Unipotent Flows and Applications
Lecture Notes for Clay Institute Summer School

Alex Eskin and Dmitry Kleinbock

Very Preliminary draft

June 18, 2007

1 General introduction

1.1 The Oppenheim Conjecture

The Oppenheim Conjecture. Let

$$Q(x_1, \ldots, x_n) = \sum_{1 \le i \le j \le n} a_{ij} x_i x_j$$

be a quadratic form in n variables. We always assume that Q is *indefinite* so that (so that there exists p with $1 \le p < n$ so that after a linear change of variables, Q can be expresses as:

$$Q_p^*(y_1, \ldots, y_n) = \sum_{i=1}^{p} y_i^2 - \sum_{i=p+1}^{n} y_i^2$$

We should think of the coefficients a_{ij} of Q as real numbers (not necessarily rational or integer). One can still ask what will happen if one substitutes *integers* for the x_i. It is easy to see that if Q is a multiple of a form with rational coefficients, then the set of values $Q(\mathbb{Z}^n)$ is a discrete subset of $\mathbb{R}$. Much deeper is the following conjecture:

Conjecture 1.1 (Oppenheim, 1929)**.** *Suppose Q is not proportional to a rational form and $n \ge 5$. Then $Q(\mathbb{Z}^n)$ is dense in the real line.*

This conjecture was extended by Davenport to $n \ge 3$.

Theorem 1.2 (Margulis, 1986). *The Oppenheim Conjecture is true as long as $n \geq 3$. Thus, if $n \geq 3$ and Q is not proportional to a rational form, then $Q(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.*

This theorem is a triumph of ergodic theory. Before Margulis, the Oppenheim Conjecture was attacked by analytic number theory methods. (In particular it was known for $n \geq 21$, and for diagonal forms with $n \geq 5$).

Failure of the Oppenheim Conjecture in dimension 2. Let $\alpha > 0$ be a quadratic irrational such that $\alpha^2 \notin \mathbb{Q}$ (e.g. $\alpha = (1+\sqrt{5})/2$), and let

$$Q(x_1, x_2) = x_1^2 - \alpha^2 x_2^2.$$

Proposition 1.3. *There exists $\epsilon > 0$ such that for all $x_1, x_2 \in \mathbb{Z}$, $|Q(x_1, x_2)| > \epsilon$.*

Proof. Suppose not. Then for any $1 > \epsilon > 0$ there exist $x_1, x_2 \in \mathbb{Z}$ such that

$$|Q(x_1, x_2)| = |x_1 - \alpha x_2||x_1 + \alpha x_2| \leq \epsilon. \tag{1}$$

We may assume $x_2 \neq 0$. If $\epsilon < \alpha^2$, one of the factors must be smaller then α. Without loss of generality, we may assume $|x_1 - \alpha x_2| < \alpha$, so $|x_1 - \alpha x_2| < \alpha |x_2|$. Then,

$$|x_1 + \alpha x_2| = |2\alpha x_2 + (x_1 - \alpha x_2)| \geq 2\alpha |x_2| - |x_1 - \alpha x_2| \geq \alpha |x_2|.$$

Substituting into (1) we get

$$\left| \frac{x_1}{x_2} - \alpha \right| \leq \frac{\epsilon}{|x_2||x_1 + \alpha x_2|} \leq \frac{\epsilon}{\alpha} \frac{1}{|x_2|^2}. \tag{2}$$

But since α is a quadratic irrational, there exists $c_0 > 0$ such that for all $p, q \in \mathbb{Z}$, $|\frac{p}{q} - \alpha| \geq \frac{c_0}{q^2}$. This is a contradiction to (2) if $\epsilon < c_0 \alpha$. $\square$

A relation to flows on homogeneous spaces. This was noticed by Raghunathan, and previously in implicit form by Cassels and Swinnerton-Dyer. However the Cassels-Swinnerton-Dyer paper was mostly forgotten. Raghunathan made clear the connection to unipotent flows, and explained from the point of view of dynamics what is different in dimension 2. See §7.1.

1.2 Some basic Ergodic Theory

Transformations, flows and Ergodic Measures. Let X be a topological space, and $T : X \to X$ a map. We assume that there is a finite measure μ on X which is preserved by T. One usually normalizes μ so that $\mu(X) = 1$, in which case μ is called a probability measure.

Sometimes, instead of a transformation T one considers a flow ϕ_t, $t \in \mathbb{R}$. For a fixed t, ϕ_t is a map from X to X. In this section we state definitions and theorems for transformations only, even though we will use them for flows later.

Definition 1.4 (Ergodic Measure)**.** An T-invariant probability measure μ is called *ergodic* for T if for every measurable T-invariant subset E of X one has $\mu(E) = 0$ or $\mu(E) = 1$.

Every measure can be written as a linear combination (possibly uncountable) of ergodic measures. This is called the "ergodic decomposition".

Ergodic measures always exist. In fact the probability measures form a convex set, and the ergodic probability measures are the extreme points of this set (cf. the Krein-Milman theorem).

Birkhoff's Ergodic Theorem

Theorem 1.5 (Birkhoff Ergodic Theorem)**.** *Suppose μ is ergodic for T, and suppose $f \in L^1(X, \mu)$. Then for μ-almost all $x \in X$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^n x) = \int_X f \, d\mu. \tag{3}$$

The sum on the left-hand side is called the "time average", and the integral on the right is the "space average". Thus the theorem says that for almost all base points x, the time average along the orbit of x converges to the space average.

This theorem is amazing in its generality: the only assumption is ergodicity of the measure μ. (This is a some sort of irreducibility assumption).

The set of $x \in X$ for which (3) holds is called the *generic* set for μ.

Mutually singular measures. Recall that two probability measures μ_1 and μ_2 are called *mutually singular* (written as $\mu_1 \perp \mu_2$ if there exists a set E such that $\mu_1(E) = 1$, $\mu_2(E) = 0$ (so $\mu_2(E^c) = 1$).

In our proofs we will use repeatedly the following:

Lemma 1.6. *Suppose μ_1 and μ_2 are distinct ergodic measures for the map $T : X \to X$. Then $\mu_1 \perp \mu_2$.*

Proof. This is an immediate consequence of the Birkhoff ergodic theorem. Since $\mu_1 \neq \mu_2$ we can find a continuous f such that $\int_X f \, d\mu_1 \neq \int_X f \, d\mu_2$. Now let E denote the set where (3) holds with $\mu = \mu_1$. $\square$

Remark. It is not difficult to give another proof of Lemma 1.6 using the Radon-Nikodym theorem.

We will need a stronger variant of Lemma 1.6 ([cf. [Ra4, Thm. 2.2], [Mor, Lem. 5.8.6]]):

Lemma 1.7. *Suppose $T : X \to X$ is preserving an ergodic measure μ. Suppose H is a group with acts continuously on X and commutes with T. Also suppose that there exists $h_0 \in H$ such that $h_0\mu \neq \mu$. Then there exists a neighborhood B of $h_0 \in H$ and a conull T-invariant subset Ω of X such that*

$$h\Omega \cap \Omega = \emptyset \quad \text{for all } h \in B.$$

Proof. Since h_0 commutes with T, the measure $h_0\mu$ is T-invariant and ergodic. Thus by Lemma 1.6, $h_0\mu \perp \mu$. This implies there is a compact subset K_0 of X, such that $\mu(K_0) > 0.99$ and $K_0 \cap h_0K_0 = \emptyset$. By continuity and compactness, there are open neighborhoods $\mathcal{U}$ and $\mathcal{U}^+$ of K_0, and a symmetric neighborhood B_e of e in H, such that $\mathcal{U}^+ \cap h_0\mathcal{U}^+ = \emptyset$ and $B_e\mathcal{U} \subset \mathcal{U}^+$. ¿From applying (3) with f the characteristic function of $\mathcal{U}$, we know there is a conull T-invariant subset Ω_{h_0} of X, such that the T-orbit of every point in Ω_{h_0} spends 99% of its life in $\mathcal{U}$. Now suppose there exists $h \in B_eh_0$, such that $\Omega_{h_0} \cap h\Omega_{h_0} \neq \emptyset$. Then there exists $x \in \Omega_{h_0}$, $n \in \mathbb{N}$, and $c \in B_e$, such that T^nx and ch_0T^nx both belong to $\mathcal{U}$. This implies that T^nx and h_0T^nx both belong to $\mathcal{U}^+$. This contradicts the fact that $\mathcal{U}^+ \cap h_0\mathcal{U}^+ = \emptyset$. $\square$

Uniquely ergodic systems. In some applications (in particular to number theory) we need some analogue of (3) for *all* points x (and not almost all). For example, we want to know if $Q(\mathbb{Z}^n)$ is dense for a specific quadratic form Q (and not for almost all forms). Then the Birkhoff ergodic theorem is not helpful. However, there is one situation where we can show that (3) holds for all x.

Definition 1.8. A map $T : X \to X$ is called *uniquely ergodic* if there exists a unique invariant probability measure μ.

Proposition 1.9. *Suppose X is compact, $T : X \to X$ is uniquely ergodic, and let μ be the invariant probability measure. Suppose $f : X \to \mathbb{R}$ is continuous. Then for all $x \in X$, (3) holds.*

Proof. This is quite easy (as opposed to the Birkhoff ergodic theorem which is hard). Let δ_n be the probability measure on X defined by

$$\delta_n(f) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^n x)$$

(we are now thinking of measures as elements of the dual space to the space $C(X)$ of continuous functions on X). Note that

$$\delta_n(f \circ T) = \frac{1}{n}\sum_{k=0}^{n-1}(f \circ T)(T^n x) = \frac{1}{n}\sum_{k=1}^{n} f(T^n x),$$

so

$$\delta_n(f \circ T) - \delta_n(f) = \frac{1}{n}(f(x) - f(T^n x)), \tag{4}$$

(since the sum telescopes). Suppose some subsequence δ_{n_j} converges to some limit δ_∞ (in the weak-* topology). Then, by (4), $\delta_\infty(f \circ T) = \delta_\infty(f)$, i.e. δ_∞ is T-invariant.

Since X is compact, δ_∞ is a probability measure, and thus by the assumption of unique ergodicity, we have $\delta_\infty = \mu$. Thus all possible limit points of the sequence δ_n are μ. Also the space of probability measures on X is compact (in the weak-* topology), so there exists a convergent subsequence. Hence $\delta_n \to \mu$, which is the same as (3). $\square$

Remarks.

- The main point of the above proof is the construction of an invariant measure (namely δ_∞) supported on the closure of the orbit of x. The same construction works with flows, or more generally with actions of amenable groups.

- We have used the compactness of X to argue that δ_∞ is a probability measure: this might fail if X is not compact. This phenomenon is called "loss of mass".

- Of course the problem with Proposition 1.9 is that most of the dynamical systems we are interested in are not uniquely ergodic. For example any system which has a closed orbit which is not the entire space is not uniquely ergodic.

- However, the proof of Proposition 1.9 suggests that (at least in the amenable case) the classification of the invariant measures is the most powerful statement one can make about a dynamical system, in the sense that it allows one to understand every orbit (and not just almost every orbit).

Exercise 1. (To be used in §3.)

(a) Show that if α is irrational then the map $T_\alpha : [0,1] \to [0,1]$ given by $T_\alpha(x) = x + \alpha \pmod 1$ is uniquely ergodic. *Hint:* Use Fourier analysis.

(b) Use part (a) to show that the flow on $\mathbb{R}^2/\mathbb{Z}^2$ given by $\phi_t(x,y) = (x + t\alpha, y + t)$ is uniquely ergodic.

1.3 Unipotent Flows.

Let G be a semisimple Lie group (I will usually assume the center of G is finite), and let Γ be a lattice in G (this means that $\Gamma \subset G$ is a discrete subgroup, and the quotient G/Γ has finite Haar measure). A lattice Γ is *uniform* if G/Γ is compact.

Let $U = \{u_t\}_{t\in\mathbb{R}}$ be a unipotent one-parameter subgroup of G. Then U acts on G/Γ by left multiplication. (Recall that in $SL(n,\mathbb{R})$ a matrix is unipotent if all its eigenvalues are 1. In a general Lie group an element is unipotent if its Adjoint (acting on the Lie algebra) is a unipotent matrix.) Examples of unipotent one parameter subgroups:

$$\left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \qquad t \in \mathbb{R} \right\},$$

and

$$\left\{ \begin{pmatrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}, \qquad t \in \mathbb{R} \right\},$$

Ratner's measure classification theorem.

Definition 1.10. A probability measure μ on G/Γ is called *algebraic* if there exists $\bar{x} \in G/\Gamma$ and a subgroup F of G such that $F\bar{x}$ is closed, and μ is the F-invariant probability measure supported on $F\bar{x}$.

Theorem 1.11 (Ratner's measure classification theorem)**.** *Let G be a Lie group, $\Gamma \subset G$ a lattice. Let U be a one-parameter unipotent subgroup of G. Then, any ergodic U-invariant measure is algebraic. (Also the group F in the definition of algebraic is generated by unipotent elements, and contains U).*

Loosely speaking, this theorem says that all U-invariant ergodic measures are very nice. The assumption that U is unipotent is crucial: if we consider instead arbitrary one-parameter subgroups, then there are ergodic invariant measures supported on Cantor sets (and worse).

Theorem 1.11 has many applications, some of which we will explore in this course. I will give some indication of the ideas which go into the proof of this theorem in the next two lectures.

Remark on algebraic measures. Let $\pi : G \to G/\Gamma$ be the projection map. Suppose $\bar{x} \in G/\Gamma$, and $F \subset G$ is a subgroup. Let $\mathrm{Stab}_F(\bar{x})$ denote the stabilizer in F of $\bar{x}$, i.e. the set of elements $g \in F$ such that $g\bar{x} = \bar{x}$. Then $\mathrm{Stab}_F(\bar{x}) = F \cap x\Gamma x^{-1}$, where $x \in G$ is any element such that $\pi(x) = \bar{x}$. Thus there is a bijection between $F\bar{x}$ and $F/(F \cap x\Gamma x^{-1})$, but this is in general not continuous.

However, in the case of algebraic measures, we are making the additional assumption that $F\bar{x}$ is closed. In this case, the above bijection is continuous, and thus μ is the image under this bijection of the Haar measure on $F/(F \cap x\Gamma x^{-1})$. The assumption that μ is a probability measure thus implies that $F \cap x\Gamma x^{-1}$ is a lattice in F. (The last condition is usually taken to be part of the definition of an algebraic measure).

Uniform Distribution and the classification of orbit closures.

Theorem 1.12 (Ratner's uniform distribution theorem)**.** *Let G be a Lie group, Γ a lattice in G, and $U = \{u_t\}_{t\in\mathbb{R}}$ a one-parameter unipotent subgroup. Then for any $\bar{x} \in G/\Gamma$ there exists a subgroup $F \supset U$ (generated by unipotents) with $F\bar{x}$ closed, and an F-invariant algebraic measure μ supported on $F\bar{x}$, such that for any $f \in C(G/\Gamma)$,*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(u_t\bar{x})\, dt = \int_{F\bar{x}} f\, d\mu \tag{5}$$

Remarks.

- It follows from (5) that the closure of the orbit $U\bar{x}$ is $F\bar{x}$. Thus Theorem 1.12 can be rephrased as "any orbit is uniformly distributed in its closure".
- Theorem 1.12 is derived from Theorem 1.11 by an argument morally similar to the proof of Proposition 1.9. There is one more ingredient: one has to show that the set of subgroups F which appear in Theorem 1.11 is countable up to conjugation (Proposition 6.1 below). For proofs of this fact see [Ra6, Theorem 1.1] and [Ra7, Cor. A(2)]), or alternatively [DM4, Proposition 2.1].

An immediate consequence of Theorem 1.12 is the following:

Theorem 1.13 (Raghunathan's topological conjecture)**.** *Let G be a Lie group, $\Gamma \subset G$ a lattice, and $U \subset G$ a one-parameter unipotent subgroup. Suppose $\bar{x} \in G/\Gamma$. Then there exists a subgroup F of G (generated by unipotents) such that the closure $\overline{U\bar{x}}$ of the orbit $U\bar{x}$ is $F\bar{x}$.*

This theorem is due to Ratner in the general case, but several cases were known previously. See §7.1 for a discussion and the relation to the Oppenheim Conjecture.

Uniformity of convergence. In many applications it is important to somehow ensure that the time averages converge to the space average uniformly in the base point $\bar{x}$ (for example we may have an additional integral over $\bar{x}$). In the context of Birkhoff's ergodic theorem, we have the following:

Lemma 1.14. *Suppose $\phi_t : X \to X$ is a flow preserving an ergodic probability measure μ. Suppose $f \in L^1(X, \mu)$. Then for any $\epsilon > 0$ and $\delta > 0$, there exists $T_0 > 0$ and a set $E \subset X$ with $\mu(E) < \epsilon$, such that for any $x \in E^c$ and any $T > T_0$ we have*

$$\left| \frac{1}{T} \int_0^T f(\phi_t(x))\, dt - \int_X f\, d\mu \right| < \delta$$

(In other words, one has uniform convergence outside of a set of small measure.)

Proof. Let E_n denote the set of $x \in X$ such that for some $T > n$,

$$\left| \frac{1}{T} \int_0^T f(\phi_t(x))\, dt - \int_X f\, d\mu \right| \geq \delta.$$

Then by the Birkhoff ergodic theorem, $\mu(\bigcap_{n=1}^{\infty} E_n) = 0$. Hence there exists $n \in \mathbb{N}$ such that $\mu(E_n) < \epsilon$. Now let $T_0 = n$, and $E = E_n$. $\square$

The uniform distribution theorem of Dani-Margulis. One problem with Lemma 1.14 is that it does not provide us with any information about the exceptional set E (other then the fact that it has small measure). In the setting of unipotent flows, Dani and Margulis proved a theorem (see §6.2 below for the precise statement) which is the analogue of Lemma 1.14, but with an explicit geometric description of the set E. This theorem is crucial for many applications. Its proof is based on the Ratner measure classification theorem (Theorem 1.11) and the "linearization" technique of Dani and Margulis (see §6).

2 The case of $\mathrm{SL}(2, \mathbb{R})/SL(2, \mathbb{Z})$

In this lecture I will be loosely following Ratner's paper [Ra8].

2.1 Basic Preliminaries.

The space of lattices. Let $G = SL(n, \mathbb{R})$, and let $\mathcal{L}_n$ denote the space of unimodular lattices in $\mathbb{R}^n$. (By definition, a lattice Δ is unimodular if an only if the volume of $\mathbb{R}^n/\Delta = 1$. G acts on $\mathcal{L}_n$ as follows: if $g \in G$ and $\Delta \in \mathcal{L}_n$ is the $\mathbb{Z}$-span of the vectors $v_1, \dots v_n$, then gv is the $\mathbb{Z}$-span of $gv_1, \dots, gv_n$. This action is clearly transitive. The stabilizer of the standard lattice $\mathbb{Z}^n$ is $\Gamma = SL(n, \mathbb{Z})$. This gives an identification of $\mathcal{L}_n$ with G/Γ. We choose a *right-invariant* metric on G; then this metric descends to G/Γ.

The set $\mathcal{L}_n(\epsilon)$. For $\epsilon > 0$ let $\mathcal{L}_n(\epsilon) \subset \mathcal{L}_n$ denote the set of lattices whose shortest non-zero vector has length at least ϵ.

Theorem 2.1 (Mahler Compactness)**.** *For any $\epsilon > 0$ the set $\mathcal{L}_n(\epsilon)$ is compact.*

The upper half plane. In the rest of this section, we set $n = 2$. Let $K = SO(2) \subset G$. Given a pair of vectors v_1, v_2 we can find a unique rotation matrix $k \in K$ so that kv_1 is pointing along the positive x-axis and kv_2 is in the upper half plane. The map $g = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \to kv_2$ gives an identification of $K\backslash G$ with the upper half plane $\mathbb{H}^2$. Now G (and in particular $\Gamma \subset G$) acts on $K\backslash G$ by multiplication on the right. Using the identification of $K\backslash G$ with $\mathbb{H}^2$ this becomes (a variant of) the usual action by fractional linear transformations.

The horocycle and geodesic flows. We use the following notation:

$$u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \qquad a_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \qquad v_t = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}.$$

Let $U = \{u_t \;:\; t \in \mathbb{R}\}$, $A = \{a_t \;:\; t \in \mathbb{R}\}$, $V = \{v_t \;:\; t \in \mathbb{R}\}$. The action of U is called the *horocycle flow* and the action of A is called the *geodesic flow*. A some basic commutation relations are the following:

$$a_t u_s a_t^{-1} = u_{e^{2t}s} \qquad a_t v_s a_t^{-1} = v_{e^{-2t}s} \tag{6}$$

Thus conjugation by a_t for $t > 0$ contracts V and expands U.

Orbits of the geodesic and horocycle flow in the upper half plane. Let $p : G \to K\backslash G$ denote the natural projection. Then for $x \in G$, $p(Ux)$ is either a horizontal line or a circle tangent to the x-axis. Also $p(Ax)$ is either a vertical line or a semicircular arc orthogonal to the x-axis.

Flowboxes. Let $W_+ \subset U$, $W_- \subset V$, $W_0 \subset A$ be intervals containing the identity (we have identified all three subgroups with $\mathbb{R}$). By a *flowbox* we mean a subset of G of the form $W_- W_0 W_+$, or one of its right translates by $g \in G$. Clearly, $W_- W_0 W_+ g$ is an open set containing g. (Recall that in our conventions, right multiplication by g is an isometry).

2.2 An elementary non-divergence result.

(Much more will be proved in Dima's lectures).

Lemma 2.2. *There exists an absolute constant $\epsilon_0 > 0$ such that the following holds: Suppose $\Delta \in \mathcal{L}_2$ is a unimodular lattice. Then Δ cannot contain two linearly independent vectors each of length less than ϵ_0.*

Proof. Let v_1 be the shortest vector in Δ, and let v_2 be the shortest vector in Δ linearly independent from v_1. Then v_1 and v_2 span a sublattice Δ' of Δ. (In fact $\Delta' = \Delta$ but this is not important for us right now). Since Δ is unimodular, this implies that $\mathrm{Vol}(\mathbb{R}^2/\Delta') \geq 1$. But $\mathrm{Vol}(\mathbb{R}^2/\Delta') = \|v_1 \times v_2\| \leq \|v_1\|\|v_2\|$. Hence $\|v_1\|\|v_2\| \geq 1$, so the lemma holds with $\epsilon_0 = 1$. $\square$

Lemma 2.3. *Suppose $\Delta \in \mathcal{L}_2$ is a unimodular lattice. Then at least one of the following holds:*

(a) *Δ contains a horizontal vector.*

(b) *There exists $t \geq 0$ such that $a_t^{-1}\Delta \in \mathcal{L}_2(\epsilon_0)$.*

Proof. Suppose Δ does not contain a horizontal vector, and $\Delta \notin \mathcal{L}_2(\epsilon_0)$. Then Δ contains a vector v with $\|v\| < \epsilon_0$. Since v is not horizontal, there exists a smallest $t_0 > 0$ such that $\|a_t^{-1}v\| = \epsilon_0$. Then by Lemma 2.2 for $t \in [0, t_0]$, $a_t^{-1}\Delta$ contains no vectors shorter then ϵ_0 (other then $a_t^{-1}v$ and possibly its multiples). In particular $a_{t_0}^{-1}\Delta$, contains no vectors shorter then ϵ_0. This means $a_{t_0}^{-1}\Delta \in \mathcal{L}_2(\epsilon_0)$. $\square$

Remark. We note that Lemma 2.2 and thus Lemma 2.3 are specific to dimension 2.

2.3 The classification of U-invariant measures.

Note that for $\Delta \in \mathcal{L}_2$, the U-orbit of Δ is closed if and only if Δ contains a horizontal vector. (The horizontal vector is fixed by the action of U). Any closed U-orbit supports a U-invariant probability measure. All such measures are ergodic.

Let ν denote the Haar measure on $\mathcal{L}_2 = G/\Gamma$. The measure ν is normalized so that $\nu(\mathcal{L}_2) = 1$. Recall that ν is ergodic for both the horocycle and the geodesic flows (this follows from the Moore ergodicity theorem).

Our main goal in this lecture is the following:

Theorem 2.4. *Suppose μ is an ergodic U-invariant probability measure on $\mathcal{L}_2$. Then either μ is supported on a closed orbit, or μ is the Haar measure ν.*

Proof. Let $\mathcal{L}_2' \subset \mathcal{L}_2$ denote the set of lattices which contain a horizontal vector. Note that the set $\mathcal{L}_2'$ is U-invariant.

Suppose μ is an ergodic U-invariant probability measure on $\mathcal{L}_2$. By ergodicity of μ, $\mu(\mathcal{L}_2') = 0$ or $\mu(\mathcal{L}_2') = 1$. If the latter holds, it is easy to show that μ is supported on a closed orbit. Thus we assume $\mu(\mathcal{L}_2') = 0$ and we must show that $\mu = \nu$.

Suppose not. Then there exists a compactly supported continuous function $f : \mathcal{L}_2 \to \mathbb{R}$ and $\epsilon > 0$ such that

$$\left| \int_{\mathcal{L}_2} f \, d\mu - \int_{\mathcal{L}_2} f \, d\nu \right| > \epsilon. \tag{7}$$

Since f is uniformly continuous, there exists a neighborhoods of the identity $W_0' \subset A$ and $W_-' \subset V$ such that such that for $a \in W_0'$, $v \in W_-'$ and $\Delta'' \in \mathcal{L}_2$,

$$|f(va\Delta'') - f(\Delta'')| < \epsilon/3. \tag{8}$$

Recall that $\pi : G \to G/\Gamma \cong \mathcal{L}_2$ denotes the natural projection. Since $\mathcal{L}_2(\epsilon_0)$ is compact the injectivity radius on $\mathcal{L}_2(\epsilon_0)$ is bounded from below, hence there exist $W_+ \subset U$, $W_0 \subset A$, $W_- \subset V$ so that for any $g \in G$ with $\pi(g) \in \mathcal{L}_2$, the restriction of π to the flowbox $W_- W_0 W_+ g$ is injective. We may also assume that $W_- \subset W_-'$ and $W_0 \subset W_0'$. Let $\delta = \nu(W_- W_0 W_+)$ denote the Lebesque measure of the flowbox.

By Lemma 1.14 applied to the Lebesque measure ν, there exists a set $E \subset \mathcal{L}_2$ with $\nu(E) < \delta$ and $T_1 > 0$ such that for any interval I with $|I| \geq T_1$ and any $\Delta' \notin E$,

$$\left| \frac{1}{|I|} \int_I f(u_t \Delta') \, dt - \int_{\mathcal{L}_2} f \, d\nu \right| < \frac{\epsilon}{3}. \tag{9}$$

Now let Δ be a generic point for U (in the sense of the Birkhoff ergodic theorem). This implies that there exists $T_2 > 0$ such that for any interval I containing the origin of length greater then T_2,

$$\left| \frac{1}{|I|} \int_I f(u_t \Delta)\, dt - \int_{\mathcal{L}_2} f\, d\mu \right| < \frac{\epsilon}{3}. \tag{10}$$

Since $\mu(\mathcal{L}_2') = 0$, we may assume that Δ does not contain any horizontal vectors. Then by repeatedly applying Lemma 2.3 we can construct arbitrarily large $t > 0$ such that

$$a_t^{-1} \Delta \in \mathcal{L}_2(\epsilon). \tag{11}$$

Now suppose t such that (11) holds, and consider the set $Q = a_t W_- W_0 W_+ a_t^{-1} \Delta$. Then Q can be rewritten as

$$Q = (a_t W_- a_t^{-1}) W_0 (a_t W_+ a_t^{-1}) \Delta$$

(so when t is large, Q is long in the U direction and short in A and V directions.) The set Q is an embedded copy of a flowbox in $\mathcal{L}_2$, and $\nu(Q) = \delta$.

We now consider the foliation of Q by the orbits of U. For any $\Delta' \in Q$, let $I(\Delta')$ denote the connected component containing the origin of the set

$$\{ t \in \mathbb{R} \ : \ u_t \Delta' \in Q \}$$

Note that the length of $I(\Delta')$ is independent of Δ' (it is just the length of W_+ multiplied by e^{2t}). By choosing t sufficiently large, we may assume that $|I(\Delta')| \geq \max(T_1, T_2)$. By (6), $a_t W_- a_t^{-1} \subset W_-'$. Also, by construction, $W_0 \subset W_0'$. Thus, by (8), we have for any $\Delta' \in Q$,

$$\left| \frac{1}{|I(\Delta')|} \int_{I(\Delta')} f(u_t \Delta')\, dt - \frac{1}{|I(\Delta)|} \int_{I(\Delta)} f(u_t \Delta)\, dt \right| < \frac{\epsilon}{3}. \tag{12}$$

(this says that Q is foliated by U-orbits, and the integral of f over each U-orbit is nearly the same).

Since $\nu(E) < \delta$ and $\nu(Q) = \delta$, there exists $\Delta' \in Q \cap E^c$. Now (9) holds with $I = I(\Delta')$, and (10) holds with $I = I(\Delta)$. These estimates together with (12) contradict (7). $\square$

Remarks.

- The above proof works with minor modifications if Γ is an arbitrary lattice in $SL(2, \mathbb{R})$ (not just $SL(2, \mathbb{Z})$).

- If Γ is a uniform lattice in $SL(2,\mathbb{R})$ then the horocycle flow on G/Γ is uniquely ergodic. This is a theorem of Furstenberg [F].
- The proof of Theorem 2.4 does not generalize to classification of measures invariant under a one-parameter unipotent subgroup on e.g. $\mathcal{L}_n$, $n \geq 3$. Completely different ideas are needed. (I will introduce some of them in the next lecture).

Horospherical subgroups and a theorem of Dani. The key property of U in dimension 2 which is used in the proof is that U is *horospherical*, i.e. that it is equal to the set contracted by a one-parameter diagonal subgroup. (One-parameter unipotent subgroups are horospherical only in $SL(2,\mathbb{R})$). An argument similar in spirit to the proof of Theorem 2.4 can be used to classify the measures invariant under the action of a horospherical subgroup. This is a theorem of Dani [Dan2] (which was proved before Ratner's measure classification theorem). However, the details, and in particular the non-divergence results needed are much more complicated.

The horospherical case also allows for an analytic approach, see e.g. [Bu].

3 The case of $SL(2,\mathbb{R}) \ltimes \mathbb{R}^2$.

In this section we will outline a proof of Ratner's measure classification theorem Theorem 1.11 in the special case $G = SL(2,\mathbb{R}) \ltimes \mathbb{R}^2$, $\Gamma = SL(2,\mathbb{Z}) \ltimes \mathbb{Z}^2$. We will be following the argument of Ratner [Ra1, Ra2, Ra3, Ra4, Ra5, Ra6] and Margulis-Tomanov [MT]. An introduction to these ideas can be found in [Mor]. Another exposition of a closely related case is in [EMaMo].

Let $X = G/\Gamma$. Then X can be viewed as a space of pairs (Δ, v), where Δ is a unimodular lattice in $\mathbb{R}^2$ and v is a marked point on the torus $\mathbb{R}^2/\Delta$. (We remove the translation invariance on the torus $\mathbb{R}^2/\Delta$ since we consider the origin as a special point). X is thus naturally a fiber bundle where the base is $\mathcal{L}_2$ and the fiber above the point $\Delta \in \mathcal{L}_2$ is the torus $\mathbb{R}^2/\Delta$. (X is also sometimes called the universal elliptic curve).

