Mixing, counting and equidistribution in Lie groups

Alex Eskin and Curt McMullen *
Princeton University, Princeton NJ 08544
University of California, Berkeley, CA 94720

8 July, 1992

1 Introduction.

Let $\Gamma \subset G = \mathrm{Aut}(\mathbb{H}^2)$ be a group of isometries of the hyperbolic plane $\mathbb{H}^2$ such that $\Sigma = \Gamma\backslash\mathbb{H}^2$ is a surface of finite area. Then:

I. The geodesic flow is mixing on the unit tangent bundle $T_1(\Sigma) = \Gamma\backslash G$.

II. The sphere $S(x, R)$ of radius R about a point $x \in \Sigma$ becomes equidistributed as $R \to \infty$.

III. The number of points $N(R)$ in an orbit Γv which lie within a hyperbolic ball $B(p, R) \subset \mathbb{H}^2$ has the asymptotic behavior

$$N(R) \sim \frac{\mathrm{area}(B(p,R))}{\mathrm{area}(\Sigma)}.$$

(See §2 for more detailed statements).

The purpose of this paper is to discuss results similar to those above where the hyperbolic plane is replaced by a general *affine symmetric space* $V = G/H$. This setting includes the classical Riemannian symmetric spaces

*Research partially supported by the NSF and the Alfred P. Sloan Foundation.

(when H is a maximal compact subgroup) as well as spaces with *indefinite* invariant metrics.

A simple non-Riemannian example is obtained by letting V be the space of oriented geodesics in the hyperbolic plane. Then $H = A$, the group of diagonal matrices in $G = PSL_2(\mathbb{R})$. In this case $\Gamma\backslash G/H$ is not even Hausdorff.

This setting includes counting theorems for integral points on a large class of homogeneous varieties (e.g. those associated to quadratic forms) and allows us to prove some of the main theorems of [DRS] by elementary arguments (see §6).

Statement of Results. Let G be a connected semisimple Lie group with finite center, and let $H \subset G$ be a closed subgroup such that G/H is an *affine symmetric space* (cf. [F-J], [Sch]). This means there is an involution $\sigma : G \to G$ such that H is the fixed-point set of σ:

$$H = \{g \ : \ \sigma(g) = g\}.$$

(By *involution* we mean a Lie group automorphism such that $\sigma^2 = \mathrm{id}$).

Let $\Gamma \subset G$ be a *lattice*, i.e. a discrete subgroup such that the volume of $X = \Gamma\backslash G$ is finite.

Assume further that Γ has dense projection to G/G' for any positive-dimensional normal *noncompact* Lie subgroup $G' \subset G$. [1]

Finally, assume that Γ meets H in a lattice: that is, the volume of $Y = (\Gamma \cap H)\backslash H$ is finite.

We may now state general results on mixing, equidistribution and counting. The mixing theorem below is standard; the aim of this paper is to deduce the equidistribution and counting theorems from it, using the geometry of affine symmetric spaces.

Theorem 1.1 (Mixing) *The action of G on $X = \Gamma\backslash G$ is mixing. That is, for any α, β in $L^2(X)$,*

$$\int_X \alpha(xg)\beta(x)dx \to \frac{\int_X \alpha \int_X \beta}{m(X)}$$

as g tends to infinity.

[1] A lattice is *irreducible* if it projects densely to G/G' for *any* positive-dimensional normal G'. Our assumption is a weak form of irreducibility which admits interesting examples when G has a compact factor.

Here the integrals and the volume $m(X)$ are taken with respect to the G-invariant Haar measure on X. A sequence of element $g_n \in G$ *tends to infinity* if any compact set $K \subset G$ contains only finitely many terms in the sequence.

Proof. This well-known result is a consequence of the Howe-Moore theorem [HM]; see also [Zim].

After passing to a finite cover we may assume that $G = G_0 \times K$, where K is compact and G_0 has no compact factors. It suffices to verify that the action of G_0 on $L^2(X)$ is mixing.

The action of G_0 on $L^2(X)$ is a unitary representation ρ such that the constant functions are the only invariant vectors. The integral above can be interpreted as a "matrix coefficient" $< \rho(g)\alpha, \beta >$ for this representation. Then mixing follows from decay of matrix coefficients for irreducible unitary representations of reductive algebraic groups [HM, Theorem 5.1]. (This reference treats only algebraic groups, but G_0 is a finite cover of such a group since its center is finite.) ∎

Theorem 1.2 (Equidistribution) *The translates Yg of the H-orbit*

$$Y = (\Gamma \cap H)\backslash H$$

become equidistributed on $X = \Gamma\backslash G$ as g tends to infinity in $H\backslash G$. That is,

$$\frac{1}{m(Y)} \int_{Yg} f(h)dh \to \frac{1}{m(X)} \int_X f(x)dx$$

as g leaves compact subsets of $H\backslash G$.

Here the measure $dh/m(Y)$ on Yg is the translate by g of the unique H-invariant probability measure on Y.

To state the theorem on counting points in an orbit, we first isolate some properties of the sets used for counting. Let $B_n \subset G/H$ be a sequence of finite volume measurable sets such that the volume of B_n tends to infinity.

Definition. The sequence B_n is *well-rounded* if for any $\epsilon > 0$ there exists an open neighborhood U of the identity in G such that

$$\frac{m(U \cdot \partial B_n)}{m(B_n)} < \epsilon$$

for all n.

It is easy to verify:

Proposition 1.3 *A sequence is well-rounded if and only if for any* $\epsilon > 0$ *there is a neighborhood* U *of* $\mathrm{id} \in G$ *such that for all* n,

$$(1-\epsilon)m\left(\bigcup_U gB_n\right) < m\left(B_n\right) < (1+\epsilon)m\left(\bigcap_U gB_n\right).$$

That is, B_n is nearly invariant under the action of a small neighborhood of the identity.

Theorem 1.4 (Counting) *Let* $V = G/H$ *be an affine symmetric space, and let* v *denote the coset* $[H]$. *For any well-rounded sequence, the cardinality of the number of points of* Γv *which lie in* B_n *grows like the volume of* B_n*: asymptotically,*

$$|\Gamma v \cap B_n| \sim \frac{m((\Gamma \cap H)\backslash H)}{m(\Gamma\backslash G)} m(B_n).$$

Normalization of measures. In the statement of the counting theorem, measures are normalized so that Haar measure on G is the product of the measure on H with that on G/H.

Outline of the paper. A crucial link in the logic above is the wavefront lemma (Theorem 3.1 below), which controls the geodesic flow on an affine symmetric space. In §2 we discuss the relationship between mixing, equidistribution and counting in the setting of the hyperbolic plane. The geometric intuition of negative curvature, transparent in this setting, leads to the wavefront lemma. In §3 we prove the wavefront lemma for $SL_n(\mathbb{R})/K$ and deduce the equidistribution theorem. The general affine symmetric space is treated in §4.

In §5 equidistribution is used to prove the counting theorem for well-rounded sets. The hypothesis of well-roundedness is implicitly verified in the course of the study of integral points on homogeneous varieties in [DRS]; this connection is made explicit in §6. Finally §7 contains some results beyond the affine symmetric setting and some open questions.

We would like to thank Marc Burger, Zeev Rudnick,Peter Sarnak and the referee for useful comments.

2 Examples in the hyperbolic plane

In this section we sketch the connection between mixing, equidistribution and counting on hyperbolic surfaces. This relation is fairly well-known and appears already in the thesis of Margulis (cf. [Mg], which contains a generalization of Theorem 2.2 below).

Let $\Sigma = \Gamma \backslash \mathbb{H}$ be a hyperbolic surface of finite volume, presented as the quotient of the hyperbolic plane $\mathbb{H}$ by a lattice

$$\Gamma \subset G$$

in the group $G = PSL_2(\mathbb{R})$ of hyperbolic isometries. Let $T_1(\Sigma)$ denote the unit tangent bundle to Σ. The *geodesic flow*

$$g_t : T_1(\Sigma) \to T_1(\Sigma)$$

transports a vector distance t along the geodesic to which it is tangent.

I. There is a natural invariant measure μ on the unit tangent bundle, which is the product of angular measure on the fiber with area measure on the base. With respect to this measure, the geodesic flow is *mixing*: for any α and β in $L^2(T_1(\Sigma), \mu)$, we have

$$\lim_{t\to\infty} \int_{T_1(\Sigma)} \alpha(x)\beta(g_t(x))\, d\mu(x) \to \frac{\int \alpha\, d\mu \int \beta\, d\mu}{\int 1 d\mu}.$$

Mixing was proved for finite volume hyperbolic surfaces by Hedlund [Hed2]; (see also [Hed1]). It is also a special case of Theorem 1.1, since $T_1(\Sigma)$ can be identified with the space $\Gamma\backslash G$, and the geodesic flow with the action of the noncompact subgroup A of diagonal matrices.

II. The equidistribution of spheres on Σ follows easily from mixing. First, consider a point p on Σ, and let $K \subset T_1(\Sigma)$ denote the preimage of x under under the fibration $T_1(\Sigma) \to \Sigma$; that is, K consists of vectors lying over x and pointing in every possible direction. Then it is clear that the image $g_t(K)$ under the geodesic flow consists of all vectors normal to the immersed sphere $S(p, t) \subset \Sigma$.

Now replace K by an open set U, consisting of the vectors lying over a small open ball $B(p, \epsilon)$ and pointing in all directions. It is easy to see that $g_t(B(p, \epsilon))$ consists of vectors (a) lying over an ϵ-neighborhood of $S(p, t)$ and (b) nearly normal to the sphere. Assertion (a) just comes from the triangle inequality, while (b) is a feature of negative curvature. Indeed, the spread of the vectors from the normal is bounded by the apparent visual angle of $B(p, \epsilon)$ as seen from distance t, which goes to zero not only in hyperbolic space but in any space of nonpositive curvature.

