Pacific Journal of Mathematics

HOROSPHERICAL COORDINATES OF LATTICE POINTS IN HYPERBOLIC SPACES: EFFECTIVE COUNTING AND EQUIDISTRIBUTION

TAL HORESH AND AMOS NEVO

Volume 324 No. 2 June 2023

PACIFIC JOURNAL OF MATHEMATICS
Vol. 324, No. 2, 2023
https://doi.org/10.2140/pjm.2023.324.265

HOROSPHERICAL COORDINATES OF LATTICE POINTS IN HYPERBOLIC SPACES: EFFECTIVE COUNTING AND EQUIDISTRIBUTION

TAL HORESH AND AMOS NEVO

We establish effective counting results for lattice points in families of domains in real, complex and quaternionic hyperbolic spaces of any dimension. The domains we focus on are defined as product sets with respect to an Iwasawa decomposition. Several natural diophantine problems can be reduced to counting lattice points in such domains. These include equidistribution of the ratio of the length of the shortest solution (x, y) to the gcd equation $bx - ay = 1$ relative to the length of (a, b), where (a, b) ranges over primitive vectors in a disc whose radius increases, the natural analog of this problem in imaginary quadratic number fields, as well as equidistribution of integral solutions to the diophantine equation defined by an integral Lorentz form in three or more variables. We establish an effective rate of convergence for these equidistribution problems, depending on the size of the spectral gap associated with a suitable lattice subgroup in the isometry group of the relevant hyperbolic space. The main result underlying our discussion amounts to establishing effective joint equidistribution for the horospherical component and the radial component in the Iwasawa decomposition of lattice elements.

1. Introduction and statement of main results

Our goal in the present paper is to establish effective counting and equidistribution results for Iwasawa components of elements of lattice subgroups of isometry groups

MSC2020: 22E30, 22E40.
Keywords: hyperbolic spaces, horospherical coordinates, equidistribution of lattice points, spectral gap.

of (real, complex or quaternionic) hyperbolic spaces. The problem is an instance of counting problems in which one seeks to study the asymptotic behavior of the number of lattice orbit points in some expanding family of regions in hyperbolic space, going beyond the classical problem of counting in hyperbolic balls. Since counting the points of a lattice orbit in regions of hyperbolic space reduces to counting the elements of the lattice subgroup in suitable lifted regions of the group of isometries G, we will focus on counting in the group itself, rather than in the symmetric space.

The domains that we consider are product sets in the Iwasawa coordinates on G. We write the Iwasawa decomposition in the form $G = NAK$, where K is maximal compact, $A \cong \mathbb{R}$, and N is the unipotent subgroup that stabilizes an ideal boundary point which we denote $\{\infty\}$. The map $N \times A \times K \to G$ given by $(n, a, k) \mapsto nak$ is a diffeomorphism, so these are indeed coordinates on G.

Let G denote a nonexceptional simple Lie group of real rank one with finite center; namely, locally isomorphic to one of the following: $\mathrm{SO}(1, n)$, $\mathrm{SU}(1, n)$, or $\mathrm{SP}(1, n)$ for some $n \geq 1$. The corresponding rank 1 symmetric spaces G/K are, respectively, the real hyperbolic space $\mathbf{H}^n_{\mathbb{R}}$, the complex hyperbolic space $\mathbf{H}^n_{\mathbb{C}}$ and the quaternionic hyperbolic space $\mathbf{H}^n_{\mathbb{H}}$. The group G acts on the corresponding space by isometries of a Riemannian distance, which we will refer to as the "hyperbolic distance" and denote by $d(\cdot\,,\cdot)$. The remaining rank one simple Lie group is $\mathrm{F}_{4(-20)}$, which corresponds to the octonionic hyperbolic plane $\mathbf{H}^2_{\mathbb{O}}$; we shall not consider this case.

A Haar measure μ on G is given in the Iwasawa coordinates as follows. We parametrize A by $A = \{a_t : t \in \mathbb{R}\}$, where $a_t = \exp tH_1$, and $H_1 \in \mathfrak{a} = \mathrm{Lie}(A)$ is the element satisfying $\alpha(H_1) = 1$, where α is the unique short positive root.

Let μ_K denote a Haar probability measure on K. The subgroup N is parametrized by a euclidean space of the appropriate dimension, see Table 1 for the different cases, and a Haar measure on N is the Lebesgue measure on this underlying euclidean space. A Haar measure on G with respect to the Iwasawa coordinates is given by

$$\mu = \mu_N \times \frac{dt}{e^{2\rho t}} \times \mu_K, \tag{1-1}$$

where ρ is a positive parameter that depends on the group G. The Iwasawa component subgroups, symmetric spaces and Haar measure of the rank 1 groups are summarized in Table 1 in Section 4A below.

Let $\Gamma \subset G$ be a lattice subgroup. We consider lattice points whose N and K components lie in given bounded subsets $\Psi \subset N$ and $\Phi \subseteq K$, and study their asymptotic behavior as their A-component a_t ranges over $(-\infty, 0]$ and tends to $-\infty$.

Define the family $\{R_T(\Psi, \Phi)\}_{T>0}$, where

$$R_T(\Psi, \Phi) := \Psi A_{[-T,0]}\Phi = \{na_tk : n \in \Psi,\ t \in [-T, 0],\ k \in \Phi\}$$

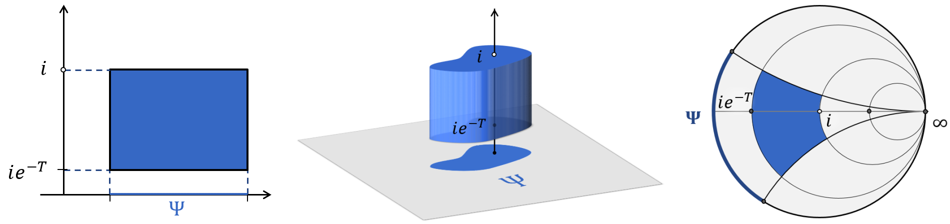

Figure 1. The domains $R_T(\Psi, \Phi)$ projected to the real hyperbolic plane in the upper half-plane model (left), the real hyperbolic 3-space in the upper half-space model (middle) and the real hyperbolic plane in the unit disc model (right).

as $T \to \infty$; see Figure 1. According to (1-1), the volume of these domains equals

$$\mu(R_T(\Psi, \Phi)) = \frac{1}{2\rho} \cdot \mu_N(\Psi)\mu_K(\Phi)(e^{2\rho T} - 1),$$

while the volume of $NA_{[0,\infty)}K$ is finite. We shall require that the domains $\Psi \subset N$ and $\Phi \subseteq K$ are *nice*: bounded, embedded smooth submanifolds of full dimension whose boundaries are piecewise smooth — namely, a finite union of smooth submanifolds of codimension 1. We allow the case where only some of these boundary submanifolds are included in the nice set, while others are not, e.g., a half open rectangle, two of whose edges are included and the remaining two are not.

We can now formulate our main result, namely an effective solution to the lattice point counting problem in the sets $R_T(\Psi, \Phi)$, for an arbitrary lattice.

Theorem 1.1. *Fix any Iwasawa decomposition in G, let $\Psi \subset N$ and $\Phi \subseteq K$ be nice domains, and consider the family $R_T(\Psi, \Phi)$ as defined above. For any lattice $\Gamma < G$, there exists a parameter $\kappa = \kappa(\Gamma) < 1$ (defined explicitly in (4-2)) such that for $T > 0$,*

$$\begin{aligned}\#(R_T(\Psi, \Phi) \cap \Gamma) &= \frac{\mu(R_T(\Psi, \Phi))}{\mu(G/\Gamma)} + O\big(\mu(R_T(\Psi, \Phi))^{\kappa} \cdot \log \mu(R_T(\Psi, \Phi))\big) \\ &= \frac{\mu_N(\Psi)\mu_K(\Phi)}{\mu(G/\Gamma)} \cdot \frac{e^{2\rho T}}{2\rho} + O_{\Gamma,\Psi,\Phi}(T(e^{2\rho T})^{\kappa}).\end{aligned}$$

For example, for the lattice $SL_2(\mathbb{Z})$ in $SL_2(\mathbb{R})$, we will see below that one can take $\kappa(SL_2(\mathbb{Z})) = \frac{7}{8}$.

This case has of course received considerable attention in the past, and we will describe it further at the end of the next section.

We now formulate our two main corollaries. The first is an effective ratio theorem for lattice points in the sets $R_T(\Psi, \Phi)$, and the second effective joint equidistribution of the N and K components in the Iwasawa decomposition, as follows.

For $H \in \{N, A, K\}$, we denote the projection to the H-component by

$$\pi_H : G = NAK \to H.$$

Corollary 1.2. (1) *Let $\Psi, \Psi' \subset N$ and $\Phi, \Phi' \subseteq K$ be nice sets, and let $\Gamma < G$ be any lattice. Then the denominator in the following ratio is eventually positive and*

$$\frac{\#(\Gamma \cap R_T(\Psi', \Phi'))}{\#(\Gamma \cap R_T(\Psi, \Phi))} = \frac{\mu_N(\Psi')\mu_K(\Phi')}{\mu_N(\Psi)\mu_K(\Phi)} + O(T(e^{2\rho T})^{-(1-\kappa)}),$$

where the implied constant depends on Ψ, Ψ', Φ, Φ' and $\kappa = \kappa(\Gamma) < 1$ is the exponent associated with Γ appearing in Theorem 1.1.

(2) *The set of N-components and K-components become jointly effectively equidistributed in $\Psi \times \Phi$ with respect to $\mu_N \times \mu_K$ as $T \to \infty$. Namely, for every Lipschitz function ψ defined on Ψ, and Lipschitz function ϕ defined on Φ,*

$$\frac{1}{\#(\Gamma \cap \Psi A_{[-T,0]}\Phi)} \sum_{\gamma \in \Psi A_{[-T,0]}\Phi} \psi(\pi_N(\gamma))\phi(\pi_K(\gamma))$$
$$= \frac{1}{\mu_N(\Psi)} \int_N \psi(n)\, d\mu_N(n) \cdot \frac{1}{\mu_K(\Phi)} \int_K \phi(k)\, d\mu_K(k) + O(Te^{-2\rho(1-\kappa)T}),$$

where the constant depends on the functions ψ and ϕ, and the sets Ψ and Φ.

The proofs of Theorem 1.1 and Corollary 1.2 are in Section 4.

Remark 1.3. When the lattice in question is cocompact, the (unbounded) cuspidal strip $\Psi A_{(0,\infty)}\Phi$ may contain infinitely many lattice points, despite its bounded volume. One expects that this fact should not change the overall asymptotics, but the irregularity caused by this cuspidal strip requires further consideration, and this is the reason why we have decided to consider here the domains $R_T(\Psi, \Phi)$ which are truncated at some fixed height, say $t = 0$. These domains are the natural ones to consider in the context of lattices with a cusp, because they account for *all but finitely many* lattice points in the strip $\Psi A_{(-\infty,\infty)}\Phi$, provided the lattice orbit points $\gamma \cdot z$ all have bounded height; namely, that the A-component a_t of $\gamma \cdot z$ satisfies $t \leq C(\Gamma, z)$. This property amounts to a generalized form of the Shimizu lemma, and it holds in all the specific examples we will consider below, namely real hyperbolic spaces of arbitrary dimension.

Remark 1.4. Iwasawa decomposition of a Lie group is used in one of two conventions: $G = NAK$ or $G = KAN$. Our results are phrased with respect to the first option, but the corresponding statements with respect to the KAN decomposition may be easily deduced. Indeed, the KAN coordinates of $g \in G$ are obtained from the NAK coordinates of g^{-1}: $g^{-1} = nak$ implies $g = k^{-1}a^{-1}n^{-1}$. In particular, the Haar measure with respect to the KAN coordinates is $\mu_K \times e^{2\rho t}\,dt \times \mu_N$, and the statement of Theorem 1.1 is replaced by

$$\#\Gamma \cap (\Phi A_{[0,T]}\Psi) = \frac{\mu_N(\Psi)\mu_K(\Phi)}{\mu(G/\Gamma)} \cdot \frac{e^{2\rho T}}{2\rho} + O(T(e^{2\rho T})^{\kappa}),$$

for $\Phi \subset K,\ \Psi \subset N,\ \kappa$ as in Theorem 1.1, and $T > 0$.

Remark 1.5. Theorem 1.1 is formulated for a family of domains in G itself, rather than in the symmetric space, and this makes it possible to analyze the distribution of the K-components of the lattice elements. As we shall see below, equidistribution of the K-components plays a key role in a number of applications, including angular equidistribution of shortest solutions to the gcd equation in $\mathbb{Z}^2$ and in imaginary quadratic number fields. The connection between the problem of equidistribution of the norms of the shortest solutions of the gcd equation in $\mathbb{Z}^2$ and the equidistribution of Iwasawa N-components in $\mathrm{SL}_2(\mathbb{Z})$ was first pointed out by Risager and Rudnick [2009], and has motivated the approach pursued in the present paper. We will first formulate and prove some applications of Theorem 1.1 and Corollary 1.2 and then comment further on the history of this problem.

2. Iwasawa components and diophantine problems

2A. ***Distribution of shortest solutions of the gcd equation.*** We begin with applications related to certain arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{C})$. In what follows, the norm we refer to is the euclidean norm on $\mathbb{R}^2$ or $\mathbb{C}^2$, denoted by $\|\cdot\|$.

For every primitive integral vector $v = (a, b)$, let w_v denote the shortest integral vector that completes v to a (positively oriented) basis of $\mathbb{Z}^2$, namely, the shortest solution to the gcd equation $bx - ay = 1$. Let θ_v denote the angle between w_v and v, where $0 < \theta_v < \pi$. We say that v is *positive* if θ_v is acute, and *negative* if θ_v is obtuse (Figure 3).