The action of $SL(2,\mathbb{R}) \subset G$ on X is by left multiplication. It amounts to $g \cdot (\Delta, v) = (g\Delta, gv)$. The action of the $\mathbb{R}^2$ part of G on X is by translating the marked point, i.e for $w \in \mathbb{R}^2$, $w \cdot (\Delta, v) = (\Delta, w + v)$. Let U be the subgroup of $SL(2,\mathbb{R})$ defined in §2.1. In this lecture our goal is the following special case of Theorem 1.11:

Theorem 3.1. *Let μ be an ergodic U-invariant measure on X. Then μ is algebraic.*

Let μ be an ergodic U-invariant measure on X. Let $\pi_1 : X \to \mathcal{L}_2$ denote the natural projection (i.e. $\pi_1(\Delta, v) = \Delta$). Then $\pi_1^*(\mu)$ is an ergodic U-invariant measure on $\mathcal{L}_2$. Thus by Theorem 2.4, either $\pi_1^*(\mu)$ is supported on a closed orbit of U, or $\pi_1^*(\mu)$ is the Haar measure ν on $\mathcal{L}_2$. The first case is easy to handle, so in the rest of this section we assume that $\pi_1^*(\mu) = \nu$. Then we can disintegrate

$$d\mu(\Delta, v) = d\nu(\Delta) d\lambda_\Delta(v)$$

where $\lambda_\Delta(v)$ is some probability measure on the torus $\mathbb{R}^2/\Delta$.

3.1 Finiteness of the fiber measures.

Many of the ideas behind the proof of Ratner's measure classification theorem Theorem 1.11 can be illustrated in the proof of the following:

Proposition 3.2. *Either μ is Haar measure on X, or for almost all $\Delta \in \mathcal{L}_2$, the measure λ_Δ is supported on a finite set of points.*

We will give an almost complete proof of Proposition 3.2 in this subsection, and then indicate how to complete the proof of Theorem 3.1 in the next subsection.

The subgroups U,V,A,H, and W. Let U, V, A be the subgroups of $SL(2, \mathbb{R})$ defined in §2.1. We also give names to certain subgroups of the $\mathbb{R}^2$ part of G. In particular, let $H = \{h_s, s \in \mathbb{R}\}$ be the subgroup of G whose action on X is given by $h_s(\Delta, v) = (\Delta, v + s \begin{pmatrix} 1 \\ 0 \end{pmatrix})$, and $W = \{w_r, r \in \mathbb{R}\}$ be the subgroup of G whose action on X is given by $w_r(\Delta, v) = (\Delta, v + r \begin{pmatrix} 0 \\ 1 \end{pmatrix})$. The action of H is called the *horizontal flow* and the action of W the *vertical flow*.

Action of the centralizer. A key observation is that H commutes with U (and so the action of H commutes with the action of U). This implies that if μ is an ergodic U-invariant measure, so is $h_s\mu$ for any $h_s \in H$.

Thus, either μ is invariant under H or there exists $s \in \mathbb{R}$ such that $h_s\mu$ is distinct from μ. Suppose μ is invariant under H. Then so are the fiber measures λ_Δ for all $\Delta \in \mathcal{L}_2$. Then by Exercise 1 (b), for ν-almost all $\Delta \in \mathcal{L}_2$, λ_Δ is the Lebesque measure on $\mathbb{R}^2/\Delta$. Thus μ coincides with Haar measure on X for almost all fibers. Then by the ergodicity of μ we can conclude that μ is the Haar measure on X.

Thus, Proposition 3.2 follows from the following:

Proposition 3.3. *Suppose μ is not H-invariant. Then for almost all $\Delta \in \mathcal{L}_2$, the measure λ_Δ is supported on a finite set of points.*

The element h and the compact set K. ¿From now on, we assume that μ is not H-invariant. Then there exists $h_{s_0} \in H$ such that $h_{s_0}\mu \neq \mu$. (We may assume that h_{s_0} is fairly close to the identity). Since $h_{s_0}\mu$ and μ are both ergodic U-invariant measures, by Lemma 1.6 we have $h_{s_0}\mu \perp \mu$. Thus the supports of μ and $h_{s_0}\mu$ are disjoint. It follows from Lemma 1.7 that there exists $\delta > 0$ and a subset $\Omega \subset X$ with $\mu(\Omega) = 1$ such that $h_s\Omega \cap \Omega = \emptyset$ for all $s \in (s_0 - \delta s_0, s_0]$. It follows that there exists a compact set K with $\mu(K) > 0.999$ such that for all $s \in [(1-\delta_0)s_0, s_0]$, $h_sK \cap K = \emptyset$. Since K is compact and the action of H is continuous, there exist $\epsilon > 0$ and $\delta > 0$ such that

$$d(h_sK, K) > \epsilon \qquad \text{for all } s \in [(1-\delta)s_0, s_0]. \tag{13}$$

The set Ω_ρ. In view of Lemma 1.14 (with f the characteristic function of K), for any $\rho > 0$ we can find a set Ω_ρ with $\mu(\Omega_\rho) > 1-\rho$ and $T_0 > 0$ such that for all $T > T_0$ and all $p \in \Omega_\rho$ we have

$$\frac{1}{T}|\{t \in [0,T] \;:\; u_t x \in K\}| \geq 1 - (0.01)\delta \tag{14}$$

Shearing. Suppose $p = (\Delta, v)$ and $p' = (\Delta, v')$ are two nearby points in the same fiber. We want to study how they diverge under the action of U. Note that u_tp and u_tp' are always in the same fiber (i.e. $\pi_1(u_tp) = \pi_1(u_tp') = u_t\Delta$), but within the fiber $\pi_1^{-1}(u_t\Delta)$ they will slowly diverge. More precisely, if we let $v = (x,y)$ and $v' = (x', y')$ we have

$$u_tv' - u_tv = (x' - x + t(y'-y), y' - y).$$

Note that if $y = y'$ (i.e. p and p' are in the same orbit of H) then u_tp and u_tp' will not diverge at all.

Now suppose $y \neq y'$. We are considering the regime where $|x'-x|$, $|y'-y|$ are very small, but t is so large that $d(p,p')$ is comparable to 1 (this amounts to $|t(y'-y)|$ comparable to 1). Under these assumptions, *the leading divergence is along H*, i.e.

$$u_tp' = h_su_tp + \text{ small error} \tag{15}$$

where $s = t(y'-y)$.

Lemma 3.4. *Suppose that for some positive measure set of $\Delta \in \mathcal{L}_2$, the support of λ_Δ is infinite. Then for any $\rho > 0$ We can find $\Delta \in \mathcal{L}_2$ and a sequence of points $p_n = (\Delta, (x_n, y_n)) \in \Omega_\rho$ which converge to $p = (\Delta, (x,y)) \in \Omega_\rho$ so that $y_n - y \neq 0$ for all n.*

We postpone the proof of this lemma (which is intuitively clear anyway).

Proof of Proposition 3.3. Suppose the conclusion of Proposition 3.3 is false, so that for some positive measure set of $\Delta \in \mathcal{L}_2$, the support of λ_Δ is infinite. Then Lemma 3.4 applies.

Let $T_n = s_0/(y_n - y)$. Then by (15) we have for $t \in [(1-\delta)T_n, T_n]$,

$$d(u_t p_n, h_s u_t p) < \epsilon_n, \qquad \text{where } s = t/(y' - y). \tag{16}$$

and $\epsilon_n \to 0$ as $n \to \infty$. If n is sufficiently large, then $T_n > T_0$ where T_0 is as in the definition of Ω_ρ. Then (14) applies to both p and p_n, and we can thus find $t \in [(1-\delta)T_n, T_n]$ such that $u_t p_n \in K$ and also $u_t p \in K$. Then $s = t/(y' - y) \in [(1-\delta_0)s_0, s_0]$, and so (16) contradicts (13). □

Proof of Lemma 3.4. Suppose that for some positive measure set of $\Delta \in \mathcal{L}_2$, the support of λ_Δ is infinite. Then (by the ergodicity of the action of U on $\mathcal{L}_2$), this the support of λ_Δ is infinite for almost all fibers Δ.

Suppose for the moment that the support of λ_Δ is countable for almost all Δ, so λ_Δ is supported on a sequence of points p_n with weights λ_n. But then the collection of points with the same weight is a U-invariant set, so by ergodicity of μ all the points must have the same weight. Thus, since λ_Δ is a probability measure if the the support of λ_Δ is countable it must be finite.

Hence we may assume that the support of λ_Δ is uncountable. Then so is $\Omega_\rho \cap \lambda_\Delta$ for almost all Δ. Since any uncountable set contains one of its accumulation points, we may construct a sequence $p_n \in \Omega_\rho$ with $p_n \to p$, where $p \in \Omega_\rho$. It only remains to verify that if we write $p_n = (\Delta, (x_n, y_n))$ and $p = (\Delta, (x, y))$ then we can ensure $y_n \neq y$.

If it is not possible to do so, then it is easy to see that the support of λ_Δ is contained in a finite union of H-orbits. Thus given $a < b$ we can define a function $u((\Delta, v)) = \lambda_\Delta(\{h_s v \ : \ s \in [a, b]\})$. This function is U-invariant hence constant for each choice of $[a, b]$. It is easy to conclude from this that the support of λ_Δ must be finite. □

3.2 Outline of the Proof of Theorem 3.1

.

The following general lemma is a stronger version of Lemma 1.14:

Lemma 3.5 (cf. [MT, Lem. 7.3])**.** *Suppose $\phi_t : X \to X$ is a flow preserving an ergodic probability measure μ. For any $\rho > 0$, there is a "uniformly generic set" Ω_ρ in X, such that*

1. *$\mu(\Omega_\rho) > 1 - \rho$,*
2. *for every $\epsilon > 0$ and every compact subset K of X, with $\mu(K) > 1 - \epsilon$, there exists $L_0 \in \mathbb{R}^+$, such that, for all $x \in \Omega_\rho$ and all $L > L_0$, we have*
$$|\{\, t \in [-L, L] \mid d(\phi_t(x), K) < \epsilon \,\} > (1 - \epsilon)(2L).$$

Outline of proof. This is similar to that of Lemma 1.14, except that one also chooses a countable basis of functions and approximates K by elements of the basis. □

We now return to the setting of §3. Let μ be an ergodic invariant measure for the action of U on $X = G/\Gamma = SL(2, R) \ltimes \mathbb{R}^2/SL(2, \mathbb{Z}) \ltimes \mathbb{Z}^2$. For any $\rho > 0$ we chose a "uniformly generic" set Ω_ρ for μ as in Lemma 3.5.

The argument of §3.1 can be summarized as the following proposition (which we state somewhat imprecisely):

Proposition 3.6. *Suppose Q is a subgroup of G normalizing U, and suppose that for any $\rho > 0$ we can find a sequences p_n and p'_n in Ω_ρ such that $d(p_n, p'_n) \to 0$, and under the action of U the leading transverse divergence of the trajectories $u_t p_n$ and $u_t p'_n$ is in the direction of Q (i.e the analogue of (15) holds with $q \in Q$ instead of $h \in H$).*

Then the measure μ is Q-invariant.

Remark. The analogous statement for unipotent flows is a cornerstone of the proof of Ratner's Measure Classification Theorem [Ra5, Lem. 3.3], [MT, Lem. 7.5], [Mor, Prop. 5.2.4′].

Remark. For two points in the same fiber, the leading divergence is always along H (if the points diverge at all). For an arbitrary pair of nearby points in X this is not the case.

Remark. It is possible that the leading direction of divergence is along U. In that case we want to consider the leading "transverse" divergence. In other words we compare $u_t p_n$ and $u_{t'} p'_n$ where t' is chosen to cancel the divergence along U (i.e. one trajectory waits for the other). In that case we say that the leading transverse divergence is along Q if for some $q \in Q$,

$$u_t p_n = q u_{t'} p'_n + \text{ small error}$$

Remark. To prove Proposition 3.6 we must use Lemma 3.5 instead of Lemma 1.14 as in §3.1 because we must choose Ω_ρ before we know what subgroup Q (and thus what compact set K) we will be dealing with.

We now continue the proof of Theorem 3.1. We assume that μ projects to Haar measure on $\mathcal{L}_2$, but that μ is not Haar measure.

Proposition 3.7. *We may assume μ is A-invariant.*

Proof. Choose Ω_ρ as in Lemma 3.5, with $\rho = 0.01$. By Proposition 3.2, the measure on each fiber is supported on a finite set. Also we are assuming that μ projects to Haar measure on $\mathcal{L}_2$. Then it is easy to see that there exist $p \in \Omega_\rho$, $\{v_n\} \subset V \smallsetminus \{e\}$, and $\{w_n\} \subset HW$, such that $p_n = v_n w_n p \in \Omega_\rho$, $v_n \to e$, and $w_n \to e$.

It is not difficult to compute that (after passing to a subsequence), the leading direction of divergence of $u_t p_n$ and $u_t p$ is a one-parameter subgroup Q which is contained in AH. Then by Proposition 3.6, μ is invariant under Q.

By §3.1, we have $Q \neq H$. Any such subgroup Q of AH is conjugate to A (via an element of H). Thus, by replacing μ with a translate under H, we may assume μ is A-invariant. □

Note. At this point we do not know that μ is A-ergodic.

Proposition 3.8 (cf. [MT, Cor. 8.4], [Mor, Cor. 5.5.2])**.** *There is a conull subset Ω of X, such that*

$$\Omega \cap VWp = \Omega \cap Vp,$$

for all $p \in \Omega$.

Proof. Let Ω be a generic set for for the action of A on X; thus, Ω is conull and, for each $p \in \Omega$,

$$a_t p \in \Omega_\rho \text{ for most } t \in \mathbb{R}^+.$$

(The existence of such a set follows e.g. from the full version of the Birkhoff ergodic theorem, in which one does not assume ergodicity). Given $p, p' \in \Omega$, such that $p' = vwp$ with $v \in V$ and $w \in W$, we wish to show $w = e$.

Choose a sequence $t_n \to \infty$, such that $a_{t_n} p$ and $a_{t_n} p'$ each belong to Ω_ρ. Because $t_n \to \infty$ and VW is the foliation that is contracted by $a_{\mathbb{R}^+}$, we know that $a_{-t_n}(vw)a_{t_n} \to e$. Furthermore, because A acts on the Lie algebra of V with twice the weight that it acts on the Lie algebra of W, we see that $\|a_{-t_n} v a_{t_n}\| / |a_{-t_n} w a_{t_n}\| \to 0$. Thus $p'_n = a_{-t_n} p' a_{t_n}$ approaches $p_n = a_{-t_n} p a_{t_n}$ from the direction of W.

If two points p_n' and p_n approach each other along W, then an easy computation shows that $u_t p_n$ and $u_t p_n'$ diverge along H. Thus by Proposition 3.6 μ must be invariant under H. But this impossible by §3.1 (since we are assuming that μ is not Haar measure). $\square$

We require the following entropy estimate.

Lemma 3.9 (cf. [MT, Thm. 9.7], [Mor, Prop. 2.5.11])**.** *Suppose $\mathcal{W}$ is a closed connected subgroup of VW that is normalized by $a \in A^+$, and let*

$$J(a^{-1}, \mathcal{W}) = \det\big((\operatorname{Ad} a^{-1})|_{\operatorname{Lie}\mathcal{W}}\big)$$

be the Jacobian of a^{-1} on $\mathcal{W}$.

1. *If μ is $\mathcal{W}$-invariant, then $h_\mu(a) \geq \log J(a^{-1}, \mathcal{W})$.*
2. *If there is a conull, Borel subset Ω of X, such that $\Omega \cap VWp \subset \mathcal{W}p$, for every $p \in \Omega$, then $h_\mu(a) \leq \log J(a^{-1}, \mathcal{W})$.*
3. *If the hypotheses of 2 are satisfied, and equality holds in its conclusion, then μ is $\mathcal{W}$-invariant.*

Proposition 3.10 (cf. [MT, Step 1 of 10.5], [Mor, Prop. 5.6.1])**.** *μ is V-invariant.*

Proof. From Lemma 3.9(1), with a^{-1} in the role of a, we have

$$\log J(a, UX) \leq h_\mu(a^{-1}).$$

From Proposition 3.8 and Lemma 3.9(2), we have

$$h_\mu(a) \leq \log J(a^{-1}, VY).$$

Combining these two inequalities with the facts that

- $h_\mu(a) = h_\mu(a^{-1})$ and
- $J(a, UX) = J(a^{-1}, VY)$,

we have

$$\log J(a, UX) \leq h_\mu(a^{-1}) = h_\mu(a) \leq \log J(a^{-1}, VY) = \log J(a, UX).$$

Thus, we must have equality throughout, so the desired conclusion follows from Lemma 3.9(3). $\square$

Proposition 3.11. *μ is the Lebesgue measure on a single orbit of $SL(2,\mathbb{R})$ on X.*

Proof We know:

- U preserves μ (by assumption),
- A preserves μ (by Proposition 3.7) and
- V preserves μ (by Proposition 3.10).

Since $SL(2,\mathbb{R})$ is generated by U, A and V, μ is $SL(2,\mathbb{R})$ invariant. Because $SL(2,\mathbb{R})$ is transitive on the quotient $\mathcal{L}_2$ and the support of μ on each fiber is finite (see Proposition 3.2), this implies that some orbit of $SL(2,\mathbb{R})$ has positive measure. By ergodicity of U, then this orbit is conull. $\square$

This completes the proof of Theorem 3.1.

4 Non-divergence of unipotent flows: the case of $\mathrm{SL}(2,\mathbb{R})$.

Here we return to the set-up of §2, that is, $U = \{u_x \ : \ x \in \mathbb{R}\}$ acting on $\mathcal{L}_2$.

4.1 Lemma 2.2 implies nondivergence

Recall that one of the lemmas from §2, namely Lemma 2.2, was a step towards "an elementary non-divergence result", Lemma 2.3. The latter essentially asserted that a trajectory $\{a_t\Lambda\}$, $\Lambda \in \mathcal{L}_2$, does not tend to infinity as $t \to -\infty$ unless Λ contains a horizontal (shrunk by a_t, $t < 0$) vector. (I know that Alex was using Δ instead of Λ, but I really prefer it this way. Maybe we will unify our notation later on.)

Here is another corollary from that lemma describing the same phenomenon for the U-action.

Corollary 4.1. *For any $\Lambda \in \mathcal{L}_2$, $u_x\Lambda$ does not tend to ∞ as $x \to \infty$.*

In other words, for any $\Lambda \in \mathcal{L}_2$ there exists a compact subset K of $\mathcal{L}_2$ such that the set $\{x > 0 \ : \ u_x\Lambda \in K\}$ is unbounded.

Proof. Assume the contrary; in view of Theorem 2.1, this would amount to assuming that the length of the shortest nonzero vector of $u_x\Lambda$ tends to zero as $x \to \infty$. Note that an obvious example of a divergent orbit would be constructed if one could find a vector $v \in \Lambda \smallsetminus \{0\}$ such that $u_x v \to 0$. But this is impossible: either v is horizontal and thus fixed by U, or its y-component is nonzero and does not change under the action. Thus the only allowed scenario for a divergent U-trajectory would be the following: for some $v \in \Lambda \smallsetminus \{0\}$, $u_x v$ gets very small, say shorter than ϵ, then starts growing but before it grows too big (longer than ϵ), another vector $v' \in \Lambda \smallsetminus \{0\}$ not proportional to v gets shrunk by u_x to the length less than ϵ. This however is prohibited by Lemma 2.2. $\square$

Remark. Observe that the analogue of this corollary is false if U is replaced by A, since a_t can contract nonzero vectors. However the same argument as above shows that for any continuous function $h : \mathbb{R}_+ \to \mathrm{SL}(2,\mathbb{R})$ and any $\Lambda \in \mathcal{L}_2$ such that $h(x)\Lambda$ diverges, it must do so in a degenerate way (Dani's terminology), that is, shrinking some nonzero vector $v \in \Lambda$. This phenomenon is specific to dimension 2: if $n > 2$ one can construct divergent trajectories $a_t\Lambda$ of diagonal one-parameter semigroups $a_t \in \mathrm{SL}(n,\mathbb{R})$ in $\mathcal{L}_n$ which diverge in a non-degenerate way (without shrinking any subpace of $\mathbb{R}^n$).

4.2 The Nondivergence Theorem of Margulis

Despite the above remark, an analogue of Corollary 4.1 holds in higher rank as well. It was conjectured by Piatetski-Shapiro in the late 1960s and showed in 1971 by Margulis [Mar1] as part of the program aimed at proving arithmeticity of lattices in higher rank algebraic groups.

Theorem 4.2. *Let $\{u_x\}$ be a one-parameter unipotent subgroup of* $\mathrm{SL}_(n,\mathbb{R})$*. Then for any $\Lambda \in \mathcal{L}_n$, $u_x\Lambda$ does not tend to ∞ as $x \to \infty$.*

An attempt to apply the proof of Corollary 4.1 verbatim fails miserably: there are no obstructions to having many short linear independent vectors. We will prove Theorem 4.2 in the next section in a much stronger (quantitative) form, which also happens to have important applications to problems arising in Diophantine approximation theory. But first, following the philosophy of these notes, we explain how one can easily establish a stronger form of Corollary 4.1, just for $n = 2$.

4.3 Quantitative nondivergence in $\mathcal{L}_2$

We are going to fix an interval $B \subset \mathbb{R}$ and $\Lambda \in \mathcal{L}_2$, and will look at the piece of trajectory $\{u_x\Lambda \;:\; x \in B\}$. Applying the philosophy of the proof of Corollary 4.1, one can see that one of the following two alternatives can take place:

Case 1. There exists a vector $v \in \Lambda \smallsetminus \{0\}$ such that $\|u_x v\|$ is small, say not greater than $\rho < 1$, for all $x \in B$. (For example this v may be fixed by U.) This case is not so interesting: again by Lemma 2.2, we know that this vector v is "the only source of trouble", namely no other vector can get small at the same time.

Case 2. The contrary, i.e.

$$\forall\, v \in \Lambda \smallsetminus \{0\} \quad \sup_{x\in B} \|u_x v\| \geq \rho\,. \tag{17}$$

In other words, every nonzero vector grows big enough at least at some point $x \in B$. This assumption turns out to be enough to conclude that for small ϵ the trajectory $\{u_x\Lambda \;:\; x \in B\}$ spends relatively small proportion of time, in terms of Lebesgue measure λ on $\mathbb{R}$, outside of $\mathcal{L}_2(\epsilon)$.

Before stating the next theorem we would like to point out one minor detail which was somewhat hidden in all the previous discussions: that the definition of the sets $\mathcal{L}_n(\epsilon)$, as well as conditions like (17), depend on the choice of the norm on $\mathbb{R}^n$, which was so far assumed to be Euclidean. Taking another norm would result in a slight change of the sets $\mathcal{L}_n(\epsilon)$ (up to a bounded distance) and some constants, such as ϵ_0 in Lemma 2.2. In particular, in the next theorem it will be more convenient to work with the supremum norm, so we are going to use an adjusted version of Lemma 2.2 with $\epsilon_0 = 1/\sqrt{2}$. We will switch back to the Euclidean norm afterwards.

Theorem 4.3. *Suppose an interval $B \subset \mathbb{R}$, $\Lambda \in \mathcal{L}_2$ and $0 < \rho < 1/\sqrt{2}$ are such that* (17) *holds. Then for any $\epsilon < \rho$,*

$$\lambda(\{x \in B \ : \ u_x\Lambda \notin \mathcal{L}_2(\epsilon)\}) \leq 2\frac{\epsilon}{\rho}\lambda(B)\,.$$

Thus, if one studies the curve $\{u_x\Lambda\}$ where x ranges from 0 to T, it suffices to look at the starting point Λ of the trajectory, find the length of its shortest vector, call it ρ, and apply the theorem to get a quantitative statement concerning the behavior of $\{u_x\Lambda \ : \ 0 \leq x \leq T\}$ for any T. Note that it is meaningful only when ϵ is small enough (not greater than $\rho/2$).

Proof. Denote by $P(\Lambda)$ the set of primitive vectors in Λ (v is said to be *primitive* in Λ if $\mathbb{R}v \cap \Lambda$ is generated by v as a $\mathbb{Z}$-module). Clearly in all the argument it will suffice to work with primitive vectors.

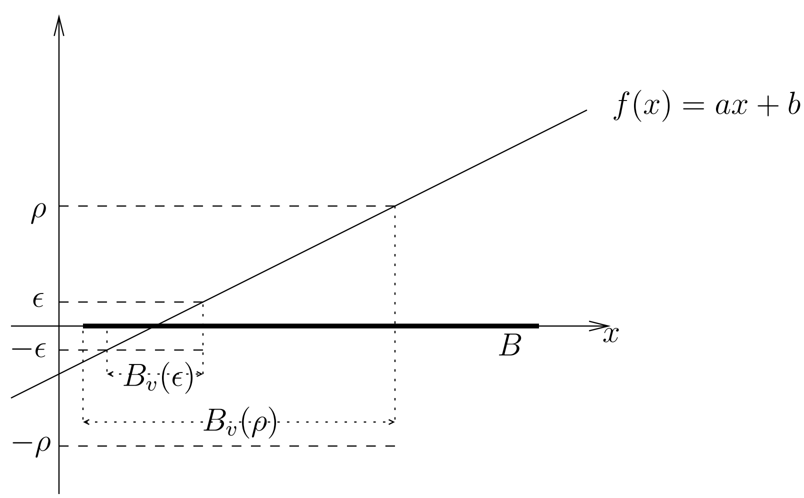

Figure 1. Proof of Theorem 4.3.

Now for each $v \in P(\Lambda)$ consider

$$B_v(\epsilon) \stackrel{\text{def}}{=} \{x \in B \ : \ \|u_xv\| < \epsilon\} \quad \text{and} \quad B_v(\rho) \stackrel{\text{def}}{=} \{x \in B \ : \ \|u_xv\| \leq \rho\}\,,$$

where $\|\cdot\|$ is the supremum norm. Let $v = \begin{pmatrix} a \\ b \end{pmatrix} \in P(\Lambda)$ be such that $B_v(\epsilon) \neq \varnothing$. Then, since $u_x v = \begin{pmatrix} a+bx \\ b \end{pmatrix}$, it follows that $|b| < \epsilon$, and (17) implies that b is nonzero. Therefore, if we denote $f(x) = a + bx$, we have

$$B_v(\epsilon) = \{x \in B \;\;:\;\; |f(x)| < \epsilon\} \quad \text{and} \quad B_v(\rho) = \{x \in B \;\;:\;\; |f(x)| < \rho\}\,.$$

Clearly the ratio of lengths of intervals $B_v(\epsilon)$ and $B_v(\rho)$ is bounded from above by $2\epsilon/\rho$ (by looking at the worst case when $B_v(\epsilon)$ is close to one of the endpoints of B). Lemma 2.2 guarantees that the sets $B_v(\rho)$ are disjoint for different $v \in P(\Lambda)$, and also that $u_x\Lambda \notin \mathcal{L}_2(\epsilon)$ whenever $x \in B_v(\rho) \smallsetminus B_v(\epsilon)$ for some $v \in P(\Lambda)$. Thus we conclude that

$$\lambda(\{x \in B \;\;:\;\; u_x\Lambda \notin \mathcal{L}_2(\epsilon)\}) \le \sum_v \lambda\big(B_v(\epsilon)\big) \le 2\frac{\epsilon}{\rho}\sum_v \lambda\big(B_v(\rho)\big) \le 2\frac{\epsilon}{\rho}\lambda(B)\,. \quad \square$$

Before proceeding to the more general case, let us summarize the main features of the argument. Each primitive vector v came with a function, $x \mapsto \|u_x v\|$, which

(1) allowed to compare measure of the subsets of B where this function is less than ϵ and ρ respectively, and

(2) attained value at least ρ on B.

Let us say that a point $x \in B$ is (ϵ/ρ)-*protected* if $x \in \overline{B_v(\rho)} \smallsetminus B_v(\epsilon)$ for some $v \in P(\Lambda)$. (1) and (2) imply that for each v, the relative measure of protected points inside $B_v(\rho)$ is big. Then Lemma 2.2 shows that protected points are safe (no other vector can cause trouble), i.e. brings us to the realm of Case 1 when restricted to $B_v(\rho)$.

In the analog of the argument for $n > 2$, (1) and (2) above will play an important role, but it will be more difficult to protect points from small vectors, and the final step using Lemma 2.2 will have to be replaced by an inductive procedure, described in the next section.

5 Quantitative non-divergence in $\mathcal{L}_n$.

5.1 The main concepts needed for the proof

The crucial idea that serves as a substitute for the absence of Lemma 2.2 in dimensions 3 and up is an observation that whenever a lattice Λ in $\mathbb{R}^n$ contains two linearly independent short vectors, one can consider a subgroup of rank two generated by them, and this subgroup will be "small", which should eventually contribute to preventing other small vectors from showing up. (Here and hereafter by the *rank* $\mathrm{rk}(\Delta)$ of a discrete subgroup Δ of $\mathbb{R}^n$ we mean its rank as a free $\mathbb{Z}$-module, or, equivalently, the dimension of the real vector space spanned by its elements.) Thus it seems to make sense to consider all subgroups of Λ, not just of rank one. In fact, similarly to the $n = 2$ case, it suffices to work with *primitive* subgroups. Namely, a subgroup Δ of Λ is called *primitive* in Λ if $\Delta = \mathbb{R}\Delta \cap \Lambda$; equivalently, if Δ admits a generating set which can be completed to a generating set of Λ. The inclusion relation makes the set $P(\Lambda)$ of all nonzero primitive subgroups of Λ a partially ordered set of length equal to $\mathrm{rk}(\Lambda)$ (any two primitive subgroups properly included in one another must have different ranks). This partial order turns out to be instrumental in creating a substitute for Lemma 2.2.

We also need a way to measure the size of a discrete subgroup Δ of $\mathbb{R}^n$. The best solution seems to be to use Euclidean norm $\|\cdot\|$ and extend it by letting $\|\Delta\|$ to be the volume of the quotient space $\mathbb{R}\Delta/\Delta$. This is clearly consistent with the one-dimensional picture, since $\|\mathbb{Z}v\| = \|v\|$. This is also consistent with the induced Euclidean structure on the exterior algebra of $\mathbb{R}^n$: if Δ is generated by $v_1, \dots, v_k$, then $\|\Delta\| = \|v_1 \wedge \cdots \wedge v_k\|$.

Our goal is to understand the trajectories $u_x\Lambda$ as in Theorem 4.2. However, observe that the group structure of U was not used at all in the proof in the previous section. Thus we are going to consider "trajectories" of a more general type. Namely, we will work with continuous functions h from an interval $B \subset \mathbb{R}$ into $\mathrm{SL}(n, \mathbb{R})$, and replace the map $x \mapsto u_x\Lambda$ with $x \mapsto h(x)\mathbb{Z}^n$ (then in the case of Theorem 4.2 we are going to have $h(x) = u_x g$ where $\Lambda = g\mathbb{Z}^n$).