The fact that $g_t(U)$ remains close to $g_t(K)$ is a special case of the wavefront lemma, to be presented in §3. From it we can deduce the equidistribution of spheres:

Theorem 2.1 *For any compactly supported continuous function α on Σ, and any point p, the average of α over the sphere $S(p,t)$ tends to the average of α over Σ as t tends to infinity.*

Here the average over the sphere is taken with respect to linear measure.

Proof. First pull α back to a function $\alpha(x)$ on the unit tangent bundle (by taking it to be constant on fibers.) Then the average of α over the sphere of radius t is the same as its average over $g_t(K)$, the lift of the sphere to the tangent bundle. By uniform continuity, this is nearly the same as the average of α over $g_t(U)$. But this second average is equal to

$$\frac{\int_{T_1(\Sigma)} \chi_U(x)\alpha(g_t(x))d\mu}{\int_{T_1(\Sigma)} \chi_U(x)d\mu}.$$

By mixing, as $t \to \infty$ the quantity above tends to the average of $\alpha(x)$ over $T_1(\Sigma)$, which is the same as the average of α over Σ. ■

Remark. This result is a special case of Theorem 1.2 with $G = PSL_2(\mathbb{R})$ and $H = K = SO_2(\mathbb{R})/\{\pm I\}$. Indeed that theorem gives the stronger conclusion that the sphere, lifted by its unit normal vectors, becomes equidistributed in $T_1(\Sigma)$. See [Ran] for another proof of the equidistribution of spheres.

III.1. Let $N(R)$ denote the number $|\Gamma v \cap B(p,R)|$ of points in the orbit Γv which lie within distance R of a point p in the hyperbolic plane. See Figure 1 for an example of such an orbit.

We now show that Theorem 2.1 easily gives the asymptotic behavior of $N(R)$.

Theorem 2.2 *As R tends to infinity,*

$$N(R) \sim \frac{\text{area}(B(p,R))}{\text{area}(\Sigma)}.$$

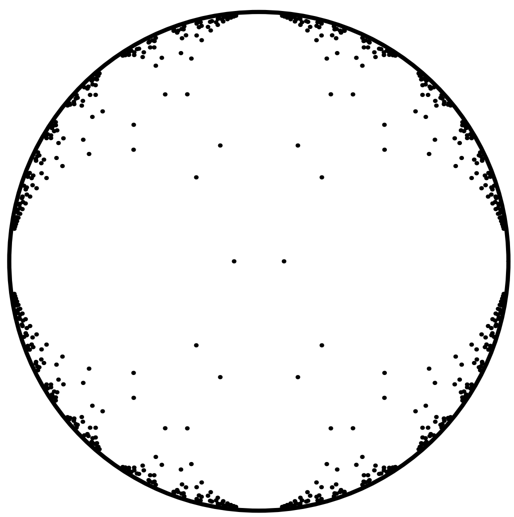

Figure 1. Orbit of a point in the hyperbolic plane.

Remark. The intuition behind the estimate for $N(R)$ is the following. Tile the hyperbolic plane by translates of a fundamental domain for the action of Γ which contains v in its interior. The tiles meeting $\Gamma v \cap B(p,R)$ form an approximate covering of the ball. Thus their number should be proportional to the area of $B(p,R)$ divided by the area of a single tile.

However, the tiling is likely to be uneven near the boundary of $B(p,R)$. In hyperbolic space, the area of a unit neighborhood of the boundary is comparable to the area of the whole ball, so these edge effects must be studied. Mixing intervenes to show that the tiles appear more or less randomly along the edge of the ball, and that the area estimate is correct.

Example. Figure 1 shows the 598 points in an orbit of $\Gamma = PSL_2(\mathbb{Z})$ which lie in B, the region farther than 0.01 from the boundary of the unit disk. For comparison, the area of $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$ is $\pi/3$, the area of B is 618.91..., so

$$\frac{\mathrm{area}(B)}{\mathrm{area}(\Sigma)} = 591.02...$$

is a reasonable estimate for the cardinality of $\Gamma v \cap B$.

Proof of Theorem 2.2. For any point q in $\mathbb{H}$, denote by $[q]$ the image of q on $\Sigma = \Gamma\backslash\mathbb{H}$. Let $\alpha(x)$ be a bump function of integral one supported in the ball $B([v],\epsilon)$, and let $\beta_R(x)$ denote the the number of distinct geodesics of length less than R joining x to $[p]$. Equivalently, β_R is the indicator function (with multiplicities) of the immersed disk of radius R about $[p]$, or the pushforward of the indicator function of $B(p,R)$.

Then it is easy to see that

$$N(R-\epsilon) \;\leq\; \int_{B(p,R)} \widetilde{\alpha}(x)dx \;=\; \int_\Sigma \alpha(x)\beta_R(x)dx \;\leq\; N(R+\epsilon),$$

where $\widetilde{\alpha}$ is the pullback of α to a function on $\mathbb{H}$. Now the measure $\beta_R(x)dx$ is a continuous convex combination of linear measures on the spheres $S([p],t)$ as t ranges from zero to R. Since the spheres are becoming equidistributed, and $\int_\Sigma \alpha = 1$, the integral above is asymptotic to

$$\frac{\mathrm{area}(B(p,R))}{\mathrm{area}(\Sigma)} \sim \frac{\pi \exp(R)}{\mathrm{area}(\Sigma)}.$$

It follows that any limit as R tends to infinity of

$$\frac{N(R)\,\mathrm{area}(\Sigma)}{\mathrm{area}(B(p,R)}$$

lies in the interval $[\exp(-\epsilon), \exp(+\epsilon)]$. Since ϵ was arbitrary, the limit exists and is equal to one. ∎

Remarks.

It is easy to show that for any $R_n \to \infty$, the balls $B_n = B_n(p, R_n)$ are well-rounded for the action of G on $\mathbb{H}$. The counting result above is thus a special case of Theorem 1.4.

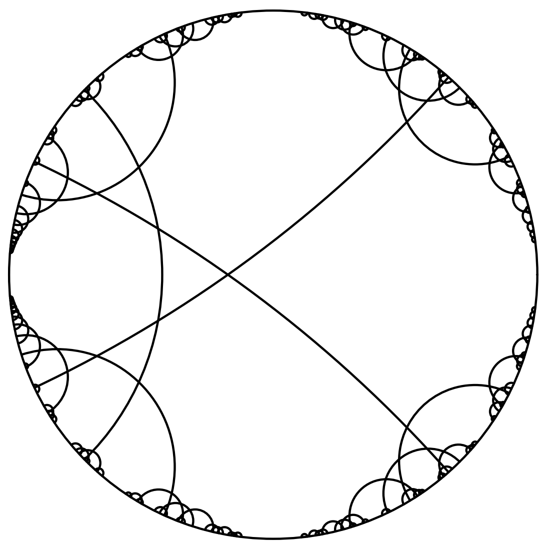

Figure 2. A closed geodesic, lifted to the hyperbolic plane.

III.2. Here is a counting problem which leads to a non-Riemannian symmetric space.

Let ℓ be a geodesic in $\mathbb{H}$ with stabilizer H, and suppose H meets Γ in a subgroup isomorphic to $\mathbb{Z}$. Equivalently, ℓ descends to a closed geodesic

$$L = (\Gamma \cap H)\backslash\ell$$

on Σ. The orbit $\Gamma\ell$ is a locally finite collection of geodesics in $\mathbb{H}$; see Figure 2 for an example.

Let $N(R)$ denote the number of geodesics in the orbit $\Gamma\ell \subset \mathbb{H}$ which intersect the ball $B(p, R)$.

Theorem 2.3 *As* $R \to \infty$,

$$N(R) \sim \frac{1}{\pi} \frac{\text{length}(L)}{\text{area}(\Sigma)} \text{area}(B(p,R))$$

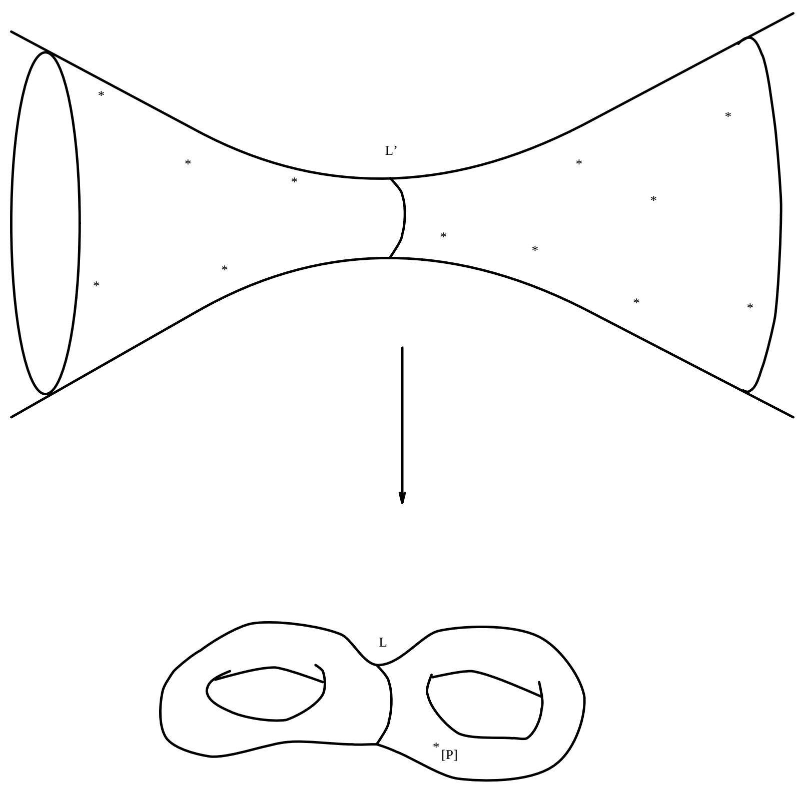

Figure 3. The covering space $\Sigma' \to \Sigma$ associated to a closed geodesic.