Let $\Theta \subseteq S^1$ be a subarc of the unit circle and $|\Theta|$ its length, $0 < |\Theta| \le 2\pi$. Let $\mathcal{S}_\Theta$ denote the corresponding sector of the plane $\mathbb{R}^2$ (see Figure 2), and

$$\mathcal{S}_\Theta(R) = \mathcal{S}_\Theta \cap B_R \quad \text{where } B_R = \{v \in \mathbb{R}^2\,;\, \|v\| \le R\}.$$

In the case of the lattice $\mathrm{SL}_2(\mathbb{Z})$, Corollary 1.2 has the following geometric interpretation.

Theorem 2.1. *For every primitive integral vector $v = (a, b)$, let w_v and θ_v be as above.*

(1) *The number of primitive vectors v in $\mathcal{S}_\Theta(R)$ is given by*

$$\frac{3}{\pi^2}|\Theta| R^2 + O(R^{7/4}\log R).$$

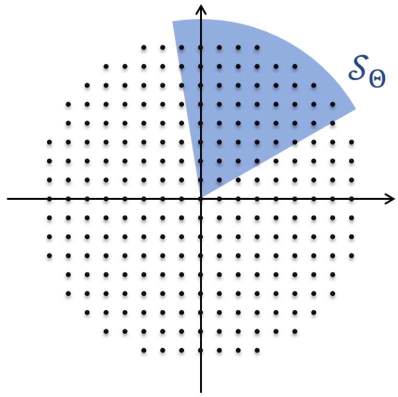

Figure 2. $\mathbb{Z}^2$-points contained in the sector $\mathcal{S}_\Theta$.

(2) *The ratios $\|w_v\|/\|v\|$ of the length of the shortest solution w_v relative to the length of v, for $v \in \mathcal{S}_\Theta(R)$, become effectively equidistributed in the interval $\left[0, \frac{1}{2}\right]$ as $R \to \infty$. The rate of convergence for a Lipschitz function f is $O_{f,\Theta}(R^{-1/4} \cdot \log R)$.*

(3) *The number of positive primitive vectors in $\mathcal{S}_\Theta(R)$ is*

$$\frac{3}{2\pi^2}|\Theta|R^2 + O(R^{7/4}\log R),$$

and the same formula holds for the number of negative primitive vectors in $\mathcal{S}_\Theta(R)$.

(4) *Part (2) holds when v is restricted to positive vectors only, or to negative vectors only.*

Remark 2.2. Note that part (1) counts the number of primitive integral vectors of norm at most R in any sector in the plane, with error estimate given by $R^{7/4}\log R$. This result is of course not new, and in fact our estimate falls short of a much better estimate for this problem that follows from the work of Selberg and Good; see [Selberg 1991; Good 1983; 1984]. We refer to Section 3C for a more detailed discussion.

Proof of Theorem 2.1. If $v = (a, b) \in \mathbb{Z}^2$ is primitive, it can be completed to countably many matrices in $\mathrm{SL}_2(\mathbb{Z})$, representing the different integral solutions to the equation $bx - ay = 1$. The NAK components of these integral matrices encode the vector v and the different solutions to $bx - ay = 1$ as follows. By the Iwasawa decomposition of $\mathrm{SL}_2(\mathbb{R})$ if (x, y) is such a solution, the corresponding matrix in

$\mathrm{SL}_2(\mathbb{Z})$ has NAK decomposition

$$\begin{pmatrix} x & y \\ a & b \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & \frac{xa+yb}{a^2+b^2} \\ & 1 \end{pmatrix}}_{N\text{-component}} \underbrace{\begin{pmatrix} \frac{1}{\sqrt{a^2+b^2}} & \\ & \sqrt{a^2+b^2} \end{pmatrix}}_{A\text{-component}} \underbrace{\frac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} b & -a \\ a & b \end{pmatrix}}_{K\text{-component}}.$$

The subgroups N, A, and K of $\mathrm{SL}_2(\mathbb{R})$ are identified with $\mathbb{R}$, $\mathbb{R}$, and S^1 respectively through $\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) \leftrightarrow x$, $\left(\begin{smallmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{smallmatrix}\right) \leftrightarrow t$, and $\left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right) \leftrightarrow \theta$. We choose here the Haar measure μ_K on the circle to have total mass 2π. Note that the A and K components of the Iwasawa decomposition depend only on the vector v: the A-component $\left(\begin{smallmatrix} 1/\|v\| & 0 \\ 0 & \|v\| \end{smallmatrix}\right)$ encodes the norm of v, and the K-component $\left(\begin{smallmatrix} v^{\perp}/\|v\| \\ v/\|v\| \end{smallmatrix}\right)$ is the rotation matrix by an angle of $\theta + \frac{1}{2}\pi$, where θ is the angle between the positive x-axis and v (counterclockwise). The N-component depends on the specific solution (x, y), namely the upper row of the matrix. If $w := (x, y)$, then the N-component is $\left(\begin{smallmatrix} 1 & \langle w,v\rangle/\|v\|^2 \\ 0 & 1 \end{smallmatrix}\right)$, namely its N-coordinate is given by the projection of w to the line $\mathrm{span}\{v\}$, divided by the norm of v.

The different solutions to $bx - ay = 1$ are $\{(x + ma, y + mb) : m \in \mathbb{Z}\}$, and they correspond to matrices $\left(\begin{smallmatrix} x+ma & y+mb \\ a & b \end{smallmatrix}\right)$ whose N-coordinate is

$$\frac{(x+ma)a + (y+mb)b}{a^2+b^2} = m + \frac{xa+yb}{a^2+b^2} = m + \frac{\langle w, v\rangle}{\|v\|^2},$$

namely all the integral translations of $\langle w, v\rangle/\|v\|^2$, the N-coordinate appearing above.

Observe that among all the integral matrices that correspond to v, the one whose N-coordinate is minimal — i.e., in the interval $\left[-\frac{1}{2}, \frac{1}{2}\right)$ — is the one that corresponds to the shortest solution to $bx - ay = 1$, namely, the one whose upper row has minimal norm. This is because the integral solutions are the integral points on the affine line $\mathrm{span}\{v\} + w$ (where w is any fixed solution of $bx - ay = 1$) which is parallel to $\mathrm{span}\{v\}$. Hence, when decomposing $\mathbb{R}^2$ as $\mathrm{span}\{v\} \oplus \mathrm{span}\{v^{\perp}\}$, all of these solutions have the same $v^{\perp}$ component, and the shortest integral solution is the one with the shortest v-component (namely, the shortest projection on $\mathrm{span}\{v\}$). The shortest integral solution w_v corresponds to the matrix $\left(\begin{smallmatrix} w_v \\ v \end{smallmatrix}\right) = \left(\begin{smallmatrix} x_v & y_v \\ a & b \end{smallmatrix}\right)$, which we denote by γ_v.

We conclude that the set $\left\{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \pi_N(\gamma) \in \left[-\frac{1}{2}, \frac{1}{2}\right)\right\}$ is in one-to-one correspondence $\gamma_v \leftrightarrow v$ with the set of primitive integral vectors. Furthermore, for each γ_v, $e^{-\pi_A(\gamma_v)/2}$ is the length of v, and the rotation angle determined by $\pi_K(\gamma_v)$ is in $\frac{1}{2}\pi + \Theta$ if and only if $v \in S_\Theta$.

Consequently, the sets

$$\left\{\gamma \in \Gamma\,;\, \pi_N(\gamma) \in \left[-\tfrac{1}{2}, \tfrac{1}{2}\right),\ -T < \pi_A(\gamma) \le 0,\ \pi_K(\gamma) \in \tfrac{1}{2}\pi + \Theta\right\}$$

and

$$\left\{v \in \mathbb{R}^2 \,;\, v \text{ is a primitive integral vector, } \|v\| < e^{T/2},\ \frac{v}{\|v\|} \in \Theta\right\}$$

are in one-to-one correspondence.

Now apply part (1) of Corollary 1.2 with $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, the interval $A_{(-T,0]} \subset A \cong \mathbb{R}$, $\Phi = \frac{1}{2}\pi + \Theta \subset S^1 \cong K$ and $\Psi = \left[-\frac{1}{2}, \frac{1}{2}\right) \subset \mathbb{R} \cong N$. Noting that $\rho = \frac{1}{2}$ for $\mathrm{SL}_2(\mathbb{R})$, the volume of the domains in question is $|\Theta| \cdot (e^T - 1)/v_\Gamma$ where v_Γ is the covolume of $\mathrm{SL}_2(\mathbb{Z})$ in $\mathrm{SL}_2(\mathbb{R})$, which is $2\zeta(2) = \frac{1}{3}\pi^2$ (with respect to the Haar measure chosen). The domain $\mathcal{S}_\Theta(R) = \mathcal{S}_\Theta \cap B_R$ is defined by $R^2 = e^T$ and hence the main term of the volume is

$$\frac{3}{\pi^2}|\Theta| R^2 = \frac{6}{\pi^2}\frac{|\Theta|}{2\pi}(\pi R^2).$$

Note that when $|\Theta| = 2\pi$ the number of primitive lattice points of norm at most R has main term $|B_R|/\zeta(2) = 6R^2/\pi$.

Furthermore, we note $\kappa(\mathrm{SL}_2(\mathbb{Z})) = \frac{7}{8}$ (see Section 4A). Hence the main factor in the error estimate is $\exp\left(\frac{7}{8}T\right) = R^{7/4}$. This proves part (1) of Theorem 2.1.

To prove part (2), recall that θ_v denotes the angle between w_v and v, so that the N-component of γ_v is given by

$$\pi_N(\gamma_v) = \frac{x_v a + y_v b}{a^2 + b^2} = \frac{\langle w_v, v\rangle}{\|v\|^2} = \frac{\|w_v\| \cos(\theta_v)}{\|v\|}. \tag{2-1}$$

Since

$$1 = \det\begin{pmatrix} x_v & y_v \\ a & b \end{pmatrix} = \det\begin{pmatrix} w_v \\ v \end{pmatrix} = \|w_v\|\,\|v\|\,|\sin(\theta_v)|,$$

and $\|w_v\| \geq 1$, it follows that $|\sin(\theta_v)| = O(\|v\|^{-1})$, or equivalently $1 - |\cos(\theta_v)| = O(\|v\|^{-2})$. From part (2) of Corollary 1.2 applied to $\Psi = \left[-\frac{1}{2}, \frac{1}{2}\right)$ and $\Phi = \frac{1}{2}\pi + \Theta$, we have that for primitive vectors v in $\mathcal{S}_\Theta(R)$, the N-components of γ_v become effectively equidistributed in $\left[-\frac{1}{2}, \frac{1}{2}\right)$ as $R \to \infty$, at rate $O(R^{-1/4} \cdot \log R)$. It follows that their absolute values become effectively equidistributed in $\left[0, \frac{1}{2}\right]$ at the same rate. These absolute values are

$$|\pi_N(\gamma_v)| = \frac{\|w_v\|}{\|v\|} \cdot |\cos(\theta_v)| = \frac{\|w_v\|}{\|v\|}(1 + O(\|v\|^{-2})) = \frac{\|w_v\|}{\|v\|} + O(\|v\|^{-2});$$

where the last equality follows since the ratio $\|w_v\|/\|v\|$ is bounded. This follows from (2-1), since $|\pi_N(\gamma_v)| \leq \frac{1}{2}$ and $|\cos\theta_v| \geq 1 - C/\|v\|^2$.

To prove part (2) it remains to show that the values $\|w_v\|/\|v\|$ for $v \in \mathcal{S}_\Theta(R)$ are also effectively equidistributed in $\left[0, \frac{1}{2}\right]$ at rate $O(R^{-1/4} \cdot \log R)$ as $R \to \infty$. Let P_R denote the primitive integral vectors $v \in \mathcal{S}_\Theta(R)$, and note that by part (1) we have $|P_R| \geq R^2$ for $R \geq R_1$. Let f be a Lipschitz function on $\left[0, \frac{1}{2}\right]$. Then for

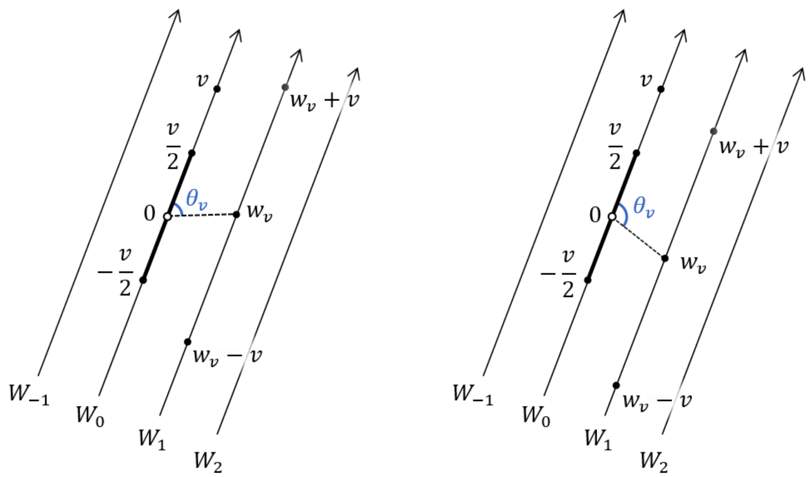

Figure 3. Integral vectors v, w_v and angle θ_v for θ_v acute and v positive (left) and θ_v obtuse and v negative (right). This figure also depicts the lines $W_m = \{w : \det\binom{w}{v} = m\}$ for $m \in \mathbb{Z}$, where $W_0 = \operatorname{span}\{v\}$ and w_v is the shortest integral vector in W_1.