Among the assumptions to be imposed on h, the central role is played by an analogue of (1) stated at the end of the previous section. This is taken care of by introducing a certain class of functions and then demanding that all functions of the form $x \mapsto \|h(x)\Delta\|$ where $\Delta \in P(\mathbb{Z}^n)$, belong to this class.

If C and α are positive numbers and B a subset of $\mathbb{R}$, let us say that a function $f : B \mapsto \mathbb{R}$ is *(C, α)-good on B* if for any open interval $J \subset B$ and any $\epsilon > 0$ one has

$$\lambda\big(\{x \in J \mid |f(x)| < \epsilon\}\big) \leq C \left(\frac{\epsilon}{\sup_{x \in J} |f(x)|}\right)^{\alpha} \lambda(J)\,. \tag{18}$$

Informally speaking, graphs of good functions are not allowed to spend a big proportion of "time" near the x-axis and then suddenly jump up. Several elementary facts about (C, α)-good functions are listed below:

Lemma 5.1. (a) *f is (C,α)-good on B $\Leftrightarrow$ so is $|f|$ $\Rightarrow$ so is cf $\forall\, c \in \mathbb{R}$;*

(b) *f_i, $i = 1,\dots,k$, are (C,α)-good on B $\Rightarrow$ so is $\sup_i |f_i|$;*

(c) *If f is (C,α)-good on B and $c_1 \le \left|\frac{f(x)}{g(x)}\right| \le c_2$ for all $x \in B$, then g is $\big(C(c_2/c_1)^\alpha, \alpha\big)$-good on B;*

The proofs are left as exercises. Another exercise is to construct a C^∞ function which is not good on (a) some interval (b) any interval.

The notion of (C,α)-good functions was introduced in [KM] in 1998, but the importance of (18) for measure estimates on the space of lattices was observed earlier. For instance, the next proposition, which describes what can be called a model example of good functions, can be traced to [DM4, Lemma 4.1]. We will prove a slightly stronger version paying more atention to the constant C (which will not really matter for the main results).

Proposition 5.2. *For any $k \in \mathbb{N}$, any polynomial of degree not greater than k is $\big(k(k+1)^{1/k}, 1/k\big)$-good on $\mathbb{R}$.*

Proof. Fix an open interval $J \subset \mathbb{R}$, a polynomial f of degree not exceeding k, and a positive ϵ. We need to show that

$$\lambda\big(\{x \in J : |f(x)| < \epsilon\}\big) \le k(k+1)^{1/k} \left(\frac{\epsilon}{\sup_{x\in J}|f(x)|}\right)^{1/k} \lambda(J)\,. \tag{19}$$

Suppose that the left hand side of (19) is strictly bigger than some number m. Then it is possible to choose $x_1,\dots,x_{k+1} \in \{x \in J : |f(x)| < \epsilon\}$ with $|x_i - x_j| \ge m/k$ for each $1 \le i \ne j \le k+1$. (Exercise.) Using Lagrange's interpolation formula one can write down the exact expression for f:

$$f(x) = \sum_{i=1}^{k+1} f(x_i) \frac{\prod_{j=1,\, j\ne i}^{k+1}(x - x_j)}{\prod_{j=1,\, j\ne i}^{k+1}(x_i - x_j)}\,.$$

Note that $|f(x_i)| < \epsilon$ for each i, $|x - x_j| < \lambda(J)$ for each j and $x \in J$, and also $|x_i - x_j| \ge m/k$. Therefore

$$\sup_{x\in J}|f(x)| < (k+1)\epsilon \frac{\lambda(J)^k}{(m/k)^k}\,.$$

which can be rewritten as

$$m < k(k+1)^{1/k} \left(\frac{\epsilon}{\sup_{x\in J} |f(x)|} \right)^{1/k} \lambda(J)\,,$$

proving (19). $\square$

Observe that in the course of the proof of Theorem 4.3 it was basically shown that linear functions are $(2,1)$-good on $\mathbb{R}$. The relevance of the above proposition for the nondivergence of unipotent flows on $\mathcal{L}_n$ is highlighted by

Corollary 5.3. *For any $n \in \mathbb{N}$ there exist (explicitly computable) $C = C(n)$, $\alpha = \alpha(n)$ such that for any one-parameter unipotent subgroup $\{u_x\}$ of $\mathrm{SL}_n(\mathbb{R})$, any $\Lambda \in \mathcal{L}_n$ and any subgroup Δ of Λ, the function $x \mapsto \|u_x\Delta\|$ is (C,α)-good.*

Proof. Represent Δ by a vector $w \in \bigwedge^k(\mathbb{R}^n)$ where k is the rank of Δ; the action of u_x on $\bigwedge^k(\mathbb{R}^n)$ is also unipotent, therefore every component of $u_x w$ (with respect to some basis) is a polynomial in x of degree uniformly bounded in terms of n. Thus the claim follows from Proposition 5.2, Lemma 5.1(b) for the supremum norm, and then Lemma 5.1(c) for the Euclidean norm. $\square$

5.2 The main nondivergence result and its history

Let us now state a generalization of Theorem 4.3 to the case of arbitrary n.

Theorem 5.4. *Suppose an interval $B \subset \mathbb{R}$, $C, \alpha > 0$, $0 < \rho < 1$ and a continuous map $h : B \to \mathrm{SL}(n,\mathbb{R})$ are given. Assume that for any $\Delta \in P(\mathbb{Z}^n)$,*

(i) *the function $x \mapsto \|h(x)\Delta\|$ is (C,α)-good on B, and*

(ii) $\sup_{x\in B} \|h(x)\Delta\| \geq \rho^{\mathrm{rk}(\Delta)}$.

Then for any $\epsilon < \rho$,

$$\lambda(\{x \in B \ : \ h(x)\mathbb{Z}^n \notin \mathcal{L}_n(\epsilon)\}) \leq n2^n C \left(\frac{\epsilon}{\rho}\right)^{\alpha} \lambda(B)\,. \tag{20}$$

This is a simplified version of a theorem from [K2], which sharpens the one proved in [KM]. The latter had a slightly stronger assumptions, with ρ in place of $\rho^{\mathrm{rk}(\Delta)}$ in the inequalities (ii) above. In most of the applications this improvement is not needed – but there are some situations in metric Diophantine approximation (which may or may not be covered in this lectures) where it becomes important. Anyway, the scheme of the proof is exactly the same for both original and new versions, and also there are some reasons why the new one appears to be more natural, as we will hopefully see below.

It is straightforward to verify that Theorem 4.2 follows from Theorem 5.4: take $B = [0, T]$ and $h(x) = u_x g$ where $\Lambda = g\mathbb{Z}^n$. Condition (i) has already been established in Corollary 5.3, and (ii) clearly holds with some ρ dependent of Λ: just put $x = 0$ and

$$\rho = \rho(\Lambda) = \inf_{\Delta \in P(\Lambda)} \|\Delta\|^{1/\operatorname{rk}(\Delta)}, \tag{21}$$

positive since Λ is discrete. Furthermore, Theorem 5.4 implies the following

Corollary 5.5. *For any $\Lambda \in \mathcal{L}_n$ and any positive δ there exists a compact subset K of $\mathcal{L}_n$ such that for any unipotent one-parameter $\{u_x\} \subset \mathrm{SL}(n, \mathbb{R})$ and any positive T one has*

$$\frac{1}{T}\lambda(\{0 \le x \le T \ : \ u_x\Lambda \notin K\}) \le \delta. \tag{22}$$

This was proved by Dani in 1979 [Dan1]. For the proof using Theorem 5.4, just take $K = \mathcal{L}_n(\epsilon)$ where ϵ is such that

$$n2^n C(n)\,(\epsilon/\rho)^{\alpha(n)} < \delta, \tag{23}$$

$C(n), \alpha(n)$ are as in Corollary 5.3 and $\rho(\Lambda)$ as defined in (21). Thus, on top of Dani's result, one can recover an expression for the "size" of K in terms of δ.

But this is not the end of the story – one can conclude much more. It immediately follows from Minkowski's Lemma that if $\operatorname{rk}(\Delta)$ is, say, k, then the intersection of Δ with any compact convex subset of $\mathbb{R}\Delta$ of volume $2^k\|\Delta\|$ contains a nonzero vector. Thus such a Δ must contain a nonzero vector of length $\le 2\|\Delta\|/\nu_k^{1/k}$, where ν_k is the volume of the unit ball in $\mathbb{R}^k$. Consequently, if we know that $\Lambda \in \mathcal{L}_n(\rho')$ for some positive ρ', then $\rho(\Lambda)$ as defined in (21) is at least $c'\rho'$ where $c' = c'(n)$ depends only on n. Thus we have proved (modulo elementary computations left as an exercise)

Corollary 5.6. *For any $\delta > 0$ there exists (explicitly computable) $c = c(n, \delta)$ such that whenever $\{u_x\Lambda \ : \ 0 \le x \le T\} \subset \mathcal{L}_n$ is a unipotent trajectory nontrivially intersecting $\mathcal{L}_n(\rho)$ for some $\rho > 0$, (22) holds with $K = \mathcal{L}_n(c\rho)$.*

We remark that that the distance between $\mathcal{L}_n(\rho)$ and the complement of $\mathcal{L}_n(c\rho)$ is uniformly bounded from above by a constant depending only on c, not on ρ. Thus the above corollary guarantees that, regardless of the size of the compact set where a unipotent trajectory begins, we only need to increase the set by a bounded distance to make sure that the trajectory spends, say, at least half the time in the bigger set. Note that for the last conclusion one really needs to have $\rho^{\mathrm{rk}(\Delta)}$ and not ρ in the right hand side of (ii) above; previously available technology forced a much more significant expansion of $\mathcal{L}_n(\rho)$.

Let us now turn our attention to another non-divergence theorem, proved by Dani in 1984 [Dan4], and later generalized by Eskin, Mozes and Shah [EMS1]:

Corollary 5.7. *For any $\delta > 0$ there exists a compact subset $K \subset \mathcal{L}_n$ such that for any unipotent one-parameter subgroup $\{u_x\} \subset \mathrm{SL}(n, \mathbb{R})$ and any $\Lambda = g\mathbb{Z}^n \in \mathcal{L}_n$, either* (22) *holds for all large T, or there exists a $(g^{-1}u_x g)$-invariant proper subspace of $\mathbb{R}^n$ defined over $\mathbb{Q}$.*

Proof. Apply Theorem 5.4 with an arbitrary $\rho < 1$ and ϵ as in (23), as before choosing K to be equal to $\mathcal{L}_n(\epsilon)$. Assume that the first alternative in the statement of the corollary is not satisfied for some $\{u_x\}$, Λ and this K. This means that there exists an unbounded sequence T_k such that for each k, the conclusion of Theorem 5.4 with $\rho = 1$, ϵ chosen as above and $h(x) = u_x g$, does not hold for $B = [0, T_k]$. Since assumption (i) of the theorem is always true, (ii) must go wrong, i.e. for each k there must exist $\Delta_k \in P(\mathbb{Z}^n)$ such that $\|u_x g \Delta_k\| < 1$ for all $0 \le x \le T_k$. However there are only finitely many choices for such subgroups, hence one of them, Δ, works for infinitely many k. But $\|u_x g \Delta\|^2$ is a polynomial, hence it must be constant, hence u_x fixes $g(\mathbb{R}\Delta) \Leftrightarrow g^{-1}u_x g$ fixes the proper rational subspace $\mathbb{R}\Delta$. $\square$

5.3 The proof

In order to prove Theorem 5.4, we are going to create a substitute for the procedure of marking points by vectors (and thereby declaring them safe from any other small vectors) used in the proof of Theorem 4.3. However now vectors will not be sufficient for our purposes, we will need to replace it with *flags*, that is, linearly ordered subsets of the partially ordered set (poset) $P(\Lambda)$, $\Lambda \in \mathcal{L}_n$. Furthermore, to set up the induction we will need to prove a version of the theorem with $P(\mathbb{Z}^n)$ repalced by its subsets (more precisely, sub-posets) P. The induction will be on the *length* of P, i.e. the number of elements in its maximal flag. In this more general theorem we will also get rid of the expressions $\rho^{\mathrm{rk}(\Delta)}$ in the right hand side of (ii), replacing them with $\eta(\Delta)$, where η is an arbitrary function $P \to (0, 1]$ (to be called the *weight function*).

Now given an interval $B \subset \mathbb{R}$, a sub-poset $P \subset P(\mathbb{Z}^n)$, a weight function η, a map $h : B \to \mathrm{SL}(n, \mathbb{R})$ and $\epsilon > 0$, say that $x \in B$ is *ϵ-protected relative to P* if there exists a flag $F \subset P$ with the following properties:

(M1) $\epsilon\eta(\Delta) \le \|h(x)\Delta\| \le \eta(\Delta) \quad \forall\, \Delta \in F$;

(M2) $\|h(x)\Delta\| \ge \eta(\Delta) \quad \forall\, \Delta \in P \smallsetminus F$ comparable with every element of F.

We are going to show that with the choice $\eta(\Delta) = \rho^{\mathrm{rk}(\Delta)}$ and $P = P(\mathbb{Z}^n)$, (ϵ/ρ)-protected points are indeed protected from vectors in $h(x)\mathbb{Z}^n$ of length less than ϵ. But first let us check that the above definition reduces to the one used for the proof of Theorem 4.3 when $P = P(\mathbb{Z}^2)$. Indeed, for $h(x) = u_x g$, $\Delta = \mathbb{Z}v$ of rank 1, $\eta(\Delta) = \rho$ and ϵ substituted with ϵ/ρ, (M1) reduces to $\epsilon \le \|u_x g v\| \le \rho$, which was exactly the condition satisfied by some vector $v \in \mathbb{Z}^2$ for $x \in \overline{B_{gv}(\rho)} \smallsetminus B_{gv}(\epsilon)$. Further, (M2) in that case holds trivially, since the only element of $P(\mathbb{Z}^2) \smallsetminus \{\Delta\}$ comparable with Δ is $\mathbb{Z}^2$ itself, and $\|g\mathbb{Z}^2\| = 1 > \rho^2$. And the conclusion was that the existence of such v forces $u_x g\mathbb{Z}^2$ to belong to $\mathcal{L}_2(\epsilon)$.

Here is a generalization:

Proposition 5.8. *Let η be given by $\eta(\Delta) = \rho^{\mathrm{rk}(\Delta)}$ for some $0 < \rho < 1$. Then for any $\epsilon < \rho$ and any $x \in B$ which is (ϵ/ρ)-protected relative to $P(\mathbb{Z}^n)$, one has $h(x)\mathbb{Z}^n \in \mathcal{L}_n(\epsilon)$.*

Proof. For x as above, let $\{0\} = \Delta_0 \subsetneq \Delta_1 \subsetneq \cdots \subsetneq \Delta_\ell = \mathbb{Z}^n$ be all the elements of $F \cup \big\{\{0\}, \mathbb{Z}^n\big\}$. Properties (M1) and (M2) translate into:

(M1) $\frac{\epsilon}{\rho} \cdot \rho^{\mathrm{rk}(\Delta_i)} \le \|h(x)\Delta_i\| \le \rho^{\mathrm{rk}(\Delta_i)} \quad \forall\, i = 0, \dots, \ell - 1$;

(M2) $\|h(x)\Delta\| \ge \rho^{\mathrm{rk}(\Delta)} \quad \forall\, \Delta \in P(\mathbb{Z}^n) \smallsetminus F$ comparable with every Δ_i.

(Even though $\Delta_0 = \{0\}$ is not in $P(\mathbb{Z}^n)$, it would also satisfy (M1) with the convention $\|\{0\}\| = 1$.)

Take any $v \in \mathbb{Z}^n \smallsetminus \{0\}$. Then there exists j, $1 \le j \le \ell$, such that $v \in \Delta_j \smallsetminus \Delta_{j-1}$. Denote $\mathbb{R}(\Delta_{j-1} + \mathbb{Z}v) \cap \Lambda$ by Δ. Clearly it is a primitive subgroup of Λ satisfying $\Delta_{j-1} \subset \Delta \subset \Delta_j$, therefore Δ is comparable with Δ_i for every i (and may or may not coincide with one of the Δ_is). Now one can use properties (M1) and (M2) to deduce that

$$\|h(x)\Delta\| \ge \min\left(\frac{\epsilon}{\rho} \cdot \rho^{\mathrm{rk}(\Delta)},\, \rho^{\mathrm{rk}(\Delta)}\right) = \epsilon\rho^{\mathrm{rk}(\Delta)-1} = \epsilon\rho^{\mathrm{rk}(\Delta_{i-1})}\,. \tag{24}$$

On the other hand, from the submultiplicativity of the covolume it follows that $\|h(x)\Delta\|$ is not greater than $\|h(x)\Delta_{i-1}\| \cdot \|v\|$ (recall a similar step in the proof of Lemma 2.2). Thus

$$\|h(x)v\| \ge \frac{\|h(x)\Delta\|}{\|h(x)\Delta_{i-1}\|} \underset{\text{by (M1) and (24)}}{\ge} \frac{\epsilon\rho^{\mathrm{rk}(\Delta_{i-1})}}{\rho^{\mathrm{rk}(\Delta_{i-1})}} = \epsilon\,.$$

Hence $\Lambda \in \mathcal{L}_n(\epsilon)$ and the proof is finished. $\square$

This is perhaps the crucial point in the proof: we showed that a flag with certain properties does exactly what a single vector was doing in the case of $\mathrm{SL}(2,\mathbb{R})$; namely, it guarantees that in the lattices corresponding to protected points, no vector can be shorter than ϵ.

Now that the above proposition is established, we will forget about the specific form of the weight function and work with an arbitrary η. Here is a more general theorem:

Theorem 5.9. *Fix $0 \le k \le n$, and suppose an interval $B \subset \mathbb{R}$, $C, \alpha > 0$, a continuous map $h : B \to \mathrm{SL}(n,\mathbb{R})$, a poset $P \subset P(\mathbb{Z}^n)$ of length k and a weight function $\eta : P \to (0,1]$ are given. Assume that for any $\Delta \in P$*

(i) *the function $x \mapsto \|h(x)\Delta\|$ is (C,α)-good on B, and*

(ii) $\sup_{x\in B} \|h(x)\Delta\| \ge \eta(\Delta)$.

Then for any $0 < \epsilon < 1$,

$$\lambda(\{x \in B \;\;:\;\; x \text{ is not } \epsilon\text{-protected relative to } P\}) \le k2^k C\epsilon^\alpha \lambda(B)\,.$$

We remark that the use of an arbitrary P in place of $P(\mathbb{Z}^n)$ is justified not only by a possibility to prove the theorem by induction, but also by some applications to Diophantine approximation where proper sub-posets of $P(\mathbb{Z}^n)$ arise naturally. Maybe they will be mentioned, or at least referred to, at the end of the lectures.

Proof. We will break the argument into several steps.

Step 0. First let us see what happens when $k = 0$, the base case of the induction. In this case P is empty, and the flag $F = \varnothing$ will satisfy both (M1) and (M2). Thus all points of B are ϵ-protected relative to P for any ϵ, which means that in the case $k = 0$ the claim is trivial. So we can take $k \ge 1$ and suppose that the theorem is proved for all the smaller lengths of P.

Step 1. For any $y \in B$ let us define

$$S(y) \stackrel{\text{def}}{=} \{\Delta \in P \;\;:\;\; \|h(y)\Delta\| < \eta(\Delta)\}\,.$$

Roughly speaking, $S(y)$ is the set of Δs which gets small enough at y, i.e. potentially could bring trouble. By the discreteness of $h(y)\mathbb{Z}^n$ in $\mathbb{R}^n$, this is a finite subset of P.

Note that if this set happens to be empty, then $\|h(y)\Delta\| \ge \eta(\Delta)$ for all $\Delta \in P$, which means that $F = \varnothing$ can be used to ϵ-protect y for any ϵ. So let us define

$$E \stackrel{\text{def}}{=} \{y \in B \;:\; S(y) \neq \varnothing\} = \{y \in B \mid \exists\, \Delta \in P \text{ with } \|h(y)\Delta\| < \eta(\Delta)\}\,;$$

then to prove the theorem it suffices to estimate the measure of the set of points $x \in E$ which are not ϵ-protected relative to P.

A flashback to the proof for $n = 2$: there $S(y)$ consisted of primitive vectors v for which $\|u_y v\|$ was less than ρ, not more than one such vector was allowed, and nonexistence of such vectors automatically placed the lattice in $\mathcal{L}_n(\epsilon)$.

Step 2. Take $y \in E$ and $\Delta \in S(y)$, and define $B_{\Delta,y}$ to be the maximal interval of the form $B \cap (y - r, y + r)$ on which the absolute value of $\|h(\cdot)\Delta\|$ is not greater than $\eta(\Delta)$. From the definition of $S(y)$ and the continuity of functions $\|h(\cdot)\Delta\|$ it follows that $B_{\Delta,y}$ contains some neighborhood of y. Further, the maximality property of $B_{\Delta,y}$ implies that

$$\sup_{x \in B_{\Delta,y}} \|h(x)\Delta\| = \eta(\Delta)\,. \tag{25}$$

Indeed, either $B_{s,y} = B$, in which case the claim follows from (ii), or at one of the endpoints of $B_{\Delta,y}$, the function $\|h(\cdot)\Delta\|$ must attain the value $\eta(\Delta)$ – otherwise one can enlarge the interval and still have $\|h(\cdot)\Delta\|$ not greater than $\eta(\Delta)$ for all its points.

Another flashback: intervals $B_{\Delta,y}$ are analogues of $B_v(\rho)$ from the proof of Theorem 4.3 – but this time there is no disjointness, since many Δs can get small simultaneously.

Step 3. For any $y \in E$ let us choose an element Δ_y of $S(y)$ such that $B_{\Delta_y,y} = \bigcup_{\Delta \in S(y)} B_{\Delta,y}$ (this can be done since $S(y)$ is finite). In other words, $B_{\Delta_y,y}$ is maximal among all $B_{\Delta,y}$. For brevity we will denote $B_{\Delta_y,y}$ by B_y. We now claim that

$$\sup_{x \in B_y} \|h(x)\Delta\| \ge \eta(\Delta) \text{ for any } y \in E \text{ and } \Delta \in P\,. \tag{26}$$

Indeed, if not, then $\|h(x)\Delta\| < \eta(\Delta)$ for all $x \in B_y$, in particular one necessarily has $\|h(y)\Delta\| < \eta(\Delta)$, hence $\Delta \in S(y)$ and $B_{\Delta,y}$ is defined. But $B_{\Delta,y}$ is contained in B_y, so (26) follows from (25).

Yet another flashback: the collection of intervals B_y looks more similar to the family $\{B_v(\rho)\}$ than that of all $B_{\Delta,y}$, $\Delta \in S(y)$; this was achieved by selecting $\Delta = \Delta_y$ which works best for every given y, that is, Δ_y's motion is the slowest among all the relevant Δs.

Step 4. Now we are ready to perform the induction step. For any $y \in E$ define

$$P_y \stackrel{\text{def}}{=} \{\Delta \in P \smallsetminus \{\Delta_y\} \;:\; \Delta \text{ is comparable with } \Delta_y\}\,.$$

We claim that P_y (a poset of length $k-1$) in place of P and B_y in place of B satisfy all the conditions of the theorem. Indeed, (i) is clear since B_y is a subset of B, and (ii) follows from (26). Therefore, by induction,

$$\lambda(\{x \in B_y \ : \ x \text{ is not } \epsilon\text{-protected relative to } P_y\}) \le (k-1)2^{k-1}C\epsilon^\alpha\lambda(B_y)\,. \tag{27}$$

Step 5. Does the previous step help us, and how? let us take x outside of this set of relatively small measure, that is, assume that x is ϵ-protected relative to P_y, and try to use this protection. By definition, there exists a flag F' inside P_y such that

$$\epsilon\eta(\Delta) \le \|h(x)\Delta\| \le \eta(\Delta) \quad \forall\,\Delta \in \ F' \tag{28}$$

and

$$\|h(x)\Delta\| \ge \eta(\Delta) \quad \forall\,\Delta \in P_y \smallsetminus F' \text{ comparable with every element of } F'\,. \tag{29}$$

However this F' will NOT protect x relative to the bigger poset P, because Δ_y, comparable with every element of F', would not satisfy (M2) – on the contrary, recall that it was chosen so that the reverse inequality, $\|h(x)\Delta_y\| \le \eta(\Delta_y)$, holds for all $x \in B_y$, see (26)! Thus our only choice seems to be to add Δ_y to F', for extra protection, and put $F \stackrel{\text{def}}{=} F' \cup \{\Delta_y\}$. Then $\Delta \in P \smallsetminus F$ is comparable with every element of F if and only if Δ is in $P_y \smallsetminus F'$, and is comparable with every element of F'. Because of that, (M2) immediately follows from (29). As for (M1), we already know it for for $\Delta \ne \Delta_y$ by (28), so it remains to put $\Delta = \Delta_y$. The upper estimate in (M1) is immediate from (26). The lower estimate, on the other hand, can fail – but only on a set of relatively small measure, because of assumption (i) which, by the way, has not been used so far at all:

$$\begin{aligned}\lambda(\{x \in B_y \ : \ \|h(x)\Delta_y\| < \epsilon\eta(\Delta_y)\}) &\le C\left(\frac{\epsilon\eta(\Delta_y)}{\sup_{x\in B_y}\|h(x)\Delta_y\|}\right)^\alpha \lambda(B_y)\\ &\underset{(25)}{\le} C(\epsilon)^\alpha\lambda(B_y)\,.\end{aligned} \tag{30}$$

The union of the two sets above, in the left hand sides of (27) and (30), has measure at most $k2^{k-1}C\epsilon^\alpha\lambda(B_y)$. We have just shown that this union exhausts all the unprotected points as long as we are restricted to B_y. Thus we have achieved an analogue of what was extremely easy for $n = 2$: bounded the measure of the set of points where things can go wrong on each of the intervals $B_v(\rho)$.

Step 6. It remains to produce a substitute for the disjointness of the intervals, that is, put together all the B_ys. For that, consider the covering $\{B_y \mid y \in E\}$ of E and choose a subcovering $\{B_i\}$ of multiplicity at most 2. (Exercise: this is always possible.) Then the measure of $\{x \in E \ : \ x \text{ is not } \epsilon\text{-protected relative to } P\}$ is not greater than

$$\sum_i \lambda(\{x \in B_i \ : \ x \text{ is not } \epsilon\text{-protected relative to } P\}) \leq k2^{k-1}C\epsilon^\alpha \sum_i \lambda(B_i) \leq k2^k C\epsilon^\alpha \lambda(B)\,,$$

and the theorem is proven. □

6 Linearization and ergodicity

6.1 Non-ergodic measures measures invariant under a unipotent.

The collection $\mathcal{H}$. (Up to conjugation, this should be the collection of groups which appear in the definition of algebraic measure).

Let G be a Lie group, Γ a discrete subgroup of G, and $\pi : G \to G/\Gamma$ the natural quotient map. Let $\mathcal{H}$ be the collection of all closed subgroups F of G such that $F \cap \Gamma$ is a lattice in F and the subgroup generated by unipotent one-parameter subgroups of G contained in F acts ergodically on $\pi(F) \cong F/(F \cap \Gamma)$ with respect to the F-invariant probability measure.

Proposition 6.1. *The collection $\mathcal{H}$ is countable.*

Proof. See [Ra6, Theorem 1.1] or [DM4, Proposition 2.1] for different proofs of this result. □

Let U be a unipotent one-parameter subgroup of G and $F \in \mathcal{H}$. Define

$$\begin{aligned} N(F,U) &= \{g \in G : U \subset gFg^{-1}\} \\ S(F,U) &= \bigcup \{N(F',U) : F' \in \mathcal{H},\ F' \subset F,\ \dim F' < \dim F\}. \end{aligned}$$

Lemma 6.2. ([MS, Lemma 2.4]) *Let $g \in G$ and $F \in \mathcal{H}$. Then $g \in N(F,U) \setminus S(F,U)$ if and only if the group gFg^{-1} is the smallest closed subgroup of G which contains U and whose orbit through $\pi(g)$ is closed in G/Γ. Moreover in this case the action of U on $g\pi(F)$ is ergodic with respect to a finite gFg^{-1}-invariant measure.*

As a consequence of this lemma,

$$\pi(N(F,U)\setminus S(F,U)) = \pi(N(F,U))\setminus \pi(S(F,U)), \qquad \forall F\in\mathcal{H}. \tag{31}$$

Ratner's theorem [Ra6] states that given any U-ergodic invariant probability measure on G/Γ, there exists $F\in\mathcal{H}$ and $g\in G$ such that μ is $g^{-1}Fg$-invariant and $\mu(\pi(F)g)=1$. Now decomposing any finite invariant measure into its ergodic component, and using Lemma 6.2, we obtain the following description for any U-invariant probability measure on G/Γ (see [MS, Theorem 2.2]).

Theorem 6.3 (Ratner)**.** *Let U be a unipotent one-parameter subgroup of G and μ be a finite U-invariant measure on G/Γ. For every $F\in\mathcal{H}$, let μ_F denote the restriction of μ on $\pi(N(F,U)\setminus S(F,U))$. Then μ_F is U-invariant and any U-ergodic component of μ_F is a gFg^{-1}-invariant measure on the closed orbit $g\pi(F)$ for some $g\in N(F,U)\setminus S(F,U)$.*

In particular, for all Borel measurable subsets A of G/Γ,

$$\mu(A) = \sum_{F\in\mathcal{H}^*}\mu_F(A),$$

where $\mathcal{H}^*\subset\mathcal{H}$ is a countable set consisting of one representative from each Γ-conjugacy class of elements in $\mathcal{H}$.

Remark. We will often use Theorem 6.3 in the following form: suppose μ is any U-invariant measure on G/Γ which is not Lebesque measure. Then there exists $F\in\mathcal{H}$ such that μ gives positive measure to some compact subset of $N(F,U)\setminus S(F,U)$.

6.2 The theorem of Dani-Margulis on uniform convergence

The "linearization" technique of Dani and Margulis was devised to understand which measures give positive weight to compact subsets subsets of $N(F,U)\setminus S(F,U)$. Using this technique Dani and Margulis proved the following theorem (which is important for many applications, in particular §7):

Theorem 6.4 ([DM4], Theorem 3)**.** *Let G be a connected Lie group and let Γ be a lattice in G. Let μ be the G-invariant probability measure on G/Γ. Let $U=\{u_t\}$ be an Ad-unipotent one-parameter subgroup of G and let f be a bounded continuous function on G/Γ. Let $\mathcal{D}$ be a compact subset of G/Γ and let $\epsilon>0$ be given. Then there exist finitely many proper closed subgroups $F_1=F_1(f,\mathcal{D},\epsilon),\cdots,F_k=F_k(f,\mathcal{D},\epsilon)$ such that*

$F_i \cap \Gamma$ is a lattice in F_i for all i, and compact subsets $C_1 = C_1(f, \mathcal{D}, \epsilon), \cdots, C_k = C_k(f, \mathcal{D}, \epsilon)$ of $N(F_1, U), \cdots, N(F_k, U)$ respectively, for which the following holds: For any compact subset K of $\mathcal{D} - \bigcup_{1 \le i \le k} \pi(C_i)$ there exists a $T_0 \ge 0$ such that for all $x \in K$ and $T > T_0$

$$\Big| \frac{1}{T} \int_0^T f(u_t x)\, dt - \int_{G/\Gamma} f\, d\mu \Big| < \epsilon. \tag{32}$$

Remarks.