Example. Let $\Gamma = \Gamma(2) \subset PSL_2(\mathbb{Z})$ be the free subgroup of index six, and consider the geodesic stabilized by $< g >$ where

$$g = \begin{pmatrix} 1 & 4 \\ 2 & 9 \end{pmatrix}.$$

The length of L is $\log(49+20\sqrt{6}) = 4.584...$, and the area of $\Sigma = \Gamma\backslash\mathbb{H}$ is 2π. Figure 2 shows the 145 lifts of L which meet B, again the region farther than 0.01 from the boundary of the unit disk. For comparison,

$$\frac{1}{\pi}\frac{\text{length}(L)}{\text{area}(\Sigma)}\,\text{area}(B) \quad = \quad 143.76...$$

Remark. There is an important difference between orbits of points and orbits of geodesics under the action of Γ. While the orbits of points are classified by the Riemannian manifold $\Sigma = \Gamma\backslash G/K$, the orbits of geodesics are classified by $Z = \Gamma\backslash G/H$ which is not even Hausdorff. This is not too surprising, since Z is intrinsically the space of geodesics on Σ, and almost every geodesic is dense in $T_1(\Sigma)$.

Nevertheless we can carry out a counting estimate for the orbit of a geodesic which is *closed* on Σ by methods similar to those of Theorem 2.2. (The counting problem does not usually make sense for geodesics which are not closed, since $\Gamma\ell$ may not be locally finite.)

We begin with an equidistribution result analogous to Theorem 2.1:

Theorem 2.4 *The parallel L_t at distance t from a closed geodesic L on Σ becomes equidistributed as $t \to \infty$.*

Proof. Let $\widetilde{L}$ denote a lift of L to a continuous family of vectors in $T_1(\Sigma)$ normal to L. Then $\widetilde{L}_t = g_t(\widetilde{L})$ is a similar lift of the parallel curve L_t at distance t from L.

As in the proof of the equidistribution of spheres, we may thicken $\widetilde{L}$ to an open set $U \subset T_1(\Sigma)$, consisting of vectors making angle ϵ with the normal to L_t for $t \in [-\epsilon, \epsilon]$. By mixing, $g_t(U)$ becomes equidistributed in $T_1(\Sigma)$ as t tends to infinity.

The main geometric point is that $g_t(U)$ lies close to $\widetilde{L}_t$ for all t. This is a property of *negative* (not just nonpositive) curvature. Namely, if a geodesic segment of length t rests with one endpoint making angle $\pi - \delta$ on L, the other endpoint rests on $L_{t+\delta'}$ where $\delta' \to 0$ as $\delta \to 0$ (independent of t).

Therefore a uniformly continuous function has nearly the same average over U as over $\widetilde{L}_t$. It follows that $\widetilde{L}_t$ and L_t both become equidistributed as t tends to infinity. ∎

Next we establish a more natural variant of Theorem 2.3.

The closed geodesic $L \subset \Sigma$ determines a cyclic subgroup of $\pi_1(\Sigma)$; let

$$\pi : \Sigma' \to \Sigma$$

denote the corresponding covering space. Then L lifts isometrically to a geodesic L' on Σ'. Let $P' \subset \Sigma'$ denote the set $\pi^{-1}([p]) = (\Gamma \cap H)\backslash\Gamma p$.

See Figure 3, in which P' is depicted by $*$'s on Σ'.

Let $B(L', R)$ denote the cylinder of points on Σ' at distance at most R from L'.

Theorem 2.5 *As $R \to \infty$,*

$$N(R) \sim \frac{\operatorname{area}(B(L', R))}{\operatorname{area}(\Sigma)}.$$

This version can also be deduced from Theorem 1.4.

Proof. It is easy to see that the following quantities are all equal to $N(R)$:

(a) the number of distinct geodesics in $\Gamma\ell \cap B(p, R)$;

(b) the number of geodesics normal to L, of length at most R, joining L to $[p]$; and

(c) the number of points in $P' \cap B(L', R)$.

For example, a shortest path connecting p to a geodesic $\gamma\ell$ as in (a) projects to a path on Σ connecting L to $[p]$ as in (b). Conversely a path on Σ as in (b) can be lifted to $\mathbb{H}$ so that p lies over $[p]$. Each lift or projection factors through Σ', proving equality with (c).

The idea of the estimate is easily explained in terms of (c). Pick a cell of full measure on Σ with $[p]$ in its interior, and consider its preimages on Σ'. These provide a tiling with one tile for each point in P'. The tiles meeting $P' \cap B(L', R)$ approximately cover $B(L', R)$, so their number should be about $\operatorname{area}(B(L', R))/\operatorname{area}(\Sigma)$.

The proof follows the same lines as Theorem 2.2, using the equidistribution of parallels of L. Let α denote a bump function on Σ supported in an ϵ-neighborhood of $[p]$. Let $\beta_R(x)$ denote the number of distinct geodesics joining x to L, perpendicular to L and of length less than or equal to R. Equivalently, $\beta_R(x)$ is the indicator function (with multiplicities) of the immersed cylinder of radius R about L, or the pushforward of the indicator function of $B(L', R)$ on Σ'.

By (b) or (c) above,

$$N(R-\epsilon) \quad \leq \quad \int_{T_1(X)} \alpha(x)\beta_R(x)dx \quad \leq \quad N(R+\epsilon).$$

The measure $\beta_R(x)dx$ can be thought of as a continuous linear combination of linear measures on the curves L_t parallel to L, for $-R \leq t \leq R$. Since the parallels are becoming equidistributed and $\int \beta_R(x)dx = \text{area}(B(L', R))$, the integral above is asymptotic to

$$\frac{\text{area}(B(L', R))}{\text{area}(\Sigma)} \sim \frac{\text{length}(L)\exp(R)}{\text{area}(\Sigma)}.$$

It follows that the estimate for $N(R)$ is correct to within a factor of $1 \pm \epsilon$. Since ϵ was arbitrary the proof is complete. ∎

Proof of Theorem 2.3. A calculation in the hyperbolic metric shows that

$$\text{area}(B(L', R)) \sim \frac{\text{length}(L)\,\text{area}(B(p, R))}{\pi}.$$

∎

Quadratic forms. To conclude, we describe the *Minkowski model* for hyperbolic space, which connects orbits with a linear representation of G and provides a common setting for the study of points and geodesics on $\mathbb{H}$. (See [GHL, p. 118].)

Let $\mathbb{R}^{2+1}$ denote a three dimensional real vector space equipped with the indefinite quadratic form

$$Q(x, y, z) \quad = \quad x^2 + y^2 - z^2.$$

This form also provides a Lorentz metric on the tangent space to each point of $\mathbb{R}^{2+1}$.

Let $SO(2, 1)$ be the group of orientation-preserving linear transformations which preserve this quadratic form, and let G be the connected component of the identity in $SO(2, 1)$. Some of the orbits of G are pictured in Figure 4.

The locus $Q(v) = -1$, sometimes called the *sphere of imaginary radius*, is a two-sheeted hyperboloid, a single sheet of which is a model for the hyperbolic plane $\mathbb{H}$. Indeed, Q induces a complete Riemannian metric of constant

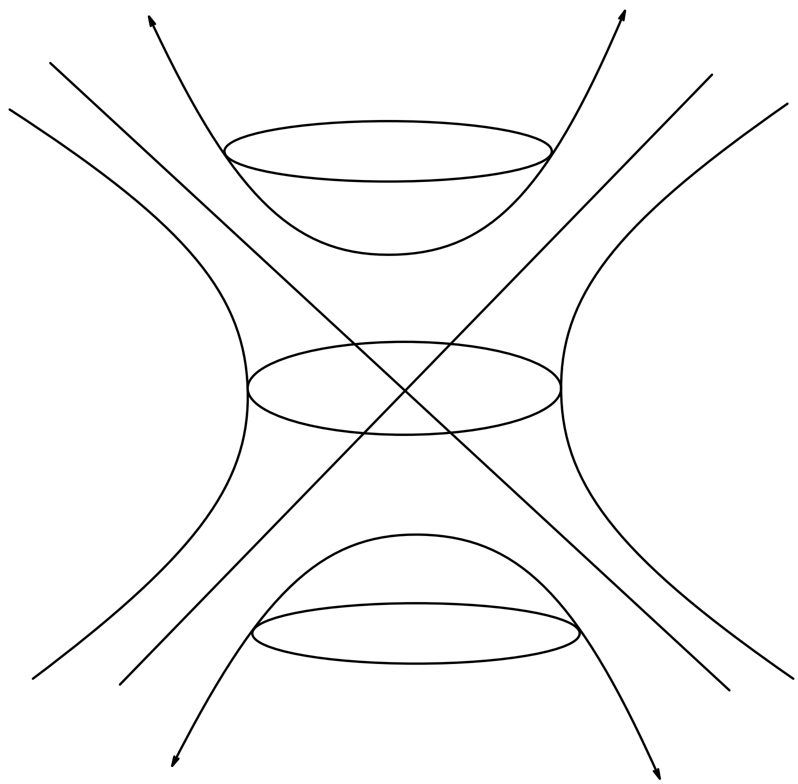

Figure 4. The light cone, and two hyperboloids

curvature -1 on each sheet, with respect to which G is the full group of orientation-preserving isometries. (Note that G, being connected, does not interchange the two sheets).

