and $\|\cdot\|$ is the Euclidean norm on V. Now the height function with respect to the anticanonical bundle $\mathcal{L}$ is then

$$h(x) = H(\eta(g_x)[v_0])$$

where $g_x \in G$ is the unique point for which $\eta(g_x)[v_0] = x$. We wish to determine the asymptotic of the function

$$N(T) = \#\{x \in X(\mathbb{Q}) \ : \ h(x) \leq T\}.$$

We will not, however, deal directly with this function. By a theorem of Borel and Harish-Chandra, $(G/P_E)(\mathbb{Q})$ can be written as a finite union of Γ-orbits. This reduces the problem to studying a single Γ orbit. Therefore we study

$$N_T = \#\{\gamma \in \Gamma/\Gamma \cap P_E \ : \ \|\eta(\gamma)v\| < T\},$$

for $v \in V$ having $\|v\| = 1$. Notice that

$$F_T(g\Gamma) = \sum_{\gamma \in \Gamma/\Gamma \cap P_E} 1_T(\eta(g\gamma)v)$$

is equal to N_T when $g = e$ and $1_T(\cdot)$ is the characteristic function of $B_T = \{v \in V \ : \ \|v\| < T\}$. Let $\tilde{B}_T$ be the corresponding subset of G, i.e.

$$\tilde{B}_T = \{g \in G \ : \ \eta(g\gamma)v \in B_T\}.$$

Let $\overline{B}_T$ be the image of $\tilde{B}_T$ in G/Q_E. If $F \subset \Delta$ and $a \in A$, then the F-projection a_F of a (defined in [13, §4]) is the unique element $a_F \in A$ such that $\lambda_\alpha(a_F) = \lambda_\alpha(a)$ for each $\alpha \in F$ and $\lambda_\alpha(a_F) = 1$ for each $\alpha \notin F$. By [13, Lemma 32] $\overline{B}_T$ can be decomposed as

$$\tilde{B}_T = KA_{E^c,T}Q_E/Q_E$$

where $A_{E^c,T} = \{a \in A \ : \ a = a_{E^c}, \ \rho_E(a) \leq T\}$. Let

$$f(T) = \int_{A^+_{E^c,T}} \rho'_E(a)da$$

where $A^+_{E^c,T} = A_{E^c,T} \cap \mathcal{C}$ and ρ'_E is a character of P_E given by

$$(\wedge^{\dim R_u P_E}\mathrm{Ad})(p)u = \rho'_E(p)u$$

for any $u \in \wedge^{\dim R_u P_E} Lie(R_u(P_E))$ where $R_u P_E$ is the unipotent radical of P_E. We will need the fact (see [13, §3]) that there is a vector $\rho_E \in \mathfrak{a}$ (in the logarithm of the convergence cone) such that

$$\rho'_E(a) = \exp(\langle \rho_E, \log(a)\rangle)$$

for any $a \in A$.

Lemma 6. *Let* $\Psi \in C^{\infty}_{comp}(G/\Gamma)$. *Then there exist constants* $C, r, \delta > 0$ *such that*

$$\left| \langle F_T, \Psi \rangle - Cf(T) \int_{\mathbf{G}/\Gamma} \Psi(g)dg \right| \ll E(\Psi)Te^{-\delta\sqrt{\log T}} + T\log(T)^{-r}\sup|\Psi|,$$

where $\tilde{E}(\Psi) = \max\{\mathrm{supp}(\Psi)\|\Psi\|_{C^\ell}, \|\Psi\|_{Lip}, \sup|\Psi|\}$.

Proof. Let $d = \dim(\mathcal{C}_{E^c})$. Suppose Ψ is supported in the ball of radius ϵ. If $\lambda_\alpha(a) < e^{-\epsilon}$, then $\int_{G/\Gamma} \Psi(kad\mu_{Q_E}) = 0$. By unfolding the F_T in the integral we obtain

$$\langle F_T, \Psi \rangle = \int_K \int_{A^{(\epsilon)}_{E^c,T}} \int_{G/\Gamma} \Psi(g\Gamma)d(ka\mu_{Q_E})(g)\rho'_E(a)dadk$$

where $A^{(\epsilon)}_{E^c,T} = \{a \in A_{E^c,T} \ : \ \lambda_\alpha(a) > e^{-\epsilon}, \text{ for each } \alpha \in E^c\}$. We estimate the integral by splitting $A^{(\epsilon)}_{E^c,T}$ into two disjoint pieces, $\Omega^{(1)}_T$ and $\Omega^{(2)}_T$. Define

$$\Omega^{(1)}_T = \left\{ a \in A^+_{E^c,T} \ : \ \mathrm{dist}\left(\frac{\log(a)}{\|\log(a)\|}, \partial\mathcal{C}_{E^c}\right) > \sqrt{\frac{1}{\log(T)}} \right\}$$

where $A^+_{E^c,T} = \{a \in A_{E^c,T} \ : \ \lambda_\alpha(a) \geq 1, \forall\alpha\}$. Then on $\Omega^{(1)}_T$ we have, by Theorem 3,