- This theorem can be informally stated as follows: Fix f and $\epsilon > 0$. Then (32) holds (i.e. the space average of f is within ϵ of the time average of f) uniformly in the base point x, as long as x is restricted to compact sets away from a finite union of "tubes" $N(F, U)$. (The $N(F, U)$ are associated with orbits which do not become equidistributed in G/Γ, because their closure is strictly smaller.)
- It is a key point that only finitely many F_k are needed in Theorem 6.4. This has the remarkable implication that if $F \in \mathcal{H}$ but not one of the F_k, then (32) holds for $x \in N(F, U)$ even though Ux is not dense in G/Γ (the closure of Ux is Fx). Informally, this means the non-dense orbits of U are themselves becoming equidistributed as they get longer.

A full proof of Theorem 6.4 is beyond the scope of this course. However, we will describe the "linearization" technique used in its proof in §6.3.

6.3 Ergodicity of limits of ergodic measures

In this subsection we are following [MS], which refers many times to [DM4].

Let $\mathcal{P}(G/\Gamma)$ be the space of all probability measures on G/Γ.

Theorem 6.5 (Mozes-Shah)**.** *Let U_i be a sequence of unipotent one-parameter subgroups of G, and for each i, let μ_i be an ergodic U_i-invariant probability measure on G/Γ. Suppose $\mu_i \to \mu$ in $\mathcal{P}(G/\Gamma)$. Then there exists a unipotent one-parameter subgroup U such that μ is an ergodic U-invariant measure on G/Γ. In particular, μ is algebraic.*

Remarks.

- Let $\mathcal{Q}(G/\Gamma) \subset \mathcal{P}(G/\Gamma)$ denote the set of measures ergodic for the action of a unipotent one-parameter subgroup of G, and let $\mathcal{Q}_0(G/\Gamma)$ denote $\mathcal{Q}(G/\Gamma)$ union the zero measure. If combined with the results of §5, Theorem 6.5 shows that $\mathcal{Q}_0(G/\Gamma)$ is *compact*.

- The theorem actually proved by Mozes and Shah in [MS] gives more information about what kind of limits of ergodic U-invariant measures are possible. Here is an easily stated consequence:

  Suppose $x_i \in G/\Gamma$ converge to $x_\infty \in G/\Gamma$, and also $x_i \in \overline{Ux_\infty}$. For $i \in \mathbb{N} \cup \{\infty\}$ let μ_i be the algebraic measures supported on $\overline{Ux_i}$, so that the trajectories Ux_i are equidistributed with respect to the measures μ_i. Then $\mu_i \to \mu_\infty$.

We now give some indication of the proof of Theorem 6.5. Let U_i, μ_i, μ be as in Theorem 6.5. Write $U_i = \{u_i(t)\}_{t \in \mathbb{R}}$.

Invariance of μ under a unipotent.

Lemma 6.6. *Suppose $U_i \neq \{e\}$ for all large $i \in \mathbb{N}$. Then μ is invariant under a one-parameter unipotent subgroup of G.*

Proof. For each $i \in \mathbb{N}$ there exists w_i in the Lie algebra $\mathfrak{g}$ of G, such that $\|w_i\| = 1$ and $U_i = \{\exp(tw_i), t \in \mathbb{R}\}$. (Here $\|\cdot\|$ is some Euclidean norm on $\mathfrak{g}$). By passing to a subsequence we may assume that $w_i \to w$ for some $w \in \mathfrak{g}$, $\|w\| = 1$. For any $t \in \mathbb{R}$ we have $Ad(\exp(tw_i)) \to Ad(\exp(tw))$ as $i \to \infty$. Note that $Ad(\exp(tw))$ is unipotent, since the set of unipotent matrices is closed (consider e.g. the characteristic polynomial). Therefore $U = \{\exp(tw) \ : \ t \in \mathbb{R}\}$ is a nontrivial unipotent subgroup of G. Since $\exp tw_i \to \exp tw$ for all t and $\mu_i \to \mu$, it follows that μ is invariant under the action of U on G/Γ. $\square$

Application of Ratner's measure classification theorem. We want to analyze the case when the limit measure μ is not the G-invariant measure. By Ratner's description of μ as in Theorem 6.3, there exists a proper subgroup $F \in \mathcal{H}$, $\epsilon_0 > 0$, and a compact set $C_1 \subset N(F,U) \setminus S(F,U)$ such that $\mu(\pi(C_1)) > \epsilon_0$. Thus for any neighborhood Φ of $\pi(C_1)$, we have $\mu_i(\Phi) > \epsilon_0$ for all large $i \in \mathbb{N}$. Thus the unipotent trajectories which are equidistributed with respect to the measures μ_i spend a fixed proportion of time in Φ.

Linearization of neighborhoods of singular subsets. Let $F \in \mathcal{H}$. Let $\mathfrak{g}$ denote the Lie algebra of G and let $\mathfrak{f}$ denote its Lie subalgebra associated to F. For $d = \dim \mathfrak{f}$, put $V_F = \wedge^d \mathfrak{f}$, the d-th exterior power, and consider the linear G-action on V_F via the representation $\wedge^d \mathrm{Ad}$, the d-th exterior power of the Adjoint representation of G on $\mathfrak{g}$. Fix $p_F \in \wedge^d \mathfrak{f} \setminus \{0\}$, and let $\eta_F : G \to V_F$ be the map defined by $\eta_F(g) = g \cdot p_F = (\wedge^d \mathrm{Ad}\, g) \cdot p_F$ for all $g \in G$. Note that

$$\eta_F{}^{-1}(p_F) = \{g \in N_G(F) : \det(\mathrm{Ad}\, g|_{\mathfrak{f}}) = 1\}.$$

Remark. The idea of Dani and Margulis is to work in the representation space V_F (or its double quotient $\bar{V}_F$) instead of G/Γ. The advantage is that F is collapsed to a point in V_F. The difficulty is that the map $\eta_F : G \to \bar{V}_F$ is not Γ-equivariant, and so becomes multivalued if considered as a map from G/Γ to V_F.

Proposition 6.7 ([DM4, Theorem 3.4])**.** *The orbit $\Gamma \cdot p_F$ is discrete in V_F.*

Remark. In the arithmetic case the above proposition is immediate.

Proposition 6.8. *([DM4, Prop. 3.2]) Let A_F be the linear span of $\eta_F(N(F,U))$ in V_F. Then*

$${\eta_F}^{-1}(A_F) = N(F,U).$$

Let $N_G(F)$ denote the normalizer in G of F. Put $\Gamma_F = N_G(F) \cap \Gamma$. Then for any $\gamma \in \Gamma_F$, we have $\gamma\pi(F) = \pi(F)$, and hence γ preserves the volume of $\pi(F)$. Therefore $|\det(\mathrm{Ad}\,\gamma|_{\mathfrak{f}})| = 1$. Hence $\gamma \cdot p_F = \pm p_F$. Now define

$$\bar{V}_F = \begin{cases} V_F/\{\text{Id,-Id}\} & \text{if } \Gamma_F \cdot p_F = \{p_F, -p_F\} \\ V_F & \text{if } \Gamma_F \cdot p_F = p_F \end{cases}$$

The action of G factors through the quotient map of V_F onto $\bar{V}_F$. Let $\bar{p}_F$ denote the image of p_F in $\bar{V}_F$, and define $\bar{\eta}_F : G \to \bar{V}_F$ as $\bar{\eta}_F(g) = g \cdot \bar{p}_F$ for all $g \in G$. Then $\Gamma_F = {\bar{\eta}_F}^{-1}(\bar{p}_F) \cap \Gamma$. Let $\bar{A}_F$ denote the image of A_F in $\bar{V}_F$. Note that the inverse image of $\bar{A}_F$ in V_F is A_F.

For every $x \in G/\Gamma$, define the set of representatives of x in $\bar{V}_F$ to be

$$\mathrm{Rep}(x) = \bar{\eta}_F(\pi^{-1}(x)) = \bar{\eta}_F(x\Gamma) \subset \bar{V}_F.$$

Remark. If one attempts to consider the map $\bar{\eta}_F : G \to V_F$ as a map from G/Γ to V_F, one obtains the multivalued map which takes $x \in G/\Gamma$ to the set $\mathrm{Rep}(x) \subset V_F$.

The following lemma allows us to understand the map Rep in a special case:

Lemma 6.9. *If $x = \pi(g)$ and $g \in N(F,U) \setminus S(F,U)$*

$$\mathrm{Rep}(x) \cap \bar{A}_F = \{g \cdot p_F\}.$$

Thus x has a single representative in $\bar{A}_F \subset V_F$.

Proof. Indeed, using Proposition 6.8,

$$\mathrm{Rep}(\pi(g)) \cap \bar{A}_F = (g\Gamma \cap N(F,U)) \cdot \bar{p}_F$$

Now suppose $\gamma \in \Gamma$ is such that $g\gamma \in N(F,U)$. Then g belongs to $N(\gamma F \gamma^{-1}, U)$ as well as $N(F,U)$. Since $g \notin S(F,U)$, we must have $\gamma F \gamma^{-1} = F$, so $\gamma \in \Gamma_F$. Then $\gamma \bar{p}_F = \bar{p}_F$, so $(g\Gamma \cap N(F,U)) \cdot \bar{p}_F = \{g \cdot \bar{p}_F\}$ as required. □

We extend this observation in the following result (cf. [Sha1, Prop. 6.5]).

Proposition 6.10 ([DM4, Corollary 3.5])**.** *Let D be a compact subset of $\bar{A}_F$. Then for any compact set $K \subset G/\Gamma \setminus \pi(S(F,U))$, there exists a neighborhood Φ of D in $\bar{V}_F$ such that any $x \in K$ has at most one representative in Φ.*

Remark. This proposition constructs a "fundamental domain" Φ around any compact subset D of $\bar{A}_F$, so that for any x in a compact subset of G/Γ away from $\pi(S(F,U))$, $\mathrm{Rep}(x)$ has at most one element in Φ. Using this proposition, one can uniquely represent in Φ the parts of the unipotent trajectories in G/Γ lying in K.

Proposition 6.11 ([DM4, Proposition 4.2])**.** *Let $n \in \mathbb{N}$, $\Lambda \geq 0$, a compact set $C \subset \bar{A}_F$ and an $\epsilon > 0$ be given. Then there exists a (larger) compact set $D \subset \bar{A}_F$ with the following property: For any neighborhood Φ of D in $\bar{V}_F$ there exists a neighborhood Ψ of C in $\bar{V}_F$ with $\Psi \subset \Phi$ such that the following holds: For any unipotent one parameter subgroup $\{u(t)\}$ of G, an element $w \in \bar{V}_H$ and and interval $I \subset \mathbb{R}$, if $u(t_0)w \notin \Phi$ for some $t_0 \in I$ then,*

$$|\{t \in I \ : \ u(t)w \in \Psi\}| \leq \epsilon \cdot |\{t \in I \ : \ u(t)w \in \Phi\}|. \tag{33}$$

Proof. This is an "polynomial divergence" estimate similar to these in §4 and §5. □

Proposition 6.12. *Let $n \in \mathbb{N}$, $\Lambda \geq 0$, $\epsilon > 0$, a compact set $K \subset G/\Gamma \setminus \pi(S(F,U))$, and a compact set $C \subset \bar{A}_F$ be given. Then there exists a neighborhood Ψ of C in $\bar{V}_F$ such that for any unipotent one-parameter subgroup $\{u(t)\}$ of G and any $x \in G/\Gamma$, at least one of the following conditions is satisfied:*

1. *There exists $w \in \mathrm{Rep}(x) \cap \overline{\Psi}$ such that $\{u(t)\} \subset G_w$, where $G_w = \{g \in G \ : \ gw = w\}$.*
2. *For all large $T > 0$,*

$$|\{t \in [0,T] : u(t)x \in K \cap \pi(\bar{\eta}_F^{-1}(\Psi))\}| \leq \epsilon T.$$

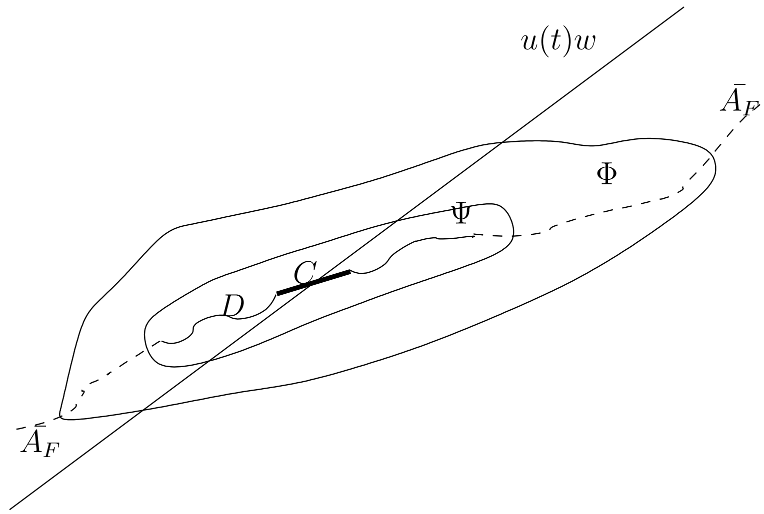

Figure 2. Proposition 6.11.

Proof. Let a compact set $D \subset \bar{A}_F$ be as in Proposition 6.11. Let Φ be a given neighborhood of D in $\bar{V}_F$. Replacing Φ by a smaller neighborhood of D, by Proposition 6.10 the set $\mathrm{Rep}(x) \cap \Phi$ contains at most one element for all $x \in K$. By the choice of D there exists a neighborhood Ψ of C contained in Φ such that equation (33) holds.

Now put $\Omega = \pi(\bar{\eta}_F^{-1}(\Psi)) \cap K$, and define

$$E = \{t \geq 0 \ : \ u(t)x \in \overline{\Omega}\}. \tag{34}$$

Let $t \in E$. By the choice of Φ, there exists a unique $w \in \bar{V}_F$ such that $\mathrm{Rep}(u(t)x) \cap \Phi = \{u(t)w\}$.

Since $s \to u(s)w$ is a polynomial function, either it is constant or it is unbounded as $s \to \pm\infty$. In the first case condition 1) is satisfied and we are done. Now suppose that condition 1 does not hold. Then for every $t \in E$, there exists a largest open interval $I(t) \subset (0, T)$ containing t such that

$$u(s)w \in \Phi \qquad \text{for all } s \in I(t). \tag{35}$$

Put $\mathcal{I} = \{I(t) : t \in E\}$, Then for any $I_1 \in \mathcal{I}$ and $s \in I_1 \cap E$, we have $I(s) = I_1$. Therefore for any $t_1, t_2 \in E$, if $t_1 < t_2$ then either $I(t_1) = I(t_2)$ or $I(t_1) \cap I(t_2) \subset (t_1, t_2)$. Hence any $t \in [0, T]$ is contained in at most two distinct elements of $\mathcal{I}$. Thus

$$\sum_{I \in \mathcal{I}} |I| \leq 2T. \tag{36}$$

Now by equations (33) and (35), for any $t \in E$,

$$|\{s \in I(t) : u(s)w \in \Psi\}| < \epsilon \cdot |I(t)|. \tag{37}$$

Therefore by equations (36) and (37), we get

$$|E| \leq \epsilon \cdot \sum_{I \in \mathcal{I}} |I| \leq (2\epsilon)T,$$

which is condition 2 for 2ϵ in place of ϵ. $\square$

Outline of Proof of Theorem 6.5. Suppose μ is not Haar measure on G/Γ. By Lemma 6.6 μ is invariant under some one-parameter unipotent subgroup μ. Then by Theorem 6.3 there exists $F \in \mathcal{H}$ such that $\mu(N(F,U)) > 0$ and $\mu(S(F,U)) = 0$. Thus there exists a compact subset C_1 of $N(F,U) \setminus S(F,U)$ and $\alpha > 0$ such that

$$\mu(\pi(C_1)) > \alpha. \tag{38}$$

Take any $y \in \pi(C_1)$. It is easy to see that for each $i \in \mathbb{N}$ there exists $y_i \in \mathrm{supp}(\mu_i)$ such that $\{u_i(t)y_i\}$ is uniformly distributed with respect to μ_i, and also $y_i \to y$ as $i \to \infty$. Let $h_i \to e$ be a sequence in G such that $h_i y_i = y$ for all $i \in \mathbb{N}$.

We now replace μ_i by $\mu_i' = h_i\mu_i$. We still have $\mu_i' \to \mu$, but now we also have $y \in \mathrm{supp}(\mu_i')$ for all i. Let $u_i'(t) = h_i u_i(t) h_i^{-1}$. Then the trajectory $\{u_i'(t)y\}$ is uniformly distributed with respect to μ_i'.

We now apply Proposition 6.12 for $C = \bar{\eta}_F(C_1)$ and $\epsilon = \alpha/2$. We can choose a compact neighborhood K of $\pi(C_1)$ such that $K \cap S(F,U) = \emptyset$. Put $\Omega = \pi(\bar{\eta}_F^{-1}(\Psi)) \cap K$. Since $\mu_i' \to \mu$, due to (38) there exists $k_0 \in \mathbb{N}$ such that $\mu_i'(\Omega) > \epsilon$ for all $i \geq k_0$. This means that Condition 2) of Proposition 6.12 is violated for all $i \geq k_0$. Therefore according to condition 1) of Proposition 6.12, for each $i \geq k_0$,

$$\{u_i'(t)y\}_{t \in \mathbb{R}} \subset G_w y,$$

where G_w is as in Proposition 6.12. By Proposition 6.7, $G_w y$ is closed in G/Γ.

The rest of the proof is by induction on $\dim G$. If $\dim G_w < \dim G$ then everything is taking place in the homogeneous space $G_w y$, and therefore μ is ergodic by the induction hypothesis. If $\dim G_w = \dim G$ then $G_w = G$ and hence F is a normal subgroup of G. In this case one can project the measures to the homogeneous space $G/(F\Gamma)$ and apply induction. $\square$

7 Oppenheim and Quantitative Oppenheim

7.1 The Oppenheim Conjecture.

Let Q be an indefinite nondegenerate quadratic form in n variables. Let $Q(\mathbb{Z}^n)$ denote the set of values of Q at integral points. The Oppenheim conjecture, proved by Margulis (cf. [Mar3]) states that if $n \geq 3$, and Q is not proportional to a form with rational coefficients, then $Q(\mathbb{Z}^n)$ is dense. The Oppenheim conjecture enjoyed attention and many studies since it was conjectured in 1929 mostly using analytic number theory methods.

In the mid seventies Raghunathan observed a remarkable connection between the Oppenheim Conjecture and unipotent flows on the space of lattices $\mathcal{L}_n = SL(n,\mathbb{R})/SL(n,\mathbb{Z})$. It can be summarized as the following:

Observation 7.1 (Raghunathan)**.** *Let Q be an indefinite quadratic form Q and let $H = SO(Q)$ denote its orthogonal group. Consider the orbit of the standard lattice $\mathbb{Z}^n \in \mathcal{L}_n$ under H. Then the following are equivalent:*

(a) *The orbit $H\mathbb{Z}^n$ is not relatively compact in $\mathcal{L}_n$.*

(b) *For all $\epsilon > 0$ there exists $u \in \mathbb{Z}^n$ such that $|Q(u)| < \epsilon$.*

(c) *The set $Q(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.*

Proof. Suppose (a) holds, so some sequence $h_k\mathbb{Z}^n$ leaves all compact sets. Then in view of the Mahler compactness criterion there exist $v_k \in h_k\mathbb{Z}^n$ such that $\|v_k\| \to 0$. Then also by continuity, $Q(v_k) \to 0$. But then $h_k^{-1}v_k \in \mathbb{Z}^n$, and $Q(h_k^{-1}v_k) = Q(v_k) \to 0$. Thus (b) holds.

It is easy to see that (b) implies (a). It is also possible to show that (b) implies (c). □

The Oppenheim Conjecture, the Raghunathan Conjecture and Unipotent Flows. Raghunathan also explained why the case $n = 2$ is different: in that case $H = SO(Q)$ is not generated by unipotent elements. Margulis's proof of the Oppenheim conjecture, given in [Mar 2-4] uses Raghunathan's observation. In fact Margulis showed that that any relatively compact orbit of $SO(2,1)$ in $SL(3,\mathbb{R})/SL(3,\mathbb{Z})$ is compact; this implies the Oppenheim Conjecture.

Raghunathan also conjectured Theorem 1.13. In the literature it was first stated in the paper [Dan2] and in a more general form in [Mar3] (when the subgroup U is not necessarily unipotent but generated by unipotent elements). Raghunathan's conjecture was eventually proved in full generality by M. Ratner (see [Ra7]). Earlier it was known in the following cases: (a) G is reductive and U is horospherical (see [Dan2]); (b) $G = SL(3,\mathbb{R})$ and $U = \{u(t)\}$ is a one-parameter unipotent subgroup of G such that $u(t) - I$ has rank 2 for all $t \neq 0$, where I is the identity matrix (see [DM2]); (c) G is solvable (see [Sta1] and [Sta2]). We remark that the proof given in [Dan2] is restricted to horospherical U and the proof given in [Sta1] and [Sta2] cannot be applied for nonsolvable G.

However the proof in [DM2] together with the methods developed in [Mar 2-4] and [DM1] suggest an approach for proving the Raghunathan conjecture in general by studying the minimal invariant sets, and the limits of orbits of sequences of points tending to a minimal invariant set. This strategy can be outlined as follows: Let x be a point in G/Γ, and U a connected unipotent subgroup of G. Denote by X the closure of Ux and consider a minimal closed U-invariant subset Y of X. Suppose that Ux is not closed (equivalently X is not equal to Ux). Then X should contain "many" translations of Y by elements from the normalizer $N(U)$ of U not belonging to U. After that one can try to prove that X contains orbits of bigger and bigger unipotent subgroups until one reaches horospherical subgroups. The basic tool in this strategy is the following fact. Let y be a point in X, and let g_n be a sequence of elements in G such that g_n converges to 1, g_n does not belong to $N(U)$, and $y_n = g_n y$ belongs to X. Then X contains AY where A is a nontrivial connected subset in $N(U)$ containing 1 and "transversal" to U. To prove this one has to observe that the orbits Uy_n and Uy are "almost parallel" in the direction of $N(U)$ most of the time in "the intermediate range". (cf. Proposition 3.6).

In fact the set AU as a subset of $N(U)/U$ is the image of a nontrivial rational map from U into $N(U)/U$. Moreover this rational map sends 1 to 1 and also comes from a polynomial map from U into the closure of G/U in the affine space V containing G/U. This affine space V is the space of the rational representation of G such that V contains a vector the stabilizer of which is U (Chevalley theorem).

This program was being actively pursued at the time Ratner's results were announced (cf. [Sha3]).

7.2 A quantitative version of the Oppenheim Conjecture.

References for this subsection are [EMM1] and [EMM2].

In this section we study some finer questions related to the distribution the values of Q at integral points.

Let ν be a continuous positive function on the sphere $\{v \in \mathbb{R}^n \mid \|v\| = 1\}$, and let $\Omega = \{v \in \mathbb{R}^n \mid \|v\| < \nu(v/\|v\|)\}$. We denote by $T\Omega$ the dilate of Ω by T. Define the following set:

$$V^Q_{(a,b)}(\mathbb{R}) = \{x \in \mathbb{R}^n \mid a < Q(x) < b\}$$

We shall use $V_{(a,b)} = V^Q_{(a,b)}$ when there is no confusion about the form Q. Also let $V_{(a,b)}(\mathbb{Z}) = V^Q_{(a,b)}(\mathbb{Z}) = \{x \in \mathbb{Z}^n \mid a < Q(x) < b\}$. The set $T\Omega \cap \mathbb{Z}^n$ consists of $O(T^n)$ points, $Q(T\Omega \cap \mathbb{Z}^n)$ is contained in an interval of the form $[-\mu T^2, \mu T^2]$, where $\mu > 0$ is a constant depending on Q and Ω. Thus one might expect that for any interval $[a, b]$, as $T \to \infty$,

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| \sim c_{Q,\Omega}(b-a)T^{n-2} \tag{39}$$

where $c_{Q,\Omega}$ is a constant depending on Q and Ω. This may be interpreted as "uniform distribution" of the sets $Q(\mathbb{Z}^n \cap T\Omega)$ in the real line. The main result of this section is that (39) holds if Q is not proportional to a rational form, and has signature (p, q) with $p \geq 3$, $q \geq 1$. We also determine the constant $c_{Q,\Omega}$.

If Q is an indefinite quadratic form in n variables, Ω is as above and (a, b) is an interval, we show that there exists a constant $\lambda = \lambda_{Q,\Omega}$ so that as $T \to \infty$,

$$\mathrm{Vol}(V_{(a,b)}(\mathbb{R}) \cap T\Omega) \sim \lambda_{Q,\Omega}(b-a)T^{n-2} \tag{40}$$

The main result is the following:

Theorem 7.2. *Let Q be an indefinite quadratic form of signature (p, q), with $p \geq 3$ and $q \geq 1$. Suppose Q is not proportional to a rational form. Then for any interval (a, b), as $T \to \infty$,*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| \sim \lambda_{Q,\Omega}(b-a)T^{n-2} \tag{41}$$

where $n = p + q$, and $\lambda_{Q,\Omega}$ is as in (40).

The asymptotically exact lower bound was proved in [DM4]. Also a lower bound with a smaller constant was obtained independently by M. Ratner, and by S. G. Dani jointly with S. Mozes (both unpublished). The upper bound was proved in [EMM1].

If the signature of Q is $(2, 1)$ or $(2, 2)$ then no universal formula like (39) holds. In fact, we have the following theorem:

Theorem 7.3. *Let Ω_0 be the unit ball, and let $q = 1$ or 2. Then for every $\epsilon > 0$ and every interval (a, b) there exists a quadratic form Q of signature $(2, q)$ not proportional to a rational form, and a constant $c > 0$ such that for an infinite sequence $T_j \to \infty$,*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega_0| > cT_j^q(\log T_j)^{1-\epsilon}.$$

The case $q=1$, $b \leq 0$ of Theorem 7.3 was noticed by P. Sarnak and worked out in detail in [Bre]. The quadratic forms constructed are of the form $x_1^2+x_2^2-\alpha x_3^2$, or $x_1^2+x_2^2-\alpha(x_3^2+x_4^2)$, where α is extremely well approximated by squares of rational numbers.

However in the $(2,1)$ and $(2,2)$ cases, one can still establish an upper bound of the form $cT^q \log T$. This upper bound is effective, and is uniform over compact sets in the set of quadratic forms. We also give an effective uniform upper bound for the case $p \geq 3$.

Theorem 7.4 ([EMM1])**.** *Let $\mathcal{O}(p,q)$ denote the space of quadratic forms of signature (p,q) and discriminant ± 1, let $n = p+q$, (a,b) be an interval, and let $\mathcal{D}$ be a compact subset of $\mathcal{O}(p,q)$. Let ν be a continuous positive function on the unit sphere and let $\Omega = \{v \in \mathbb{R}^n \mid \|v\| < \nu(v/\|v\|)\}$. Then, if $p \geq 3$ there exists a constant c depending only on $\mathcal{D}$, (a,b) and Ω such that for any $Q \in \mathcal{D}$ and all $T > 1$,*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| < cT^{n-2}$$

If $p=2$ and $q=1$ or $q=2$, then there exists a constant $c>0$ depending only on $\mathcal{D}$, (a,b) and Ω such that for any $Q \in \mathcal{D}$ and all $T>2$,

$$|V_{(a,b)} \cap T\Omega \cap \mathbb{Z}^n| < cT^{n-2}\log T$$

Also, for the $(2,1)$ and $(2,2)$ cases, we have the following "almost everywhere" result:

Theorem 7.5. *For almost all quadratic forms Q of signature $(p,q)=(2,1)$ or $(2,2)$*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| \sim \lambda_{Q,\Omega}(b-a)T^{n-2}$$

where $n=p+q$, and $\lambda_{Q,\Omega}$ is as in (40).

Theorem 7.5 may be proved using a recent general result of Nevo and Stein [NS]; see also [EMM1].

It is also possible to give a "uniform" version of Theorem 7.2, following [DM4]:

Theorem 7.6. *Let $\mathcal{D}$ be a compact subset of $\mathcal{O}(p,q)$, with $p \geq 3$. Let $n=p+q$, and let Ω be as in Theorem 7.4. Then for every interval $[a,b]$ and every $\theta>0$, there exists a finite subset $\mathcal{P}$ of $\mathcal{D}$ such that each $Q \in \mathcal{P}$ is a scalar multiple of a rational form and for any compact subset $\mathcal{F}$ of $\mathcal{D}-\mathcal{P}$ there exists T_0 such that for all Q in $\mathcal{F}$ and $T \geq T_0$,*

$$(1-\theta)\lambda_{Q,\Omega}(b-a)T^{n-2} \leq |V_{(a,b)}(\mathbb{Z}) \cap T\Omega| \leq (1+\theta)\lambda_{Q,\Omega}(b-a)T^{n-2}$$

where $\lambda_{Q,\Omega}$ is as in (40).

As in Theorem 7.2 the upper bound is from [EMM1]; the asymptotically exact lower bound, which holds even for $SO(2,1)$ and $SO(2,2)$, was proved in [DM4].

Remark 7.7. If we consider $|V_{(a,b)}(\mathbb{R}) \cap T\Omega \cap \mathcal{P}(\mathbb{Z}^n)|$ instead of $|V_{(a,b)}(\mathbb{Z}) \cap T\Omega|$ (where $\mathcal{P}(\mathbb{Z}^n)$ denotes the set of primitive lattice points, then Theorem 7.2 and Theorem 7.6 hold provided one replaces $\lambda_{Q,\Omega}$ by $\lambda'_{Q,\Omega} = \lambda_{Q,\Omega}/\zeta(n)$, where ζ is the Riemann zeta function.

More on signature (2,2). Recall that a subspace is called isotropic if the restriction of the quadratic form to the subspace is identically zero. Observe also that whenever a form of signature $(2,2)$ has a rational isotropic subspace L then $L \cap T\Omega$ contains on the order of T^2 integral points x for which $Q(x) = 0$, hence $N_{Q,\Omega}(-\epsilon, \epsilon, T) \geq cT^2$, independently of the choice of ϵ. Thus to obtain an asymptotic formula similar to (41) in the signature $(2,2)$ case, we must exclude the contribution of the rational isotropic subspaces. We remark that an irrational quadratic form of signature $(2,2)$ may have at most 4 rational isotropic subspaces (see [EMM2, Lemma 10.3]).

The space of quadratic forms in 4 variables is a linear space of dimension 10. Fix a norm $\|\cdot\|$ on this space.