The locus $Q(v) = 1$, the *one-sheeted hyperboloid*, is naturally identified with the space of all (oriented) geodesics in $\mathbb{H}$; we will denote it by $\mathbb{G}$. Geodesics are parameterized by $\mathbb{G}$ as follows. Let $v^\perp$ be the orthogonal subspace of v with respect to the inner product $Q(v, w)$ associated to Q. Then a point $v \in \mathbb{G}$ determines a hyperplane $v^\perp$ through the origin, which meets $\mathbb{H}$ in a unique geodesic $\ell(v)$. All geodesics are so obtained.

The form Q induces a Lorentz metric of type $(1, 1)$ on $\mathbb{G}$ which is invariant under the transitive action of G. Since this metric is indefinite, there is no reason that a discrete subgroup of G should act properly discontinuously on $\mathbb{G}$, and indeed almost every Γ-orbit on $\mathbb{G}$ is dense.

The one-sheeted hyperboloid is the simplest example of a non-Riemannian symmetric space. It can be presented as G/H where H is the stabilizer of a geodesic ℓ in $\mathbb{H}$. Since H consists exactly of those isometries which commute with reflection through ℓ, G/H is an affine symmetric space.

For completeness, we remark that the locus $Q(v) = 0$ is called the *light cone*, since light rays move along null geodesics in special relativity. With its vertex removed, the upper half of the light cone is also a homogeneous space for G; it parameterizes horocycles in the hyperbolic plane, by letting v correspond to

$$\{w \,:\, Q(v, w) = -1\} \cap \mathbb{H}.$$

(See §7 for counting and mixing on the light cone, which is *not* an affine symmetric space.)

By symmetry considerations, the Euclidean ball

$$B(R) = \{(x, y, z) \,:\, x^2 + y^2 + z^2 < R^2\}$$

meets $\mathbb{H}$ in a hyperbolic ball $B(p, t(R))$ centered at $p = (0, 0, 1)$. Similarly, a point v on the one-sheeted hyperboloid $Q(v) = 1$ lies in the Euclidean ball $B(R)$ if and only if the geodesic $\ell(v)$ passes through the hyperbolic ball $B(p, t(R))$. Thus the counting theorems 2.2 and 2.3 also address instances of the following:

> Problem: Estimate the number of points in an orbit Γv which meet the Euclidean ball $B(R)$, where Γ is a lattice in a Lie group G acting linearly on a real vector space.

We will return to this problem (which forms the subject of [DRS]) in §6.

3 Equidistribution: the wavefront lemma

As we will see below, associated to an affine symmetric space G/H is a decomposition $G = HAK$ generalizing the polar decomposition KAK for a Riemannian symmetric space. Given this decomposition, the following lemma carries most of the proof of the equidistribution theorem.

Theorem 3.1 (The wavefront lemma) *For any open neighborhood U of the identity in G, there exists an open set $V \subset G$ such that*

$$HVg \subset HgU$$

for all g in AK.

Geometrically, this lemma asserts that the translate of a slightly thickened copy of H remains, like a focused wavefront, near a single H-orbit. Assuming this, we can complete the:

Proof of Theorem 1.2(Equidistribution). Let $X = \Gamma\backslash G$, $Y = (\Gamma \cap H)\backslash H$, and let $\alpha(g)$ be a compactly supported continuous function on X. Let g_n be a sequence of elements of G tending to infinity in $H\backslash G$. We need to show that

$$\frac{1}{m(Yg_n)} \int_Y \alpha(h)dh \to \frac{1}{m(X)} \int_X \alpha(g)dg$$

as n tends to infinity.

Since $G = HAK$, we can assume that the g_n lie in AK. Given $\epsilon > 0$, choose an open neighborhood U of the identity in G such that $|\alpha(gu) - \alpha(g)| < \epsilon$ for all u in U. By the wavefront lemma, there is an open neighborhood V of the identity in G such that $HVg \subset HgU$ for all g in AK.

By mixing (Theorem 1.1),

$$\frac{1}{m(YV)} \int_{YVg_n} \alpha(g)dg \;=\; \frac{1}{m(YV)} \int_{\Gamma\backslash G} \chi_{YV}(g)\alpha(gg_n)dg \to \frac{1}{m(X)} \int_X \alpha(g)dg$$

as g_n tends to infinity, where χ_{YV} is the indicator function of YV. Thus there is an N such that the integrals above differ by at most ϵ for all $n > N$.

We now analyze the integral over YVg_n in light of the wavefront lemma. Since Y is an H-orbit on X, the restriction of Haar measure on X to YVg_n is a continuous linear combination of translates of Haar measure on Y by elements of Vg_n. By the wavefront lemma, $YVg_n \subset Yg_nU$, so the integral above lies within the convex hull of the quantities

$$\frac{1}{m(Y)} \int_{Yg_n u} \alpha(h)dh \quad = \quad \frac{1}{m(Y)} \int_{Yg_n} \alpha(hu^{-1})dh$$

as u ranges over U. By the choice of U (i.e. by uniform continuity of α), each integral above differs by at most ϵ from

$$\frac{1}{m(Y)} \int_{Yg_n} \alpha(h)dh,$$

so

$$\left| \frac{1}{m(Y)} \int_{Yg_n} \alpha(h)dh - \frac{1}{m(X)} \int_X \alpha(g)dg \right| < 2\epsilon$$

for all $n > N$. These two quantities therefore converge as $n \to \infty$. ∎

To explain the proof of the wavefront lemma, we first treat a simple case that carries all the main ideas.

Let e_i be a basis for $\mathbb{R}^n$. Denote by

G : the group $SL_n(\mathbb{R})$ of orientation-preserving linear transformations of $\mathbb{R}^n$;

K : the maximal compact subgroup $SO_n(\mathbb{R})$ of G, consisting of transformations preserving the Euclidean norm

$$\left|\left|\sum \alpha_i e_i\right|\right|^2 \quad = \quad \sum \alpha_i^2;$$

A : the maximal abelian subgroup consisting of diagonal matrices with respect to this basis; and by

N : the maximal unipotent subgroup consisting of upper-triangular matrices with $n_{ii} = 1$ along the diagonal.

We recall two structure theorems for G (see [Kn, V.2, V.4]):

(a) The polar decomposition $G = HAK$. This decomposition is not unique in general (consider $K = HK$), but every element of G can be so factored; this is the property we will use. Even though $H = K$, we have denoted them by separate letters because the HAK decomposition will generalize to affine symmetric spaces.

(b) The Iwasawa decomposition $G = HAN$. We will use the fact that the multiplication map $H \times A \times N \to G$ is a diffeomorphism near the identity, so that every small g can be factored as han with h, a and n small. This is immediate from the the fact that

$$\mathfrak{h} \oplus \mathfrak{a} \oplus \mathfrak{n} \;=\; \mathfrak{g},$$

i.e. the Lie algebras of H, A and N span that of G. In reality every element of G admits a unique HAN decomposition, as follows from the Gram-Schmidt process for constructing an orthonormal basis.

Crucial to the proof is the following dynamical relation between A and N.

Lemma 3.2 (Contraction of N) . *Let $a \in A$ be a diagonal matrix with decreasing entries ($|a_{jj}| \le |a_{ii}|$ whenever $j > i$). Then conjugation by a contracts N, in the sense that*

$$|(a^{-1}na)_{ij}| \le |n_{ij}|$$

for any n in N.

Proof. If $j \ge i$, then $|a_{jj}/a_{ii}| \le 1$, so

$$|(a^{-1}na)_{ij}| \;=\; |a_{ii}^{-1}n_{ij}a_{jj}| \;\le\; |n_{ij}|;$$

while if $j < i$, $n_{ij} = 0$. ■

Corollary 3.3 *There are arbitrarily small neighborhoods U of the identity in N such that $a^{-1}Ua \subset U$.*

We now prove the wavefront lemma for the special case $H = K$. Since K is the fixed-point set of the Cartan involution $\theta(g) = (g^t)^{-1}$, $K\backslash G$ is a symmetric space.

Proof of Theorem 3.1 for $G = SL_n(\mathbb{R})$, $H = K$.

Let g be an arbitrary element of AK. For the moment, assume that g is an element of A with decreasing diagonal entries, as in the lemma above.

Choose neighborhoods V_a and V_n in A and N such that $V_aV_n \subset U$ and such that $g^{-1}V_ng \subset V_n$. Let $V = HV_aV_n$. Then

$$HVg \;=\; HV_agg^{-1}V_ng \subset HgV_aV_n \subset HgU$$

as required. (Note that g commutes with V_a since A is abelian.)

This argument produces a V which works for g in A with decreasing diagonal entries. Now for an arbitrary g in A, there is a permutation of the basis for $\mathbb{R}^n$ such that the diagonal entries of g are decreasing; all other considerations being natural, there is a V which works for this type of g as well. Since the number of such permutations is finite, we can intersect these V to obtain a neighborhood which works simultaneously for all elements of A.

To complete the proof, we must treat the case of an arbitrary element of AK. This case follows easily from compactness of K. First, choose $U' \subset U$ such that $k^{-1}U'k \subset U$ for all element of K. Choose V such that $HVa \subset HaU'$ for all a in A. Then for any $g = ak$,

$$HVg \;=\; HVak \subset HaU'k \;=\; Hakk^{-1}U'k \subset HakU \;=\; HgU.$$

∎

Remark on unipotent actions. The equidistribution theorem of this section can also be studied in light of the general theory of invariant measures on homogeneous spaces. The natural algebraic measure μ_g on $(\Gamma \cap H)\backslash Hg \subset \Gamma\backslash G$ is invariant under $g^{-1}Hg$, so (informally speaking) as $g \to \infty$ any limiting measure ν is invariant under a limiting Lie subgroup $\overline{H} = \lim g^{-1}Hg$. If $\overline{H}$ contains a unipotent subgroup, then recent work of Ratner places strong restrictions on the possibilities for ν (see [Rat1], [Rat3], [Rat2], [Rat4]).