$R \geq R_1$,

$$\frac{1}{|P_R|}\sum_{v\in P_R}\left|f\left(\frac{\|w_v\|}{\|v\|}\right) - f(|\pi_N(\gamma_v)|)\right| \leq \frac{C_f}{|P_R|}\sum_{v\in P_R}\left|\frac{\|w_v\|}{\|v\|} - |\pi_N(\gamma_v)|\right|$$
$$\leq \frac{C_f}{R^2}\sum_{v\in B_R}\frac{C}{\|v\|^2} = \frac{C_f C}{R^2}\sum_{n=1}^{R^2}\sum_{\substack{v\in B_R\\ \|v\|^2=n}}\frac{1}{n} \leq \frac{C_f C}{R^2}\sum_{n=1}^{R^2}\frac{r_2(n)}{n} = O_{f,\Theta}(R^{-2}\log R).$$

Here $r_2(n)$ is the number of representations of an integer n as a sum of two squares, and the last estimate follows using Abel's partial summation formula. Since $\sum_{k=1}^{K} r_2(k) = \pi K + O(\sqrt{K})$,

$$\sum_{n=1}^{R^2}\frac{r_2(n)}{n} = \frac{1}{R^2+1}\sum_{n=1}^{R^2} r_2(n) + \sum_{n=1}^{R^2}(\pi n + O(\sqrt{n}))\left(\frac{1}{n} - \frac{1}{n+1}\right) = O(\log R).$$

Since the N-components equidistribute at rate $R^{-1/4}\log R$, and the term just estimated vanishes faster, this proves part (2) of Theorem 2.1.

By the expression (2-1) for the N-component of γ_v, the vector v is positive if and only if $\cos\theta_v > 0$, i.e., if and only if $\pi_N(\gamma_v) > 0$. Similarly, v is negative if and only if $\pi_N(\gamma_v) < 0$. Thus, by applying Corollary 1.2 to $\Psi_+ = \left[0, \frac{1}{2}\right]$, we obtain part (2) for the positive vectors, and similarly when $\Psi_- = \left[-\frac{1}{2}, 0\right]$, we obtain part (2) for the negative vectors. This proves part (3), and part (4) follows by applying the argument used in the proof of part (1) to the choices Ψ_- or Ψ_+ as the set of N-components. □

2B. ***The gcd equation in imaginary quadratic number fields.*** Theorem 2.1 extends to rings of integers in imaginary quadratic number fields, as follows. Let d be a negative squarefree integer, and let $\mathcal{O}_d$ denote the ring of integers in the quadratic number field $\mathbb{Q}[\sqrt{d}]$. The ring $\mathcal{O}_d$ is a lattice in $\mathbb{C}$, which is a rectangular lattice when d is congruent to 1 or 3 modulo 4, and an "isosceles-triangular" lattice when d is congruent to 1 modulo 4. The Dirichlet fundamental domain $\mathcal{D}_d$ of the lattice $\mathcal{O}_d$ is defined as the (closed) set of all points whose distance to 0 is less than or equal to their distance to any other lattice point. The *strict* Dirichlet fundamental domain $\mathcal{D}'_d$ is defined as the interior of $\mathcal{D}_d$ together with a union of intervals contained in the boundary $\partial\mathcal{D}_d$, chosen so that every orbit has a unique point in $\mathcal{D}'_d$. We let $[0, r_d]$ denote the image of the norm function $\|\cdot\| : \mathcal{D}_d \to \mathbb{R}$, and let ν_d denote the probability measure on $[0, r_d]$ which is the distribution of the norm of a random uniform point in $\mathcal{D}_d$. Equivalently ν_d is the push-forward of Lebesgue measure on $\mathcal{D}_d$, normalized to have total mass 1. Note that ν_d is equivalent but not equal to Lebesgue measure on the interval.

We refer to $v = (\alpha, \beta)$ in $\mathcal{O}_d^2$ as *primitive* if the ideals $\langle\alpha\rangle$ and $\langle\beta\rangle$ are coprime; namely, if there exists a solution (ξ, η) in $\mathcal{O}_d^2$ to $\beta\xi - \alpha\eta = 1$. Consider a shortest vector that completes v to a basis of $\mathcal{O}_d^2$, namely, a shortest $\mathcal{O}_d$-integral solution to the equation $\beta\xi - \alpha\eta = 1$. We will presently explain how to specify a unique such shortest solution, which we denote by w_v.

Let $\Theta \subseteq S^3$ be a spherical cap in the unit sphere, let $\mathcal{S}_\Theta$ denote the corresponding sector of $\mathbb{R}^4 \cong \mathbb{C}^2$, and $\mathcal{S}_\Theta(R) = \mathcal{S}_\Theta \cap B_R$ where $B_R = \{v \in \mathbb{C}^2 ; \|v\| \le R\}$.

Theorem 2.3. *Consider primitive vectors $v = (\alpha, \beta) \in \mathcal{O}_d^2$ in $\mathcal{S}_\Theta(R) \subset \mathbb{C}^2$.*

(1) *The number of primitive vectors in $\mathcal{S}_\Theta(R)$ is given by $c_d|\Theta|R^4 + O(R^{4\kappa_d} \log R)$, where c_d is a positive constant depending only on d.*

(2) *The ratios $\|w_v\|/\|v\|$ of the length of the shortest solution w_v relative to the length of v, for $v \in \mathcal{S}_\Theta(R)$, become effectively equidistributed in the interval $[0, r_d]$ as $R \to \infty$, with respect to the measure ν_d. The rate of convergence for a Lipschitz function f is $O_{f,\Theta}(R^{-4(1-\kappa_d)} \cdot \log R)$.*

Here κ_d is the exponent that corresponds to the lattice $\mathrm{PSL}_2(\mathcal{O}_d)$ *of* $\mathrm{PSL}_2(\mathbb{C})$ *in Theorem 1.1.*

Observe that when $\mathcal{O}_d$ is a euclidean domain with respect to the usual norm, namely when $d \in \{-1, -2, -3, -7, -11\}$, and α and β are coprime, the equation $\beta\xi - \alpha\eta = 1$ is their gcd equation. Then w_v is the shortest solution to the gcd equation defined by v, as in the case of $\mathbb{Z}$.

The arguments in the proof of Theorem 2.1 can be applied, with some modifications, to prove Theorem 2.3. In the present context, Corollary 1.2 is applied for the Iwasawa components of the lattice $\mathrm{SL}_2(\mathcal{O}_d)$ in $\mathrm{SL}_2(\mathbb{C})$. We briefly describe the necessary adjustments, see [Elstrodt et al. 1998] for further details.

Proof. Recall the Iwasawa decomposition of $\mathrm{SL}_2(\mathbb{C})$ consists of the subgroups

$$N=\left\{\begin{pmatrix}1 & z\\ & 1\end{pmatrix} : z\in\mathbb{C}\right\},$$
$$A=\left\{\begin{pmatrix}e^{t/2} & 0\\ 0 & e^{-t/2}\end{pmatrix} : t\in\mathbb{R}\right\},$$
$$K=\left\{\begin{pmatrix}\bar b & -\bar a\\ a & b\end{pmatrix} : |a|^2+|b|^2=1\right\}=\mathrm{SU}(2).$$

Clearly, K is isomorphic to the unit sphere S^3 in $\mathbb{C}^2$.

A primitive pair $(\alpha,\beta)\in\mathcal{O}_d^2$ can be completed to a matrix $\left(\begin{smallmatrix}\xi & \eta\\ \alpha & \beta\end{smallmatrix}\right)$ in $\mathrm{SL}_2(\mathcal{O}_d)$, and the Iwasawa coordinates of such a matrix are

$$\begin{pmatrix}\xi & \eta\\ \alpha & \beta\end{pmatrix}=\underbrace{\begin{pmatrix}1 & \frac{\xi\bar\alpha+\eta\bar\beta}{\|\alpha\|^2+\|\beta\|^2}\\ & 1\end{pmatrix}}_{N\text{-component}}\underbrace{\begin{pmatrix}\frac{1}{\sqrt{\|\alpha\|^2+\|\beta\|^2}} & \\ & \sqrt{\|\alpha\|^2+\|\beta\|^2}\end{pmatrix}}_{A\text{-component}}\underbrace{\frac{1}{\sqrt{\|\alpha\|^2+\|\beta\|^2}}\begin{pmatrix}\bar\beta & -\bar\alpha\\ \alpha & \beta\end{pmatrix}}_{K\text{-component}}.$$

The A and K components encode the vector v: the A-component encodes its norm, and the bottom row of the K-component encodes the unit vector $v/\|v\|$ in the unit sphere S^3. The N-component encodes the upper row of the matrix: if $w=(\xi,\eta)$, this component equals $\langle w,v\rangle/\|v\|^2$. The set of solutions to $\beta\xi-\alpha\eta=1$ is $\{(\xi+m\alpha,\eta+m\beta) : m\in\mathcal{O}_d\}$, and the matrices in $\mathrm{SL}_2(\mathcal{O}_d)$ that correspond to these solutions differ only by their N-components; these components are

$$\begin{pmatrix}1 & \frac{(\xi+m\alpha)\bar\alpha+(\eta+m\beta)\bar\beta}{\|\alpha\|^2+\|\beta\|^2}\\ & 1\end{pmatrix}=\begin{pmatrix}1 & m+\frac{\xi\bar\alpha+\eta\bar\beta}{\|\alpha\|^2+\|\beta\|^2}\\ & 1\end{pmatrix}=\begin{pmatrix}1 & m+\frac{\langle w,v\rangle}{\|v\|^2}\\ & 1\end{pmatrix},$$

where $m\in\mathcal{O}_d$. Using v and $v^\perp$ as orthogonal coordinates in the plane $\mathbb{C}^2$, by the same argument that was used in the real case (Theorem 2.1), the shortest $\mathcal{O}_d$-integral solution $w_v=(\xi_v,\eta_v)$ to $\beta\xi-\alpha\eta=\det\binom{w}{v}=1$ corresponds to the matrix $\gamma_v\in\mathrm{SL}_2(\mathcal{O}_d)$ whose N-coordinate has minimal norm.

Clearly, the set of N-components $\{m+\langle w,v\rangle/\|v\|^2 : m\in\mathcal{O}_d\}$ is the orbit of $\langle w,v\rangle/\|v\|^2$ under translations by the lattice $\mathcal{O}_d$ in $\mathbb{C}$. Hence there is a unique element in this orbit in every $\mathcal{O}_d$-integral translation of the *strict* Dirichlet fundamental domain $\mathcal{D}'_d$. The representative which is of minimal norm is the one that lies in $\mathcal{D}'_d$ itself. Thus, γ_v is the unique matrix in $\mathrm{SL}_2(\mathcal{O}_d)$, among the matrices whose bottom row is v, whose N-component lies in $\mathcal{D}'_d$.

Let $\mathrm{s}(v)$ and $\mathrm{c}(v)$ be such that

$$1 = \det\begin{pmatrix} \xi_v & \eta_v \\ \alpha & \beta \end{pmatrix} = \det\begin{pmatrix} w_v \\ v \end{pmatrix} = \|w_v\|\|v\| \cdot \mathrm{s}(v),$$

and $\langle w_v, v\rangle = \|w_v\|\|v\| \cdot \mathrm{c}(v)$; these are the analogs for $\sin\theta_v$ and $\cos\theta_v$: see (2-1).

Taking the absolute value squared of the previous two equations, the identity

$$|\mathrm{s}(v)|^2 + |\mathrm{c}(v)|^2 = 1,$$

is easily seen to be equivalent to $1 + |\langle w_v, v\rangle|^2 = \|w_v\|^2\|v\|^2$, and this equation is a consequence of $\beta\xi_v - \alpha\eta_v = 1$, as can be verified directly (starting with $\langle w_v, v\rangle = \xi_v\bar{\alpha} + \eta_v\bar{\beta}$).

In particular, when $|v|$ is large, $\mathrm{s}(v)$ is small (since $\|w_v\|$ is bounded below), and in fact $\mathrm{s}(v) = O(\|v\|^{-1})$. Now $|\mathrm{s}(v)|^2 = 1 - |\mathrm{c}(v)|^2 \geq 1 - |\mathrm{c}(v)|$ so $1 - |\mathrm{c}(v)| = O(\|v\|^{-2})$, and

$$\frac{|\langle w_v, v\rangle|}{\|v\|^2} = \frac{\|w_v\|}{\|v\|} \cdot |\mathrm{c}(v)| = \frac{\|w_v\|}{\|v\|}(1 + O(\|v\|^{-2})).$$

As above, the last estimate follows since $|\langle w_v, v\rangle|/\|v\|^2$ is the N-component which is in $\mathcal{D}'_d$ and hence bounded, and $|\mathrm{c}(v)| \geq 1 - C/\|v\|^2$.

The proof now proceeds in a manner analogous to the proof of Theorem 2.3. Namely, we apply Corollary 1.2 to $\Gamma = \mathrm{SL}_2(\mathcal{O}_d)$, the interval $A_{[-T,0]}$ in A with $e^{T/2} = R$, and to $\Psi = \mathcal{D}'_d$, $\Phi = \Theta$. Using also that $\rho = 1$ for $\mathrm{SL}_2(\mathbb{C})$, the measure of $R_T(\Phi, \Psi)$ here is $c'_d|\Theta| \cdot e^{2T}$. Therefore, the error estimate for the number of lattice points in the set is bounded by a multiple of $e^{2T\kappa_d}$.

We conclude that the number of primitive lattice points $(\alpha, \beta) \in \mathcal{S}_\Theta(R)$ is then given by $c_d|\Theta|R^4 + O(R^{4\kappa_d}\log R)$. This completes the proof of part (1) of Theorem 2.3.