Definition 7.8. (EWAS) A quadratic form Q is called *extremely well approximable by split forms (EWAS)* if for any $N > 0$ there exists a split integral form Q' and $2 \leq k \in \mathbb{R}$ such that

$$\left\| Q - \frac{1}{k} Q' \right\| \leq \frac{1}{k^N}.$$

The main result of [EMM2] is:

Theorem 7.9. *Suppose Ω is as above. Let Q be an indefinite quadratic form of signature $(2,2)$ which is not EWAS. Then for any interval (a,b), as $T \to \infty$,*

$$\tilde{N}_{Q,\Omega}(a,b,T) \sim \lambda_{Q,\Omega}(b-a)T^2, \tag{42}$$

where the constant $\lambda_{Q,\Omega}$ is as in (40), and $\tilde{N}_{Q,\Omega}$ counts the points not contained in isotropic subspaces.

Open Problem. State and prove a result similar to Theorem 7.9 for the signature $(2,1)$ case.

Eigenvalue spacings on flat 2-tori

It has been suggested by Berry and Tabor that the eigenvalues of the quantization of a completely integrable Hamiltonian follow the statistics of a Poisson point-process, which means their consecutive spacings should be i.i.d. exponentially distributed. For the Hamiltonian which is the geodesic flow on the flat 2-torus, it was noted by P. Sarnak [Sar] that this problem translates to one of the spacing between the values at integers of a binary quadratic form, and is related to the quantitative Oppenheim problem in the signature $(2, 2)$ case. We briefly recall the connection following [Sar].

Let $\Delta \subset \mathbb{R}^2$ be a lattice and let $M = \mathbb{R}^2/\Delta$ denote the associated flat torus. The eigenfunctions of the Laplacian on M are of the form $f_v(\cdot) = e^{2\pi i\langle v,\cdot\rangle}$, where v belongs to the dual lattice Δ^*. The corresponding eigenvalues are $4\pi^2\|v\|^2$, $v \in \Delta^*$. These are the values at integral points of the binary quadratic $B(m, n) = 4\pi^2\|mv_1 + nv_2\|^2$, where $\{v_1, v_2\}$ is a $\mathbb{Z}$-basis for Δ^*. We will identify Δ^* with $\mathbb{Z}^2$ using this basis.

We label the eigenvalues (with multiplicity) by

$$0 = \lambda_0(M) < \lambda_1(M) \le \lambda_2(M) \dots$$

It is easy to see that Weyl's law holds, i.e.

$$|\{j \;\;:\;\; \lambda_j(M) \le T\}| \sim c_M T,$$

where $c_M = (\operatorname{area} M)/(4\pi)$. We are interested in the distribution of the local spacings $\lambda_j(M) - \lambda_k(M)$. In particular, for $0 \notin (a, b)$, set

$$R_M(a, b, T) = \frac{|\{(j, k) \;\;:\;\; \lambda_j(M) \le T, \lambda_k(M) \le T, a \le \lambda_j(M) - \lambda_k(M) \le b\}|}{T}.$$

The statistic R_M is called the pair correlation. The Poisson-random model predicts, in particular, that

$$\lim_{T\to\infty} R_M(a, b, T) = c_M^2(b - a). \tag{43}$$

Note that the differences $\lambda_j(M) - \lambda_k(M)$ are precisely the integral values of the quadratic form $Q_M(x_1, x_2, x_3, x_4) = B(x_1, x_2) - B(x_3, x_4)$.

P. Sarnak showed in [Sar] a that (43) holds on a set of full measure in the space of tori. Some remarkable related results for forms of higher degree and higher dimensional tori were proved in [V1], [V2] and [V3]. These methods, however, cannot be used to explicitly construct a specific torus for which (43) holds. A corollary of Theorem 7.9 is the following:

Theorem 7.10. *Let M be a 2 dimensional flat torus rescaled so that one of the coefficients in the associated binary quadratic form B is 1. Let A_1, A_2 denote the two other coefficients of B. Suppose that there exists $N > 0$ such that for all triples of integers (p_1, p_2, q) with $q \geq 2$,*

$$\max_{i=1,2} \left| A_i - \frac{p_i}{q} \right| > \frac{1}{q^N}.$$

Then, for any interval (a, b) not containing 0, (43) holds, i.e.

$$\lim_{T \to \infty} R_M(a, b, T) = c_M^2 (b - a).$$

In particular, the set of $(A_1, A_2) \subset \mathbb{R}^2$ for which (43) does not hold has zero Hausdorff dimension.

Thus, if one of the A_i is Diophantine's (e.g. algebraic), then M has a spectrum whose pair correlation satisfies the Berry-Tabor conjecture.

This establishes the pair correlation for the flat torus or "boxed oscillator" considered numerically by Berry and Tabor. We note that without some diophantine condition, (43) may fail.

7.3 Passage to the space of lattices.

We now relate the counting problem of Theorem 7.2 to a certain integral expression involving the orthogonal group of the quadratic form and the space of lattices $SL(n, \mathbb{R})/SL(n, \mathbb{Z})$. Roughly this is done as follows. Let f be a bounded function on $\mathbb{R}^n - \{0\}$ vanishing outside a compact subset. For a lattice $\Delta \in \mathcal{L}_n$ let

$$\tilde{f}(\Delta) = \sum_{v \in \Delta \setminus \{0\}} f(\Delta) \tag{44}$$

(the function $\tilde{f}$ is called the "Siegel Transform" of f). The proof is based on the identity of the form

$$\int_K \tilde{f}(a_t k \Delta)\, dk = \sum_{v \in \Delta \setminus \{0\}} \int_K f(a_t k v)\, dk \tag{45}$$

obtained by integrating (44). In (45) $\{a_t\}$ is a certain diagonal subgroup of the orthogonal group of Q, and K is a maximal compact subgroup of the orthogonal group of Q. Then for an appropriate function f, the right hand side is then related to the number of lattice points $v \in [e^t/2, e^t]\partial\Omega$ with $a < Q(v) < b$. The asymptotics of the left-hand side is then established using the ergodic theory of unipotent flows and some other techniques.

Quadratic Forms, and the lattice Δ_Q. Let $n \geq 3$, and let $p \geq 2$. We denote $n-p$ by q, and assume $q > 0$. Let $\{e_1, e_2, \dots e_n\}$ be the standard basis of $\mathbb{R}^n$. Let Q_0 be the quadratic form defined by

$$Q_0\left(\sum_{i=1}^{n} v_i e_i\right) = 2v_1 v_n + \sum_{i=2}^{p} v_i^2 - \sum_{i=p+1}^{n-1} v_i^2 \quad \text{for all } v_1, \dots, v_n \in \mathbb{R}. \tag{46}$$

It is straightforward to verify that Q_0 has signature (p, q). Let $G = SL(n, \mathbb{R})$, the group of $n \times n$ matrices of determinant 1. For each quadratic form Q and $g \in G$, let Q^g denote the quadratic form defined by $Q^g(v) = Q(gv)$ for all $v \in \mathbb{R}^n$. By the well known classification of quadratic forms over $\mathbb{R}$, for each $Q \in \mathcal{O}(p, q)$ there exists $g \in G$ such that $Q = Q_0^g$. Then let Δ_Q denote the lattice $g\mathbb{Z}^n$, so that $Q_0(\Delta_Q) = Q(\mathbb{Z}^n)$.

For any quadratic form Q let $SO(Q)$ denote the special orthogonal group corresponding to Q; namely $\{g \in G \mid Q^g = Q\}$. Let $H = SO(Q_0)$. Then the map $H\backslash G \to \mathcal{O}(p, q)$ given by $Hg \to Q_0^g$ is a homeomorphism.

The map a_t and the group K. For $t \in \mathbb{R}$, let a_t be the linear map so that $a_t e_1 = e^{-t} e_1$, $a_t e_n = e^t e_n$, and $a_t e_i = e_i$, $2 \leq i \leq n-1$. Then the one-parameter group $\{a_t\}$ is contained in H. Let $\hat{K}$ be the subgroup of G consisting of orthogonal matrices, and let $K = H \cap \hat{K}$. It is easy to check that K is a maximal compact subgroup of H, and consists of all $h \in H$ leaving invariant the subspace spanned by $\{e_1 + e_n, e_2, \dots, e_p\}$. We denote by m the normalized Haar measure on K.

A Lemma about vectors in $\mathbb{R}^n$. In this section we will be somewhat informal. For a completely rigorous argument see [EMM1, §§3.4-3.5]. Also for simplicity we let $\nu = 1$ in this section.

Let $W \subset \mathbb{R}^n$ be the characteristic function of the region defined by the inequalities on $x = (x_1, \dots, x_n)$:

$$a \leq Q_0(x) \leq b, \quad (1/2) \leq \|x\| \leq 2,$$
$$x_1 > 0, \quad (1/2)x_1 \leq |x_i| \leq (1/2)x_1 \text{ for } 2 \leq i \leq n-1.$$

Let f be the characteristic function of W.

Lemma 7.11. *There exists $T_0 > 0$ such that for every t with $e^t > T_0$, and every $v \in \mathbb{R}^n$ with $\|v\| > T_0$,*

$$c_{p,q} e^{(n-2)t} \int_K f(a_t k v)\, dm(k) \approx \begin{cases} 1 \text{ if } a \leq Q_0(x) \leq b \text{ and } \frac{e^t}{2} \leq \|v\| \leq e^t, \\ 0 \text{ otherwise} \end{cases} \tag{47}$$

where $c_{p,q}$ is a constant depending only on p and q.

Proof. This is a direct calculation. □

Remark. The $\approx$ in (47) is essentially equality up to "edge effects". These edge effects can be overcome if one approximated f from above and below by continuous functions f_+ and f_- in such a way that the L^1 norm of $f_+ - f_-$ is small. We choose not to do this here in order to not clutter the notation.

In (47), we let $T = e^t$ and sum over $v \in \Delta_Q$. We obtain:

Proposition 7.12. *As $T \to \infty$,*

$$c_{p,q} T^{n-2} \int_K \tilde{f}(a_t k \Delta_Q) \approx |\{v \in \Delta_Q \ : \ a < Q_0(v) < b \text{ and } \tfrac{1}{2}T \leq \|v\| \leq T\}|,$$

where $t = \log T$. Note that the right-hand side is by definition $|V^Q_{(a,b)}(\mathbb{Z}) \cap [T/2, T]\Omega_0|$, where Ω_0 is the unit ball.

We also note without proof the following lemma:

Lemma 7.13. *Let ρ be a continuous positive function on the sphere, and let $\Omega = \{v \in \mathbb{R}^n \mid \|v\| < \rho(v/\|v\|)\}$. Then there exists a constant $\lambda = \lambda_{Q,\Omega}$ so that as $T \to \infty$,*

$$\mathrm{Vol}(V^Q_{(a,b)}(\mathbb{R}) \cap T\Omega) \sim \lambda_{Q,\Omega}(b-a)T^{n-2}.$$

Also (using Siegel's formula), $c_{p,q} \int_{\mathcal{L}_n} \tilde{f} = c_{p,q} \int_{\mathbb{R}^n} f = (1 - 2^{2-n})\lambda_{Q,\Omega}$.

Remark. One can verify that:

$$\lambda_{Q,\Omega} = \int_{L \cap \Omega} \frac{dA}{\|\nabla Q\|},$$

where L is the lightcone $Q = 0$ and dA is the area element on L.

The main theorems. In view of Proposition 7.12 and Lemma 7.13, to prove Theorem 7.2 one may use the following theorem:

Theorem 7.14. *Suppose $p \geq 3$, $q \geq 1$. Let $\Lambda \in \mathcal{L}_n$ be a unimodular lattice such that $H\Lambda$ is not closed. Let ν be any continuous function on K. Then*

$$\lim_{t\to+\infty} \int_K \tilde{f}(a_t k \Lambda)\nu(k)\, dm(k) = \int_K \nu\, dm \int_{\mathcal{L}_n} \tilde{f}(\Delta)\, d\mu(\Delta). \tag{48}$$

To prove Theorem 7.6 we use the following generalization:

Theorem 7.15. *Suppose $p \geq 3$, $q \geq 1$. Let ν be as in Theorem 7.14, and let $\mathcal{C}$ be any compact set in $\mathcal{L}_n$. Then for any $\epsilon > 0$ there exist finitely many points $\Lambda_1, \ldots, \Lambda_\ell \in \mathcal{L}_n$ such that*

(i) *The orbits $H\Lambda_1, \ldots, H\Lambda_\ell$ are closed and have finite H-invariant measure.*

(ii) *For any compact subset F of $\mathcal{C} \setminus \bigcup_{1\leq i\leq \ell} H\Lambda_i$, there exists $t_0 > 0$, so that for all $\Lambda \in F$ and $t > t_0$,*

$$\left| \int_K \tilde{f}(a_t k\Lambda)\nu(k)\, dm(k) - \int_{\mathcal{L}_n} \tilde{f}\, d\mu \int_K \nu\, dm \right| \leq \epsilon \tag{49}$$

Theorem 7.14 and Theorem 7.15 if $\tilde{f}$ is replaced by a bounded function ϕ. If we replace $\tilde{f}$ by a bounded continuous function ϕ then (48) and (49) follow easily from Theorem 6.4. (This was the original motivation for Theorem 6.4). The fact that Theorem 6.4 deals with unipotents and Theorem 7.15 deals with large spheres is not a serious obstacle, since large spheres can be approximated by unipotents. In fact, the integral in (48) can be rewritten as

$$\int_B \left(\frac{1}{T(x)} \int_0^{T(x)} \phi(u_t x)\, dm(k) \right) dx,$$

where B is a suitable subset of G and U is a suitable unipotent. Now by Theorem 6.4, the inner integral tends to $\int_{G/\Gamma} \phi$ uniformly as long as x is in a compact set away from an explicitly described set E, where E is a finite union of neighborhoods of sets of the form $\pi(C)$ where C is a compact subset of some $N(F, U)$. By direct calculation one can show that only a small part of B is near E, hence Theorem 7.14 and Theorem 7.15 both hold.

Remark. Both Theorem 6.4 and Ratner's uniform distribution theorem Theorem 1.12 hold for bounded continuous functions, but not for arbitrary continuous functions from $L^1(G/\Gamma)$. However, for a non-negative bounded continuous function

f on $\mathbb{R}^n$, the function $\tilde{f}$ defined in (44) is non-negative, continuous, and L^1 but unbounded (it is in $L^s(G/\Gamma)$ for $1 \leq s < n$, where $G = SL(n, \mathbb{R})$, and $\Gamma = SL(n, \mathbb{Z})$).

The lower bounds. As it was done in [DM4] it is possible to obtain asymptotically exact lower bounds by considering bounded continuous functions $\phi \leq \tilde{f}$. However, to prove the upper bounds in the theorems stated above we need to examine carefully the situation at the "cusp" of G/Γ, i.e outside of compact sets. This will be done in the next section.

8 Quantitative Oppenheim (upper bounds)

The references for this section are [EMM1] and [EMM2].

Lattices. Let Δ be a lattice in $\mathbb{R}^n$. We say that a subspace L of $\mathbb{R}^n$ is Δ-*rational* if $L \cap \Delta$ is a lattice in L. For any Δ-rational subspace L, we denote by $d_\Delta(L)$ or simply by $d(L)$ the volume of $L/(L \cap \Delta)$. Let us note that $d(L)$ is equal to the norm of $e_1 \wedge \cdots \wedge e_\ell$ in the exterior power $\bigwedge^\ell(\mathbb{R}^n)$ where $\ell = \dim L$ and $(e_1, \cdots, e_\ell)$ is a basis over $\mathbb{Z}$ of $L \cap \Delta$. If $L = \{0\}$ we write $d(L) = 1$. A lattice is Δ *unimodular* if $d_\Delta(\mathbb{R}^n) = 1$. The space of unimodular lattices is isomorphic to $SL(n, \mathbb{R})/SL(n, \mathbb{Z})$.

Let us introduce the following notation:

$$\alpha_i(\Delta) = \sup\left\{ \frac{1}{d(L)} \,\middle|\, L \text{ is a } \Delta\text{-rational subspace of dimension } i \right\}, \quad 0 \leq i \leq n,$$
$$\alpha(\Delta) = \max_{0 \leq i \leq n} \alpha_i(\Delta). \tag{50}$$

The following lemma is known as the "Lipshitz Principle":

Lemma 8.1 ([Sch, Lemma 2])**.** *Let f be a bounded function on $\mathbb{R}^n$ vanishing outside a compact subset. Then there exists a positive constant $c = c(f)$ such that*

$$\tilde{f}(\Delta) < c\alpha(\Delta)$$

for any lattice Δ in $\mathbb{R}^n$. Here $\tilde{f}$ is the function on the space of lattices defined in (44).

Replacing $\tilde{f}$ by α. By Lemma 8.1, the function $\tilde{f}(g)$ on the space of unimodular lattices $\mathcal{L}_n$ is majorized by the function $\alpha(g)$. The function α is more convenient since it is invariant under the left action of the maximal compact subgroup $\hat{K}$ of G, and its growth rate at infinity is known explicitly. Theorems 7.2 and 7.6 are proved by combining Theorem 6.4 with the following integrability estimate:

Theorem 8.2 ([EMM1]). *If $p \geq 3$, $q \geq 1$ and $0 < s < 2$, or if $p = 2$, $q \geq 1$ and $0 < s < 1$, then for any lattice Δ in $\mathbb{R}^n$*

$$\sup_{t>0} \int_K \alpha(a_t k \Delta)^s \, dm(k) < \infty.$$

The upper bound is uniform as Δ varies over compact sets in the space of lattices.

This result can be interpreted as follows. For a lattice Δ in $\mathcal{L}_n$ and for $h \in H$, let $f(h) = \alpha(h\Delta)$. Since α is left-$\hat{K}$ invariant, f is a function on the symmetric space $X = K \backslash H$. Theorem 8.2 is the statement that if if $p \geq 3$, then the averages of f^s, $0 < s < 2$ over the sets $K a_t K$ in X remain bounded as $t \to \infty$, and the bound is uniform as one varies the base point Δ over compact sets. We remark that in the case $q = 1$, the rank of X is 1, and the sets $K a_t K$ are metric spheres of radius t, centered at the origin.

If $(p, q) = (2, 1)$ or $(2, 2)$, Theorem 8.2 does not hold even for $s = 1$. The following result is, in general, best possible:

Theorem 8.3 ([EMM1]). *If $p = 2$ and $q = 2$, or if $p = 2$ and $q = 1$, then for any lattice Δ in $\mathbb{R}^n$,*

$$\sup_{t>1} \frac{1}{t} \int_K \alpha(a_t k \Delta) \, dm(k) < \infty, \tag{51}$$

The upper bound is uniform as Δ varies over compact sets in the space of lattices.

In this section we prove Theorem 8.2 and Theorem 8.3. We recall the notation from §7: G is $SL(n, \mathbb{R})$, $\Gamma = SL(n, \mathbb{Z})$, $\hat{K} \cong SO(n)$ is a maximal compact subgroup of G, $H \cong SO(p, q) \subset G$, $K = H \cap \hat{K}$ is a maximal compact subgroup of H, and X is the symmetric space $K \backslash H$. From its definition (50), the function $\alpha(\Delta)$ is the maximum over $1 \leq i \leq n$ of $\hat{K}$ invariant functions $\alpha_i(\Delta)$. The main idea of the proof is to show that the α_i^s satisfy a certain system of integral inequalities which imply the desired bounds.

If $p \geq 3$ and $0 < s < 2$, or if $(p, q) = (2, 1)$ or $(2, 2)$ and $0 < s < 1$, we show that for any $c > 0$ there exist $t > 0$, and $\omega > 1$ so that the the functions α_i^s satisfy the following system of integral inequalities in the space of lattices:

$$A_t \alpha_i^s \leq c_i \alpha_i^s + \omega^2 \max_{0 < j \leq \min(n-i, i)} \sqrt{\alpha_{i+j}^s \alpha_{i-j}^s}, \tag{52}$$

where A_t is the averaging operator $(A_t f)(\Delta) = \int_K f(a_t k \Delta)$, and $c_i \leq c$ (Lemma 8.7). If $(p, q) = (2, 1)$ or $(2, 2)$ and $s = 1$, then (52) also holds (for suitably modified functions α_i), but some of the constants c_i cannot be made smaller than 1.

Let $f_i(h) = \alpha_i(h\Delta)$, so that each f_i is a function on the symmetric space X. When one restricts to an orbit of H, (52) becomes:

$$A_t f_i^s \leq c_i f_i^s + \omega^2 \max_{0<j\leq \min(n-i,i)} \sqrt{f_{i+j}^s f_{i-j}^s}. \tag{53}$$

If $\operatorname{rank} X = 1$, then $(A_t f)(h)$ can be interpreted as the average of f over the sphere of radius $2t$ in X, centered at h. In §8.4 we show that if the f_i satisfy (53) then for any $\epsilon > 0$, the function $f = f_{\epsilon,s} = \sum_{0\leq i\leq n} \epsilon^{i(n-i)} f_i^s$ satisfies the scalar inequality:

$$A_t f \leq cf + b, \tag{54}$$

where t, c and b are constants. This inequality is studied in §8.3. We show that if c is sufficiently small, then (54) for a fixed t together with the uniform continuity of $\log f$ imply that $(A_r f)(1)$ is bounded as a function of r, which is the conclusion of Theorem 8.2. If $c = 1$, which will occur in the $SO(2,1)$ and $SO(2,2)$ cases, then (54) implies that $(A_r f)(1)$ is growing at most linearly with the radius. In §8.4, we complete the proof of Theorem 8.2, and also prove Theorems 8.3 and 7.15.

Throughout the proof we consider the functions $\alpha(g)^s$ for $0 < s < 2$ even though for the application to quadratic forms we only need $s = 1 + \delta$. This yields a better integrability result, and is also necessary for the proof of Theorem 7.14 and Theorem 7.15.

8.1 Averages of the functions $1/d_i^s$ over spheres.

Recall that the function d_i is the norm of a certain vector in the exterior power $\bigwedge^i(\mathbb{R}^n)$. We have the following:

Proposition 8.4. *Let $\{a_t \mid t \in \mathbb{R}\}$ be a self-adjoint one-parameter subgroup of $SO(2,1)$. Let p and q be positive integers and let $0 < i < p+q$. Let*

$$F(i) = \{x_1 \wedge x_2 \wedge \cdots \wedge x_i \mid x_1, x_2, \cdots, x_i \in \mathbb{R}^{p+q}\} \subset \textstyle\bigwedge^i(\mathbb{R}^{p+q}).$$

Then, if $p \geq 3$, or if $p = 2$, $q = 2$ and $i \neq 2$, then for any s, $0 < s < 2$,

$$\lim_{t\to\infty} \sup_{v\in F(i),\ \|v\|=1} \int_K \frac{dm(k)}{\|a_t k v\|^s} = 0. \tag{55}$$

where $K = SO(p) \times SO(q)$ and $SO(2,1)$ is embedded into $SO(p,q)$. If $p = 2$ and $q = 1$, or if $p = 2$, $q = 2$ and $i = 2$, then (55) holds for any s, $0 < s < 1$.

Proof. This is a direct calculation. $\square$

The next lemma we obtain an analogous result for the case $(p,q) = (2,1)$, $s = 1$.

Lemma 8.5. *Let $H \cong SO(2,1)$ be the orthogonal group of the quadratic form $x^2 + y^2 - z^2$. Let $\{a_t \mid t \in \mathbb{R}\}$ be a self-adjoint one-parameter subgroup of H, and let $K = H \cap O(3)$ denote the maximal compact of H. We define another norm $\|\cdot\|^*$ on $\mathbb{R}^3$ by*

$$\|(x,y,z)\|^* = \max(\sqrt{x^2+y^2}, |z|). \tag{56}$$

Then, for any $v \in \mathbb{R}^3$, $v \neq 0$, and any $t > 0$,

$$\int_K \frac{dm(k)}{\|a_t k v\|^*} \leq \frac{1}{\|v\|^*}. \tag{57}$$

8.2 A system of inequalities

Lemma 8.6. *For any two Δ-rational subspaces L and M*

$$d(L)d(M) \geq d(L \cap M)d(L+M). \tag{58}$$

Proof. Let $\pi : \mathbb{R}^n \to \mathbb{R}^n/(L \cap M)$ denote the natural projection. Then $d(L) = d(\pi(L))d(L \cap M)$, $d(M) = d(\pi(M))d(L \cap M)$ and $d(L+M) = d(\pi(L+M))d(L \cap M)$. On the other hand the inequality (58) is equivalent to the inequality

$$\frac{d(L)}{d(L \cap M)} \frac{d(M)}{d(L \cap M)} \geq \frac{d(L+M)}{d(L \cap M)}.$$

Therefore replacing L, M and $L+M$ by $\pi(L)$, $\pi(M)$ and $\pi(L+M)$ we can assume that $L \cap M = \{0\}$. Let $(e_1, \cdots, e_\ell)$, $\ell = \dim L$, and $(e_{\ell+1}, \cdots, e_{\ell+m})$, $m = \dim M$, be bases in L and M respectively. Then

$$\begin{aligned} d(L)d(M) &= \|e_1 \wedge \cdots \wedge e_\ell\| \, \|e_{\ell+1} \wedge \cdots \wedge e_{\ell+m}\| \\ &\geq \|e_1 \wedge \cdots \wedge e_\ell \wedge e_{\ell+1} \wedge \cdots \wedge e_{\ell+m}\| \geq d(L+M) \end{aligned} \tag{59}$$

that proves (58) (the second inequality in (59) is true because $(L \cap \Delta) + (M \cap \Delta) \subset (L+M) \cap \Delta$. $\square$

Lemma 8.7. *Let $\{a_t \mid t \in \mathbb{R}\}$ be a self-adjoint one-parameter subgroup of $SO(2,1)$. Let p and q be positive integers, and denote $p+q$ by n. Denote $SO(p) \times SO(q)$ by K. Suppose $p \geq 3$, $q \geq 1$ and $0 < i < n$, or $p = 2$, $q = 2$ and $i = 1$ or 3. Then for any s, $0 < s < 2$, and any $c > 0$ there exist $t > 0$ and $\omega > 1$ such that for any lattice Λ in $\mathbb{R}^n$*

$$\int_K \alpha_i(a_t k \Lambda)^s \, dm(k) < \frac{c}{2}\alpha_i(\Lambda)^s + \omega^2 \max_{0<j\leq \min\{n-i,i\}} \left(\sqrt{\alpha_{i+j}(\Lambda)\alpha_{i-j}(\Lambda)} \right)^s. \qquad (60)$$

If $p = 2$, $q = 1$ and $i = 1, 2$, or if $p = 2$, $q = 2$ and $i = 2$, then for any s, $0 < s < 1$, and any $c > 0$ there exist $t > 0$ and $\omega > 1$ such that (60) holds.

Proof. Fix $c > 0$. In view of Proposition 8.4 one can find $t > 0$ such that

$$\int_K \frac{dm(k)}{\|a_t k v\|^s} < \frac{c}{2},$$

whenever $v \in F(i), \|v\| = 1$. It follows that

$$\int_K \frac{dm(k)}{\|a_t k v\|^s} < \frac{c}{2} \cdot \frac{1}{\|v\|^s}, \qquad (61)$$

for any $v \in F(i), v \neq 0$. Let Λ be a lattice in $\mathbb{R}^n$. There exists a Λ-rational subspace L_i of dimension i such that

$$\frac{1}{d_\Lambda(L_i)} = \alpha_i(\Lambda). \qquad (62)$$

The inequality (61) implies

$$\int_K \frac{dm(k)}{d_{a_t k \Lambda}(a_t k L_i)^s} < \frac{c}{2} \frac{1}{d_\Lambda(L_i)^s}. \qquad (63)$$

Let $\omega = \max_{0<j<n} \| \bigwedge^j (a_t) \|$. (In fact $\omega = e^t$). We have that

$$\omega^{-1} \leq \frac{\|a_t v\|}{\|v\|} \leq \omega, \; 0 < j < n, \; v \in F(j). \qquad (64)$$

Let us denote the set of Λ-rational subspaces L of dimension i with $d_\Lambda(L) < \omega^2 d_\Lambda(L_i)$ by Ψ_i. We get from (64) that for a Λ-rational i-dimensional subspace $L \notin \Psi_i$

$$d_{a_t k \Lambda}(a_t k L) > d_{a_t k \Lambda}(a_t k L_i), \qquad k \in K. \qquad (65)$$

It follows from (63), (65) and the definition of α_i that

$$\int_K \alpha_i(a_t k\Lambda)^s \, dm(k) < \frac{c}{2}\alpha_i(\Lambda)^s \text{ if } \Psi_i = \{L_i\}. \tag{66}$$

Assume now that $\Psi_i \neq \{L_i\}$. Let $M \in \Psi_i,\ M \neq L_i$. Then $\dim(M+L_i) = i+j,\ j > 0$. Now using (62), (64) and Lemma 8.6 we get that for any $k \in K$

$$\begin{aligned} \alpha_i(a_t k\Lambda) < \omega\alpha_i(\Lambda) = \frac{\omega}{d_\Lambda(L_i)} &< \frac{\omega^2}{\sqrt{d_\Lambda(L_i)d_\Lambda(M)}} \\ &\leq \frac{\omega^2}{\sqrt{d_\Lambda(L_i \cap M)d_\Lambda(L_i + M)}} \\ &\leq \omega^2\sqrt{\alpha_{i+j}(\Lambda)\alpha_{i-j}(\Lambda)}. \end{aligned} \tag{67}$$

Hence if $\Psi_i \neq \{L_i\}$

$$\int_K \alpha_i(a_t k\Lambda)^s \, dm(k) \leq \omega^2 \max_{0<j\leq\min\{n-i,i\}} \left(\sqrt{\alpha_{i+j}(\Lambda)\alpha_{i-j}(\Lambda)}\,\right)^s. \tag{68}$$

Combining (66) and (68) we get that for any lattice $\Lambda \subset \mathbb{R}^n$, (60) holds. $\square$

In the rest of this subsection we obtain similar systems of inequalities for the $SO(2,1)$ and $SO(2,2)$ cases, with $s = 1$. For $H = SO(2,1)$, Δ a lattice in $\mathbb{R}^3$, and L a Δ-rational subspace of $\mathbb{R}^3$, let $d^*_\Delta(L) = \|e_1 \wedge \dots e_\ell\|^*$ where $(e_1, \dots e_l)$ is a basis for $\Delta \cap L$. (The norm $\|\cdot\|^*$ defined in (56) on $\mathbb{R}^3 = \bigwedge^1(\mathbb{R}^3)$ can be extended to $\bigwedge^2(\mathbb{R}^3)$ by duality.) For $1 \leq i \leq 2$, let

$$\alpha^*_i(\Delta) = \sup\left\{\frac{1}{d^*_\Delta(L)} \middle| L \text{ is a } \Delta\text{-rational subspace of dimension } i\right\}. \tag{69}$$

Clearly for any Δ,

$$(1/2)\alpha_i(\Delta) < \alpha^*_i(\Delta) < 2\alpha_i(\Delta). \tag{70}$$

Lemma 8.8. *Let $\{a_t \mid t \in \mathbb{R}\}$ be a self-adjoint one-parameter subgroup of $H = SO(2,1)$, and denote $SO(2)$ by K. Then there exist $t_0 > 0$ and $\omega > 1$, such that for any $t < t_0$, for any unimodular lattice Λ in $\mathbb{R}^3$, and $1 \leq i \leq 2$,*

$$\int_K \alpha^*_i(a_t k\Lambda) \, dm(k) < \alpha^*_i(\Lambda) + \omega^2\sqrt{\alpha_{3-i}(\Lambda)}. \tag{71}$$

Proof. The argument is identical to the proof of Lemma 8.7 except that one uses Lemma 8.5 instead of Proposition 8.4. $\square$

Now let $H = SO(2,2)$. The space $V = \bigwedge^2(\mathbb{R}^4)$ splits as a direct sum $V_1 \oplus V_2$ of two invariant subspaces, where on each V_i, H preserves a quadratic form Q_i of signature $(2,1)$. We define on each V_i a Euclidean norm $\|\cdot\|_i^*$ by (56) (adapted to Q_i). Let π_i denote the orthogonal projections from V to V_i. Now let Δ be a lattice in $\mathbb{R}^4$, and let L be a two-dimensional Δ-rational subspace of $\mathbb{R}^4$. For $1 \le i \le 2$, let

$$d_\Delta^{i,\#}(L) = \|\pi_i(e_1 \wedge e_2)\|_i^*, \tag{72}$$

where $\{e_1, e_2\}$ is a basis over $\mathbb{Z}$ for $\Delta \cap L$. Then let

$$\alpha_2^\#(\Delta) = \sup_L \left\{ \min\left(\frac{1}{d_\Delta^{1,\#}(L)}, \frac{1}{d_\Delta^{2,\#}(L)} \right) \right\}. \tag{73}$$

The supremum is taken over Δ-rational two dimensional subspaces L. By construction, for any Δ,

$$C^{-1}\alpha_2^\#(\Delta) < \alpha_2(\Delta) < C\alpha_2^\#(\Delta), \tag{74}$$

where C is an absolute constant.