This idea is quite transparent in the hyperbolic plane: while a large sphere is invariant under a conjugate $g^{-1}Kg$ of a fixed compact group K, as the center tends to infinity the sphere converges to a horocycle, invariant under the unipotent subgroup N.

Rather than pursuing this direction, we have relied on the simpler mixing result and the geometry of affine symmetric spaces.

4 Structure of affine symmetric spaces

In this section we establish the HAK decomposition and the wavefront lemma for general affine symmetric spaces G/H. We begin with some structure theorems, following [Sch].

Let $\mathfrak{g}$ denote the Lie algebra of G, and let $\sigma : G \to G$ be the involution whose fixed points are H. The differential of σ at the identity gives a linear involution (which we denote by the same letter) $\sigma : \mathfrak{g} \to \mathfrak{g}$. Writing $\mathfrak{g}$ as a direct sum of σ-eigenspaces, we obtain the decomposition $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{q}$, where $\sigma|_{\mathfrak{h}} = +1$ and $\sigma|_{\mathfrak{q}} = -1$. Then $\mathfrak{h}$ is the Lie algebra of H.

One can construct a Cartan involution θ of G which commutes with the affine symmetric involution σ [Sch, Prop 7.1.1]. Then the Lie algebra of G may also be written $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, where $\theta|_{\mathfrak{k}} = +1$ and $\theta|_{\mathfrak{p}} = -1$. Since θ is a Cartan involution, $\mathfrak{k}$ is the Lie algebra of a maximal compact subgroup K.

The linear map

$$\mathrm{ad}_X : \mathfrak{g} \to \mathfrak{g}$$

is defined, for each X in $\mathfrak{g}$, by $\mathrm{ad}_X(Y) = [X, Y]$. For semisimple Lie algebras, it is a standard fact [Kn, Section 1.2.] that

$$< X, Y > \ = \ -\operatorname{tr}(\mathrm{ad}_X \, \mathrm{ad}_{\theta(Y)})$$

is an inner product on $\mathfrak{g}$, with respect to which ad_X is self-adjoint for all X in $\mathfrak{p}$.

To proceed further, we briefly recall the root space decomposition of $\mathfrak{g}$ (cf. [Kn, Ch.4]).

Choose a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p} \cap \mathfrak{q}$. Then $\mathfrak{a}$ is the Lie algebra of an abelian subgroup A, and the exponential map $\mathfrak{a} \to A$ is a diffeomorphism. The mappings ad_Z for Z in $\mathfrak{a}$ are commuting and self-adjoint. Therefore there is a basis for $\mathfrak{g}$ with respect to which all ad_Z are diagonal. The *roots* are a finite set $\Sigma_{\mathfrak{a}} \subset \mathfrak{a}^*$ (the dual space of $\mathfrak{a}$) such that for each Z,

$$< \alpha(Z) \ : \ \alpha \in \Sigma_{\mathfrak{a}} >$$

enumerates the eigenvalues of ad_Z. The *root space* (eigenspace) of the root α is denoted $\mathfrak{g}^\alpha$. Roots and root spaces are characterized by the equation:

$$[Z, X_\alpha] \ = \ \alpha(Z) X_\alpha$$

for all $X_\alpha \in \mathfrak{g}^\alpha$ and for all $Z \in \mathfrak{a}$. The Lie algebra $\mathfrak{g}$ is a direct sum of root spaces.

Next we choose a system of positive roots. The hyperplanes $\{Z | \alpha(Z) = 0\}$ for $\alpha \in \Sigma_{\mathfrak{a}}$ divide $\mathfrak{a}$ into finitely many open regions called *Weyl chambers*. Pick a Weyl chamber $\mathcal{C}$. The *positive roots* $\Sigma_{\mathfrak{a}}^+$ consist of those α for which $\alpha(\mathcal{C}) > 0$. (Of course the space of positive roots depends on the choice of $\mathcal{C}$.)

Let $\mathfrak{n}$ be the linear span of the positive root spaces; it is the Lie algebra of a (unipotent) subgroup $N \subset G$. Let $\bar{\mathfrak{n}}$ denote the span of the negative root spaces, and let $\mathfrak{g}^0$ be the zero eigenspace of the ad_Z. Then

$$\mathfrak{g} \;=\; \mathfrak{n} \oplus \mathfrak{g}^0 \oplus \bar{\mathfrak{n}},$$

because every root is either positive, negative or zero on a given Weyl chamber.

Remark. In the case of $SL_n(\mathbb{R})$, A can be taken to be the group of diagonal matrices. There are $n!$ Weyl chambers, each corresponding to an ordering of the standard basis for $\mathbb{R}^n$. The matrices of A with strictly decreasing diagonal entries form the exponential of a Weyl chamber, for which the corresponding unipotent subgroup N consists of upper triangular matrices.

Proposition 4.1 (Contraction of N) *Let $\mathcal{C}$ be a Weyl chamber and let N be the corresponding unipotent subgroup. Then there exist arbitrarily small neighborhoods U of the identity in N such that $a^{-1}Ua \subset U$ for all a in $\exp(\overline{\mathcal{C}})$.*

Proof. Let $c_a : N \to N$ be the conjugation map $n \to ana^{-1}$; this is an automorphism of N. Since $\exp : \mathfrak{n} \to N$ is a group isomorphism, it suffices to verify contraction on the level of the Lie algebra $\mathfrak{n}$ of N.

To this end, let

$$\mathrm{Ad}(a) : \mathfrak{n} \to \mathfrak{n}$$

denote the differential of c_a at the identity. For any $X \in \mathfrak{n}$, we may write

$$X \;=\; \sum_{\alpha \in \Sigma_{\mathfrak{a}}^+} x_\alpha X_\alpha$$

where X_α lies in $\mathfrak{g}^\alpha$.

Now write $a = \exp(Z)$ where Z lies in the closure of the Weyl chamber $\mathcal{C}$. By the well-known identity $\mathrm{Ad}(\exp(Z)) = \exp(\mathrm{ad}_Z)$ [Kn, Prop. A.111.], we may write

$$\begin{aligned}
\mathrm{Ad}(a^{-1})X \;&=\; \sum_{\alpha \in \Sigma_{\mathfrak{a}}^+} x_\alpha \,\mathrm{Ad}(a^{-1})(X_\alpha) \\
&=\; \sum_{\alpha \in \Sigma_{\mathfrak{a}}^+} x_\alpha \exp(-\,\mathrm{ad}_Z) X_\alpha \\
&=\; \sum_{\alpha \in \Sigma_{\mathfrak{a}}^+} x_\alpha \exp(-\alpha(Z)) X_\alpha.
\end{aligned}$$

But $\alpha(Z) \geq 0$ for all $\alpha \in \Sigma^+_{\mathfrak{a}}$. Therefore $\mathrm{Ad}(a^{-1})$ contracts a product neighborhood U' of the identity in $\mathfrak{n}$, which can be taken to be arbitrarily small.

Since $c_{a^{-1}}(\exp(X)) = \exp(\mathrm{Ad}(a^{-1})X))$, we have $a^{-1}Ua \subset U$. ■

Next we state two structure theorems for G.

Proposition 4.2 (HAK decomposition) *The map*

$$H \times A \times K \to G$$

given by $(h, a, k) \to hak$ *is surjective.*

This proposition is well-known; see [Sch, Proposition 7.1.3.] and [F-J, Corollary 1.4.] for mild variants.

Now let

$$M = \{m \in K \ : \ ma = am \ \text{ for all } a \text{ in } A\}$$

denote the centralizer of A in K.

Proposition 4.3 (HMAN decomposition) *The map*

$$H \times M \times A \times N \longmapsto G$$

given by $(h, m, a, n) \mapsto hman$ *is an open mapping in a neighborhood of the identity.*

This proposition is a local version of the Iwasawa decomposition and it is also well-known; see [Sch, Prop. 7.1.8(ii)]. Global properties of this decomposition are discussed in [Mat1] and [Mat2].

The HAK decomposition was assumed above to deduce Theorem 1.2 (Equidistribution) from the wavefront lemma. The $HMAN$ decomposition will be used below to complete the general proof of the wavefront lemma.

For completeness we sketch the proof of these two propositions.

Definition. A connected subgroup G_0 of a Lie group G is *reductive* if it is stable under a Cartan involution θ of G.

Let $\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0$ denote the decomposition of the Lie algebra of G_0 into $+1$ and -1 eigenspaces of θ respectively. Then $\mathfrak{k}_0$ is the Lie algebra of a maximal compact subgroup K_0 of G_0.

Proposition 4.4 (KAK decomposition) *Let G_0 be a reductive group, $\mathfrak{a}_0$ a maximal abelian subspace of $\mathfrak{p}_0$, and let $A_0 = \exp(\mathfrak{a}_0)$. Then*

$$G_0 \quad = \quad K_0 A_0 K_0.$$

See [Kn, Theorem 5.20.]

Sketch of the HAK decomposition. Let g be an element of G. By [Sch, Prop. 7.1.2.], the map

$$(\mathfrak{p} \cap \mathfrak{h}) \times (\mathfrak{p} \cap \mathfrak{q}) \times K \to G$$

given by

$$(X, Y, k) \mapsto \exp(X) \exp(Y) k$$

is surjective. Thus we can write

$$g \quad = \quad \exp(X) \exp(Y) k$$

where $\exp(X) \in H$ and $k \in K$. It remains to express $\exp(Y)$ in the form $h_0 a_0 k_0$.

Let $\mathfrak{g}_0 = (\mathfrak{h} \cap \mathfrak{k}) \oplus (\mathfrak{p} \cap \mathfrak{q})$. Then Y lies in $\mathfrak{g}_0$. Since θ and σ commute, θ stabilizes $\mathfrak{g}_0$, so $\mathfrak{g}_0$ is the Lie algebra of a connected reductive subgroup G_0 of G.