To prove part (2), note by Corollary 1.2, part (2), that the N-components of γ_v, where $v \in \mathcal{S}_\Theta(R)$, equidistribute effectively in Ψ with respect to the Lebesgue measure, at the rate $O_{f,\Theta}(R^{-4(1-\kappa_d)} \cdot \log R)$ for any Lipschitz function f. Therefore, their norms effectively equidistribute in $[0, r_d]$ with respect to ν_d, at the same rate. To show that the ratios $\|w_v\|/\|v\|$, with $v \in \mathcal{S}_\Theta(R)$, also have that property, we consider the difference. Let $Q_R = \{v = (\alpha, \beta) \in \mathcal{S}_\Theta(R),\ \alpha, \beta \in \mathcal{O}_d \text{ coprime}\}$. By part (1) of the theorem, Q_R satisfies $|Q_R| \geq R^4/\tilde{c}_d$. Then

$$\frac{1}{|Q_R|}\sum_{v\in Q_R}\left| f\left(\frac{\|w_v\|}{\|v\|}\right) - f(|\pi_N(\gamma_v)|)\right| \leq \frac{C_f}{|Q_R|}\sum_{v\in Q_R}\left|\frac{\|w_v\|}{\|v\|} - |\pi_N(\gamma_v)|\right|$$

$$\leq \frac{\tilde{c}_d C_f}{R^4}\sum_{v\in B_R}\frac{C}{\|v\|^2} = \frac{\tilde{c}_d C_f C}{R^4}\sum_{n=1}^{R^2}\sum_{\substack{v\in B_R\\ \|v\|^2=n}}\frac{1}{n} = \frac{\tilde{c}_d C_f C}{R^4}\sum_{n=1}^{R^2}\frac{r_4(n)}{n} = O_{f,\Theta}(R^{-2}).$$

Here $r_4(n)$ is the number of representations of an integer n as a sum of four squares. To establish this estimate, note that $\sum_{k=1}^{K} r_4(k) = |B_1|K^2 + O(K^{3/2})$, and using Abel's partial summation formula,

$$\sum_{n=1}^{R^2} \frac{r_4(n)}{n} = \frac{1}{R^2+1} \sum_{n=1}^{R^2} r_4(n) + \sum_{n=1}^{R^2} (|B_1|n^2 + O(n^{3/2}))\Big(\frac{1}{n} - \frac{1}{n+1}\Big) = O(R^2).$$

Finally, since the N-components equidistribute at rate $R^{-4(1-\kappa_d)} \log R$, and the term just estimated vanishes faster, the conclusion of part (2) holds. □

3. Further diophantine and geometric applications

Let us now demonstrate some further applications of Theorem 1.1 and Corollary 1.2.

3A. *Lifts of horospheres.* Let Γ be a nonuniform lattice in G, with a cusp at the point σ at the boundary of the associated hyperbolic space. Let H_σ be the unipotent subgroup in G which stabilizes σ (in particular, it is conjugate to N). We consider the case in which $\Gamma \cap H_\sigma$ is a lattice in H_σ. Let $\mathcal{H}$ be a horosphere in the hyperbolic space of G which is based at σ; in other words, $\mathcal{H}$ is an orbit of H_σ. Observe that $\mathcal{H}$ projects to a closed horosphere $\overline{\mathcal{H}}$ in the space $\Gamma \backslash G$. Let $B_T(z)$ denote a hyperbolic ball of radius T that is centered at z, and let $N(T)$ denote the number of horospheres of the form $\gamma\mathcal{H}$, with $\gamma \in \Gamma$, that meet the ball $B_T(z)$. Eskin and McMullen [1993, Theorem 7.2] have considered the counting function $N(T)$ and discussed the case of $G = \mathrm{PSL}_2(\mathbb{R})$. This problem can be formulated for a simple Lie group of any real rank, see [Mohammadi and Salehi Golsefidy 2014]; we will provide an effective estimate for real rank 1.

Theorem 3.1. *Let $\Gamma < G$ be a nonuniform lattice, and let σ, H_σ, $\mathcal{H}$ be as above. If $\Gamma \cap H_\sigma$ is a lattice in H_σ, then*

$$N(T) = \frac{\mathrm{Vol}_{G/\Gamma}(\overline{\mathcal{H}})}{\mu(G/\Gamma)} \cdot \frac{e^{2\rho T}}{2\rho} + O_{z,\Gamma,\mathcal{H}}(T(e^{2\rho T})^{\kappa}),$$

where $\kappa = \kappa(\Gamma)$ is the exponent associated with Γ.

Proof. Assume first that $z = i$ (where i is the point stabilized by K) and that $\sigma = \infty$ (where ∞ is the point stabilized by N), or equivalently $H_\sigma = N$ and $N \cap \Gamma$ is a lattice in N. Then $\mathcal{H}$ is a horizontal horosphere, i.e., it is orthogonal to the geodesic $A \cdot i$, and we may write $\mathcal{H} = Na_y \cdot i$ for some $y \in \mathbb{R}$. Then the set of horospheres $\gamma\mathcal{H}$ that meet the ball $B_T(i)$ is in one-to-one correspondence with the elements of the set

$$\{\gamma N : d(i, \gamma\mathcal{H}) < T\} = \{\gamma N : d(i, \gamma N a_y \cdot i) < T\}.$$

We write the elements of Γ in their KAN coordinates, and denote $\gamma = k_\gamma a_{t(\gamma)} n_\gamma$. Then

$$\begin{aligned}\{\gamma N : d(i, k_\gamma a_{t(\gamma)} n_\gamma N a_y \cdot i) < T\} &= \{\gamma N : d(i, a_{t(\gamma)} N a_y \cdot i) < T\}\\ &= \{\gamma N : d(i, a_{t(\gamma)} N \cdot a_{-t(\gamma)} a_{t(\gamma)} \cdot a_y \cdot i) < T\}\\ &= \{\gamma N : d(i, N a_{t(\gamma)+y} \cdot i) < T\}\\ &= \{\gamma N : d(i, a_{t(\gamma)+y} \cdot i) < T\},\end{aligned}$$

where the last equality follows since the horosphere $Na_{y+t(\gamma)} \cdot i$ is orthogonal to the geodesic $A \cdot i$, and thus the point nearest to i on this horosphere is its meeting point with the geodesic, $a_{y+t(\gamma)} \cdot i$.

Now, $d(i, a_{t(\gamma)+y} \cdot i) = |t(\gamma) + y|$, so $d(i, a_{t(\gamma)+y} \cdot i) < T$ if and only if $-T - y \le t(\gamma) \le T - y$. Moreover, the cosets γN are in one-to-one correspondence with the lattice elements $\gamma = k_\gamma a_{t(\gamma)} n_\gamma$ such that $n_\gamma \in \Psi(\Gamma)$, for a choice $\Psi(\Gamma)$ of a fundamental domain for $\Gamma \cap N$ in N. Then,

$$\begin{aligned}N(T) &= \#\{\gamma = k_\gamma a_{t(\gamma)} n_\gamma : n_\gamma \in \Psi(\Gamma),\ -T - y \le t(\gamma) \le T - y\}\\ &= \#\Gamma \cap (K A_{[-T-y, T-y]} \Psi(\Gamma)).\end{aligned}$$

Now the desired result follows from Theorem 1.1 and Remark 1.4:

$$\begin{aligned}N(T) &= \frac{\mu_N(\Psi(\Gamma))}{\mu(G/\Gamma)} \cdot \frac{e^{2\rho(T-y)}}{2\rho} + O((T-y)(e^{2\rho(T-y)})^\kappa)\\ &= \frac{\mu_N(\Psi(\Gamma))}{\mu(G/\Gamma)} \cdot e^{-2\rho y} \cdot \frac{e^{2\rho T}}{2\rho} + O(T(e^{2\rho(T-y)})^\kappa)\\ &= \frac{\mathrm{Vol}(\overline{\mathcal{H}})}{\mu(G/\Gamma)} \cdot \frac{e^{2\rho T}}{2\rho} + O(T(e^{2\rho T})^\kappa).\end{aligned}$$

The foregoing result is valid for any lattice Γ satisfying that $N \cap \Gamma$ is a lattice in N. Recall that $S = AN$ stabilizes ∞, normalizes N, and acts transitively on the symmetric space. For any z we can choose $s \in S$ such that $s \cdot z = i$, and then $sNs^{-1} \cap s\Gamma s^{-1} = N \cap \Gamma^s$ is a lattice in $N^s = N$ and the previous result applies to Γ^s. Writing $s = a_s n_s$, we have

$$d(i, s\gamma s^{-1} N a_y \cdot i) = d(s^{-1} \cdot i, \gamma s^{-1} N s n_s^{-1} a_s^{-1} a_y \cdot i) = d(z, \gamma N a_{y-s} \cdot i),$$

and we can conclude that the result holds for the lattice Γ and any choice of origin z.

Finally, if the cusp σ of a lattice Γ' is arbitrary, we can conjugate the lattice Γ' and obtain a lattice Γ containing a lattice in N, for which the result applies with any choice of origin. This implies its validity for the lattice Γ', with any choice of origin. $\square$

3B. ***Diophantine equation associated with a Lorentz form.*** When $G = \mathrm{SO}^0(1, n)$, the elements of the subgroups A and N of G can also be given explicitly as

$$a_t = \begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & I_{n-1} & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix} \tag{3-1}$$

and

$$n_v = \begin{pmatrix} 1+\frac{1}{2}\|v\|^2 & v^* & -\frac{1}{2}\|v\|^2 \\ v & I_{n-1} & -v \\ \frac{1}{2}\|v\|^2 & v^* & 1-\frac{1}{2}\|v\|^2 \end{pmatrix}, \tag{3-2}$$

see, e.g., [Faraut 1983, pp. 373 and 375].

The explicit N and A Iwasawa components of a given $g \in G$ can be deduced as follows.

Claim 3.2. *Let*

$$g = \begin{pmatrix} g_{0,0} & \cdots & g_{0,n} \\ \vdots & & \vdots \\ g_{n,0} & \cdots & g_{n,n} \end{pmatrix} \in G.$$

If $g = n_v a_t k$, *then*

$$e^t = (g_{0,0} - g_{n,0})^{-1} \quad \textit{and} \quad v = \frac{1}{g_{0,0}-g_{n,0}} \begin{pmatrix} g_{1,0} \\ \vdots \\ g_{n-1,0} \end{pmatrix}.$$

Proof. On the one hand,

$$g \cdot i = \begin{pmatrix} g_{0,0} & \cdots & g_{0,n} \\ \vdots & & \vdots \\ g_{n,0} & \cdots & g_{n,n} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} g_{0,0} \\ \vdots \\ g_{n,0} \end{pmatrix}.$$

On the other hand,

$$g \cdot i = n_v a_t k \cdot i = n_v a_t \cdot i,$$

where

$$n_v a_t \cdot i = \begin{pmatrix} 1+\frac{1}{2}\|v\|^2 & v^* & -\frac{1}{2}\|v\|^2 \\ v & I_{n-1} & -v \\ \frac{1}{2}\|v\|^2 & v^* & 1-\frac{1}{2}\|v\|^2 \end{pmatrix} \begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & I_{n-1} & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \cosh t + \frac{1}{2}e^{-t}\|v\|^2 \\ e^{-t} \cdot v \\ \sinh t + \frac{1}{2}e^{-t}\|v\|^2 \end{pmatrix}.$$

Namely,

$$\begin{pmatrix} g_{0,0} \\ \vdots \\ g_{n,0} \end{pmatrix} = \begin{pmatrix} \cosh t + \frac{1}{2} e^{-t} \|v\|^2 \\ e^{-t} \cdot v \\ \sinh t + \frac{1}{2} e^{-t} \|v\|^2 \end{pmatrix}.$$

In particular,

$$g_{0,0} - g_{n,0} = \cosh t - \sinh t = e^{-t},$$

and

$$\begin{pmatrix} g_{1,0} \\ \vdots \\ g_{n-1,0} \end{pmatrix} = e^{-t} \cdot v.$$

The expression for e^t and v easily follows from the foregoing equations. $\square$

$\mathrm{SO}^0(1,n)$ is the isometry group of the Lorentz form, and each element $g = (g_{i,j})_{0 \le i,j \le n}$ satisfies $g_{0,0} > 0$ and $g_{0,0}^2 - g_{1,0}^2 - \cdots - g_{n,0}^2 = 1$. Namely, the first column of g satisfies the Lorentz equation

$$x_0^2 - x_1^2 - \cdots - x_n^2 = 1, \quad x_0 > 0. \tag{3-3}$$

Conversely, every such vector $(x_0, \ldots, x_n)$ appears as the first column of a matrix $(g_{i,j})$ belonging to the connected component of the isometry groups of the Lorentz form.

By Claim 3.2 above, the N and A components of g depend only on the first column of g. Hence, Corollary 1.2 concerning the equidistribution of the N-components as the A-components approach ∞ can be used to study the behavior of the corresponding parameters of solutions to the Lorentz equation (3-3).

For every $\underline{x} = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$ that satisfies (3-3) with $x_0 > 0$, note that $x_0 > x_n$ and define the height function $h(\underline{x}) = \log(1/(x_0 - x_n))$ (corresponding to the A-component). Define also the vector $v(\underline{x}) = (1/(x_0 - x_n))(x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$ (corresponding to the N-component). Assume $\Psi \subset \mathbb{R}^{n-1}$ is nice, and consider the family $R_T(\Psi, K) \subset G$. By applying Corollary 1.2 (part (2)) to the lattice $\mathrm{SO}^0(1,n)(\mathbb{Z})$ in $\mathrm{SO}^0(1,n)$, we conclude:

Corollary 3.3. *Let S_T denote the set of integral solutions $\underline{x}$ to the Lorentz equation* (3-3) *with $-T \le h(\underline{x}) \le 0$. The rational vectors $v(\underline{x})$, $\underline{x} \in S_T$ become effectively equidistributed in Ψ as $T \to \infty$, at rate $O(e^{2\rho T(\kappa_\Gamma - 1)})$.*

3C. ***Equidistribution of Iwasawa coordinates: history of the problem.*** The problem of joint equidistribution of the Iwasawa coordinates of nonuniform lattices has quite a long history. We summarize the relevant results we are aware of as follows.