Lemma 8.9. *Let $\{a_t \mid t \in \mathbb{R}\}$ be a self-adjoint one-parameter subgroup of $SO(2,1)$, where $SO(2,1)$ is diagonally embedded in $H = SO(2,2)$, under its local identification with $SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$. Denote $SO(2) \times SO(2)$ by K, and the maximal compact of $SO(2,1)$ by $\tilde{K}$. Then there exist $t_0 > 0$ and $\omega > 1$, such that for any $t < t_0$ and for any unimodular lattice Λ in $\mathbb{R}^4$,*

$$\int_{\tilde{K}} \alpha_2^\#(a_t \tilde{k} \Lambda)\, dm(\tilde{k}) < \alpha_2^\#(\Lambda) + \omega^2 \sqrt{\alpha_1(\Lambda)\alpha_3(\Lambda)}. \tag{75}$$

Proof. The group $\tilde{K}$ is diagonally embedded in K. Recall that each $SO(2,2)$ invariant subspace V_i of $\bigwedge^2(\mathbb{R}^4)$ is fixed pointwise by one of the $SL(2,\mathbb{R})$ factors, while the other fixes a quadratic form of signature $(2,1)$. Thus, for $1 \le i \le 2$, the inequalities:

$$\int_{\tilde{K}} \frac{dm(\tilde{k})}{\|\pi_i(a_t \tilde{k} v)\|_i^*} \le \frac{1}{\|\pi_i(v)\|_i^*} \tag{76}$$

follow immediately from Lemma 8.5. Hence,

$$\begin{aligned}\int_{\tilde{K}} \min\left(\frac{1}{\|\pi_1(a_t \tilde{k} v)\|_1^*}, \frac{1}{\|\pi_2(a_t \tilde{k} v)\|_2^*} \right) dm(k) &\le \min\left(\int_{\tilde{K}} \frac{dm(\tilde{k})}{\|\pi_1(a_t \tilde{k} v)\|_1^*}, \int_{\tilde{K}} \frac{dm(\tilde{k})}{\|\pi_2(a_t \tilde{k} v)\|_2^*} \right) \\ &\le \min\left(\frac{1}{\|\pi_1(v)\|_1^*}, \frac{1}{\|\pi_2(v)\|_2^*} \right). \end{aligned} \tag{77}$$

The rest of the proof is identical to that of Lemma 8.7 except that (77) is used in place of Proposition 8.4. □

8.3 Coarsely Superharmonic Functions.

Let $n \in \mathbb{N}^+$ and let D_n^+ denote the set of diagonal matrices $d(\lambda_1, \cdots, \lambda_n) \in GL_n(\mathbb{R})$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$. For any $g \in GL_n(\mathbb{R})$, consider the Cartan decomposition $g = k_1(g)d(g)k_2(g), k_1(g), k_2(g) \in K = O_n(\mathbb{R}),\ d(g) \in D_n^+$ and denote by $\lambda_1(g) \geq \lambda_2(g) \geq \cdots \geq \lambda_n(g)$ the eigenvalues of $d(g)$.

Lemma 8.10. *For every $\epsilon > 0$ there exists a neighborhood U of e in $O_n(\mathbb{R})$ such that*

$$\Big|\frac{\lambda_i(d_1kd_2)}{\lambda_i(d_1)\lambda_i(d_2)} - 1\Big| < \epsilon \tag{78}$$

for any $d_1, d_2 \in D_n^+,\ k \in U$ and $1 \leq i \leq n$.

Proof. Let $(e_1, \cdots, e_n)$ be the standard orthonormal basis in $\mathbb{R}^n$. If $k \in O_n(\mathbb{R})$ and $\langle ke_1, e_1\rangle > 1 - \epsilon$ then

$$\|d_1kd_2e_1\| > (1-\epsilon)\lambda_1(d_1)\lambda_1(d_2). \tag{79}$$

On the other hand, for any $g \in GL_n(\mathbb{R})$.

$$\lambda_1(g) = \|g\| \geq \|ge_1\|. \tag{80}$$

Since $\|d_1kd_2\| \leq \|d_1\|\ \|d_2\|$ it follows from (79) and (80) that

$$1 \geq \frac{\lambda_1(d_1kd_2)}{\lambda_1(d_1)\lambda_1(d_2)} > 1 - \epsilon, \tag{81}$$

if $\langle ke_1, e_1\rangle > 1 - \epsilon$. Analogously considering the representation of $GL_n(\mathbb{R})$ in the i-th exterior product $\bigwedge^i(\mathbb{R}^n)$ of $\mathbb{R}^n$ we get that

$$1 \geq \frac{(\lambda_1\lambda_2\cdots\lambda_i)(d_1kd_2)}{(\lambda_1\lambda_2\cdots\lambda_i)(d_1d_2)} > 1 - \epsilon, \tag{82}$$

if $k \in O_n(\mathbb{R})$ and $\langle\bigwedge^i(k)(e_1\wedge\cdots\wedge e_i), e_1\wedge\cdots\wedge e_i\rangle > 1-\epsilon$. It is clear that there exists a neighborhood U of identity in $O_n(\mathbb{R})$ such that $\langle\bigwedge^i(k)(e_1\wedge\cdots\wedge e_i),\ e_1\wedge\cdots\wedge e_i\rangle > \sqrt{1-\epsilon}$ for every $k \in U$ and $1 \leq i \leq n$. But

$$\lambda_i(g) = \frac{(\lambda_1\lambda_2\cdots\lambda_i)(g)}{(\lambda_1\lambda_2\cdots\lambda_{i-1})(g)}.$$

Therefore (78) follows from (82). □

Lemma 8.11. *(cf. the "wavefront lemma" [EMc, Theorem 3.1]) Let H be a self-adjoint connected reductive subgroup of $GL_n(\mathbb{R})$, let $K = O_n(\mathbb{R}) \cap H$ be a maximal compact subgroup of H and let $\{a_t \mid t \in \mathbb{R}\}$ be a self-adjoint one-parameter subgroup of H. Then for every neighborhood V of e in H there exists a neighborhood U of e in K such that*

$$a_t U a_s \subset K V a_t a_s K \tag{83}$$

for any $t \geq 0$ and $s \geq 0$.

Proof. Conjugating a_t by an element of K we can assume that $\{a_t \mid t \geq 0\} \subset D_n^+$. It is easy to see that there exists $\epsilon > 0$ such that $h_1 \in V h_2$ whenever $h_1, h_2 \in D_n^+$ and $\left|\frac{\lambda_i(h_1)}{\lambda_i(h_2)} - 1\right| < \epsilon$ for every $1 \leq i \leq n$. Take a neighborhood U such that (78) is satisfied. Then (83) is true for this U. □

Proposition 8.12. *Let H be a self-adjoint reductive subgroup of $GL_n(\mathbb{R})$, let $K = O_n(\mathbb{R}) \cap H$, let m denote the normalized measure on K, and let $A = \{a_t \mid t \in \mathbb{R}\}$ be a self-adjoint one-parameter subgroup of H. Let $\mathcal{F}$ be a family of strictly positive functions on H having the following properties:*

(a) *The logarithms $\log f$ for $f \in \mathcal{F}$ are equicontinuous with respect to a left-invariant uniform structure on H or, equivalently, for any $\epsilon > 0$ there exists a neighborhood $V(\epsilon)$ of 1 in H such that for any $f \in \mathcal{F}$,*

$$(1 - \epsilon) f(h) < f(uh) < (1 + \epsilon) f(h)$$

for any $h \in H$ and $u \in V(\epsilon)$;

(b) *The functions $f \in \mathcal{F}$ are left K-invariant, that is $f(Kh) = f(h)$, $h \in H$,*

(c) $\sup_{f \in \mathcal{F}} f(1) < \infty$.

Then there exists $0 < c = c(\mathcal{F}) < 1$ such that for any $t > 0$ and $b > 0$ there exists $B = B(t, b) < \infty$ with the following property: If $f \in \mathcal{F}$ and

$$\int_K f(a_t k h)\, dm(k) < c f(h) + b \tag{84}$$

for any $h \in KAK \subset H$, then

$$\int_K f(a_\tau k)\, dm(k) < B$$

for any $\tau > 0$.

Proof. Fix $f \in \mathcal{F}$, and let

$$\tilde{f}(h) = \int_K f(hk)\, dm(k).$$

Properties (a), (b), (c) of the function f imply that $\tilde{f}$ has the same properties. Hence it suffices to show that the conclusion of the proposition holds for $\tilde{f}$. Therefore we can assume that

$$f(KhK) = f(h),\ h \in H, \tag{85}$$

and we have to prove that

$$\sup_{\tau>0} f(a_\tau) < B < \infty. \tag{86}$$

It follows from property (a) that

$$\frac{1}{2}f(h) < f(uh) < 2f(h),\ h \in H,\ u \in V = V(\frac{1}{2}). \tag{87}$$

According to Lemma 8.11 there exists a neighborhood U of 1 in H such that $a_t U a_\tau \in KV a_t a_\tau K$ for any $t \geq 0$ and $\tau \geq 0$. Then we get from (85) and (87) that

$$\int_K f(a_t k a_\tau)\, dm(k) \geq \int_{U\cap K} f(a_t k a_\tau)\, dm(k) > \frac{1}{2} m(U \cap K) f(a_t a_\tau). \tag{88}$$

Suppose for some $t > 0$ and $b > 0$

$$\int_K f(a_t k h)\, dm(k) < \frac{1}{4} m(U \cap K) f(h) + b,\ h \in H. \tag{89}$$

It follows from (88) and (89) that for some $b' > 0$,

$$f(a_t a_\tau) < \frac{1}{2} f(a_\tau) + b',\ \text{for all}\ \tau > 0. \tag{90}$$

Using induction on ℓ we get from (90) that

$$f(a_{\ell t}) < 2\max\{f(1), b'\},\ ; \ell \in \mathbb{N}^+. \tag{91}$$

Since $\{a_r \mid 0 \leq r \leq t\}$ belongs to V^i for some i where $V^1 = V,\ V^i = V\ V^{i-1}$, it follows that $\sup_{h\in H,\, 0\leq r\leq t} \quad \frac{f(a_r h)}{f(h)} < \infty$. Therefore (91) and property (c) imply (86). $\square$

8.4 Averages over large spheres.

In this subsection we complete the proofs of Theorem 8.2, Theorem 8.3 and Theorem 7.15.

Proof of Theorem 8.2. Define functions $f_0, f_1, \cdots, f_n$ on $H = SO(p,q)$ by the following equalities

$$f_i(h) = \alpha_i(h\Delta),\ h \in H,\ 0 \le i \le n.$$

Since $\alpha(a_t k\Delta)^s = \max_{0\le i\le n} f_i(a_t k)^s < \sum_{0\le i\le n} f_i(a_t k)^s$ it is enough to show that

$$\sup_{t>0,\ 0\le i\le n} \int_K f_i^s(a_t k)\, dm(k) < \infty. \tag{92}$$

Let A_t denote the averaging operator defined by

$$(A_t f)(h) = \int_K f(a_t k h)\, dm(k), \qquad h \in H.$$

As in Proposition 8.4, let

$$F(i) = \{x_1 \wedge x_2 \wedge \cdots \wedge x_i \mid x_1, x_2, \cdots, x_i \in \mathbb{R}^n\} \subset \bigwedge^i(\mathbb{R}^n).$$

Since $\|Kv\| = \|v\|$ and $\frac{\|hv\|}{\|v\|} \le \|\bigwedge^i(h)\|$, for any $v \in F(i)$ and $h \in H$, each f_i has properties (a) and (b) of Proposition 8.12. Applying Lemma 8.7 to $\Lambda = h\Delta$ we see that for any $i, 0 < i < n$, and $h \in H$

$$A_t f_i^s < \frac{c}{2} f_i^s + \omega^2 \max_{0<j\le \min\{n-i,i\}} \sqrt{f_{i+j}^s f_{i-j}^s}. \tag{93}$$

Let us denote $q(i) = i(n-i)$. Then by direct computations $2q(i) - q(i+j) - q(i-j) = 2j^2$. Therefore we get from (93) that for any $i, 0 < i < n$, and any positive $\epsilon < 1$

$$\begin{aligned} A_t(\epsilon^{q(i)} f_i^s) &< \frac{c}{2}\epsilon^{q(i)} f_i^s + \omega^2 \max_{0<j\le\min\{n-i,i\}} \epsilon^{q(i) - \frac{q(i+j)+q(i-j)}{2}} \sqrt{\epsilon^{q(i+j)} f_{i+j}^s \epsilon^{q(i-j)} f_{i-j}^s} \\ &\le \frac{c}{2}\epsilon^{q(i)} f_i^s + \epsilon\omega^2 \max_{0<j\le\min\{n-i,i\}} \sqrt{\epsilon^{q(i+j)} f_{i+j}^s \epsilon^{q(i-j)} f_{i-j}^s}. \end{aligned} \tag{94}$$

Consider the linear combination

$$f_{\epsilon,s} = \sum_{0\le i\le n} \epsilon^{q(i)}\ f_i^s.$$

The function $f_{\epsilon,s}$ then also has properties (a) and (b) of Proposition 8.12. Since $\epsilon^{q(i)} f_i^s < f_{\epsilon,s}, f_0 = 1$ and $f_n = 1/d(\Delta)$, the inequalities (94) imply the following inequality:

$$A_t f_{\epsilon,s} < 1 + d(\Delta)^{-s} + \frac{c}{2} f_{\epsilon,s} + n\epsilon\omega^2 f_{\epsilon,s}. \tag{95}$$

Taking $\epsilon = \frac{c}{2n\omega^2}$ we see that (84) from Proposition 8.12 also holds. Furthermore property (a) and (84) of Proposition 8.12 hold with the same constants for any unimodular lattice $\Delta \in \mathbb{R}^n$. Since $f_{\epsilon,s}(1) \leq n\alpha(\Delta)^s$, $f_{\epsilon,s}(1)$ is uniformly bounded as Δ varies over a compact set $\mathcal{C}$ of unimodular lattices. Hence the family $\mathcal{F}$ of functions $f_{\epsilon,s}$ obtained as Δ varies over $\mathcal{C}$ satisfies all the conditions of Proposition 8.12. Since $\alpha_i(h\Delta)^s = f_i(h)^s \leq \epsilon^{-q(i)} f_{\epsilon,s}(h)$, Proposition 8.12 implies that there exists a constant $B > 0$ so that for each i, all $t > 0$, and all $\Delta \in \mathcal{C}$,

$$\int_K \alpha_i(a_t k\Delta)^s \, dm(k) < B.$$

From this the theorem follows. $\square$

Proof of Theorem 7.15. We can assume that ϕ is nonnegative. Let $A(r) = \{x \in G/\Gamma \ : \ \alpha(x) > r\}$. Choose a continuous nonnegative function g_r on G/Γ such that $g_r(x) = 1$ if $x \in A(r+1)$, $g_r(x) = 0$ if $x \notin A(r)$ and $0 \leq g_r(x) \leq 1$ if $x \in A(r) - A(r+1)$. Then

$$\begin{aligned} &\int_K \phi(a_t kx)\nu(k)\, dm(k) = \\ &= \int_K (\phi g_r)(a_t kx)\nu(k)\, dm(k) + \int_K (\phi - \phi g_r)(a_t kx)\nu(k)\, dm(k). \end{aligned} \tag{96}$$

But (letting $\beta = 2 - s$), $(\phi g_r)(y) \leq B_1\alpha(y)^{2-\beta} g_r(y) = B_1\alpha(y)^{2-\frac{\beta}{2}} g_r(y)\alpha(y)^{-\frac{\beta}{2}} \leq B_1 r^{-\frac{\beta}{2}}\alpha(y)^{2-\frac{\beta}{2}}$ (the last inequality is true because $g_r(y) = 0$ if $\alpha(y) \leq r$). Therefore

$$\int_K (\phi g_r)(a_t kx)\nu(k)\, dm(k) \leq B_1 r^{-\frac{\beta}{2}} \int_K \alpha(a_t kx)^{2-\frac{\beta}{2}} \nu(k)\, dm(k). \tag{97}$$

According to Theorem 8.2 there exists B such that

$$\int_K \alpha(a_t kx)^{2-\frac{\beta}{2}} \, dm(k) < B$$

for any $t \geq 0$ and uniformly over $x \in \mathcal{C}$. Then (97) implies that

$$\int_K (\phi g_r)(a_t kx)\nu(k)\, dm(k) \leq BB_1(\sup \nu) r^{-\frac{\beta}{2}}. \tag{98}$$

Since the function $\phi - \phi g_r$ is continuous and has a compact support, the "bounded function" case of Theorem 7.6 implies that for every $\epsilon > 0$ there exists a finite set of points $x_1, \dots, x_\ell$ with Hx_i closed for each i so that for every compact subset F of $\mathcal{C} \setminus \bigcup_{i=1}^{\ell} Hx_i$ there exists $t_0 > 0$ such that for every $t > t_0$ and every $x \in F$,

$$\left| \int_K (\phi - \phi g_r)(a_t k x)\nu(k)\, dm(k) - \int_{G/\Gamma} (\phi - \phi g_r)(y)\, d\mu(y) \int_K \nu(k)\, dm(k) \right| < \frac{\epsilon}{2}. \quad (99)$$

It is easy to see that (96), (98) and (99) imply (49) if r is sufficiently large. $\square$

9 Connections to dynamics of rational billiards

For references to this section see [E2].

In this lecture, we describe some counting problems on translation surfaces and outline their connection to the dynamics of the $SL(2,\mathbb{R})$ action on the moduli space of translation surfaces. Much of this is presented in analogy with the quantitative Oppenheim conjecture (see §7 and §8).

Recall that $\mathcal{L}_n = SL(n,\mathbb{R})/SL(n,\mathbb{Z})$ is the space of covolume 1 lattices in $\mathbb{R}^n$. This space is non-compact, since we can have arbitrarily short vectors in a lattice.

The strata and the measure μ. Let $\beta = \beta_1, \dots, \beta_m$ be a partition of $2g-2$. Let $\mathcal{H}(\beta)$ denote the moduli space of translation surfaces with conical singularities of total angles $2\pi(\beta_1+1), \dots, 2\pi(\beta_m+1)$. (I am using the notation from [Zor]: Jean-Christophe is using $\mathcal{M}(\cdot)$.) We will sometimes call $\mathcal{H}(\beta)$ a *stratum*. Let $\mathcal{H}_1(\beta) \subset \mathcal{H}(\beta)$ denote the subset consisting of surfaces of area 1. Let μ be the normalized Lebesque measure on $\mathcal{H}_\beta$ (as defined by Jean-Christophe). We will use the same letter to denote the restriction of μ to $\mathcal{H}_1(\beta)$. A theorem of Masur and Veech (proved in Jean-Christophe's lectures) states that $\mu(\mathcal{H}_1(\beta)) < \infty$. In §9.5 we will describe how to evaluate the numbers $\mu(\mathcal{H}_1(\beta))$.

Note that the case of $n = 2$ in the space of lattices $\mathcal{L}_2$ and the case of stratum $\mathcal{H}_1(\emptyset)$ boil down to the same thing, since we are considering the space of unit volume tori (or more precisely, the space of 1-forms on unit volume tori), which is given by $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$.

Note. I will use the term *saddle connection* to denote what Jean-Christophe is calling a *connection*.

Holonomy and the sets $V_{sc}(S)$ and $V(S)$. Recall that a point $S \in \mathcal{H}(\beta)$ can be viewed as a pair (M, ω) where M is a Riemann surface and ω is a holomorphic 1-form on M. Recall that the holonomy of a curve γ on S is given by

$$hol(\gamma) = \int_\gamma \omega.$$

Let

$$V_{sc}(S) = \{hol(\gamma) \quad : \quad \gamma \text{ is a saddle connection on } S\},$$

so that $V_{sc}(S) \subset \mathbb{C} \simeq \mathbb{R}^2$. Note that $V_{sc}(S)$ is a discrete subset of $\mathbb{R}^2$, but it is not, in general, a subgroup. We also define the analogous set:

$$V(S) = \{hol(\gamma) : \gamma \text{ is a closed geodesic on } S \text{ not passing through singularities}\}.$$

Note that any such closed geodesic is part of a cylinder and all the closed geodesics in the cylinder have the same holonomy. (If $S = \mathbb{R}^2/\mathbb{Z}^2$ is the standard torus with the standard flat structure, then $V(S) = \mathbb{Z}^2$).

9.1 Counting cylinders and saddle connections

Let $B(R)$ denote a ball of radius R. Then, $|V(S) \cap B(R)|$ is the number of cylinders on S of length at most R. Masur proved the following:

Theorem 9.1. *For all flat surfaces S in a compact set, there are constants c_1 and c_2 so that for $R \gg 1$*

$$c_1 R^2 < |V(S) \cap B(R)| \leq |V_{sc}(S) \cap B(R)| < c_2 R^2.$$

The upper bound is proved in [Mas2] and the lower bound is proved in [Mas3]. The proof of the lower bound depends on the proof of the upper bound. Another proof of both the upper and lower bounds with explicit constants was given by Vorobets in [Vo1] and [Vo2]. We will sketch below yet another proof of the upper bound, using the ideas of §8. (See [EM] for the details).

We also note that there is a dense set of directions with a closed trajectory and thus a cylinder.

The following theorem, gives asymptotic formulas for the number of saddle connections and cylinders of closed geodesics on a generic surface. It was first proved in this form in [EM], but many of the ideas came from [Ve], where a slightly weaker version was proved.

Theorem 9.2. *For a.e. $S \in \mathcal{H}_1(\beta)$, we have*

$$|V_{sc}(S) \cap B(R)| \sim \pi b(\beta) R^2,$$

where $V_{sc}(S)$ is the collection of vectors in $\mathbb{R}^2$ given by holonomy of saddle connections on S, and $b(\beta)$ is the Siegel-Veech constant defined in §9.2 (see also (103)).

Similarly, for closed geodesics, we have that there is a constant $b_1(\beta)$ so that

$$|V(S) \cap B(R)| \sim \pi b_1(\beta) R^2$$

where $V(S)$ is the collection of vectors given by holonomy along (imprimitive) closed geodesics not passing through singularities, and $b_1(\beta)$ is the associated Siegel-Veech constant.

It will turn out that the problem of counting saddle connections or cylinders closed geodesics on a flat surface is analogous to the quantitative Oppenheim problem (§7 and §8).

9.2 The Siegel-Veech formula

The following construction and its analogues play a key role. For any function of compact support $f \in C_c(\mathbb{R}^n)$, let $\hat{f}(\Delta) = \sum_{v \in \Delta \setminus 0} f(v)$. Note that if $f = \chi_{B(1)}$, we get $\hat{f}(\Delta) = |\Delta \cap B(1)|$. We have the *Siegel formula*: For any $f \in C_c(\mathbb{R}^n)$,

$$\frac{1}{\mu(\mathcal{L}_n)} \int_{\mathcal{L}_n} \hat{f}(\Delta) d\mu(\Delta) = \int_{\mathbb{R}^n} f d\lambda, \tag{100}$$

where μ is Haar measure on $\mathcal{L}_n = SL(n,\mathbb{R})/SL(n,\mathbb{Z})$, and λ is Lebesgue measure on $\mathbb{R}^n$.

The generalization of this formula to moduli space was developed, so the legend goes, by Veech while he listened to Margulis lecture on the Oppenheim conjecture. For $f \in C_c(\mathbb{R}^2)$ we define the Siegel-Veech transform $\hat{f}(S) = \sum_{v \in V_{sc}(S)} f(v)$. Just as above, if $f = \chi_{B(1)}$, $\hat{f}$ counts the number of saddle connections of length ≤ 1.

Just as we had the Siegel formula for lattices, here we have the *Siegel-Veech formula*: There is a constant $b(\beta)$, called the *Siegel-Veech constant*, such that for any $f \in C_c(\mathbb{R}^2)$, we have

$$\frac{1}{\mu(\mathcal{H}_1(\beta))} \int_{\mathcal{H}_1(\beta)} \hat{f}(S)\, d\mu(S) = b(\beta) \int_{\mathbb{R}^2} f, \tag{101}$$

where μ is the natural $SL(2,\mathbb{R})$ invariant measue on $\mathcal{H}_1(\beta)$.

Let us sketch the proof of this result (essentially from [Ve], also reproduced in [EM]). The first step (which is by far the most technical) is to show that $\hat{f} \in L^1(\mathcal{H}_1(\beta))$, so that the left hand side is finite. This can be deduced e.g. from (108) below. Having done this, we denote the quantity on the left hand side of (101) by $\varphi(f)$.

Thus we have a linear functional $\varphi : C_c(\mathbb{R}^2) \to \mathbb{R}$, i.e. a measure. But it also has to be $SL(2,\mathbb{R})$ invariant. Only Lebesgue measue and δ_0, the delta measure at 0 are $SL(2,\mathbb{R})$ invariant. Thus we have $\varphi(f) = af(0) + b\int_{\mathbb{R}^2} f$. It remains to show $a = 0$. Consider the limit of indicator functions $f = \chi_{B(R)}$ as $R \to 0$. Both sides of the equation tend to 0, so we have that $a = 0$, and thus our result.

Returning to lattices, we can apply literally the same arguments to prove the Siegel formula (100). Note that nothing was special about dimension 2 in the above proof sketch. Thus, we have almost proved (100) as well. To be precise, we currently have:

$$\frac{1}{\mu(\mathcal{L}_n)} \int_{\mathcal{L}_n} \hat{f}(\Delta) d\mu(\Delta) = b \int_{\mathbb{R}^n} f d\lambda,$$

for some constant b. We need to show $b = 1$. Here, we once again use $f = \chi_{B(R)}$, but this time consider $R \to \infty$. Recall that $\hat{f}(\Delta) = |\Delta \cap B(R)| \sim \mathrm{Vol}(B(R))$, for $R \to \infty$ and Δ fixed. Thus, we get $b = 1$, and the Siegel formula.

We should remark that for the space of lattices the proof of the Siegel formula indicated above is not the easiest available. In fact, it is possible to avoid proving *apriori* that $\hat{f} \in L^1(\mathcal{L}_n)$. See [Sie] or [Cas] or [Ter] for the details. A well known consequence of the Siegel formula is the following:

$$\mu(\mathcal{L}_n) = \frac{1}{n}\zeta(2)\zeta(3)\ldots\zeta(n). \tag{102}$$

For the stata $\mathcal{H}(\beta)$, this method of evaluating $b(\beta)$ (i.e. considering $f = \chi_{B(R)}$ and taking $R \to \infty$) is not avaliable. Essentially the problem is that we do not have an alternative expression for the constant in Theorem 7.5.

Another approach is to let $f = \chi_{B(\epsilon)}$, send $\epsilon \to 0$ and keep track of the leading term in the asymptotics of both sides. This was done in [EMZ] where we obtained the following result: For any stratum $\mathcal{H}_1(\beta)$ in the moduli space of translation surfaces the coefficient $b(\beta)$ involved in (101) can be expressed in the following form:

$$b(\beta) = \sum_{\alpha<\beta} c(\alpha,\beta) \frac{\mu(\mathcal{H}_1(\alpha))}{\mu(\mathcal{H}_1(\beta))}, \tag{103}$$

where the sum is over lower dimensional strata α (which lie at the "boundary" of $\mathcal{H}(\beta)$), and $c(\alpha, \beta)$ are explicitly known rational numbers.

We note that (103) fails as a method for calculating the volumes, since (unlike the lattice case) we do not have an independent formula for $b(\beta)$. In §9.5 we will show that the volumes can be computed in a different way; then (103) can be used to evaluate the Siegel-Veech constants $b(\beta)$. These numbers appear in some other contexts as well, in particular in connection with the Lyapunov exponents of the geodesic flow.

9.3 Counting using the $SL(2, \mathbb{R})$ action

This subsection is closely parallel to §7.3. The following exposition will be along the lines of [EM], which was heavily influenced by [Ve]. To simplify the notation, we only deal with the case of saddle connections. Define $g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$ and $r_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. Let f be the indicator function of the trapezoid defined by the points

$$(1, 1), (-1, 1), (-1/2, 1/2), (1/2, 1/2).$$

Lemma 9.3. *We have* $\int_0^{2\pi} f(g_t r_\theta v)\, d\theta \approx \begin{cases} 2e^{-2t} & \text{if } e^t/2 \leq ||v|| \leq e^t, \\ 0 & \text{otherwise.} \end{cases}$

Proof. Let U denote the trapezoid. Note that

$$f(g_t r_\theta v) \neq 0 \Leftrightarrow g_t r_\theta v \in U \Leftrightarrow r_\theta v \in g_t^{-1} U. \qquad (104)$$

The set $g_t^{-1}U$ is the shaded region in Figure 3. From (104) it is clear that the integral in Lemma 9.3 is equal to (2π times) the fraction of the circle which lies inside the shaded region $g_t^{-1}U$. If v is too long or too short (not drawn), then the circle would completely miss the shaded region, and the integral would be zero. If it does not miss, then (2π times) the fraction of the circle in the shaded region is approximately $2e^{-2t}$, independent of $\|v\|$. $\square$

We now prove Theorem 9.2. Summing our formula from Lemma 9.3 over all $v \in V_{sc}(S)$ and recalling the definition of the *Siegel-Veech transform* $\hat{f}(S) = \sum_{v \in V_{sc}(S)} f(v)$, we get

$$\tfrac{1}{2} e^{2t} \int_0^{2\pi} \hat{f}(g_t r_\theta S)\, d\theta \approx |V_{sc}(S) \cap B(e^t)| - |V_{sc}(S) \cap B(e^t/2)|.$$

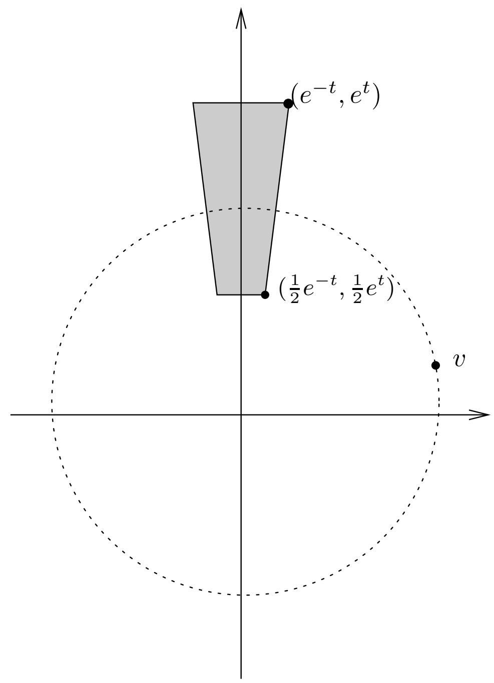

Figure 3. Lemma 9.3.