The eigenspace decomposition of $\mathfrak{g}_0$ with respect to θ is just the restriction of that of $\mathfrak{g}$, so

$$\mathfrak{k}_0 \quad = \quad \mathfrak{k} \cap \mathfrak{g}_0 \quad = \quad \mathfrak{h} \cap \mathfrak{k}$$

and

$$\mathfrak{p}_0 \quad = \quad \mathfrak{p} \cap \mathfrak{g}_0 \quad = \quad \mathfrak{p} \cap \mathfrak{q}.$$

Thus we may take $K_0 = H \cap K$ and $A_0 = A$ in the KAK decomposition $G_0 = K_0 A_0 K_0$.

We may therefore write

$$\exp(Y) \quad = \quad h_0 a_0 k_0$$

where a_0 lies in A and both h_0 and k_0 lie in $H \cap K$. Then

$$g \quad = \quad \exp(X) h_0 a_0 a k_0 k$$

expresses g in the form HAK. ∎

Sketch of the HMAN decomposition.

It suffices to show that

$$\mathfrak{h} + \mathfrak{m} + \mathfrak{a} + \mathfrak{n} \;=\; \mathfrak{g},$$

since this implies the map $H \times M \times A \times N \to G$ is a submersion at the identity, and therefore open.

To prove this, rewrite $\mathfrak{g}$ as

$$\mathfrak{g} \;=\; \mathfrak{n} \oplus \mathfrak{g}^0 \oplus \bar{\mathfrak{n}}.$$

Since $\mathfrak{n}$ already appears, it suffices to show that $\bar{\mathfrak{n}}$ and $\mathfrak{g}^0$ lie in the span of $\mathfrak{n}$, $\mathfrak{h}$, $\mathfrak{m}$ and $\mathfrak{a}$.

First note that

$$\sigma(\mathfrak{n}) \;=\; \bar{\mathfrak{n}}.$$

Indeed, $\sigma(Z) = -Z$ for all Z in $\mathfrak{a}$, so $\sigma(\mathfrak{g}^\alpha) = \mathfrak{g}^{-\alpha}$. Thus σ exchanges the positive and negative root spaces, and therefore $\mathfrak{n}$ and $\bar{\mathfrak{n}}$.

From this it follows that

$$\bar{\mathfrak{n}} \subset \mathfrak{h} + \mathfrak{n},$$

for if X lies in $\bar{\mathfrak{n}}$, then

$$X \;=\; (X + \sigma(X)) - \sigma(X),$$

and $X + \sigma(X)$ is in $\mathfrak{h}$, while $\sigma(X) \in \mathfrak{n}$.

It remains to show that

$$\mathfrak{g}^0 \subset \mathfrak{m} + \mathfrak{a} + \mathfrak{h}.$$

Recall that $\mathfrak{g}^0$ consists exactly of those X with $[Z, X] = 0$ for all Z in $\mathfrak{a}$. Given X in $\mathfrak{g}^0$, we may write

$$X \;=\; \frac{1}{2}(X + \theta(X)) + \frac{1}{2}(X - \theta(X))$$

where $\theta(X)$ and $\sigma(X)$ are also in $\mathfrak{g}^0$, since each involution stabilizes $\mathfrak{a}$. Then $X + \theta(X)$ lies in $\mathfrak{k} \cap \mathfrak{g}^0$, which is exactly the Lie algebra $\mathfrak{m}$ of the centralizer of A in K.

We claim that

$$Y = \frac{1}{2}(X - \theta(X))$$

lies in $\mathfrak{h} + \mathfrak{a}$. To see this, write

$$Y = \frac{1}{2}(Y + \sigma(Y)) + \frac{1}{2}(Y - \sigma(Y)).$$

Then $Y+\sigma(Y)$ lies in $\mathfrak{h}$, and $Y-\sigma(Y)$ lies in $\mathfrak{p}\cap\mathfrak{q}\cap\mathfrak{g}^0$; in particular $Y-\sigma(Y)$ commutes with all of $\mathfrak{a}$. But $\mathfrak{a}$ is a maximal abelian subspace of $\mathfrak{p}\cap\mathfrak{q}$, so $Y - \sigma(Y) \in \mathfrak{a}$. ∎

We can now complete the:

Proof of Theorem 3.1 (The wavefront lemma). Given the preliminaries above, the proof follows the same lines as that for $SL_n(\mathbb{R})$.

We are given that g lies in AK; for the moment assume g lies in A. Then g belongs to $\exp(\overline{\mathcal{C}})$ for some Weyl chamber $\mathcal{C}$. Let N be the corresponding unipotent subgroup, so that the contraction Lemma 4.1 holds. By the HMAN decomposition, there exist neighborhoods V_m, V_a and V_n in M, A and N respectively, such that $V_mV_aV_n \subset U$ and $g^{-1}V_ng \subset V_n$ for $g \in \exp(\bar{\mathcal{C}})$.

Let $V = HV_mV_aV_n$. Since M and A commute,

$$\begin{aligned} HVg &= HV_mV_aV_ng \\ &= HgV_mV_a(g^{-1}V_ng) \\ &\subset HgV_mV_aV_n \\ &\subset HgU. \end{aligned}$$

Thus we have produced a V which works for all g in $\exp(\bar{\mathcal{C}})$. Since the number of Weyl chambers is finite, we may intersect these V's to obtain a neighborhood which works simultaneously for all $g \in A$.

We now treat the general case of an element $g = ak$ lying in AK. Because K is compact, we can find $U' \subset U$ such that $k^{-1}U'k \subset U$ for all $k \in K$. Then choose V so that $HVa \subset HaU'$ for all $a \in A$. It follows that

$$HVg = HVak \subset HaU'k = Hakk^{-1}U'k = Hgk^{-1}Uk \subset HgU$$

as desired. ∎

5 Counting

Our approach to counting is along the same lines as §2 of [DRS], with an emphasis on axiomatics.

Given a sequence of sets of finite measure B_n in the affine symmetric space $V = G/H$ such that the measure $m(B_n) \to \infty$, we will show under suitable hypotheses that

$$|\Gamma v \cap B_n| \sim Cm(B_n)$$

for an explicit constant $C > 0$. Here v is the coset $[H]$.

Aside from working in the affine symmetric setting, there are two crucial hypotheses leading to this asymptotic estimate:

(1) Γ meets H in a lattice; and

(2) the sets B_n are well-rounded.

We will see that even without (2) the asymptotic estimate holds in a weaker sense.

Fibrations and integration. As a preliminary, suppose $A \subset B \subset G$ is a chain of closed subgroups of a Lie group G. Then there is a fibration

$$A\backslash B @>>> A\backslash G @>\pi>> B\backslash G.$$

More precisely, $A\backslash G$ fibers over $B\backslash G$ with $A\backslash Bg$ as the fiber over Bg.

Now assume A, B and G are unimodular. (Any group which contains a lattice is unimodular, so this condition will be satisfied in our applications below.) Then $A\backslash B$ admits a B-invariant measure, and similarly for $A\backslash G$ and $B\backslash G$. We may normalize so the measure on $A\backslash G$ is the product of the measures on $B\backslash G$ and $A\backslash B$. (Compare [Wl, Ch. II, §9].)

If β is in $L^1(A\backslash G)$, then the *pushforward*

$$(\pi_*\beta)(g) \;=\; \int_{A\backslash Bg} \beta(b)db$$

is finite almost everywhere, $\pi_*\beta \in L^1(B\backslash G)$ and

$$\int_{B\backslash G} \pi_*\beta \;=\; \int_{A\backslash G} \beta.$$

In addition, if $m(A\backslash B) < \infty$, then the *pullback* $\pi^*\alpha$ is in $L^1(A\backslash G)$ for any $\alpha \in L^1(B\backslash G)$.

Weak convergence. While we are interested in studying the number of points in $\Gamma v \cap B_n$, it proves fruitful to consider more generally the count

$$F_n(g) \quad = \quad |\Gamma v \cap g B_n|$$

giving the part of the orbit in B_n shifted by g. It is clear that F_n descends to a function

$$F_n : \Gamma \backslash G \to \mathbb{R} \cup \{\infty\},$$

which we denote by the same letter.

The function $F_n(g)$ can be built from

$$\chi_n(g) \quad = \quad \begin{cases} 1 & \text{if } v \in g B_n \\ 0 & \text{otherwise.} \end{cases}$$

Since H stabilizes v, the function χ_n descends to

$$\chi_n : H \backslash G \to \mathbb{R}$$

which is simply the indicator function of B_n^{-1}. In particular

$$\int_{H \backslash G} \chi_n \quad = \quad m(B_n) < \infty.$$

$$\begin{array}{ccccc} & & (\Gamma \cap H)\backslash \Gamma & & \\ & & \big\downarrow & & \\ (\Gamma \cap H)\backslash H & \longrightarrow & (\Gamma \cap H)\backslash G & \xrightarrow{\ \phi\ } & H \backslash G \\ & & \pi \big\downarrow & & \\ & & \Gamma \backslash G & & \end{array}$$

Figure 5. Pullbacks and pushforwards.

To describe the relationship between χ_n and F_n, it is useful to refer to Figure 5, where each vertical and horizontal triple is a fibration. Then F_n can be expressed as

$$F_n(g) \quad = \quad \sum_{\gamma \in (\Gamma \cap H) \backslash \Gamma} \begin{cases} 1 & \text{if } v \in \gamma g B_n \\ 0 & \text{otherwise} \end{cases} \quad = \quad \sum_{(\Gamma \cap H) \backslash \Gamma} \chi_n(\gamma g) \quad = \quad \pi_* \phi^*(\chi_n).$$

Note that the fibers of ϕ have finite volume, so integrability of χ_n implies the same for $\phi^*\chi_n$ and F_n.