The analysis by Selberg and Good of the $\mathrm{SL}_2(\mathbb{R})$ *case.* Consider a nonuniform lattice $\Gamma \subset \mathrm{SL}_2(\mathbb{R}) = NAK$, with a maximal parabolic subgroup Γ_∞ equal to $N(\mathbb{Z})$, so that its fundamental domain in N is (say) [0, 1]. The joint equidistribution (in $[0, 1] \times K$) of the N component and the K components of the lattice elements was established by Selberg [1991]. This result was also proved by Good [1983, Corollary p. 119; 1984, p. 101].

In particular, the problem of counting the number of points $\gamma \in \Gamma/\Gamma_\infty$ with $\mathrm{Im}(\gamma i) \geq T^{-1}$ corresponds to taking Ψ to be a fundamental domain of $\Gamma_\infty \subset N$ in Theorem 1.1, and $\Phi = K$, and amounts to establishing equidistribution for the real parts of the orbit points. This problem was considered in detail by Good [1983, § 11], using a thorough analysis of generalized Kloosterman sums. The error term in this problem stated by Good, who attributes it also to Selberg, is superior to the one which follows from Theorem 1.1 in this case. For a tempered lattice $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$ as above, the error exponent stated by Good [1983, § 11, Theorem 4] amounts to $\frac{2}{3}$, as opposed to $\frac{7}{8}$ which follows from Theorem 1.1.

Good's analysis [1983] can be elaborated further and used to derive effective joint equidistribution of both the N and the K components of lattice elements, which again will be superior to Theorem 1.1 in this case. Apparently, this elaboration has not been carried out in that paper and it is not clear what is the error estimate it provides.

The analysis of Selberg and Good is based on intricate estimates in harmonic analysis of automorphic forms and generalized Kloosterman sums, and was confined to the $\mathrm{SL}_2(\mathbb{R})$ case only. We note that Selberg [1991, § 3] raises specifically the problem of extending the joint equidistribution statement to higher-dimensional hyperbolic spaces. This problem is given an effective solution in Corollary 1.2 of the present paper. A superior error estimate could in principle be derived by extending the analysis of Selberg and Good to higher-dimensional hyperbolic spaces, but judging by [Good 1983], this is quite a formidable task.

From equidistribution of real parts of lattice orbits to the gcd equation in $\mathbb{Z}^2$. The problem of analyzing the distribution of the shortest solution to the gcd equation in $\mathbb{Z}^2$ was considered by Dinaburg and Sinai [1990] who measured the size of the shortest solution by the maximum norm, and used the theory of continued fractions. It was subsequently observed by Risager and Rudnick [2009] that when the size of the smallest solution is measured using the euclidean norm, equidistribution of shortest solutions is equivalent to the problem of equidistribution of real parts of the points in the orbit of i under $\mathrm{SL}_2(\mathbb{Z})$ in the upper half-plane, and the latter result has already been established by Good [1983]. This elegant observation has provided the motivation for the present paper. Following Risager and Rudnick, Truelsen [2013] has established a quantitative form for the equidistribution of real parts for any lattice with a standard cusp in $\mathrm{SL}_2(\mathbb{R})$. In particular, Truelsen established a rate of convergence in the equidistribution of the lengths of shortest solutions of the gcd equation, which is $R^{-1/4+\epsilon}$ in the case of tempered lattices. Theorem 2.1 provides the error estimate of $R^{-1/4}\log R$, and shows that effective equidistribution at this rate is valid with v varying in any given fixed angular sector of radius R. Note that Truelsen's results are confined to lattices in $\mathrm{SL}_2(\mathbb{R})$, and do not consider lattices in $\mathrm{SL}_2(\mathbb{C})$, for example.

Counting in general domains. Chamizo [2011] and Truelsen [2013] have established some further lattice point counting results for a variety of other families of sets in the upper half-plane. Theorem 1.1 also allows for a large collection of families of increasing domains in hyperbolic spaces, namely $R_T(\Psi, \Phi)$, with $\Psi \subset N$ and $\Phi \subset K$ any nice sets. These families are also called bisectors associated with the decomposition $G = NAK$. Mohammadi and Oh [2015, Theorem 7.21] have considered the problem of counting the points of a Zariski-dense geometrically finite discrete subgroup Λ in such sets. Under some restrictions they have obtained an asymptotic estimate for the count, with unspecified rate. A generalization of this discussion to the case of discrete geometrically finite groups in general real, complex and quaternionic hyperbolic spaces was established by Kim [2015].

Lifts of horospheres in hyperbolic space. Eskin and McMullen [1993] have raised the problem of counting the number of lifts of a closed horosphere $\mathcal{H}$ in G/Γ which intersect a ball of radius T in hyperbolic space, and have established the main term for this counting problem. As shown in Theorem 3.1, in the case of hyperbolic space, the problem amounts to counting the points of a nonuniform lattice lying in the sets $R_T(\Psi, K)$, where $\Psi \subset \mathbb{R}^{n-1}$ is the fundamental domain of $\Gamma_\infty = \Gamma \cap N$, and thus is solved effectively by Theorem 1.1.

The problem can be formulated also for higher-rank symmetric space, and the main term of the asymptotics for counting such lifts has been established recently by Mohammadi and Salehi Golsefidy [2014]. Further work on the subject has been recently carried out by Dabbs, Kelly and Li in [Dabbs et al. 2016], where an effective count for the lift of horospheres in certain higher rank locally symmetric spaces was established.

Local statistics of the Iwasawa N-component. Marklof and Vinogradov [2018] have recently considered, among other things, the projection of lattice orbit points to a neighborhood of a horizontal horosphere tending to the boundary, namely the sets given by $R_T(\Psi, K) \setminus R_{T-c}(\Psi, K)$. Theorem 1.1 implies without difficulty, as we shall see below, an effective counting result for these sets as well. In their work, they have analyzed the local statistics of the Iwasawa N-components in Ψ as $T \to \infty$. This problem is more delicate than just the equidistribution of the N-component, and it was established by them in real hyperbolic space of any dimension, but not in an effective form.

4. Proof of the main theorem

4A. ***A spectral method for counting lattice points.*** In the discussion of the present section, G is a connected almost simple Lie group, not necessarily of real rank 1. The lattice point counting method in the family of domains $\{\mathcal{B}_T\} \subset G$ that we will use [Gorodnik and Nevo 2010; 2012] has two ingredients: a spectral estimate and a regularity property. The crucial spectral estimate requires bounding the norm of the averaging operators defined by $\mathcal{B}_T$ in the representation on $L^2_0(\Gamma\backslash G)$. Let us recall the fact that there exists $m \in \mathbb{N}$ such that the unitary representation of G in $L^2_0(\Gamma\backslash G)$, when taken to the m-th tensor power, is weakly contained in the regular representation of G. The essential property of such m is that $m \geq \frac{1}{2}p$, where p satisfies that the K-finite matrix coefficients of $\pi^0_{\Gamma\backslash G}$ are in $L^{p+\zeta}(G)$ for every $\zeta > 0$. We define $m(\Gamma)$ to be the least even integer with this property if $p > 2$, or 1 if $p = 2$; see [Gorodnik and Nevo 2012, Definition 3.1]. One of the remarkable features of harmonic analysis on simple Lie groups is that then for any measurable set of positive finite measure B in G, if we denote by β the Haar uniform measure on B divided by $\mu(B)$, the estimate

$$\|\pi^0_{\Gamma\backslash G}(\beta)\| \leq C_{G,\zeta} \cdot \mu(B)^{-1/(2m(\Gamma))+\zeta} \quad \text{for every } \zeta > 0 \tag{4-1}$$

holds [Nevo 1998]. Thus, $m(\Gamma)$ measures the size of the spectral gap in $L^2(\Gamma\backslash G)$. The lattice Γ is called *tempered* if the representation $\pi^0_{\Gamma\backslash G}$ is already weakly contained in regular representation, namely if $m(\Gamma) = 1$.

We now turn to the second ingredient, which is the Lipschitz property of the domains $\mathcal{B}_T$.

Definition 4.1 [Gorodnik and Nevo 2012]. Let G be a Lie group with Haar measure m_G. Assume $\{\mathcal{B}_T\} \subset G$ is a family of bounded measurable sets such that $\mu(\mathcal{B}_T) \to \infty$ as $T \to \infty$. Let $\mathcal{O}_\epsilon \subset G$ be the image of a euclidean ball of radius ϵ in the Lie algebra under the exponential map. Denote

$$\mathcal{B}^+_T(\epsilon) := \mathcal{O}_\epsilon \mathcal{B}_T \mathcal{O}_\epsilon = \bigcup_{u,v\in\mathcal{O}_\epsilon} u\mathcal{B}_T v \quad \text{and} \quad \mathcal{B}^-_T(\epsilon) := \bigcap_{u,v\in\mathcal{O}_\epsilon} u\mathcal{B}_T v.$$

The family $\{\mathcal{B}_T\}$ is *Lipschitz well-rounded* if there exist $\epsilon_0 > 0$ and $T_0 \geq 0$ such that for every $0 < \epsilon \leq \epsilon_0$ and $T \geq T_0$,

$$\mu(\mathcal{B}^+_T(\epsilon)) \leq (1 + C\epsilon)\mu(\mathcal{B}^-_T(\epsilon)),$$

where $C > 0$ is a constant that does not depend on ϵ or T.

The concept of well-roundedness appeared first in [Duke et al. 1993; Eskin and McMullen 1993], and was used later also in [Gorodnik and Weiss 2007]. Lipschitz well-roundedness was applied extensively in [Gorodnik and Nevo 2012].

Theorem 4.2 [Gorodnik and Nevo 2012]. *Let G be a connected almost simple Lie group with Haar measure m_G, and let $\Gamma < G$ be a lattice. Assume $\{\mathcal{B}_T\} \subset G$ is a family of bounded measurable sets which satisfies $\mu(\mathcal{B}_T) \to \infty$ as $T \to \infty$. If the family $\{\mathcal{B}_T\}$ is Lipschitz well-rounded, then*

$$\#(\mathcal{B}_T \cap \Gamma) = \frac{1}{\mu(G/\Gamma)}\mu(\mathcal{B}_T) + O(\mu(\mathcal{B}_T) \cdot E(T)^{1/(1+\dim G)})$$

as $T \to \infty$, where $\mu(G/\Gamma)$ is the measure of a fundamental domain of Γ in G, and $E(T)$ is (a bound on) the rate of decay of operator norm $\|\pi^0_{\Gamma\backslash G}(\beta_T)\|$.

Note that the above theorem applies to every lattice Γ.

When using the estimate (4-1) for $\|\pi^0_{\Gamma\backslash G}(\beta_T)\|$, the error term obtained in Theorem 4.2 is

$$O(\mu(\mathcal{B}_T)^{\kappa(\Gamma)+\zeta}) \quad \text{for every } \zeta > 0,$$

where

$$\kappa(\Gamma) = 1 - \frac{1}{2m(\Gamma)(1+\dim G)} \in (0, 1). \tag{4-2}$$

In our case, where G is of real rank 1 and the family of domains is $R_T(\Psi, \Phi)$, the estimate (4-1) may be very slightly improved so that the error term is given by

$$O\big((\log \mu(R_T(\Psi, \Phi)) \cdot \mu(R_T(\Psi, \Phi)))^{\kappa(\Gamma)}\big),$$

as we now explain. Assume that a set $B \subset G$ of positive finite measure satisfies that

$$\mu(K \cdot B \cdot K) \le \text{const} \cdot \mu(B).$$

This property is called K-radializability; see [Gorodnik and Nevo 2010, Definition 3.21]. When B is radializable, then it is a consequence of the spectral transfer principle [Nevo 1998] and of estimates of the Harish-Chandra function in real rank 1 that

$$\|\pi^0_{\Gamma\backslash G}(\beta)\| \le C_G \cdot (\log \mu(B))^{\frac{1}{m(\Gamma)}} \cdot \mu(B)^{-\frac{1}{2m(\Gamma)}} \le C_G \cdot (\log \mu(B)) \cdot \mu(B)^{-\frac{1}{2m(\Gamma)}};$$

see [Gorodnik and Nevo 2010, Proposition 5.9; Nevo 1998, §2.2, Theorem 6]. The sets $R_T(\Psi, \Phi)$ are indeed radializable, with constant that depends on Ξ and Φ but does not depend on T. In particular, when β_T are the probability measures that correspond to $R_T = R_T(\Psi, \Phi)$, then

$$E(T) = \|\pi^0_{\Gamma\backslash G}(\beta_T)\| \le C_G \cdot (\log \mu(R_T)) \cdot \mu(R_T)^{-\frac{1}{2m(\Gamma)}},$$

as claimed.

From the above discussion it follows that *in order to prove Theorem 1.1, it suffices to show that the family $\{R_T(\Psi, \Phi)\}$ is Lipschitz well-rounded.*

G	$\mathrm{SO}^0(1,n)$	$\mathrm{SU}(1,n)$	$\mathrm{SP}(1,n)$	$\mathrm{F}_{4(-20)}$
$\mathbb{K}$	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	$\mathbb{O}$
N (as a manifold)	$\mathbb{R}^{n-1}$	$\mathbb{C}^{n-1}\oplus\mathbb{R}$	$\mathbb{H}^{n-1}\oplus\mathbb{R}^3$	$\mathbb{O}\oplus\mathbb{R}^7$
K	$\mathrm{SO}(n)$	$\mathrm{S}(\mathrm{U}(1)\times\mathrm{U}(n))$	$\mathrm{SP}(1)\times\mathrm{SP}(n)$	$\mathrm{Spin}(9)$
G/K	$\mathbf{H}^n_{\mathbb{R}}$	$\mathbf{H}^n_{\mathbb{C}}$	$\mathbf{H}^n_{\mathbb{H}}$	$\mathbf{H}^2_{\mathbb{O}}$
(p,q)	$(n-1,0)$	$(2n-2,1)$	$(4n-4,3)$	$(8,7)$
$dt/e^{2\rho t}$	$dt/e^{(n-1)t}$	dt/e^{2nt}	$dt/e^{(4n+2)t}$	dt/e^{22t}

Table 1. Simple rank 1 Lie groups: Iwasawa subgroups, symmetric spaces and Haar measure $\mu=\mu_N\times(dt/e^{2\rho t})\times\mu_K$.