Writing $R = e^t$, we can rewrite this as

$$\tfrac{1}{2}R^2 \int_0^{2\pi} \hat{f}(g_t r_\theta S) d\theta \approx |V_{sc}(S) \cap B(R)| - |V_{sc}(S) \cap B(R/2)|. \qquad (105)$$

This equation is key to the counting problem, since the right hand side counts saddle connections in an annulus, and the left hand side is an integral over (part of) an $SL(2,\mathbb{R})$ orbit. (The fact that we only have approximate equality does not affect the leading order asymptotics.) Now we are supposed to use some sort of ergodic theory to analyze the behavoir of integral on the left-hand-side of (105) as $t \to \infty$ (or equivalently as $R \to \infty$).

There is an ergodic theorem of Nevo [Ne] which implies that[1] for almost all $S \in \mathcal{H}_1(\beta)$, and provided that $\hat{f} \in L^{1+\epsilon}(\mathcal{H}_1(\beta))$, the integral converges to $2\pi \int_{\mathcal{H}_1(\beta)} \hat{f}(S)\, dS = 2\pi b(\beta) \int_{\mathbb{R}^2} f$. The assertion that $\hat{f} \in L^{1+\epsilon}$ can be verified using (108). This immediately implies Theorem 9.2. $\square$

[1]The theorem of Nevo used here is about a general $SL(2,\mathbb{R})$ action, and uses nothing about the geometry of the moduli space.

However, this approach is a *failure* if one wants to prove things about billiards: our theorems hold for almost every point S, and the set of translation surfaces arising from rational billiards has measure zero.

One eventual goal is to prove analogues of Ratner's theorems on unipotent flows for the $SL(2,\mathbb{R})$ action on $\mathcal{H}_1(\beta)$. That is, we would like to classify invariant measures, orbit closures, and prove uniform distribution, for both the full $SL(2,\mathbb{R})$ action, and for the horocycle flow. One partial result in this direction is due to McMullen [Mc]: he has classified the $SL(2,\mathbb{R})$ orbit closures and invariant measures for the moduli space of genus 2 surfaces (i.e., the strata $\mathcal{H}(1,1)$ and $\mathcal{H}(2)$). Note that the integral in (105) is over large circles in $SL(2,\mathbb{R})$, which can be approximated well by horocycles. Thus the action of the horocycle flow (i.e. the action of $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$) is directly relevant. For other very partial results in this direction see [EMaMo], [EMS] and [CW].

9.4 The upper bounds.

In this subsection, we will outline a proof of the upper bound in Theorem 9.1, following the scheme of §8.

Let $B(R)$ be the ball of radius R centered at 0 in $\mathbb{R}^n$. For a given lattice $\Delta \in \mathcal{L}_n$. we would like to find out how many lattice points, that is, how many points of Δ are contained in $B(R)$.

It is immediately clear that for a fixed lattice Δ, as $R \to \infty$,

$$|\Delta \cap B(R)| \sim \mathrm{Vol}(B(R)) = \mathrm{Vol}(B(1))R^n. \tag{106}$$

(i.e. the number of lattice points is asymptotic to the volume). However, this is not uniform in Δ. A uniform upper bound has been given in Lemma 8.1, in particular:

$$|\Delta \cap B(1)| < C\alpha(\Delta). \tag{107}$$

The analagous problem in moduli space is as follows: We are interested in $|V_{sc}(S) \cap B(1)|$, i.e. the number of saddle connections of length at most 1 on S.

The result is as follows: Fix $\epsilon > 0$. Then there is a constant $c = c(\beta, \epsilon)$ such that for all $S \in \mathcal{H}(\beta)$ of area 1,

$$|V_{sc}(S) \cap B(1)| \leq \frac{c}{\ell(S)^{1+\epsilon}}, \tag{108}$$

where $\ell(S)$ is the length of the shortest saddle connection on S.

Assuming (108), the proof of the upper bound in Theorem 9.1 can be following the scheme of §8 (with a suitable definition for the functions α_i).

However, it turns out that the proof of (108) is more difficult that that of (107); it itself uses the system of inequalities along the line of §8, as well as induction on the genus.

9.5 Evaluation of the volumes

In this lecture we describe briefly another strategy for calculating volumes of strata, which also has a parallel for the space of lattices. Recall that we are considering the spaces $\mathcal{H}(\beta)$ of flat structures with singularity structure $\beta = (\beta_1, \beta_2, \ldots \beta_n)$, where $\beta_i \in \mathbb{N}$, $\sum \beta_i = 2g-2$. Let the set of singularities be denoted by Σ. We have $|\Sigma| = n$, and we have

$$H_1(S, \Sigma; \mathbb{Z}) = \mathbb{Z}^{2g+n-1}.$$

We can pick a basis by selecting g a-cycles, g b-cycles (from absolute homology), and $n-1$ relative cycles.

Fix a $\mathbb{Z}$-basis $\gamma_1, \gamma_2, \ldots \gamma_k$ of $H_1(S, \Sigma; \mathbb{Z})$, where $k = 2g + n - 1$. We recall the following fact (see [K]):

Theorem 9.4. *The map $(X, \omega) \to (hol(\gamma_1), \ldots, hol(\gamma_k))$ from $\mathcal{H}(\beta) \to (\mathbb{R}^2)^k$ is a local coordinate system.*

By pulling back Lebesgue measure on $(\mathbb{R}^2)^k$, we obtain a normalized measure ν on $\mathcal{H}(\beta)$. (For more details on the above constuction, see [Mas1, §3].) Now, we would like to define a measure on the hypersurface $\mathcal{H}_1(\beta)$.

This is similar to the lattice setting, where if we pick a basis $v_1, v_2, \ldots v_n$ for our lattice $\Delta \subset \mathbb{R}^n$, we get a matrix in $M_n(\mathbb{R})$ by letting v_i be the ith column. Note that since our lattice is unit volume, our matrix has determinant 1. We have a natural (Lebesque) measure ν on $M_n(\mathbb{R})$. Consider the det $= 1$ hypersurface Ω_1 (i.e., $SL(n, \mathbb{R})$). We define a measure μ on this space as follows: let $E \subset \Omega_1$, and let $C_1(E)$ be the cone over E (i.e. the union of all line segments which start at the origin and end at a point of E). We define $\mu(E) = \nu(C_1(E))$. This yields a finite measure since we are considering a fundamental domain under the $SL(n, \mathbb{Z})$-action. This is in fact the measure used in the previous section.

Returning to the setting of surfaces, recall that the area of our surface $S = (X, \omega)$ is given by

$$Area(S) = \frac{1}{2i} \int_X \omega \wedge \bar{\omega} = \frac{1}{2i} \sum_{i=1}^{g} \int_{a_i} \bar{\omega} \int_{b_i} \omega - \int_{b_i} \bar{\omega} \int_{a_i} \omega$$

where a_i and b_i are the a- and b-cycles on X respectively.

This gives that the area is a quadratic form in the coordinate sytem, i.e.,

$$Area(X, \omega) = Q(hol(\gamma_1), \ldots, hol(\gamma_k)).$$

However, it is a degenerate form, since it only depends on the absolute cycles a_i and b_i. We can mimic the lattice picture now: we define $\mu(E) = \nu(C_1(E))$ for any subset $E \subset \mathcal{H}_1(\beta)$. Thus,

$$\mu(\mathcal{H}_1(\beta)) = \mu(\mathcal{F}) = \nu(C_1(\mathcal{F})),$$

where $\mathcal{F}$ is a fundamental domain.

We now make a cosmetic step. Let $C_R(\mathcal{F})$ denote the cone of $\mathcal{F}$ extended to the hypersurface of area R-surfaces. Clearly

$$\mu(\mathcal{H}_1(\beta)) = \nu(C_1(\mathcal{F})) = \frac{\nu(C_R(\mathcal{F}))}{R^k}.$$

We have the following fact:

$$|C_R(\mathcal{F}) \cap (\mathbb{Z}^2)^k| \sim \nu(C_R(\mathcal{F}))$$

as $R \to \infty$, i.e. the number of lattice points in a cone is asymptotic to the volume. Ususally this is used to estimate the number of lattice points, but here we use this in reverse and estimate the volume by the number of lattice points. Thus, we get that

$$\mu(\mathcal{H}_1(\beta)) = \frac{\nu(C_R(\mathcal{F}))}{R^k} \sim \frac{|C_R(\mathcal{F}) \cap (\mathbb{Z}^2)^k|}{R^k},$$

or, equivalently,

$$|C_R(\mathcal{F}) \cap (\mathbb{Z}^2)^k| \sim \mu(\mathcal{H}_1(\beta))R^k. \tag{109}$$

The equation (109) is not useful unless we can find an interpretation of the points of $C_R(\mathcal{F}) \cap (\mathbb{Z}^2)^k$. This is given by the following:

Lemma 9.5. *$S = (X, \omega) \in C_R(\mathcal{F}) \cap (\mathbb{Z}^2)^k$ if and only if X is a holomorphic branched cover of the standard torus of degree $\leq R$, ω is the pullback of dz under the covering map, and all singularities branch over the same point.*

Proof: Since $S \in C_R(\mathcal{F})$, $area(S) \leq R$. By definition, $S \in (\mathbb{Z}^2)^k$ is equivalent to $hol(\gamma_1), \ldots, hol(\gamma_k) \in \mathbb{Z}^2$. Fix a non-singular point z_0 on S, and define $\pi : S \to T$, where T is the standard torus, by $\pi(z) = \int_{z_0}^{z} \omega$. Since $\int_\gamma \omega \in \mathbb{Z} + i\mathbb{Z}$ for any closed curve or saddle connection γ, this is a well defined covering map with all singularities branching over the same point. Since the torus is unit volume, the area of S is equal to the degree of the covering. $\square$

Let $N_\beta(d)$ denote the number[2] of branched covers of T of degree d with branching type β. (Note that $N_\beta(d)$ is defined in purely combinatorial terms).

Combining Lemma 9.5 with (109), we obtain the following: as $R \to \infty$,

$$\sum_{d=1}^{R} N_\beta(d) \sim \mu(\mathcal{H}_1(\beta)) R^k. \tag{110}$$

(This relation was discovered by Kontsevich and Zorich, and independently by Masur and the author.) Thus, we can compute $\mu(\mathcal{H}_1(\beta))$ if we can compute the asymptotics of the left-hand-side of (110). This is a purely combinatorial problem.

Suppose we are considering a degree d cover of the torus. Consider the standard basis a and b of curves on the torus (when the torus is viewed as the unit square, the curves correspond to the sides of the square). They give rise to permutations of the sheets, that is, elements of the symmetric group S_d. We will abuse notation by denoting these permutations also by a and b. Singularity types of covers correspond to different conjugacy classes of the commutator $aba^{-1}b^{-1}$. A simple zero is a transposition, a double zero a three cycle, a two simple zeroes is a product of two transpositions, etc. (So for example, if we are considering the stratum $\mathcal{H}(1,1)$, the commutator will be in the same conjugacy class as a product of two transpositions.) The number of pairs $(a, b) \in S_d \times S_d$ satisfying such a commutation relation can be expressed as a sum over the characters of the symmetric group S_d.

However, simply looking at the conjugacy class of the commutator permutation does not guarantee that the resulting surface is connected. We wish to count only the connected covers. However, the disconnected ones dominate the count. If one knows the number of disconnected covers exactly, one can compute the number of connected covers (by using inclusion/exclusion to subtract off all the possible ways a cover can disconnect). Unfortunately, as one does that, the first n terms in the asymptotic formula cancel. Still, it is possible, using the exact formula for the number of disconnected covers in [BO], to carry out the computation (see [EO]). The result, is a fairly messy but computable formula for $\mu(\mathcal{H}_1(\beta))$.

There are two consequences of the above computations worth mentioning:

[2] In order for Theorem 9.6 below to hold, we should, when defining $N_\beta(d)$, weigh each cover by the inverse of its automorphism group. However this does not affect the asymptotics and can be ignored for most purposes.

Theorem 9.6. *The generating function $F_\beta(q) = \sum_{d=0}^{\infty} N_\beta(d)q^d$ is a quasi-modular form, that is, it is a polynomial in the Eisenstein series $G_k(q)$, $k = 2, 4, 6$.*

Theorem 9.7. *$\pi^{-2g}\mu(\mathcal{H}_1(\beta)) \in \mathbb{Q}$, where g is the genus of any surface in $\mathcal{H}(\beta)$.*

Both of the above theorems were conjectured by Kontsevich. Further work showed that they hold also for the connected components of strata, and that similar results hold for spaces of quadratic differentials. We remark that Theorem 9.7 implies that the Siegel-Veech constants are rational.

For the space of lattices, one can carry out the same construction. The main difference is that one ends up counting *unbranched* covers of the standard torus T^n, or what is equivalent, sublattices of the standard lattice $\mathbb{Z}^n$. By computing the number of sublattices of $\mathbb{Z}^n$ of index at most R, and sending $R \to \infty$, it is not difficult to reproduce (102).

10 Equidistribution of translates and applications to Diophantine equations

We will follow parts of [EMc] and [EMS1].

In this section, using ergodic properties of subgroup actions on homogeneous spaces of Lie groups, we study asymptotic behavior of number of lattice points on certain affine varieties. Consider for instance the following.

Example 1 Let $p(\lambda)$ be a monic polynomial of degree $n \geq 2$ with integer coefficients and irreducible over $\mathbb{Q}$. Let $\mathrm{M}_n(\mathbb{Z})$ denote the set of $n \times n$ integer matrices, and put

$$V_p(\mathbb{Z}) = \{A \in \mathrm{M}_n(\mathbb{Z}) \ : \ \det(\lambda I - A) = p(\lambda)\}.$$

Hence $V_p(\mathbb{Z})$ is the set of integral matrices with characteristic polynomial $p(\lambda)$. Consider the norm on $n \times n$ real matrices given by $\|(x_{ij})\| = \sqrt{\sum_{ij} x_{ij}^2}$, and let $N(T, V_p)$ denote the number of elements of $V_p(\mathbb{Z})$ with norm less than T.

Theorem 10.1. *Suppose further that $p(\lambda)$ splits over $\mathbb{R}$, and for a root α of $p(\lambda)$ the ring of algebraic integers in $\mathbb{Q}(\alpha)$ is $\mathbb{Z}[\alpha]$. Then, asymptotically as $T \to \infty$,*

$$N(T, V_p) \sim \frac{2^{n-1} h R \omega_n}{\sqrt{D} \cdot \prod_{k=2}^{n} \Lambda(k/2)} T^{n(n-1)/2}$$

where h is the class number of $\mathbb{Z}[\alpha]$, R is the regulator of $\mathbb{Q}(\alpha)$, D is the discriminant of $p(\lambda)$, ω_n is the volume of the unit ball in $\mathbb{R}^{n(n-1)/2}$, and $\Lambda(s) = \pi^{-s}\Gamma(s)\zeta(2s)$.

Example 1 is a special case of the following counting problem which was first studied in [DRS] and [EMc].

The counting problem: Let W be a real finite dimensional vector space with a $\mathbb{Q}$ structure and V a Zariski closed real subvariety of W defined over $\mathbb{Q}$. Let G be a reductive real algebraic group defined over $\mathbb{Q}$, which acts on W via a $\mathbb{Q}$-representation $\rho : G \to \mathrm{GL}(W)$. Suppose that G acts transitively on V. Let $\|\cdot\|$ denote a Euclidean norm on W. Let B_T denote the ball of radius $T > 0$ in W around the origin, and define

$$N(T, V) = |V \cap B_T \cap \mathbb{Z}^n|,$$

the number of integral points on V with norm less than T. We are interested in the asymptotics of $N(T, V)$ as $T \to \infty$.

Let Γ be a subgroup of finite index in $G(\mathbb{Z})$ such that $W(\mathbb{Z})\Gamma \subset W(\mathbb{Z})$. By a theorem of Borel and Harish-Chandra [BH-C], $V(\mathbb{Z})$ is a union of finitely many Γ-orbits. Therefore to compute the asymptotics of $N(T, V)$ it is enough to consider each Γ-orbit, say $\mathcal{O}$, separately and compute the asymptotics of

$$N(T, V, \mathcal{O}) = |\mathcal{O} \cap B_T|.$$

Suppose that $\mathcal{O} = \Gamma \cdot v_0$ for some $v_0 \in V(\mathbb{Z})$. Then the stabilizer $H = \{g \in G : gv_0 = v_0\}$ is a reductive real algebraic $\mathbb{Q}$-subgroup, and $V \cong G/H$. Define

$$R_T = \{gH \in G/H : gv_0 \in B_T\},$$

the pullback of the ball B_T to G/H.

Assume that G^0 and H^0 do not admit nontrivial $\mathbb{Q}$-characters. Then by the theorem of Borel and Harish-Chandra, G/Γ admits a G-invariant (Borel) probability measure, say μ_G, and $H/(\Gamma \cap H)$ admits an H-invariant probability measure, say μ_H. Now the natural inclusion $H/(\Gamma \cap H) \hookrightarrow G/\Gamma$ is an H-equivariant proper map. Let $\pi : G \to G/\Gamma$ be the natural quotient map. Then the orbit $\pi(H)$ is closed, $H/(\Gamma \cap H) \cong \pi(H)$, and μ_H can be treated as a measure on G/Γ supported on $\pi(H)$. Such finite invariant measures supported on closed orbits of subgroups are called *algebraic measures*. Let $\lambda_{G/H}$ denote the (unique) G-invariant measure on G/H induced by the normalization of the Haar measures on G and H.

The following result was proved in [DRS]; subsequently a simpler proof appeared in [EMc].

Theorem 10.2. *Suppose that V is affine symmetric and Γ is irreducible (equivalently, H is the set of fixed points of an involution of G, and G is $\mathbb{Q}$-simple). Then asymptotically as $T \to \infty$,*

$$N(T, V, \mathcal{O}) \sim \lambda_{G/H}(R_T).$$

Translates of algebraic measures. For any $g \in G$, let $g\mu_H$ denote the translated measure defined as

$$g\mu_H(E) = \mu_H(g^{-1}E), \qquad \forall \text{ Borel sets } E \subset G/\Gamma.$$

Note that $g\mu_H$ is supported on $g\pi(H)$. A key ingredient in the proofs of Theorem 10.2 in [DRS] and [EMc] is showing that if H is the set of fixed points of an involution of G then for any sequence $\{g_i\} \subset G$, such that $\{g_i H\}$ has no convergent subsequence in G/H, the translated measures $g_i\mu_H$ get 'equidistributed' on G/Γ as $i \to \infty$; that is, the sequence $\{g_i\mu_H\}$ weakly converges to μ_G. The method of [DRS] uses spectral analysis on G/Γ, while the argument of [EMc] uses the mixing property of the geodesic flow. However, both methods seem limited essentially to the affine symmetric case. It should be remarked that for the proof of Theorem 10.2 one needs only certain averages of translates of the form $g\mu_H$ to become equidistributed.

One can show that under certain conditions if for some sequence $\{g_i\}$ we have $\lim g_i\mu_H = \nu$ then the measure ν is again algebraic. We give exact algebraic conditions on the sequence $\{g_i\}$ relating it to the limit measure ν. Using this analysis, we show that the counting estimates as in Theorem 10.2 hold for a large class of homogeneous varieties. The following particular cases of homogeneous varieties, which are not affine symmetric, are of interest. We first place Example 1 in this context.

Example 1 continued. Note that $V_p(\mathbb{Z})$ is the set of integral points on the real subvariety $V_p = \{A \in \mathrm{M}_n(\mathbb{R}) : \det(\lambda I - A) = p(\lambda)\}$ contained in the vector space $W = \mathrm{M}_n(\mathbb{R})$. Let $G = \{g \in \mathrm{GL}_n(\mathbb{R}) : \det g = \pm 1\}$. Then G acts on W via conjugations, and V_p is a closed orbit of G (see [New, Theorem III.7]). Put $\Gamma = G(\mathbb{Z}) = \mathrm{GL}_n(\mathbb{Z})$. The companion matrix of $p(\lambda)$ is

$$v_0 = \begin{pmatrix} 0 & & 0 & -a_n \\ 1 & & 0 & -a_{n-1} \\ 0 & \cdots & \vdots & \vdots \\ \vdots & & 0 & \\ 0 & & 1 & -a_1 \end{pmatrix} \in V_p(\mathbb{Z}). \tag{111}$$

The centralizer H of v_0 is a maximal $\mathbb{Q}$-torus and H^0 has no nontrivial $\mathbb{Q}$-characters. Note that H is not the set of fixed points of an involution, and the variety $V_p = H\backslash G$ is *not* affine symmetric. Nevertheless, we show that $N(T, V_p, \Gamma v_0) \sim \lambda_{G/H}(R_T)$. By computing the volumes, we obtain the following estimate.

Theorem 10.3. *Let $N(T, V_p)$ be the number of points on $V_p(\mathbb{Z})$ of norm less than T. Then asymptotically as $T \to \infty$,*

$$N(T, V_p) \sim c_p T^{n(n-1)/2},$$

where $c_p > 0$ is an explicitly computable constant.

We obtain a 'formula' for calculating c_p; for the sake of simplicity we calculate it explicitly only under the additional assumptions on $p(\lambda)$ of Theorem 10.1.

See [BR] for some deeper consequences of the above result.

Example 2. Let A be a nondegenerate indefinite integral quadratic form in $n \geq 3$ variables and of signature (p, q), where $p \geq q$, and B a definite integral quadratic form in $m \leq p$ variables. Let $W = \mathrm{M}_{m\times n}(\mathbb{R})$ be the space of $m \times n$ matrices. Consider the norm on W given by $\|(x_{ij})\| = \sqrt{\sum_{i,j} x_{ij}^2}$. Define

$$V_{A,B} = \{X \in \mathrm{M}_{m\times n}(\mathbb{R}) \ : \ XA\,{}^{\mathrm{t}}X = B\}.$$

Thus a point on $V_{A,B}(\mathbb{Z})$ corresponds to a way of representing B by A over $\mathbb{Z}$. We assume that $V_{A,B}(\mathbb{Z})$ is not empty.

The group $G = \mathrm{SO}(A)$ acts on W via right multiplication, and the action is transitive on $V_{A,B}$. The stabilizer of a point $\xi \in V_{A,B}$ is an orthogonal group H_ξ in $n - m$ variables. Let $\Gamma = G(\mathbb{Z})$. Then the number of Γ-orbits on $V_{A,B}(\mathbb{Z})$ is finite. Let $\xi_1, \ldots, \xi_h$ be the representatives for the orbits.

Theorem 10.4. *Let $N(T, V_{A,B})$ denote the number of points on $V_{A,B}(\mathbb{Z})$ with norm less than T. Then asymptotically as $T \to \infty$,*

$$N(T, V_{A,B}) \sim \sum_{i=1}^{h} \frac{\mathrm{vol}(\Gamma \cap H_{\xi_i} \backslash H_{\xi_i})}{\mathrm{vol}(\Gamma \backslash G)} c_{A,B} T^{r(n-r-1)}$$

where $r = \min(m, q)$, and $c_{A,B} > 0$ is an explicitly computable constant.

Remark 10.5. In some ranges of p, q, m, n this formula may be proved by the Hardy-Littlewood circle method, or by Θ-function techniques. Using our method one also obtains asymptotic formulas for the number of points in the individual orbits $\Gamma\xi_i$.

Remark 10.6. In the case $m > q$, the asymptotics of the number of integer points does not agree with the heuristic of the Hardy-Littlewood circle method, even if the number of variables mn is very large compared to the number of quadratic equations $m(m+1)/2$. The discrepancy occurs because the null locus $\{X : XA\,{}^{\mathrm{t}}X = 0\}$ does not contain a non-singular real point (cf. [Bir, Theorem 1]) and so the 'singular integral' vanishes.

10.1 Connection between counting and translates of measures

We recall some observations from [DRS, Sect. 2]; see also [EMc]. Let the notation be as in the counting problem stated in the introduction. For $T > 0$, define a function F_T on G by

$$F_T(g) = \sum_{\gamma \in \Gamma/(H\cap\Gamma)} \chi_T(g\gamma \cdot v_0),$$

where χ_T is the characteristic function of B_T. By construction F_T is left Γ-invariant, and hence it will be treated as a function on G/Γ. Note that

$$F_T(e) = \sum_{\gamma \in \Gamma/(H\cap\Gamma)} \chi_T(\gamma \cdot v_0) = N(T, V, \mathcal{O}).$$

Since we expect, as in Theorem 10.2, that

$$N(T, V, \mathcal{O}) \sim \lambda_{H\backslash G}(R_T),$$

we define

$$\hat{F}_T(g) = \frac{1}{\lambda_{G/H}(R_T)} F_T(g).$$

Thus the asymptotics in Theorem 10.2 is the assertion

$$\hat{F}_T(e) \to 1 \qquad \text{as } T \to \infty. \tag{112}$$

Proposition 10.7 ([DRS, Sect. 2])**.** *For any compactly supported function ψ on G/Γ,*

$$\langle \hat{F}_T, \psi \rangle = \frac{1}{\lambda_{G/H}(R_T)} \int_{R_T} \overline{\psi^H}\, d\lambda_{G/H},$$

where

$$\psi^H(gH) = \int_{G/\Gamma} \psi\, d(g\mu_H)$$

is a function on G/H.

Proof. Let $\mathcal{F}$ be a fundamental domain for G/Γ. By definition,

$$\begin{aligned}
\langle F_T, \psi \rangle &= \sum_{\gamma \in \Gamma/(H\cap\Gamma)} \int_{\mathcal{F}} \chi_T(g\gamma)\psi(g)\, d\mu_G(g) \\
&= \sum_{\gamma \in \Gamma/(H\cap\Gamma)} \int_{\mathcal{F}\gamma} \chi_T(g)\psi(g)\, d\mu_G(g) \\
&= \int_{G/(H\cap\Gamma)} \chi_T(g)\psi(g)\, d\mu_G(g) \\
&= \int_{G/H} \int_{H/(H\cap\Gamma)} \chi_T(\bar{g})\psi(\bar{g}h)\, d\mu_H(h)\, d\lambda_{G/H}(\bar{g}) \\
&\int_{R_T} \left(\int_{G/\Gamma} \psi\, d_{\bar{g}\mu_H} \right) \lambda_{G/H}(\bar{g})
\end{aligned}$$

□

10.2 Limiting distributions of translates of algebraic measures.

The following is the main result of this section which allows us to investigate the counting problems.

Theorem 10.8. *Let G be a connected real algebraic group defined over $\mathbb{Q}$, $\Gamma \subset G(\mathbb{Q})$ an arithmetic lattice in G with respect to the $\mathbb{Q}$-structure on G, and $\pi : G \to G/\Gamma$ the natural quotient map. Let $H \subset G$ be a connected real algebraic $\mathbb{Q}$-subgroup admitting no nontrivial $\mathbb{Q}$-characters. Let μ_H denote the H-invariant probability measure on the closed orbit $\pi(H)$. For a sequence $\{g_i\} \subset G$, suppose that the translated measures $g_i\mu_H$ converge to a probability measure μ on G/Γ. Then there exists a connected real algebraic $\mathbb{Q}$-subgroup L of G containing H such that the following holds:*

(i) There exists $c_0 \in G$ such that μ is a ${c_0}L{c_0}^{-1}$-invariant measure supported on $c_0\pi(L)$.

In particular, μ is a algebraic measure.

(ii) There exist sequences $\{\gamma_i\} \subset \Gamma$ and $c_i \to c_0$ in G such that $\gamma_i H {\gamma_i}^{-1} \subset L$ and $g_iH = c_i\gamma_iH$ for all but finitely many $i \in \mathbb{N}$.

The proof of this theorem is based on the following observation.

Proposition 10.9. *Let the notation be as in Theorem 10.8. Then either there exists a sequence $c_i \to c$ in G such that $c_i\mu_i = \mu_H$ for all $i \in \mathbb{N}$ (in which case $\mu = c\mu_H$), or μ is invariant under the action of a nontrivial unipotent one-parameter subgroup of G.*

In order to be able to apply Theorem 10.8 to the problem of counting, we need to know some conditions under which the sequence $\{g_i\mu_H\}$ of probability measures does not escape to infinity. Suppose further that G and H are reductive. Let $Z(H)$ be the centralizer of H in G. By rationality $\pi(Z(H))$ is closed in G/Γ. Now if $\pi(Z(H))$ is noncompact, there exits a sequence $\{z_i\} \subset Z(H)$ such that $\{\pi(z_i)\}$ is divergent; that is, it has no convergent subsequence. Then $z_i\mu_H$ escapes to the infinity; that is $(z_i\mu_H)(K) \to 0$ for any compact set $K \subset G/\Gamma$. The condition that $\pi(Z(H))$ is noncompact is equivalent to the condition that H is contained in a proper parabolic $\mathbb{Q}$-subgroup of G. In the converse direction we have the following (see [EMS2]).

Theorem 10.10. *Let G be a connected real reductive algebraic group defined over $\mathbb{Q}$, and H a connected real reductive $\mathbb{Q}$-subgroup of G, both admitting no nontrivial $\mathbb{Q}$-characters. Suppose that H is not contained in any proper parabolic $\mathbb{Q}$-subgroup of G defined over $\mathbb{Q}$. Let $\Gamma \subset G(\mathbb{Q})$ be an arithmetic lattice in G and $\pi : G \to G/\Gamma$ the natural quotient map. Let μ_H denote the H-invariant probability measure on $\pi(H)$. Then given an $\epsilon > 0$ there exists a compact set $K \subset G/\Gamma$ such that $(g\mu_H)(K) > 1-\epsilon$, $\forall g \in G$.*

The proof of this result uses generalizations of some results of Dani and Margulis [DM3]. Combining this theorem with Theorem 10.8, we deduce the following consequence.

Corollary 10.11. *Suppose that H is reductive and a proper maximal connected real algebraic $\mathbb{Q}$-subgroup of G. Then for any sequence $\{g_i\} \subset G$, if the sequence $\{g_iH\}$ is divergent (that is, it has no convergent subsequence) in G/H, then the sequence $\{g_i\mu_H\}$ gets equidistributed with respect to μ_G as $i \to \infty$ (that is, $g_i\mu_H \to \mu_G$ weakly).*

In the general case, one obtains the following analogue of Corollary 10.11. We note that the condition that H is not contained in any proper $\mathbb{Q}$-parabolic subgroup of G, is also equivalent to saying that any real algebraic $\mathbb{Q}$-subgroup L of G containing H is reductive.