Our first result requires only measure-theoretic assumptions on B_n.

Theorem 5.1 *If $m(B_n) \to \infty$, then the function $F_n(g)/m(B_n)$ tends weakly to a constant function C on $X = \Gamma\backslash G$. More precisely, as $n \to \infty$,*

$$\frac{1}{m(B_n)} \int_X F_n(g)\alpha(g)dg \to C \int_X \alpha(g)dg$$

for any compactly supported continuous function α, where

$$C = \frac{m((\Gamma \cap H)\backslash H)}{m(\Gamma\backslash G)}.$$

Question. Can weak convergence be replaced by pointwise convergence almost everywhere?

Proof. The idea of the proof is to transfer the integral of F_n against α to an integral against χ_n on $H\backslash G$, again making reference to Figure 5. Thus

$$\int_X F_n\alpha = \int_{\Gamma\backslash G} (\pi_*\phi^*\chi_n)(g)\alpha(g)dg = \int_{H\backslash G} \chi_n(g)(\phi_*\pi^*\alpha)(g)dg = \int_{H\backslash G} \chi_n(g)\beta(g),$$

where ϕ_* is defined by integration over the fibers of ϕ. Thus

$$\beta(g) = \int_{(\Gamma\cap H)\backslash Hg} \alpha(h)dh$$

(which clearly lives on $H\backslash G$).

By Theorem 1.2(Equidistribution),

$$\beta(g) \to \frac{m((\Gamma \cap H)\backslash H)}{m(\Gamma\backslash G)} \int_X \alpha$$

as g tends to infinity in $H\backslash G$. On the other hand,

$$\frac{1}{m(B_n)} \int_X F_n\alpha = \frac{\int_{H\backslash G} \chi_n\beta}{\int_{H\backslash G} \chi_n}$$

is just the average of β over the set $B_n^{-1} \subset H\backslash G$, whose measure is tending to infinity. Thus

$$\frac{1}{m(B_n)} \int_X F_n\alpha \to \frac{m((\Gamma \cap H)\backslash H)}{m(\Gamma\backslash G)} \int_X \alpha = C \int_X \alpha.$$

∎

Remark. The argument above requires only that $\beta(g)$ tends to a constant (i.e. $(\Gamma\cap H)\backslash Hg$ becomes equidistributed) as g tends to infinity in a *measure theoretic* sense. That is, we used only that $\beta(g)$ can be made as close as one likes to a constant by neglecting a set of g of finite measure.

We now impose the additional topological assumption that B_n is a well-rounded sequence (see §1 for the definition).

Theorem 5.2 *If $m(B_n) \to \infty$ and B_n is a well-rounded sequence, then $F_n(g)/m(B_n) \to C$ pointwise as $n \to \infty$.*

Corollary 5.3 *For a well-rounded sequence,*

$$F_n(\mathrm{id}) \quad = \quad |\Gamma v \cap B_n| \sim \frac{m((\Gamma \cap H)\backslash H)}{m(\Gamma\backslash G)} m(B_n).$$

This corollary is Theorem 1.4(Counting).

Proof of the theorem. To simplify notation, we prove that $F_n(\mathrm{id})/m(B_n) \to C$, this being the main case of interest.

By Proposition 1.3, for any $\epsilon > 0$ we can find a symmetric neighborhood U of the identity such that $m(B_n') > (1-\epsilon)m(B_n)$, where

$$B_n' = \bigcap_{g\in U} gB_n.$$

Let $F_n'(g) = |\Gamma v \cap g B_n'|$. Then $F_n'(g) \leq F_n(\mathrm{id})$ for all g in U. But $m(B_n') \to \infty$, so by Theorem 5.1 $F_n'(g)/m(B_n')$ tends weakly to the constant function C on $\Gamma\backslash G$. Pairing F_n' with a bump function α supported in $\Gamma\backslash U$ such that $\int \alpha = 1$, we find:

$$\frac{1}{m(B_n')}\int_{\Gamma\backslash G} F_n'(g)\alpha(g)dg \quad \leq \quad \frac{1}{(1-\epsilon)m(B_n)}\int_{\Gamma\backslash G} F_n(\mathrm{id})\alpha(g)dg \quad = \quad \frac{F_n(\mathrm{id})}{(1-\epsilon)m(B_n)}.$$

Consequently

$$C = \lim \frac{1}{m(B_n')}\int_{\Gamma\backslash G} F_n'\alpha \quad \leq \quad \liminf \frac{F_n(\mathrm{id})}{m(B_n)}.$$

Replacing B_n' by

$$\bigcup_{g\in U} B_n g$$

yields the upper bound, showing that $F_n(\mathrm{id})/m(B_n) \to C$.

The argument for convergence of $F_n(g)/m(B_n)$ is similar. ∎

6 Representations of G and integral points on homogeneous varieties.

For completeness, we connect the results above with some of the central theorems obtained in [DRS] by very different means. These ideas can be used to count integral points on affine homogeneous varieties, and coupled with the circle method of Hardy, Littlewood and Ramanujan, they lead to a proof of Siegel's mass formula [ERS].

Let G be a connected semisimple Lie group with finite center and maximal compact subgroup K. Let $\rho : G \to GL(S)$ be a representation of G acting on the left on a finite-dimensional real vector space S. Let V be an affine symmetric orbit of G in S; this means $V = Gv$ for some v in S, and the stabilizer H of v is the fixed point set of an involution on G.

For convenience, in this section we replace the sequence B_n by a continuous family of sets $B_t \subset G/H$, defined as follows.

Let $||\cdot||$ be a K-invariant Euclidean norm on S; this means $||\sum \alpha_i s_i||^2 = \sum \alpha_i^2$ for a suitable basis s_i. Let

$$B = \{s \;:\; ||s|| < 1\} \subset S$$

be the unit ball in this norm. For $t > 0$ define

$$B_t \;\; = \;\; \{[gH] : ||gv|| < t\} \subset G/H.$$

Then B_t corresponds to the dilation of B by t, under the identification of V with G/H.

Theorem 6.1 *Given $\epsilon > 0$, there is a neighborhood U of the identity in G and a $T > 0$ such that*

$$\frac{m(U \cdot \partial B_t)}{m(B_t)} < \epsilon$$

for all $t > T$. In other words, B_t is a continuous family of well-rounded subsets of G/H.

The proof relies on an elementary fact about linear maps and an estimate for the measure $m(B_t)$ as a function of t.

Proposition 6.2 *For any $\epsilon > 0$, there is a neighborhood U of the identity in G such that*

$$B_{(1-\epsilon)t} \subset gB_t \subset B_{(1+\epsilon)t}$$

for all $g \in U$ and for all $t > 0$.

Proof. It is easy to see that

$$(1-\epsilon)B \subset gB \subset (1+\epsilon)B$$

for any linear map $g : S \to S$ sufficiently close to the identity. Since linear maps commute with dilations, the proposition follows. (Compare [DRS, Lemma 2.2].) ■

Proposition 6.3 *There are constants $a(\epsilon)$, $b(\epsilon)$ tending to 1 as $\epsilon \to 0$ such that*

$$b(\epsilon) \leq \liminf_{t\to\infty} \frac{m(B_{(1-\epsilon)t})}{m(B_t)} \leq \limsup_{t\to\infty} \frac{m(B_{(1+\epsilon)t})}{m(B_t)} \leq a(\epsilon).$$

For a proof, see [DRS, Appendix 1].

Proof of Theorem 6.1. Given $\epsilon > 0$, choose $\delta > 0$ and $T > 0$ by Proposition 6.3 so that

$$\frac{m(B_{(1+\delta)t})}{m(B_{(1-\delta)t})} < 1+\epsilon$$

for all $t > T$. By Proposition 6.2, we can find a neighborhood U of the identity in G such that $U \cdot \partial B_t \subset B_{(1+\delta)t}$ and $(U \cdot \partial B_t) \cap B_{(1-\delta)t} = \emptyset$. Theorem 6.1 follows immediately. ■

Now let $\Gamma \subset G$ be a lattice satisfying the conditions of §1; that is, $\Gamma \cap H$ is a lattice in H, and Γ has dense projection to G/G' for any noncompact normal subgroup G' of positive dimension.

Applying Theorem 1.4 (Counting), we deduce:

Theorem 6.4 *As $t \to \infty$,*

$$|\Gamma v \cap tB| \;\; = \;\; |[\Gamma H] \cap B_t| \sim \frac{m((\Gamma \cap H)\backslash H)}{m(\Gamma \backslash G)} m(B_t).$$

Remarks. Since $|\Gamma v \cap tB|$ is simply the number of points in the orbit of v with norm less than t, we have obtained a new proof of the central counting result (Theorem 1.2) of [DRS].

To count integral points on affine symmetric varieties defined over $\mathbb{Z}$, one may appeal to a result of Borel and Harish-Chandra which states that $V(\mathbb{Z})$ falls into finitely many orbits under the action of $\Gamma = G(\mathbb{Z})$. This reduces the problem to the case of a single orbit, which is handled by the theorem above. For details, see [DRS].

It seems likely that the sets B_t are well-rounded for much more general choices of B, and therefore the counting result above holds for these sets as well. For example, Proposition 6.2 holds when B is any open bounded convex neighborhood of the origin in S, and the methods of [DRS, Appendix 1] can probably be adapted so that Proposition 6.3 covers this case too.

7 Beyond affine symmetric spaces

To conclude, we present a few examples, counterexamples and open questions connected with spaces G/H which are not affine symmetric.

Horocycles. The simplest such example occurs when $H = N$, a maximal unipotent subgroup of $G = PSL_2(\mathbb{R})$. The space G/N is not affine symmetric; geometrically this is reflected in the fact that a horocycle is not the fixed-point set of any isometric involution of the hyperbolic plane.