4B. ***Lipschitz property for Iwasawa coordinates in the negative direction of A.*** In order to show that the family $R_T(\Psi,\Phi)$ is Lipschitz well-rounded, we introduce coordinates on N, in addition to the parametrization we have already introduced above for A, namely $A=\{a_t : t\in\mathbb{R}\}$, where $a_t=\exp tH_1$, and $H_1\in\mathfrak{a}=\mathrm{Lie}(A)$ is the element satisfying $\alpha(H_1)=1$, with α the unique short positive root.

Let $\mathbb{K}\in\{\mathbb{R},\mathbb{C},\mathbb{H}\}$ be the (possibly noncommutative) field over which the matrices in G are defined, and n the dimension (over $\mathbb{K}$) of the corresponding hyperbolic space. The group N is of Heisenberg type (see [Cowling et al. 1991; 1998]), and in particular it is parametrized by the space $\mathbb{K}^n\oplus\Im(\mathbb{K})$, where $\Im(\mathbb{K})$ is the subspace of "pure imaginary" numbers in $\mathbb{K}$, namely of elements w such that $w+\mathrm{conj}(w)=0$. A parametrization may be chosen such that

$$N=\{n_{v,z} : v\in\mathbb{K}^n,\ z\in\Im(\mathbb{K})\},$$

with the group multiplication

$$n_{v_1,z_1}n_{v_2,z_2}=n_{v_1+v_2,z_1+z_2+\Im(\langle v_2,v_1\rangle)},$$

where $\langle v_2,v_1\rangle=v_1^*v_2$. The subspaces $\mathbb{K}^n$ and $\Im(\mathbb{K})$ correspond to subsets of N that are invariant under conjugation by A, and specifically,

$$a_t n_{v,z} a_{-t}=n_{e^t v,e^{2t}z}. \tag{4-3}$$

As a result, if $p:=\dim_{\mathbb{R}}(\mathbb{K}^n)$ and $q:=\dim_{\mathbb{R}}(\Im(\mathbb{K}))=\dim_{\mathbb{R}}(\mathbb{K})-1$, then μ_N is the Lebesgue measure on $\mathbb{R}^{p+q}$, and the parameter ρ that appears in (1-1) for the Haar measure equals $\frac{1}{2}(p+2q)$.

Let $\overline{N}$ denote the opposite unipotent group, namely the one that corresponds to the negative roots

$$a_t\bar{n}_{v,z}a_{-t}=n_{e^{-t}v,e^{-2t}z}. \tag{4-4}$$

On the subgroups $H\in\{A,K\}$ we consider the metric d_H induced by the riemannian metric on G. We denote by $K_{(\phi,\delta)}$ the open ball in K with center $\phi\in K$ and radius δ, and by $A_{(t,\delta)}$ the open ball in A, with center t and radius δ (these are simply the elements that correspond to the interval $(t-\delta, t+\delta)$, since d_A is the euclidean metric on $\mathbb{R}$). On the product space $N = \mathbb{R}^p \times \mathbb{R}^q$ we let d_N denote the maximum of the two euclidean metrics $d_N^{(1)}, d_N^{(2)}$ on the components $\mathbb{K}^n \cong \mathbb{R}^p$ and $\Im(\mathbb{K}) \cong \mathbb{R}^q$, and let $N_{(v,\delta_1)\times(z,\delta_2)}$ be the domain in N parametrized by the product of open euclidean balls in $\mathbb{K}^n \cong \mathbb{R}^p$ and $\Im(\mathbb{K}) \cong \mathbb{R}^q$ with centers v, z and radii δ_1, δ_2 respectively. When a ball is centered at the identity we omit the center and denote it by $K_{(\delta)}$, $A_{(\delta)}$, and $N_{(\delta_1)\times(\delta_2)}$.

In what follows, $\|\cdot\|_{\mathrm{CK}} = \|\cdot\|$ is the Cartan–Killing norm on the Lie algebra $\mathrm{Lie}(G)$ of G, and $\|\cdot\|_{\mathrm{op}}$ is the norm on the space of linear operators on $\mathrm{Lie}(G)$.

Lemma 4.3. *Let $G \subset \mathrm{SL}_N(\mathbb{R})$ be a connected semisimple linear Lie group. Let $\mathcal{O}_\epsilon = \exp(B_\epsilon)$, where $B_\epsilon = \{X \in \mathrm{Lie}(G) : \|X\|_{CK} < \epsilon\}$. For every $g \in G$, the following inclusion holds*:

$$g\mathcal{O}_\epsilon g^{-1} \subseteq \mathcal{O}_{\epsilon\cdot\|\mathrm{Ad}\, g\|_{\mathrm{op}}} = \exp\{X \in \mathrm{Lie}(G) : \|X\|_{\mathrm{CK}} < \epsilon \cdot \|\mathrm{Ad}\, g\|_{\mathrm{op}}\}.$$

Proof. The operator norm is defined by

$$\|\mathrm{Ad}\, g\|_{\mathrm{op}} = \max_{\|X\|\le 1} \|\mathrm{Ad}\, g(X)\| = \max_{\|X\|\le 1} \|gXg^{-1}\|.$$

Therefore $\mathrm{Ad}\, g(B_\epsilon) \subset \mathrm{Lie}(G)$ is contained in a ball of radius

$$\max_{X\in B_\epsilon} \|\mathrm{Ad}\, g(X)\| = \max_{X\in B_\epsilon} \|gXg^{-1}\| = \max_{X\in\epsilon B_1} \|gXg^{-1}\| = \max_{X\in B_1} \|g\epsilon Xg^{-1}\| = \epsilon \|\mathrm{Ad}\, g\|_{\mathrm{op}}.$$

Since the exponential function $e^X : M_N(\mathbb{R}) \to M_N(\mathbb{R})$ has a convergent power series expansion at every point X, it follows that $ge^Xg^{-1} = e^{gXg^{-1}}$ for every $g \in \mathrm{GL}_N(\mathbb{R})$.

Therefore

$$\begin{aligned} g\mathcal{O}_\epsilon g^{-1} = g\exp(B_\epsilon)g^{-1} = \exp(gB_\epsilon g^{-1}) &= \exp(\mathrm{Ad}\, g(B_\epsilon)) \\ &\subseteq \exp(B_{\epsilon\cdot\|\mathrm{Ad}\, g\|_{\mathrm{op}}}) = \mathcal{O}_{\epsilon\cdot\|\mathrm{Ad}\, g\|_{\mathrm{op}}}. \qquad \square \end{aligned}$$

Let M denote the centralizer of A in K. We will use that there exists $\delta_0 > 0$ and positive constants c_1, c_2, c_1', c_2' such that for all $0 < \delta < \delta_0$,

$$\mathcal{O}_\delta \subseteq N_{(c_1\delta)\times(c_1\delta)} A_{(c_1\delta)} K_{(c_1\delta)} \subseteq \mathcal{O}_{c_1' c_1 \delta}, \tag{4-5}$$

$$\mathcal{O}_\delta \subseteq N_{(c_2\delta)\times(c_2\delta)} A_{(c_2\delta)} M_{(c_2\delta)} \overline{N}_{(c_2\delta)\times(c_2\delta)} \subseteq \mathcal{O}_{c_2' c_2 \delta}, \tag{4-6}$$

the latter being the Bruhat coordinates on a neighborhood of the identity in G.

The foregoing inclusions are consequences of the following fact. Let there be given a decomposition of a Lie algebra as a direct sum of Lie subalgebras, and let there be given any basis of the Lie algebra. Then there exists a closed ball of fixed size centered at 0 in the Lie algebra satisfying the following property. The map assigning the canonical coordinates associated with the first decomposition (of either the first or the second kind) to the canonical coordinates associated with the given basis (of either the first or the second kind), is an invertible smooth map with all its derivatives (and the derivatives of its inverse) bounded. In particular, it is a bi-Lipschitz map.

Proposition 4.4 (effective Iwasawa decomposition). *Let $n_{v,z} \in N$, $\phi \in K$, $a_t \in A$ with $t \leq 0$. There exists $\epsilon_1 > 0$ such that for every $0 < \epsilon < \epsilon_1$ there are positive constants C'_N, C''_N, C_A, C_K that depend only on $n_{v,z}$ and ϕ (in particular, independent of t) such that*

$$\mathcal{O}_\epsilon \cdot n_{v,z} a_t \phi \cdot \mathcal{O}_\epsilon \subseteq N_{(v, C'_N \epsilon)\times(z, C''_N \epsilon)} A_{(t, C_A \epsilon)} K_{(\phi, C_K \epsilon)}.$$

Furthermore, when $n_{v,z}$ varies over a compact set Ψ, and ϕ varies over K, these constants can be taken to be uniform.

Proof. Observe that

$$N_{(\delta_1)\times(\delta_2)} N_{(\rho_1)\times(\rho_2)} \subseteq N_{(\delta_1+\rho_1)\times(\delta_2+\rho_2+\rho_1\delta_1)}$$

and

$$n_{v,z} N_{(\rho_1)\times(\rho_2)} \subseteq N_{(v,\rho_1)\times(z,\rho_2+\|v\|\rho_1)}. \tag{4-7}$$

In particular,

$$\begin{aligned} n_{v,z} N_{(\delta_1)\times(\delta_2)} N_{(\rho_1)\times(\rho_2)} &\subseteq n_{v,z} N_{(\delta_1+\rho_1)\times(\delta_2+\rho_2+\rho_1\delta_1)} \\ &\subseteq N_{(v,\delta_1+\rho_1)\times(z,\delta_2+\rho_2+\rho_1\delta_1+\|v\|(\delta_1+\rho_1))}. \end{aligned} \tag{4-8}$$

Finally, note that

$$K_{(\delta)}\phi \subset K_{(\phi,\delta)}. \tag{4-9}$$

Step 1: Right perturbations. We show that

$$n_{v,z} a_t \phi \cdot \mathcal{O}_\epsilon \subseteq N_{(v,r_1\epsilon)\times(z,r_2\epsilon)} A_{(t,r_3\epsilon)} K_{(\phi,r_4\epsilon)}$$

where $r_i = r_i(v, z)$ is independent of $t \leq 0$. Recall $\|\mathrm{Ad}\,\phi\| = 1$. Then, by Lemma 4.3,

$$\begin{aligned} n_{v,z} a_t \phi \cdot \mathcal{O}_\epsilon \subseteq n_{v,z} a_t \mathcal{O}_\epsilon \phi &\overset{(4\text{-}5)}{\subseteq} n_{v,z} a_t (N_{(c_1\epsilon)\times(c_1\epsilon)} A_{(c_1\epsilon)} K_{(c_1\epsilon)})\phi \\ &\overset{(4\text{-}3)}{\subseteq} n_{v,z} N_{(e^t c_1\epsilon)\times(e^{2t} c_1\epsilon)} a_t A_{(c_1\epsilon)} K_{(c_1\epsilon)} \phi \\ &\overset{(4\text{-}7),(4\text{-}9)}{\subseteq} N_{(v, c_1 e^t\epsilon)\times(z, c_1 e^{2t}\epsilon + c_1\|v\| e^t\epsilon)} A_{(t,c_1\epsilon)} K_{(\phi,c_1\epsilon)} \\ &\subseteq N_{(v,c_1\epsilon)\times(z,c_1\epsilon+c_1\|v\|\epsilon)} A_{(t,c_1\epsilon)} K_{(\phi,c_1\epsilon)}, \end{aligned}$$

the latter inclusion holding since $e^t \leq 1$.

Step 2: Left perturbations. We show that

$$\mathcal{O}_\epsilon \cdot n_{v,z} a_t \phi \subseteq N_{(v,\ell_1\epsilon)\times(z,\ell_2\epsilon)} A_{(t,\ell_3\epsilon)} K_{(\phi,\ell_4\epsilon)}$$

where $\ell_i = \ell_i(v,z)$ is independent of $t \le 0$.

Denote $\eta = \|\mathrm{Ad}\, n_{v,z}^{-1}\|_{\mathrm{op}}$. By Lemma 4.3,

$$\mathcal{O}_\epsilon \cdot n_{v,z} a_t \phi \subseteq n_{v,z} \mathcal{O}_{\eta\epsilon} a_t \phi.$$

(We note that we will apply this argument below to $n_{v,z} \in \Psi \subset N$ for a fixed bounded set Ψ).