$$\begin{aligned}
&\int_K \int_{\Omega^{(1)}_T} \int_{G/\Gamma} \Psi(g\Gamma)((ka)d\mu_{Q_E})(g)\rho'_E(a)dadk \\
= &\int_K \int_{S^{(1)}} \int_0^{\log(T)/\langle\rho_E,\theta\rangle} \int_{G/\Gamma} \Psi(g\Gamma)((k\exp(R\theta))d\mu_{Q_E})(g)e^{\langle\rho'_E,\theta\rangle R}R^{n-2}dRd\sigma(\theta)dk \\
= &\ \mathrm{vol}(K)\int_{\Omega^{(1)}_T} \rho'_E(a)da \int_{G/\Gamma} \Psi(g\Gamma)dg + O\left(\tilde{E}(\Psi)Te^{-\delta\sqrt{\log(T)}}\right)
\end{aligned}$$

where $S^{(1)}$ is the intersection of $\Omega^{(1)}_T$ with the unit sphere in $\mathfrak{a}$ and we have used the dependence on the implied constant on Ψ in Theorem 3 by following its proof. Now we estimate the integral on $\Omega^{(2)}_T$ trivially to obtain

$$\left| \int_K \int_{\Omega^{(2)}_T} \int_{G/\Gamma} \Psi(g\Gamma)((ka)d\mu_{Q_E})(g)\rho'_E(a)dadk - \mathrm{vol}(K)\mathrm{vol}(\Omega^{(2)}_T)\int_{G/\Gamma}\Psi \right| \ll \mathrm{vol}(\Omega^{(2)}_T)\sup|\Psi|.$$

But $\mathrm{vol}(\Omega^{(2)}_T) \ll Tp(\log(T))\log(T)^{-r}$ for some $r > 0$ depending only on the dimension. □

Proposition 4. *Let* d *be the dimension of* $A_{E^c,T}$ *defined above. Then there exists a polynomial* $p(s)$ *of degree* $d-1$ *such that*

$$f(T) = \int_{A_{E^c,T}} \rho'_E(a)da = Tp(\log T).$$

Lemma 7. *Suppose* $\Psi_\epsilon \in C^\infty_c(G/\Gamma)$ *is supported in the ball of radius* $\epsilon > 0$ *about* Γ *and that* $B_{T+\epsilon} \subset \mathrm{supp}(\Psi_\epsilon)B_T$ *and* $\mathrm{supp}(\Psi_\epsilon)B_{T-\epsilon} \subset B_T$. *Then*

$$F_{T-\epsilon}(g) \leq F_T(e) \leq F_{T+\epsilon}(g)$$

for each $g \in \mathrm{supp}(\Psi_\epsilon)$, *and if* $\epsilon = \log(T)^{-c}$, *then*

$$|\langle F_T, \Psi_\epsilon \rangle - F_T(e)| \ll \tilde{E}(\Psi_\epsilon)Te^{-\delta\sqrt{\log(T)}} + \log(T)^{d-1-c} + (\sup|\Psi_\epsilon|)Tp(\log(T))\log(T)^{-r},$$

where $r > 0$ *is the exponent coming from the previous lemma.*

Proof. Observe that

$$\begin{aligned}|\langle F_T, \Psi_\epsilon\rangle - F_T(e)| &\leq \langle F_{T+\epsilon}, \Psi_\epsilon\rangle - \langle F_{T-\epsilon}, \Psi_\epsilon\rangle \\ &\ll f(T+\epsilon) - f(T-\epsilon) \\ &\quad + \tilde{E}(\Psi_\epsilon) T e^{-\delta\sqrt{\log(T)}} + (\sup|\Psi_\epsilon|) T p(\log(T)) \log(T)^{-r}.\end{aligned}$$

But $f(T) = Tp(\log(T))$ for some polynomial $p(x)$ of degree $d-1$, so $f(T+\epsilon) - f(T-\epsilon) = 2\epsilon f'(T) + o(\epsilon) = 2(p(\log(T)) + p'(\log(T)))\log(T)^{-c} + o(\log(T)^{-c})$. ☐

Proof of Theorem 6. We let Ψ_ϵ be the approximate identity given in Proposition 3 that is supported in the ball of radius $\epsilon = \log(T)^{-c} > 0$ about Γ (for some $c > 0$ to be chosen later) and that $B_{T+\epsilon} \subset \mathrm{supp}(\Psi_\epsilon) B_T$ and $\mathrm{supp}(\Psi_\epsilon) B_{T-\epsilon} \subset B_T$. Then by Proposition 3, Lemma 7 and Proposition 4 we have for some $s, \delta > 0$

$$\begin{aligned}|N(T) - \text{main term}| &= |F_T(\Gamma) - \text{main term}| \\ &\leq |F_T(\Gamma) - \langle F_T, \Psi_\epsilon\rangle| + |\langle F_T, \Psi_\epsilon\rangle - \text{main term}| \\ &\ll \tilde{E}(\Psi_\epsilon) T e^{-\delta\sqrt{\log(T)}} + \log(T)^{d-1-c} + (\sup|\Psi_\epsilon|) T p(\log(T)) \log(T)^{-r} \\ &\ll T p(\log(T)) \log(T)^{cs-r}.\end{aligned}$$