Nevertheless the following conditional equidistribution result holds.

Theorem 7.1 *Let C be a closed horocycle on a hyperbolic surface Σ of finite volume, and let C_t denote the parallel horocycle of length* $\exp(t)\,\mathrm{length}(C)$*. Then C_t becomes equidistributed on Σ as t tends to $+\infty$.*

Sketch of the proof. One may use the same line of argument as the proof of Theorems 2.1 and 2.4. Thicken the horocycle to an open set U of vectors nearly normal to C and pointing away from the cusp. Then $g_t(U)$ lies close to C_t for $t > 0$, and $g_t(U)$ becomes equidistributed by mixing. ■

Remark. On the other hand, $g_t(U)$ and C_t diverge for negative t, and C_t is asymptotic to a cusp of Σ as $t \to -\infty$.

To state a counting theorem, we fix a horocycle C on Σ with a lift c to $\mathbb{H}$, and let $N(R)$ denote the number of horocycles in the orbit Γc which meet $B(p, R)$.

Theorem 7.2 *As $R \to \infty$,*

$$N(R) \sim \frac{1}{\pi}\frac{\mathrm{length}(C)}{\mathrm{area}(\Sigma)}\,\mathrm{area}(B(p, R)).$$

Sketch of the proof. Consider the lift C' of C to the covering space Σ' induced by $\pi_1(C)$. The surface Σ' has two ends: a finite volume cusp, and an annular end with exponential volume growth. To estimate $N(R)$, one may mimic the proof of Theorem 2.3, using that fact that an R-neighborhood of C' on Σ'has

$$\operatorname{area}(B(C', R)) \sim \operatorname{length}(C) \exp(R).$$

∎

Unfortunately, this discussion does not seem to extend in a straightforward way to maximal unipotent subgroups in higher rank groups. Compare, however, [FMT] for the case of H a maximal parabolic subgroup.

Counterexamples. In §6 we discussed the number of points in a Γ-orbit on an affine symmetric variety V presented as a G-orbit for a linear representation ρ. One might hope that this asymptotic count also holds without the assumption that $V = G/H$ is an affine symmetric space. This is not the case; below we sketch a (non-affine symmetric) example where the conclusion of Theorem 6.4 fails to hold. For more details see [EMS].

Let $G = SL(2, \mathbb{C})$, let H be the subgroup of real diagonal matrices in G, and let $\Gamma = gSL(2, \mathbb{Z}[i])g^{-1}$ where g is a real matrix chosen to conjugate a hyperbolic element of $SL_2(\mathbb{Z})$ into H. Then Γ meets H is a lattice, but the space G/H is *not* affine symmetric.

Here is a representation of G with an orbit isomorphic to G/H. Fix a large positive number N, and let (z_1, z_2) be coordinates on $\mathbb{C}^2$. Let S be the vector space of polynomial functions $f(z_1, z_2, \bar{z_1}, \bar{z_2})$ on $\mathbb{C}^2$ which are homogeneous of degree N in z_1 and z_2, and also homogeneous of degree N in $\bar{z_1}$ and $\bar{z_2}$. The monomials $z_1^m z_2^{N-m} \bar{z_1}^n \bar{z_2}^{N-n}$, where $0 \le m, n \le N$, are a basis for S.

Since G acts linearly on $\mathbb{C}^2$, it also acts on S by substitution. This determines a representation $\rho : G \to GL(S)$ of the form $\rho(g)f(z) = f(g^{-1}z)$ for z in $\mathbb{C}^2$.

Let v be the polynomial

$$v(z_1, z_2) \quad = \quad \left[\frac{z_1\bar{z}_2 + \bar{z}_1 z_2}{2}\right]^2 \left[\frac{-z_1\bar{z}_2 + \bar{z}_1 z_2}{2i}\right]^{N-2}.$$

Then stabilizer of v is exactly H, so the variety $V = Gv$ is naturally identified with G/H.

Next we consider the distribution of the subset $\Gamma v \subset V$. Let $K = SU(2,\mathbb{C})$ be a maximal compact subgroup of G, let $\|\cdot\|$ be a K-invariant norm on S, and let B be the open unit ball in this norm. Following §§5 and 6, define $B_t \subset G/H$ by $B_t = \{[gH] : gv \in tB\}$, and let $F_t(g) = |\Gamma v \cap g(tB)|$.

Theorem 7.3 *There exists a* nonconstant *smooth positive function* $\Lambda(g)$ *on* $\Gamma\backslash G$*, such that*

$$F_t(g) \sim \Lambda(g)\frac{m((\Gamma \cap H)\backslash H)}{m(\Gamma\backslash G)}m(B_t)$$

as $t \to \infty$*.*

Remarks. The asymptotic count above would have the same form as that of Theorems 1.4 and 6.4 only if $\Lambda(g)$ were identically equal to one. (Compare Theorem 5.2, which asserts that $F_n(g)/m(B_n)$ tends to a constant independent of g in the affine symmetric setting.) Here the number of lattice points still grows like the volume, but the constant of proportionality is subtle (it depends on g).

This dependence results from a failure of the equidistribution Theorem 1.2. To explain this, let $L = SL(2,\mathbb{R}) \subset G$ denote the subgroup of real matrices. In the example above we have

$$H \subset L \subset G$$

where $\Gamma \cap L$, being a conjugate of $SL_2(\mathbb{Z})$, is a lattice in L. Thus if g tends to infinity in L, the translates Yg of the H-orbit

$$Y = (\Gamma \cap H)\backslash H$$

lie in

$$Z = (\Gamma \cap L)\backslash L,$$

so they cannot become equidistributed in $\Gamma\backslash G$. Roughly speaking, the count in Theorem 7.3 differs from that of Theorem 6.4 because most of the measure of the B_t is a finite distance from L.

The sets B_t above are nevertheless well-rounded, as can be seen by the methods of §6.

This example still supports the conjecture that $F_n(g)$ always converges pointwise as $n \to \infty$, in the notation of §5. And since

$$\frac{1}{m(\Gamma\backslash G)}\int_{\Gamma\backslash G}\Lambda(g)\,dg = 1$$

the count of Theorem 6.4 is still correct on average.

Open questions.

1. It seems likely that Theorem 1.4 (Counting) can be strengthened to include an error term of the form:

$$|\Gamma v \cap B_n| \sim Cm(B_n) + O(m(B_n)^\alpha)$$

for some $\alpha < 1$. Indeed, such an error term might be obtained by an analysis of the proof offered herein.

2. As remarked above, Theorem 1.2 (Equidistribution) fails to generalize when there are subgroups L between H and G which also meet Γ in a lattice. However one can hope to establish a more general equidistribution result for $(\Gamma \cap H)\backslash Hg$ by either controlling the groups which occur between H and G, or by requiring that g tend to infinity in a stronger sense.

Progress on this question appears in [EMS].

References

[DRS] W. Duke, Z. Rudnick, and P. Sarnak. Density of integer points on affine homogeneous varieties. *Preprint.*

[EMS] A. Eskin, S. Mozes, and N. Shah. Unipotent flows and counting lattice points on homogeneous spaces. *In preparation.*

[ERS] A. Eskin, Z. Rudnick, and P. Sarnak. A proof of Siegel's weight formula. *Preprint.*

[F-J] M. Flensted-Jensen. *Analysis on non-Riemannian Symmetric Spaces.* Amer. Math. Soc., 1986.

[FMT] J. Franke, Yu. Manin, and Yu. Tschinkel. Rational points of bounded height on Fano varieties. *Invent. math.* **95**(1989), 421–435.

[GHL] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry.* Spinger-Verlag, 1987.

[Hed1] G. A. Hedlund. The dynamics of geodesic flows. *Bull. Amer. Math. Soc.* **45**(1939), 241—260.

[Hed2] G. A. Hedlund. Fuchsian groups and mixtures. *Annals of Math.* **40**(1939), 370–383.

[HM] R. E. Howe and C. C. Moore. Asymptotic properties of unitary representations. *J. Functional Analysis* **32**(1979), 72–96.

[Kn] A. W. Knapp. *Representation Theory of Semisimple Groups.* Princeton University Press, 1986.

[Mg] G. A. Margulis. Applications of ergodic theory to the investigation of manifolds of negative curvature. *Funct. Anal. Appl.* **4**(1969), 335.

[Mat1] T. Matsuki. The orbits of affine symmetric spaces under the action of minimal parabolic subgroups. *J. Math. Soc. Japan* **31**(1979), 331–357.

[Mat2] T. Matsuki. Orbits on affine symmetric spaces under the action of parabolic subgroups. *Hiroshima Math. J.* **12**(1982), 307–320.

[Ran] B. Randol. The behavior under projection of dilating sets in a covering space. *Trans. Amer. Math. Soc.* **285**(1984), 855–859.

[Rat1] M. Ratner. Invariant measures for unipotent translations on homogeneous spaces. *Proc. Natl. Acad. Sci. USA* **87**(1990), 4309–4311.

[Rat2] M. Ratner. On measure rigidity of unipotent subgroups of semisimple groups. *Acta. Math.* **165**(1990), 229–309.

[Rat3] M. Ratner. Strict measure rigidity for nilpotent subgroups of solvable groups. *Invent. math.* **101**(1990), 449–482.

[Rat4] M. Ratner. On Raghunathan's measure conjecture. *Annals of Math.* **134**(1991), 545–607.

[Sch] H. Schlichtkrull. *Hyperfunctions and Harmonic Analysis on Symmetric Spaces.* Birkhäuser, 1984.

[Wl] A. Weil. *L'Intégration dans les Groupes Topologiques et ses Applications.* Hermann, 1951.

[Zim] R. Zimmer. *Ergodic Theory and Semisimple Groups.* Birkhäuser, 1984.