Assume $0 < \epsilon < \min\{1, \delta_0/\eta\}$. Then

$$\begin{aligned} n_{v,z}\mathcal{O}_{\eta\epsilon} a_t \phi &\overset{(4\text{-}6)}{\subseteq} n_{v,z}(N_{(c_2\eta\epsilon)\times(c_2\eta\epsilon)} A_{(c_2\eta\epsilon)} M_{(c_2\eta\epsilon)} \overline{N}_{(c_2\eta\epsilon)\times(c_2\eta\epsilon)}) a_t \phi \\ &\overset{(4\text{-}4)}{\subseteq} n_{v,z} N_{(c_2\eta\epsilon)\times(c_2\eta\epsilon)} A_{(c_2\eta\epsilon)} a_t M_{(c_2\eta\epsilon)} \overline{N}_{(c_2 e^t\eta\epsilon)\times(c_2 e^{2t}\eta\epsilon)} \phi \\ &\subseteq n_{v,z} N_{(c_2\eta\epsilon)\times(c_2\eta\epsilon)} A_{(c_2\eta\epsilon)} a_t (\mathcal{O}_{c_2' c_2\eta\epsilon})\phi, \end{aligned}$$

the latter inclusion coming from the second inclusion in (4-6), since $1 \ge e^t \ge e^{2t}$ and $M_{(c_2\eta\epsilon)} \overline{N}_{(c_2 e^t\eta\epsilon)\times(c_2 e^{2t}\eta\epsilon)} \subset \mathcal{O}_{c_2' c_2\eta\epsilon}$. By the first inclusion in (4-5), provided $0 < \epsilon < \delta_0/(c_2' c_2 \eta)$, this is included in

$$n_{v,z} N_{(c_2\eta\epsilon)\times(c_2\eta\epsilon)} A_{(t,c_2\eta\epsilon)} (N_{(c_1c_2'c_2\eta\epsilon)\times(c_1c_2'c_2\eta\epsilon)} A_{(c_1c_2'c_2\eta\epsilon)} K_{(c_2'c_1c_2\eta\epsilon)})\phi,$$

and by (4-3) and (4-9), the above set is included in

$$n_{v,z} N_{(c_2\eta\epsilon)\times(c_2\eta\epsilon)} N_{(e^{t+c_2\eta\epsilon}\cdot c_1c_2'c_2\eta\epsilon)\times(e^{2(t+c_2\eta\epsilon)}\cdot c_1c_2'c_2\eta\epsilon)} A_{(t,c_2\eta\epsilon)} A_{(c_1c_2'c_2\eta\epsilon)} K_{(\phi,c_1c_2'c_2\eta\epsilon)}.$$

Hence by (4-8), using $e^t \le 1$, for every $\epsilon < \epsilon_1$ where ϵ_1 satisfies the two foregoing conditions, we obtain that the above set is included in

$$N_{(v,(1+c_1c_2'e^{c_2\eta})c_2\eta\epsilon)\times(z,(1+c_1c_2'e^{2c_2\eta\epsilon}(1+c_2\eta\epsilon)+\|v\|(1+e^{2c_2\eta\epsilon}c_1c_2')c_2\eta\epsilon))} A_{(t,(1+c_1c_2')c_2\eta\epsilon)} K_{(\phi,c_1c_2'c_2\eta\epsilon)}.$$

Step 3: Combining left and right perturbations. Let $g := n_{v,z}a_t\phi$ with $t \le 0$ and let $\epsilon < \epsilon_1$. Fix positive constants $\bar{\ell}_i = \max\{\ell_i(v',z') : n_{v',z'} \in \pi_N(g\cdot\mathcal{O}_1)\}$, which are uniform, namely independent of t. Since $g\cdot\mathcal{O}_\epsilon \subset g\cdot\mathcal{O}_1$, it follows from Step 2 that for every

$$g_0 = n_{v_0,z_0} a_{t_0} \phi_0 \in g\cdot\mathcal{O}_\epsilon,$$

it holds that

$$\mathcal{O}_\epsilon \cdot g_0 \subset N_{(v_0,\bar{\ell}_1\epsilon)\times(z_0,\bar{\ell}_2\epsilon)} A_{(t_0,\bar{\ell}_3\epsilon)} K_{(\phi_0,\bar{\ell}_4\epsilon)}.$$

But, as was shown in Step 1, $d_N^{(1)}(v_0,v) < r_1\epsilon$, $d_N^{(2)}(z_0,z) < r_2\epsilon$, $d_A(t_0,t) < r_3\epsilon$ and $d_K(\phi_0,\phi) < r_4\epsilon$. Then by the triangle inequality for the metrics $d_N^{(1)}, d_N^{(2)}, d_A, d_K$,

$$\mathcal{O}_\epsilon \cdot g \cdot \mathcal{O}_\epsilon \subset N_{(v,r_1\epsilon+\bar{\ell}_1\epsilon)\times(z,r_2\epsilon+\bar{\ell}_2\epsilon)} A_{(t,r_3\epsilon+\bar{\ell}_3\epsilon)} K_{(\phi,r_4\epsilon+\bar{\ell}_4\epsilon)}. \qquad \square$$

4C. ***Lipschitz regularity of the domains $R_T(\Psi, \Phi)$.*** Recall that we wish to show that the family $\{R_T(\Psi, \Phi)\}_{T>0}$ is Lipschitz well-rounded (Definition 4.1). Since we have already established the Lipschitz property for the Iwasawa coordinates in the negative direction of A, all that remains is to bound the quotient of the measures of $R_T(\Psi, \Phi)^+(\epsilon)$ and $R_T(\Psi, \Phi)^-(\epsilon)$, which we perform below.

Proof of Theorem 1.1. Throughout this proof, it will be convenient to parametrize N as $\mathbb{R}^{p+q}$ instead of $\mathbb{R}^p \oplus \mathbb{R}^q$. We will write $n_{\underline{x}}$ instead of $n_{v,z}$, and $N_{(\underline{x},\delta)} = N_{(v,\delta),(z,\delta)}$ for a ball of radius δ centered at $\underline{x} = (v, z)$. It will also be convenient to denote $\mu_A = dt/e^{2\rho t}$, and then $\mu = \mu_N \times \mu_A \times \mu_K$.

The proof will proceed by showing that there exists $\epsilon_0 > 0$, which will be described explicitly below, such that for $0 < \epsilon < \epsilon_0$, $R_T(\Psi, \Phi)^+(\epsilon)$ is contained in a product set of the form $\Psi^+ A_{-T^+,S^+} \Phi^+$, and $R_T(\Psi, \Phi)^-(\epsilon)$ contains a product set of the form $\Psi^- A_{-T^-,S^-} \Phi^-$, with the following property. The ratio of the measure of the three components of $\Psi^+ A_{-T^+,S^+} \Phi^+$ to the corresponding components of $\Psi^- A_{-T^-,S^-} \Phi^-$ is bounded by $1 + C\epsilon$, for $0 < \epsilon < \epsilon_0$, $T \geq T_0$. It then follows immediately that

$$\frac{\mu(R_T(\Psi, \Phi)^+(\epsilon))}{\mu(R_T(\Psi, \Phi)^-(\epsilon))} \leq 1 + C'\epsilon.$$

To construct the sets alluded to above, recall first that for every $H \in \{N, A, K\}$, $\xi \in H$ and $\delta > 0$, $H_{(\xi,\delta)}$ denotes the open ball of radius δ centered at ξ with respect to the metric d_H on H. By Proposition 4.4 there exist positive constants C_N, C_A, C_K that depend on Ψ and Φ alone such that for every $\underline{x} \in \Psi$, $\phi \in \Phi$, $0 < \epsilon < \epsilon_1$ and $t \leq 0$,

$$\mathcal{O}_\epsilon \cdot n_{\underline{x}} a_t k_\phi \cdot \mathcal{O}_\epsilon \subseteq N_{(\underline{x}, C_N\epsilon)} A_{(t, C_A\epsilon)} K_{(\phi, C_K\epsilon)}.$$

We now claim that the inclusions

$$\pi_H(R_T(\Psi, \Phi)^+(\epsilon)) \subseteq \bigcup_{\xi\in\Xi} H_{(\xi, C_H\epsilon)}, \tag{4-10}$$

$$\pi_H(R_T(\Psi, \Phi)^-(\epsilon)) \supseteq \bigcup_{\xi\in\Xi} H_{(\xi, C_H\epsilon)} \setminus \bigcup_{\xi\in\partial\Xi} H_{(\xi, C_H\epsilon)} \tag{4-11}$$

hold for every $H \in \{N, A, K\}$ and the corresponding $\Xi \in \{\Psi, [-T, 0], \Phi\}$ in H. The sets appearing in the right-hand side of the first inclusion are the sets Ψ^+, $[-T^+, S^+]$, Φ^+, and the sets appearing in the right-hand side of the second inclusion are the sets Ψ^-, $[-T^-, S^-]$, Φ^-, alluded to above, for $H = N, A, K$.

Note that the set on the right-hand side of (4-11) is the set of points in Ξ whose distance from the complement of Ξ is at least $C_H\epsilon$. Namely it is the set of points such that an open $C_H\epsilon$-ball centered around them is fully contained in $\Xi = \pi_H(R_T(\Psi, \Phi))$. This follows from the following fact. If $\xi \in \Xi$, $\xi' \notin \Xi$ and the distance between them is less than some $\eta > 0$, then ξ has distance less than η

from some point in $\partial\Xi$. Conversely, if $\xi \in \Xi$ has distance less than η from $\partial\Xi$, then it has distance less than η to some point $\xi' \notin \Xi$.

The inclusion (4-10) is a straightforward consequence of Proposition 4.4. To prove the inclusion (4-11), we note that

$$
\begin{aligned}
\text{(4-12)}\quad g \in R_T(\Psi,\Phi)^-(\epsilon) &\iff g \in u R_T(\Psi,\Phi) v && \forall u, v \in \mathcal{O}_\epsilon\\
&\iff ugv \in R_T(\Psi,\Phi) && \forall u, v \in \mathcal{O}_\epsilon\\
&\iff \pi_H(ugv) \in \pi_H(R_T(\Psi,\Phi)) && \forall u, v \in \mathcal{O}_\epsilon,\ \forall H,
\end{aligned}
$$

since $R_T(\Psi,\Phi)$ is a product set, and $\mathcal{O}_\epsilon = \mathcal{O}_\epsilon^{-1}$.

Now consider g such that $\pi_H(g) \in \bigcup_{\xi\in\Xi} H_{(\xi,C_H\epsilon)} \setminus \bigcup_{\xi\in\partial\Xi} H_{(\xi,C_H\epsilon)}$. Then as just noted above, $\pi_H(g) \in \Xi$ and $H_{(\pi_H(g),C_H\epsilon)} \subset \Xi$, and therefore by the version of Proposition 4.4 stated just before (4-10),

$$\pi_H(ugv) \in H_{(\pi_H(g),C_H\epsilon)} \subset \pi_H(R_T(\Psi,\Phi)).$$

Thus every such g is contained in $R_T(\Psi,\Phi)^-(\epsilon)$ by (4-12). Letting g range over the product set, we conclude that $\Psi^- A_{-T^-,S^-} \Phi^- \subseteq R_T(\Psi,\Phi)^-(\epsilon)$, as stated.

As to the volume estimate, we begin with the N-component. Since Ψ is assumed to be nice, there exists a constant α_1 which depends on Ψ and C_N such that

$$\mu_N\biggl(\bigcup_{\underline{x}\in\partial\Psi} N_{(\underline{x},C_N\epsilon)}\biggr) \le \alpha_1\epsilon.$$

Note that $\bigcup_{\underline{x}\in\Psi} N_{(\underline{x},C_N\epsilon)} \subset \Psi \cup \bigcup_{\underline{x}\in\partial\Psi} N_{(\underline{x},C_N\epsilon)}$. Therefore, by definition of Ψ^+ in (4-10),

$$\mu_N(\Psi^+) \le \mu_N\biggl(\bigcup_{\underline{x}\in\Psi} N_{(\underline{x},C_N\epsilon)}\biggr) \le \mu_N(\Psi) + \mu_N\biggl(\bigcup_{\underline{x}\in\partial\Psi} N_{(\underline{x},C_N\epsilon)}\biggr) \le \mu_N(\Psi) + \alpha_1\epsilon.$$

On the other hand, by (4-11),

$$
\begin{aligned}
\mu_N(\Psi^-) &\ge \mu_N\biggl(\biggl(\bigcup_{\underline{x}\in\Psi} N_{(\underline{x},C_N\epsilon)} \cup \bigcup_{\underline{x}\in\partial\Psi} N_{(\underline{x},C_N\epsilon)}\biggr) \setminus \bigcup_{\underline{x}\in\partial\Psi} N_{(\underline{x},C_N\epsilon)}\biggr)\\
&\ge \mu_N(\Psi) - \mu_N\biggl(\bigcup_{\underline{x}\in\partial\Psi} N_{(\underline{x},C_N\epsilon)}\biggr)\\
&\ge \mu_N(\Psi) - \alpha_1\epsilon.
\end{aligned}
$$

By assuming ϵ is small enough such that $\alpha_1\epsilon \le \frac{1}{2}\mu_N(\Psi)$, the last two inequalities imply

$$\frac{\mu_N(\Psi^+)}{\mu_N(\Psi^-)} - 1 \le \frac{\mu_N(\Psi) + \alpha_1\epsilon - (\mu_N(\Psi) - \alpha_1\epsilon)}{\frac{1}{2}\mu_N(\Psi)} = \frac{2\alpha_1}{\frac{1}{2}\mu_N(\Psi)} \cdot \epsilon.$$

The same considerations apply for $\Phi \subseteq K$. Since it is also assumed to be nice, there exists $\alpha_2 > 0$ that depends on $\partial\Phi$ and C_K such that

$$\mu_K(\Phi^+) = \mu_K\Bigg(\bigcup_{k\in\partial\Phi} K_{(k,C_K\epsilon)}\Bigg) \le \alpha_2\epsilon,$$

and, similarly to the N case, by assuming $\alpha_2\epsilon \le \frac{1}{2}\mu_K(\Phi)$,

$$\frac{\mu_K(\Phi^+)}{\mu_K(\Phi^-)} - 1 \le \frac{2\alpha_2}{\frac{1}{2}\mu_K(\Phi)} \cdot \epsilon.$$

Finally, for the A-component, it follows from (4-10) and (4-11) that

$$\pi_A(R_T(\Psi,\Phi)^+(\epsilon)) \subseteq [-T - C_A\epsilon, 0 + C_A\epsilon] = [-T^+, S^+]$$

and

$$\pi_A(R_T(\Psi,\Phi)^-(\epsilon)) \supseteq [-T + C_A\epsilon, 0 - C_A\epsilon] = [-T^-, S^-].$$