We choose c such that $0 < c < r/s$. The remainder of the proof regarding the volume estimate is similar to the end of the proof of Theorem 5. ☐

Acknowledgements We wish to sincerely thank the anonymous referees for pointing out mistakes and inconsistencies in the preliminary versions of this paper and for many other helpful suggestions. Many thanks go to our editor Dmitry Kleinbock for his care, his insights, and his attention to detail. Finally, we most sincerely thank Amir Mohammadi for suggesting this problem and generously sharing his ideas with us. During the course of this work Michael Kelly was partially supported by NSF grants DMS-1045119, DMS-0943832, and DMS-1101326; and Han Li was partially supported by an AMS Simons Travel Grant.

References

[1] V. Bernik, D. Kleinbock, and G. A. Margulis. Khintchine-type theorems on manifolds: the convergence case for standard and multiplicative versions. *Internat. Math. Res. Notices*, (9):453–486, 2001.
[2] Armand Borel and Jacques Tits. Groupes réductifs. *Inst. Hautes Études Sci. Publ. Math.*, (27):55–150, 1965.
[3] S. G. Dani. Invariant measures of horospherical flows on noncompact homogeneous spaces. *Inventiones Mathematicae*, 47(2):101–138, 1978.
[4] W. Duke, Z. Rudnick, and P. Sarnak. Density of integer points on affine homogeneous varieties. *Duke Math. J.*, 71(1):143–179, 1993.
[5] Alex Eskin and Curt McMullen. Mixing, counting, and equidistribution in Lie groups. *Duke Math. J.*, 71(1):181–209, 1993.
[6] Jens Franke, Yuri I. Manin, and Yuri Tschinkel. Rational points of bounded height on Fano varieties. *Invent. Math.*, 95(2):421–435, 1989.
[7] James E Humphreys. *Linear algebraic groups*, volume 21. Springer Science & Business Media, 2012.
[8] D. Y. Kleinbock and G. A. Margulis. Flows on homogeneous spaces and Diophantine approximation on manifolds. *Ann. of Math. (2)*, 148(1):339–360, 1998.
[9] D. Y. Kleinbock and G. A. Margulis. On effective equidistribution of expanding translates of certain orbits in the space of lattices. In *Number theory, analysis and geometry*, pages 385–396. Springer, New York, 2012.
[10] Dmitry Kleinbock, Ronggang Shi, and Barak Weiss. Pointwise equidistribution with an error rate and with respect to unbounded functions. *arxiv*, (1505.06717), 2015.
[11] Dmitry Kleinbock and Barak Weiss. Dirichlet's theorem on Diophantine approximation and homogeneous flows. *J. Mod. Dyn.*, 2(1):43–62, 2008.
[12] G. A. Margulis. On some aspects of the theory of Anosov systems. *Springer-Verlag, Berlin*, 2004.
[13] Amir Mohammadi and Alireza Salehi Golsefidy. Translate of horospheres and counting problems. *American Journal of Mathematics*, 136.5:1301–1346, 2014.
[14] Hee Oh. Orbital counting via mixing and unipotent flows. In *Homogeneous flows, moduli spaces and arithmetic*, volume 10 of *Clay Math. Proc.*, pages 339–375. Amer. Math. Soc., Providence, RI, 2010.

[15] Peter Sarnak. Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series. *Comm. Pure Appl. Math.*, 34(6):719–739, 1981.

[16] Atle Selberg. Harmonic analysis and discontinuous groups in weakly symmetric riemannian spaces with applications to Dirichlet series. *J. of the Indian Math. Soc. (N.S.)*, 20:47–87, 1956.

[17] Nimesh Shah and Barak Weiss. On actions of epimorphic subgroups on homogeneous spaces. *Ergodic Theory and Dynamical Systems*, 20(02):567–592, 2000.

[18] Nimish A. Shah. Limit distributions of expanding translates of certain orbits on homogeneous spaces. In *Proceedings of the Indian Academy of Sciences-Mathematical Sciences*, volume 106, pages 105–125. Springer, 1996.

[19] Ronggang Shi. Expanding cone and applications to homogeneous dynamics. *arxiv 1510.05256*, 2015.

[20] D. Zagier. Eisenstein series and the riemann zeta-function. In *Automorphic Forms, Representation Theory and Arithmetic*, Tata Institute of Fundamental Research Studies in Mathematics, pages 275–301. Springer Berlin Heidelberg, 1981.

Department of Mathematics, Univerisity of Texas, 1 University Station, Austin, TX 78712
E-mail address: kdabbs@math.utexas.edu

Department of Mathematics, Univerisity of Michigan, 530 Church St., Ann Arbor, MI 48109
E-mail address: michaesk@umich.edu

Department of Mathematics and Computer Science, Wesleyan Univerisity, 265 Church Street, Middletown, CT 06459
E-mail address: hli03@wesleyan.edu