Clearly,

$$\mu_A([-T^+, S^+]) \le \int_{-T-C_A\epsilon}^{tC_A\epsilon} \frac{dt}{e^{2\rho t}} = \frac{1}{2\rho}(e^{2\rho(T+C_A\epsilon)} - e^{-2\rho C_A\epsilon})$$

and

$$\mu_A([-T^-, S^-]) \ge \int_{-T+C_A\epsilon}^{-C_A\epsilon} \frac{dt}{e^{2\rho t}} = \frac{1}{2\rho}(e^{2\rho(T-C_A\epsilon)} - e^{2\rho C_A\epsilon}).$$

As a result,

$$\begin{aligned}\frac{\mu_A([-T^+, S^+])}{\mu_A([-T^-, S^-])} - 1 &\le \frac{e^{2\rho(T+C_A\epsilon)} - e^{-2\rho C_A\epsilon} - (e^{2\rho(T-C_A\epsilon)} - e^{2\rho C_A\epsilon})}{e^{2\rho(T-C_A\epsilon)} - e^{2\rho C_A\epsilon}}\\ &= \frac{(e^{2\rho T}+1)}{e^{2\rho T}} \cdot \frac{(e^{2\rho C_A\epsilon} - e^{-2\rho C_A\epsilon})}{e^{-2\rho C_A\epsilon} - e^{-2\rho T}e^{2\rho C_A\epsilon}}.\end{aligned}$$

For $\epsilon \le (4\rho C_A)^{-1}$ and $T \ge 2\rho^{-1}$ it holds that $e^{2\rho C_A\epsilon} - e^{-2\rho C_A\epsilon} \le 3 \cdot 2\rho C_A\epsilon$ and $e^{-2\rho C_A\epsilon} - e^{-2\rho T}e^{2\rho C_A\epsilon} \ge \frac{1}{2}$; therefore,

$$\frac{\mu_A([-T^+, S^+])}{\mu_A([-T^-, S^-])} - 1 \le 2 \cdot \frac{6\rho C_A\epsilon}{1/2} = 24\rho C_A\epsilon.$$

Now since

$$\frac{\mu(R_T(\Psi,\Phi)^+(\epsilon))}{\mu(R_T(\Psi,\Phi)^-(\epsilon))} \le \frac{\mu_N(\Psi^+)\mu_A([-T^+, S^+])\mu_K(\Phi^+)}{\mu_N(\Psi^-)\mu_A([-T^-, S^-])\mu_K(\Phi^-)},$$

by choosing $T_0 = 2\rho^{-1}$ and $\epsilon_0 = \min\{\epsilon_1, \mu_N(\Psi)/(2\alpha_1), \mu_K(\Phi)/(2\alpha_2), 1/(4\rho C_A)\}$ we conclude that the family $\{R_T(\Psi,\Phi)\}_{T>T_0}$ is Lipschitz well-rounded for $0 < \epsilon < \epsilon_0$, and by Theorem 4.2 (and the discussion in Section 4A) we are done. $\square$

4D. ***Proof of Corollary 1.2.*** Let Ψ, Ψ', Φ, Φ' and κ as in the statement of the corollary. By Theorem 1.1, the denominator in the following ratio is eventually positive and the following estimate holds:

$$\begin{aligned}\frac{\#(\Gamma \cap R_T(\Psi', \Phi'))}{\#(\Gamma \cap R_T(\Psi, \Phi))} &= \frac{\mu_N(\Psi')\mu_K(\Phi')e^{2\rho T} + O(Te^{2\rho\kappa T})}{\mu_N(\Psi)\mu_K(\Phi)e^{2\rho T} + O(Te^{2\rho\kappa T})} \\ &= \frac{\mu_N(\Psi')\mu_K(\Phi')}{\mu_N(\Psi)\mu_K(\Phi)} + O(T(e^{2\rho T})^{-(1-\kappa)}).\end{aligned}$$

The limit of the foregoing expression is $(\mu_N(\Psi')\mu_K(\Phi'))/(\mu_N(\Psi)\mu_K(\Phi))$ as $T \to \infty$, since $\kappa < 1$. The implied constant depends on Ψ, Ψ', Φ, Φ'.

Let now ψ and ϕ be nonnegative Lipschitz functions with positive integral, with ψ defined on Ψ, and ϕ defined on Φ. We also view ψ as a (measurable bounded) function on N by defining it to be zero outside Ψ, and we extend ϕ to K similarly. Let $R_T(\psi, \phi)$ be the measure on G whose density with respect to Haar measure on G (written in Iwasawa coordinates as in (1-1)) is given by the function $D_T(na_tk) = \psi(n)\chi_{[-T,0]}(a_t)\phi(k)$. Equivalently, the measure is given by the following formula: for $F \in C_c(G)$,

$$R_T(\psi, \phi)(F) = \int_\Psi \int_{-T}^0 \int_\Phi F(na_tk)\psi(n)\phi(k)\, d\mu_N(n)\, \frac{dt}{e^{2\rho t}}\, d\mu_K(k).$$

The family of measures $R_T(\psi, \phi)$ is Lipschitz well-rounded, in the following sense. Defining

$$D_T^{+,\epsilon}(g) = \sup_{u,v \in \mathcal{O}_\epsilon} D_T(ugv), \qquad D_T^{-,\epsilon}(g) = \inf_{u,v \in \mathcal{O}_\epsilon} D_T(ugv),$$

we have

$$\int_G D_T^{+,\epsilon}(g)\, d\mu(g) \le (1 + C\epsilon) \int_G D_T^{-,\epsilon}(g)\, d\mu(g).$$

The family $R_T(\psi, \phi)$ satisfies a weighted version of the lattice point counting result which the sets $R_T(\Psi, \Phi)$ satisfy, namely

$$\sum_{\gamma \in \Gamma} D_T(\gamma) = \int_G D_T(g) d\mu(g) + O_{\phi,\psi}\Big(\Big(\int_G D_T(g) d\mu(g)\Big)^{\kappa(\Gamma)} \cdot \log \int_G D_T(g) d\mu(g)\Big),$$

so that in the present case,

$$\begin{aligned}&\sum_{\gamma \in \Gamma} \psi(\pi_N(\gamma))\chi_{[-T,0]}(\pi_A(\gamma))\phi(\pi_K(\gamma)) \\ &\qquad = e^{2\rho T} \int_N \psi(n)\, d\mu_N(n) \cdot \int_K \phi(k)\, d\mu_K(k) + O_{\phi,\psi}(Te^{2\rho T\kappa(\Gamma)}).\end{aligned}$$

The proof of the weighted version of the lattice point problem stated above under the assumption of Lipschitz well-roundedness is a straightforward modification of the arguments that appear in [Gorodnik and Nevo 2012]. The fact that when ψ and

ϕ are Lipschitz functions on N and K the measures $R_T(\psi, \phi)$ defined above are Lipschitz well-rounded is a straightforward modification of the arguments in the present paper. Note that it suffices to consider nonnegative Lipschitz functions on N and K, and the case of general Lipschitz functions follows, since $\max(f, 0)$ and $\max(-f, 0)$ are nonnegative Lipschitz functions and f is their difference. Finally, the statement of Corollary 1.2 part (2) follows by considering a Lipschitz function ψ defined on $\Psi \subset N$, a Lipschitz function ϕ defined on $\Phi \subset K$, defining D_T using ψ and ϕ, and estimating the ratios as

$$\begin{aligned}\frac{\sum_{\gamma\in\Gamma} D_T(\gamma)}{\sum_{\gamma\in\Gamma} \chi_{R_T(\Psi,\Phi)}(\gamma)} \\ = \frac{1}{\mu_N(\Psi)}\int_N \psi(n)\,d\mu_N(n)\cdot\frac{1}{\mu_K(\Phi)}\int_K \phi(k)\,d\mu_K(k) + O(Te^{-2\rho(1-\kappa)T}),\end{aligned}$$

where the implied constant depends on Φ, Ψ, ϕ, ψ. □

Acknowledgments

The authors thank the referee for important comments which led to significant improvements is the presentation of several results in the paper. They also thank Ami Paz for preparing the figures for this paper. Horesh thanks Ami Paz and Yakov Karasik for helpful discussions. Nevo thanks John Parker and Rene Rühr for providing some very useful references. Nevo is supported by ISF Grant No. 2095/15.

References

[Chamizo 2011] F. Chamizo, "Hyperbolic lattice point problems", *Proc. Amer. Math. Soc.* **139**:2 (2011), 451–459. MR Zbl

[Cowling et al. 1991] M. Cowling, A. H. Dooley, A. Korányi, and F. Ricci, "H-type groups and Iwasawa decompositions", *Adv. Math.* **87**:1 (1991), 1–41. MR Zbl

[Cowling et al. 1998] M. Cowling, A. Dooley, A. Korányi, and F. Ricci, "An approach to symmetric spaces of rank one via groups of Heisenberg type", *J. Geom. Anal.* **8**:2 (1998), 199–237. MR Zbl

[Dabbs et al. 2016] K. Dabbs, M. Kelly, and H. Li, "Effective equidistribution of translates of maximal horospherical measures in the space of lattices", *J. Mod. Dyn.* **10** (2016), 229–254. MR Zbl

[Dinaburg and Sinai 1990] E. I. Dinaburg and Y. G. Sinai, "The statistics of the solutions of the integer equation $ax - by = \pm 1$", *Funktsional. Anal. i Prilozhen.* **24**:3 (1990), 1–8. MR Zbl

[Duke et al. 1993] W. Duke, Z. Rudnick, and P. Sarnak, "Density of integer points on affine homogeneous varieties", *Duke Math. J.* **71**:1 (1993), 143–179. MR Zbl

[Elstrodt et al. 1998] J. Elstrodt, F. Grunewald, and J. Mennicke, *Groups acting on hyperbolic space: harmonic analysis and number theory*, Springer, 1998. MR Zbl

[Eskin and McMullen 1993] A. Eskin and C. McMullen, "Mixing, counting, and equidistribution in Lie groups", *Duke Math. J.* **71**:1 (1993), 181–209. MR Zbl

[Faraut 1983] J. Faraut, "Analyse harmonique sur les paires de Guelfand et les espaces hyperboliques", pp. 315–446 in *Analyse harmonique* (Nancy, France, 1980), edited by J.-L. Clerc et al., Les Cours du CIMPA, Nice, 1983.

[Good 1983] A. Good, *Local analysis of Selberg's trace formula*, Lecture Notes in Mathematics **1040**, Springer, 1983. MR Zbl

[Good 1984] A. Good, "Analytical and arithmetical methods in the theory of Fuchsian groups", pp. 86–103 in *Number theory* (Noordwijkerhout, The Netherlands, 1983), edited by H. Jager, Lecture Notes in Math. **1068**, Springer, 1984. MR Zbl

[Gorodnik and Nevo 2010] A. Gorodnik and A. Nevo, *The ergodic theory of lattice subgroups*, Annals of Mathematics Studies **172**, Princeton University Press, 2010. MR Zbl

[Gorodnik and Nevo 2012] A. Gorodnik and A. Nevo, "Counting lattice points", *J. Reine Angew. Math.* **663** (2012), 127–176. MR Zbl

[Gorodnik and Weiss 2007] A. Gorodnik and B. Weiss, "Distribution of lattice orbits on homogeneous varieties", *Geom. Funct. Anal.* **17**:1 (2007), 58–115. MR Zbl

[Kim 2015] I. Kim, "Counting, mixing and equidistribution of horospheres in geometrically finite rank one locally symmetric manifolds", *J. Reine Angew. Math.* **704** (2015), 85–133. MR Zbl

[Marklof and Vinogradov 2018] J. Marklof and I. Vinogradov, "Directions in hyperbolic lattices", *J. Reine Angew. Math.* **740** (2018), 161–186. MR Zbl

[Mohammadi and Oh 2015] A. Mohammadi and H. Oh, "Matrix coefficients, counting and primes for orbits of geometrically finite groups", *J. Eur. Math. Soc.* (*JEMS*) **17**:4 (2015), 837–897. MR Zbl

[Mohammadi and Salehi Golsefidy 2014] A. Mohammadi and A. Salehi Golsefidy, "Translate of horospheres and counting problems", *Amer. J. Math.* **136**:5 (2014), 1301–1346. MR Zbl

[Nevo 1998] A. Nevo, "Spectral transfer and pointwise ergodic theorems for semi-simple Kazhdan groups", *Math. Res. Lett.* **5**:3 (1998), 305–325. MR Zbl

[Risager and Rudnick 2009] M. S. Risager and Z. Rudnick, "On the statistics of the minimal solution of a linear Diophantine equation and uniform distribution of the real part of orbits in hyperbolic spaces", pp. 187–194 in *Spectral analysis in geometry and number theory*, edited by M. Kotani et al., Contemp. Math. **484**, Amer. Math. Soc., Providence, RI, 2009. MR Zbl

[Selberg 1991] A. Selberg, "Abstracts of lectures given by A. Selberg at Tel Aviv University", 1991, available at http://publications.ias.edu/node/2482.

[Truelsen 2013] J. L. Truelsen, "Effective equidistribution of the real part of orbits on hyperbolic surfaces", *Proc. Amer. Math. Soc.* **141**:2 (2013), 505–514. MR Zbl

Received October 24, 2021. Revised December 29, 2022.

TAL HORESH
IST AUSTRIA
KLOSTERNEUBURG
AUSTRIA
talhoresh@gmail.com

AMOS NEVO
DEPARTMENT OF MATHEMATICS
TECHNION — ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA
ISRAEL
anevo@tx.technion.ac.il

PACIFIC JOURNAL OF MATHEMATICS

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

msp.org/pjm

See inside back cover or msp.org/pjm for submission instructions.

The subscription price for 2023 is US $605/year for the electronic version, and $820/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 1945-5844 electronic, 0030-8730 printed) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS

Volume 324 No. 2 June 2023