Corollary 10.12. *Let G be a connected real reductive algebraic group defined over $\mathbb{Q}$, and H a connected real reductive $\mathbb{Q}$-subgroup of G not contained in any proper parabolic $\mathbb{Q}$-subgroup of G. Let $\Gamma \subset G(\mathbb{Q})$ be an arithmetic lattice in G. Suppose* *that a sequence $\{g_i\} \subset G$ is such that the sequence $\{g_i \mu_H\}$ does not converge to the G-invariant probability measure. Then after passing to a subsequence, there exist a proper connected real reductive $\mathbb{Q}$-subgroup L of G containing H and a compact set $C \subset G$ such that*

$$\{g_i\} \subset CL(Z(H) \cap \Gamma)$$

10.3 Applications to the counting problem.

The case where H is maximal. The following is a consequence of Corollary 10.11:

Theorem 10.13. *Let G and H be as in the counting problem. Suppose that H^0 is reductive and a proper maximal connected real algebraic $\mathbb{Q}$-subgroup of G, where H^0 denotes the connected component of identity in H. Then asymptotically as $T \to \infty$*

$$N(T, V, \mathcal{O}) \sim \lambda_{G/H}(R_T).$$

Remark 10.14. Suppose that H is the set of fixed point of an involution of G. Let L be a connected real reductive $\mathbb{Q}$-subgroup of G containing H^0. Then there exists a normal $\mathbb{Q}$-subgroup N of G such that $L = H^0 N$. Now if G is $\mathbb{Q}$-simple, then H^0 is a maximal proper connected $\mathbb{Q}$-subgroup of G (see [Bor, Lemma 8.0]). Hence Theorem 10.2 follows from Theorem 10.13.

The general case. We now use Corollary 10.12. For applying this result to the counting problem, we need to know that averages of translates of the measure μ_H along the sets R_T become equidistributed as T tends to infinity. I.e., we want the set of 'singular sequences', for which the limit measure is not G-invariant, to have negligible 'measure' in the sets R_T as $T \to \infty$. This does not hold when the sets R_T are 'focused' along $L/H (\subset G/H)$:

Definition 10.15. Let G and H be as in the counting problem. For a sequence $T_n \to \infty$, the sequence $\{R_{T_n}\}$ of open sets in G/H is said to be *focused*, if there exist a proper connected reductive real algebraic $\mathbb{Q}$-subgroup L of G containing H^0 and a compact set $C \subset G$ such that

$$\limsup_{n\to\infty} \frac{\lambda_{G/H}(q_H(CL(Z(H^0) \cap \Gamma)) \cap R_{T_n})}{\lambda_{G/H}(R_{T_n})} > 0,$$

where $q_H : G \to G/H$ is the natural quotient map.

Note that since L is reductive and defined over $\mathbb{Q}$, we have that $\pi(L)$ is closed in G/Γ. In particular, $L(Z(H^0)\cap\Gamma)$ is closed in G. Also $LzH^0 = Lz$ for any $z \in Z(H^0)$. Now since C is compact, the set $q_H(CL(Z(H^0)\cap\Gamma))$ is closed in G/H.

Now if the focusing of $\{R_{T_n}\}$ does not occur, then using Corollary 10.12 we can obtain the following analogue of Corollary 10.11.

Corollary 10.16. *Let G and H be as in the counting problem. Suppose that H^0 is not contained in any proper $\mathbb{Q}$-parabolic subgroup of G^0, and for some sequence $T_n \to \infty$, the sequence $\{R_{T_n}\}$ is not focused. Then given $\epsilon > 0$ there exists an open set $\mathcal{A} \subset G/H$ with the following properties:*

$$\liminf_{n\to\infty} \frac{\lambda_{G/H}(\mathcal{A}\cap R_{T_n})}{\lambda_{G/H}(R_{T_n})} > 1-\epsilon \tag{113}$$

and given any sequence $\{g_i\} \subset {q_H}^{-1}(\mathcal{A})$, if the sequence $\{q_H(g_i)\}$ is divergent in G/H then the sequence $\{g_i\mu_H\}$ converges to μ_G.

This corollary allows us to obtain the counting estimates like in Theorem 10.2 and Theorem 10.13 for a large class of homogeneous varieties.

Theorem 10.17. *Let G and H be as in the counting problem. Suppose that H^0 is not contained in any proper $\mathbb{Q}$-parabolic subgroup of G^0 (equivalently, $Z(H)/(Z(H)\cap\Gamma)$ is compact), and for some sequence $T_n \to \infty$ with bounded gaps, the sequence $\{R_{T_n}\}$ is not focused. Then asymptotically*

$$N(T,V,\mathcal{O}) \sim \lambda_{G/H}(R_T).$$

Remark. The non-focusing assumption in Theorem 10.17 is not vacuous. In the above setup one is required to verify the condition of nonfocusing in Theorem 10.17 separately for each application of the result.

Outline of the proof of Theorem 10.17, assuming Corollary 10.16.

Proposition 10.18. *Let the notation and conditions be as in Theorem 10.17. Then $\hat{F}_{T_n} \to 1$ in the weak-star topology on $L^\infty(G/\Gamma,\mu_G)$; that is, $\langle \hat{F}_{T_n},\psi\rangle \to \langle 1,\psi\rangle$ for any compactly supported continuous function ψ on G/Γ.*

Proof. As in Proposition 10.7,

$$\langle \hat{F}_T,\psi\rangle = \frac{1}{\lambda_{G/H}(R_T)} \int_{R_T} \overline{\psi^H}\, d\lambda_{G/H},$$

where

$$\psi^H(gH) = \int_{H\Gamma/\Gamma} \psi(gh\Gamma)\, d\mu_H(h\Gamma) = \int_{G/\Gamma} \psi\, d(g\mu_H)$$

is a function on G/H.

Let $\epsilon > 0$ be given. Since the sequence $\{R_{T_n}\}$ is not focused, we obtain a set $\mathcal{A} \subset G/H$ as in Corollary 10.16. Break up the integral over R_{T_n} into the integrals over $R_{T_n} \cap \mathcal{A}$ and $R_{T_n} \setminus \mathcal{A}$. By equation (113) and the boundedness of ψ, the second integral is $O(\epsilon)$. By Corollary 10.16, for any sequence $\{g_i\} \subset {q_H}^{-1}(\mathcal{A})$, if $\{q_H(g_i)\}$ has no convergent subsequence in G/H, then $g_i \cdot \mu_H \to \mu_G$. Hence

$$\psi^H(g_iH) \to \int_{G/\Gamma} \psi\, d\mu_G = \langle \psi, 1\rangle.$$

We use dominated convergence theorem to justify the interchange of limits. Now

$$\begin{aligned}
\lim_{n\to\infty} \langle \hat{F}_{T_n}, \psi\rangle &= \lim_{n\to\infty} \frac{1}{\lambda_{G/H}(R_{T_n})} \int_{R_{T_n}\cap\mathcal{A}} \overline{\psi^H}\, d\lambda_{G/H} + O(\epsilon) \\
&= \lim_{n\to\infty} \frac{1}{\lambda_{G/H}(R_{T_n})} \int_{R_{T_n}\cap\mathcal{A}} \overline{\langle\psi, 1\rangle}\, d\lambda_{G/H} + O(\epsilon) \\
&= \lim_{n\to\infty} \frac{\lambda_{G/H}(R_{T_n}\cap\mathcal{A})}{\lambda_{G/H}(R_{T_n})} \langle 1, \psi\rangle + O(\epsilon) \\
&= \langle 1, \psi\rangle + O(\epsilon)
\end{aligned}$$

Since ϵ is arbitrary, the proof is complete. □

Proposition 10.19 ([EMS1])**.** *There are constants $a(\delta)$ and $b(\delta)$ tending to 1 as $\delta \to 0$ such that*

$$b(\delta) \le \liminf_{T\to\infty} \frac{\lambda_{G/H}(R_{(1-\delta)T})}{\lambda_{G/H}(R_T)} \le \limsup_{T\to\infty} \frac{\lambda_{G/H}(R_{(1+\delta)T})}{\lambda_{G/H}(R_T)} \le a(\delta).$$

Proof of Theorem 10.17. Let ψ in Proposition 10.18 tend to a δ-function at the origin. Then, combining Proposition 10.18 and Proposition 10.19, we obtain that $\hat{F}_{T_i} \to 1$ pointwise on G/Γ as $i \to \infty$. (See [DRS, Lemma 2.3] for the details). Thus (112) holds. This completes the proof. □

10.4 Invariance under unipotents

Proposition 10.20. *Let G be a semisimple Lie group, Γ be a discrete subgroup of G, and $\pi : G \to G/\Gamma$ be the natural quotient map. Let H be a nontrivial reductive subgroup of G and Ω be a relatively compact neighborhood of identity in H. Let μ_Ω be the probability measure on $\pi(\Omega)$ which is the pushforward under π of the restriction to Ω of a Haar measure on H.*

Suppose that for a sequence $\{g_i\}_{i\in\mathbb{N}} \subset G$, the sequence $\{g_i \cdot \mu_\Omega\}_{i\in\mathbb{N}} \subset \mathcal{P}(G/\Gamma)$ converges weakly to a nonzero measure μ on G/Γ. Then one of the following holds:

1. *There exists a compact set $C \subset G$ such that $\{g_i\}_{i\in\mathbb{N}} \subset C Z_G(H)$.*
2. *μ is invariant under a nontrivial unipotent one-parameter subgroup of G.*

Proof. (Cf. [Moz, Lemma ??]) Let $\mathfrak{g}$ be the Lie algebra of G and $\mathfrak{h} \subset \mathfrak{g}$ be the Lie subalgebra corresponding to H. Equip $\mathfrak{g}$ with a Euclidean norm, say $\|\cdot\|$.

Claim 1. *If the Condition 1 above does not hold then there exists a sequence $X_i \to 0$ in $\mathfrak{h}$ as $i \to \infty$, such that a subsequence of $\{\mathrm{Ad}\, g_i \cdot X_i\}_{i\in\mathbb{N}}$ converges to a nonzero element $Y \in \mathfrak{g}$.*

To prove the claim there is no loss of generality if we pass to a subsequence of $\{g_i\}_{i\in\mathbb{N}}$, or replace $\{g_i\}_{i\in\mathbb{N}}$ by $\{g_i c_i\}_{i\in\mathbb{N}}$, where $\{c_i\}_{i\in\mathbb{N}}$ is contained in a compact subset of G.

Since H is reductive, there is a Cartan involution θ of G such that $\theta(H) = H$. Let K be the set of fixed points of θ. Then K is a maximal compact subset of G. There exists a maximal $\mathbb{R}$-split torus A in G such that

$$\theta(a) = a^{-1}, \quad \forall a \in A. \tag{114}$$

Choose an order on the system of $\mathbb{R}$-roots of A for G and let Δ be the set of simple roots. Let A_+ be the exponential of the closure of the positive Weyl chamber. Then by Cartan decomposition we have

$$G = K A_+ K.$$

Hence without loss of generality we can assume that $g_i = a_i k_i$ for all $i \in \mathbb{N}$, where $k_i \to k$ in K as $i \to \infty$ and $\{a_i\}_{i\in\mathbb{N}} \subset A_+$.

Let

$$\Phi = \{\alpha \in \Delta : \sup_{i\in\mathbb{N}} \alpha(a_i) < \infty\}.$$

Then by modifying the sequence $\{a_i\}_{i\in\mathbb{N}}$ from the left by multiplications by elements from a compact set in $A_+ \cap (\cap_{\beta\in\Delta\setminus\Phi} \ker\beta)$, we may assume that

$$\alpha(a_i) = 1, \quad \forall\alpha \in \Phi. \tag{115}$$

By passing to a subsequence, we may also assume that

$$\lim_{i\to\infty} \alpha(a_i) = \infty, \quad \forall\alpha \in \Delta \setminus \Phi. \tag{116}$$

Let P be the standard parabolic subgroup of G associated to Φ. Let $\mathfrak{p}$ be the Lie algebra of P, and $\mathfrak{n}$ be the Lie algebra of the unipotent radical N of P. Due to (114), we have

$$\mathfrak{g} = \theta(\mathfrak{p}) \oplus \mathfrak{n}.$$

Let $\pi_{\mathfrak{n}}$ denote the projection onto $\mathfrak{n}$ with $\ker(\pi_{\mathfrak{n}}) = \sigma(\mathfrak{p})$.

Suppose that the claim fails to hold. Then

$$\sup_{i\in\mathbb{N}} \|\operatorname{Ad} g_i \cdot X\| < \infty, \quad \forall X \in \mathfrak{h}. \tag{117}$$

Hence by (116),

$$\lim_{i\to\infty} \pi_{\mathfrak{n}}(\operatorname{Ad} k_i \cdot X) = 0, \quad \forall X \in \mathfrak{h}.$$

Therefore $kHk^{-1} \subset \theta(P)$. Since $\theta(H) = H$ and $\theta(k) = k$, we have that $kHk^{-1} \subset P \cap \theta(P)$. Hence due to (115),

$$\{a_i\}_{i\in\mathbb{N}} \subset Z_G(P \cap \theta(P)) \subset kZ_G(H)k^{-1}.$$

Since $\mathfrak{g} = \theta(\mathfrak{p}) + \mathfrak{n}$ and $k_ik^{-1} \to e$ as $i \to \infty$, by passing to subsequences, there exist sequences $b_i \to e$ in $\theta(P)$ and $n_i \to e$ in N such that

$$k_ik^{-1} = b_in_i, \quad \forall i \in \mathbb{N}.$$

Let $\{X_1, \ldots, X_m\}$ be a basis of $\mathfrak{h}$ and put $\mathbf{q} = (X_1, \ldots, X_m) \in \oplus_{i=1}^m \mathfrak{g}$. Consider the action of G on $\oplus_{i=1}^m \mathfrak{g}$ via the Adjoint action on each of the summands. Then

$$g_i \cdot \mathbf{q} = (g_ik^{-1})(k \cdot \mathbf{q}) = (a_ik_ik^{-1})(k \cdot \mathbf{q}) = (a_ib_ia_i^{-1})(a_in_i{a_i}^{-1})(k \cdot \mathbf{q})$$

By (117), $\{g_i \cdot \mathbf{q}\}_{i\in\mathbb{N}}$ is a bounded sequence. By (114) and (116), $a_ib_i{a_i}^{-1} \to e$ as $i \to \infty$. Therefore $(a_in_i{a_i}^{-1})(k \cdot \mathbf{q}) : i \in \mathbb{N}\}$ is a bounded sequence. Since N is a unipotent group, the orbit $N(k \cdot \mathbf{q})$ is closed. Therefore there exists a compact set $C_1 \subset N$ such that

$${a_i}^{-1}n_ia_i \in C_1(kZ_G(H)k^{-1} \cap N).$$

Therefore, since $\{a_i\} \subset kZ_G(H)k^{-1}$ and $a_i b_i a_i{}^{-1} \to e$ as $i \to \infty$, there exists a compact set $C \subset G$, such that

$$g_i k^{-1} = a_i k_i k^{-1} = (a_i b_i a_i{}^{-1})(a_i n_i a_i{}^{-1}) \in CZ_G(H)k^{-1}, \quad \forall i \in \mathbb{N}.$$

This contradicts the hypothesis of the claim, and hence the proof of Claim 1 is complete.

Now we can assume that there exists a sequence $X_i \to 0$ in $\mathfrak{h}$ and a nonzero elements $Y \in \mathfrak{g}$ such that

$$\lim_{i\to\infty} \operatorname{Ad} g_i X_i = Y.$$

Consider the one-parameter subgroup $u : \mathbb{R} \to G$ defined as $u(t) = \exp(tY)$ for all $t \in \mathbb{R}$. Since $X_i \to 0$, all the eigenvalues of $\operatorname{Ad} tX_i$ converge to 1 as $i \to \infty$. Since $u(t) = \lim_{i\to\infty} g_i{}^{-1}(\exp tX_i)g_i$ and the eigenvalues are invariant under conjugation, we have that 1 is the only eigenvalue of $\operatorname{Ad} u(t)$ for all $t \in \mathbb{R}$. Therefore u is a unipotent one-parameter subgroup of G.

Claim 2. *The measure μ is invariant under the action of $\{u(t) : t \in \mathbb{R}\}$.*

To prove the claim let $t \in \mathbb{R}$ and put $\delta = \exp(tX_i)$ for all $i \in \mathbb{N}$. Then by the definition of μ_Ω, for any $\psi \in \mathrm{C_c}(G/\Gamma)$,

$$\left| \int_{G/\Gamma} \psi(x)\, d\mu_\Omega(x) - \int_{G/\Gamma} \psi(\delta_i x)\, d\mu_\Omega(x) \right| \le \epsilon_i \cdot \sup|\psi|, \tag{118}$$

where ϵ_i depends only on δ_i, and $\epsilon_i \to 0$ as $\delta_i \to 0$. Let $i \in \mathbb{N}$. Applying Eq. 118 for $\psi_i(x) := \psi(g_i x)$ for all $x \in X$, we get

$$\left| \int_{G/\Gamma} \psi(g_i x)\, d\mu_\Omega(x) - \int_{G/\Gamma} \psi((g_i \delta_i g_i{}^{-1}) g_i x)\, d\mu_\Omega(x) \right| \le \epsilon_i \cdot \sup|\psi|.$$

We have $g_i \cdot \mu_\Omega \to \mu$ weakly as $i \to \infty$, $g_i{}^{-1}\delta_i g_i \to u(t)$ as $i \to \infty$, and ψ is uniformly continuous. Therefore

$$\int_{G/\Gamma} \psi(x)\, d\mu(x) = \int_{G/\Gamma} \psi(xu(g))\, d\mu(x).$$

This shows that μ is invariant under $\{u(t) : t \in \mathbb{R}\}$. This completes the proof of the theorem. $\square$

10.5 Proving Ergodicity

In view of Proposition 10.20 and the measure classification theorem, Theorem 10.8 would follow immediately if we knew that μ was ergodic. In general the ergodicity of μ does not follow from Theorem 6.5 since we are not assuming that H contains unipotents.

The next part of the proof of Theorem 10.8 parallels §6.3. One applies the measure classification theorem followed by linearization. The analysis is somewhat more complicated then that of §6.3 because of the multi-dimensional situation, and the fact that we have a map only from a compact subset of H. The end result is:

Proposition 10.21. *Let $B \subset H$ be a ball of diameter at most δ_0 in H around e. Let g_i be a sequence of elements in G, and let λ_i be the probability measure on $\pi(g_i(B))$ which is the pushforward under g_i of the normalized Lebesgue measure on B. Suppose that $\lambda_i \to \lambda$ weakly in the space of probability measures on G/Γ. Suppose there exist a unipotent one-parameter subgroup U of G and $F \in \mathcal{H}$ such that $\lambda(\pi(N(F,U))) > 0$ and $\lambda(\pi(S(F,U))) = 0$. Then there exists a compact set $D \subset \mathcal{A}_F$ such that the following holds: For any sequence of neighborhoods $\{\Phi_i\}$ of D in $\bar{V}_F$, there exists a sequence $\{\gamma_i\} \subset \Gamma$ such that for all large $i \in \mathbb{N}$,*

$$g_i(B)\gamma_i \cdot \bar{p}_F \subset \Phi_i. \tag{119}$$

In general the condition (119) is difficult to analyze using linear algebra methods. The idea of the proof of Theorem 10.8 is the following: Since we are assuming that $g_i B$ return to a compact set in G/Γ, we may write $g_i = c_i \gamma_i' h_i$, where c_i is in a compact set, $\gamma_i' \in \Gamma$ and $h_i \in B \subset H$. Without loss of generality, we may then replace g_i by $\gamma_i' h_i$. Consider rational points h_j in BB. The orbit of each rational point under Γ is discrete, so there are only finitely many possibilities for $\gamma_i' h_j \gamma_i \cdot \bar{p}_F$. By passing to a subsequence one can assume that $\gamma_i' h_j \gamma_i \cdot \bar{p}_F$ is constant, which eventually yields the proof of Theorem 10.8.

11 Applications of non-divergence to metric Diophantine approximation

12 The work of Goetze-Margulis on quadratic forms.

References

[BM] M. Bekka and M. Mayer, **Ergodic theory and topological dynamics of group actions on homogeneous spaces**, London Math. Soc. Lecture Note Series, Vol. 269, Cambridge University Press, Cambridge, 2000.

[Bir] B. J. Birch. Forms in many variables. *Proc. Roy. Soc. London* **265** (1962), 245–263.

[Bre] T. Brennan. . Princeton University undergraduate thesis, 1994.

[Bor] A. Borel. Values of indefinite quadratic form at integral points and flows on spaces of lattices. *Bull. AMS. (N.S.)* **32** (1995), 184–204.

[BH-C] A. Borel and Harish-Chandra. Arithmetic subgroups of algebraic groups. *Annals of Math* **75** (1962), 485–535.

[BO] S. Bloch, A. Okounkov, *The Character of the Infinite Wedge Representation*, Adv. Math. 149 (2000), no. 1, 1–60

[BR] M. Borovoi and Z. Rudnick. Hardy–Littlewood varieties and semisimple groups. *Inv. Math.* (1994)

[Bu] M. Burger. *Horocycle flow on geometrically finite surfaces.* Duke Math. J. 61 (1990), no. 3, 779–803.

[Cas] J.W.S Cassels. *An introduction to the geometry of numbers.* Springer, ?????.

[CW] K. Calta, K. Wortman. *On unipotent flows in H(1,1).* arXiv:math/0702238.

[Dan1] S.G. Dani. On invariant measures, minimal sets and a lemma of Margulis. *Invent. Math* **51**(1979), 239–260.

[Dan2] S.G. Dani. Invariant measures and minimal sets of horoshperical flows. *Invent. Math* **64**(1981), 357–385.

[Dan3] S.G. Dani. On orbits of unipotent flows on homogeneous spaces. *Ergod. Theor. Dynam. Syst.* **4**(1984), 25–34.

[Dan4] S.G. Dani. On orbits of unipotent flows on homogenous spaces II. *Ergod. Theor. Dynam. Syst* **6**(1986), 167–182.

[DM1] S.G. Dani and G.A. Margulis. Values of quadratic forms at primitive integral points. *Invent. Math.* **98**(1989), 405–424.

[DM2] S.G. Dani and G.A. Margulis. Orbit closures of generic unipotent flows on homogeneous spaces of $SL(3, \mathbb{R})$. *Math. Ann.* **286**(1990), 101–128.

[DM3] S.G. Dani and G.A. Margulis. Asymptotic behaviour of trajectories of unipotent flows on homogeneous spaces. *Indian. Acad. Sci. J.* **101** (1991), 1–17.

[DM4] S. G. Dani and G. ,A. Margulis, *Limit distributions of orbits of unipotent flows and values of quadratic forms*, in: **I. M. Gelfand Seminar**, Amer. Math. Soc., Providence, RI, 1993, pp. 91–137.

[DRS] W. Duke, Z. Rudnick, and P. Sarnak. Density of integer points on affine homogeneous varieties. *Duke Math. J.* **71** (1993), 181–209.

[E1] A. Eskin, *Counting problems and semisimple groups*, in: **Proceedings of the International Congress of Mathematicians**, Vol. II (Berlin, 1998), Doc. Math., 1998, pp. 539–552.

[E2] A. Eskin, Handbook of Dynamical Systems.

[EMM1] A. Eskin, G. A. Margulis, and S. Mozes, *Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture*,Ann. Math. (2), **147** (1998), 93–141.

[EMM2] A. Eskin, G. Margulis, S. Mozes, *Quadratic forms of signature* $(2, 2)$ *and eigenvalue spacings on rectangular 2-tori.* Ann. of Math. (2) **161** (2005), no. 2, 679–725.

[EM] A. Eskin and H. Masur, *Asymptotic formulas on flat surfaces*, Ergodic Th. and Dynam. Syst, v.21, 443-478, 2001.

[EMS] A. Eskin, H. Masur, M. Schmoll, *Billiards in rectangles with barriers.* Duke Math. J. 118 (2003), no. 3, 427–463.

[EMZ] A. Eskin, H. Masur and A. Zorich. *Moduli Spaces of Abelian Differentials: The Principal Boundary, Counting Problems and the Siegel–Veech Constants.* Publ. Math. Inst. Hautes Etudes Sci. No. 97, (2003), 61–179.

[EO] A. Eskin and A. Okounkov, *Asymptotics of numers of branched covers of a torus and volumes of moduli spaces of holomorphic differentials*, Invent. Math., 145 (1) (2001), 59–104.

[EMaMo] A. Eskin, J. Marklof and D. Morris, *Unipotent flows on the space of branched covers of Veech surfaces*, Ergodic Theory Dynam. Systems **26** (2006), no. 1, 129–162.

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

[EMS1] A. Eskin, S. Mozes and N. Shah, *Unipotent flows and counting lattice points on homogeneous varieties*, Ann. Math. **143** (1996), 253–299.

[EMS2] A. Eskin, S. Mozes, and N. Shah. Nondivergence of translates of certain algebraic measures Geom. Funct. Anal. **7** (1997), no. 1, 48–80.

[K] M. Kontsevich, *Lyapunov Exponents and Hodge Theory*, The mathematical beauty of physics (Saclay, 1996), 318–332, Adv. Ser. Math. Phys., **24**, World Sci. Publishing, River Edge, NJ, 1997.

[F] H. Furstenberg. *The unique ergodicity of the horocycle flow.* in: **Recent advances in topological dynamics** (Proc. Conf., Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund), pp. 95–115. Lecture Notes in Math., Vol. 318, Springer, Berlin, 1973.

[K1] D. Kleinbock, *Some applications of homogeneous dynamics to number theory*, in: **Smooth ergodic theory and its applications**,Proc. Sympos. Pure Math., 69, Amer. Math. Soc., Providence, RI, 2001, pp. 639–660.

[K2] D. Kleinbock *An extension of quantitative nondivergence and applications to Diophantine exponents*,Preprint (2005), `arXiv:math/0508517`.

[KLW] D. Kleinbock, E. Lindenstrauss and B. Weiss, *On fractal measures and diophantine approximation*, Selecta Math. **10** (2004), 479–523.

[KM] D. Kleinbock and G.A. Margulis, *Flows on homogeneous spaces and Diophantine approximationon manifolds*, Ann. Math. **148** (1998), 339–360.

[KSS] D. Kleinbock, N. Shah and A. Starkov, *Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory*, in:

Handbook on Dynamical Systems, Volume 1A, Elsevier Science, North Holland, 2002, pp. 813–930.

[KW] D. Kleinbock and B. Weiss, *Dirichlet's theorem on diophantine approximation and homogeneous flows,* Preprint (2006), `arXiv:math/0612171`.

[Mas1] H. Masur. *Ergodic theory of translation surfaces,* Handbook of Dynamical Systems.

[Mas2] H. Masur. *The growth rate of trajectories of a quadratic differential.*, Ergodic Th. and Dynam. Syst, v.10, 151-176, 1990.

[Mas3] H. Masur. *Lower bounds for the number of saddle connections and closed trajectories of a quadratic differential.* In Holomorphic Functions and Moduli, Vol.1 D.Drasin ed. Springer-Verlag 1988, pp. 215-228

[Mc] C. McMullen. *Dynamics of $SL_2(\mathbb{R})$ actions in genus* 2. Preprint.

[Mar1] G. A. Margulis. On the action of unipotent groups in the space of lattices. In *Lie Groups and their representations, (Proc. of Summer School in Group Representations, Bolyai Janos Math. Soc., Akademai Kiado, Budapest, 1971)*, pages 365–370. Halsted, New York, 1975.

[Mar2] G. A. Margulis. Formes quadratiques ind'efinies et flots unipotents sur les spaces homog'enes. *C. R. Acad. Sci. Paris Ser. I* **304**(1987), 247–253.

[Mar3] G. A. Margulis. Discrete Subgroups and Ergodic Theory. In *Number theory, trace formulas and discrete subgroups (a symposium in honor of A. Selberg)*, pages 377–398. Academic Press, Boston, MA, 1989.

[Mar4] G. A. Margulis. Indefinite quadratic forms and unipotent flows on homogeneous spaces. In *Dynamical systems and ergodic theory*, volume 23, pages 399–409. Banach Center Publ., PWN — Polish Scientific Publ., Warsaw, 1989.

[Mar5] G. A. Margulis. Dynamical and ergodic properties of subgroup actions on homogeneous spaces with applications to number theory. In *Proc. of ICM (Kyoto, 1990)*, pages 193–215. Math. Soc. of Japan and Springer, 1991.

[Mar6] G. A. Margulis, *Oppenheim conjecture*, in: **Fields Medalists' lectures**, WorldSci. Publishing, River Edge, NJ, 1997, pp. 272–327.

[Mar7] G. A. Margulis, *Diophantine approximation, lattices and flows on homogeneous spaces*, in: **A panorama of number theory or the view from Baker's garden**, Cambridge Univ. Press, Cambridge, 2002, pp. 280–310.

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

[Mor] D. Morris, **Ratner's theorems on unipotent flows**, Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 2005.

[Moz] S. Mozes. Mixing of all orders of Lie group actions. *Invent. Math.* **107** (1992), 235–241.

[MS] S. Mozes, N. Shah. *On the space of ergodic invariant measures of unipotent flows.* Ergodic Theory Dynam. Systems **15** (1995), no. 1, 149–159.

[New] M. Newman. *Integral Matrices.* Academic Press, New York, 1972.

[Ne] A. Nevo, *Equidistribution in measure preserving actions of semisimple groups*, preprint.

[NS] A. Nevo and E. Stein. A generalization of Wiener's pointwise ergodic theorem. Preprint. FIXME

[Ra1] M. Ratner. Rigidity of horocycle flows. *Ann. Math.* **115** (1982), 597–614.

[Ra2] M. Ratner. Factors of horocycle flows. *Ergodic Theory Dynam. Systems* **2** (1982), 465–489.

[Ra3] M. Ratner. Horocycle flows, joinings and rigidity of products. *Ann. Math.* **118** (1983), 277–313.

[Ra4] M. Ratner. Strict measure rigidity for unipotent subgroups of solvable groups. *Invent. Math.* **101** (1990), 449–482.

[Ra5] M. Ratner. On measure rigidity of unipotent subgroups of semisimple groups. *Acta Math.* **165** (1990), 229–309.

[Ra6] M. Ratner. On Raghunathan's measure conjecture. *Ann. Math.* **134** (1991), 545–607.

[Ra7] M. Ratner. Raghunathan's topological conjecture and distributions of unipotent flows. *Duke Math. J.* **63** (1991), no. 1, 235–280.

[Ra8] M. Ratner. *Raghunathan's conjectures for* $\mathrm{SL}(2, \mathbf{R})$, Israel J. Math. **80** (1992), no. 1-2, 1–31.

[Sar] P. Sarnak. Values at integers of binary quadratic forms. *To appear in a volume in memory of C. Herz. S. Drury, editor.*

[Sch] W. Schmidt. Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height. *Duke Math. J.* **35**(1968), 327–339.

[Sie] C. L. Siegel. *Lectures on the geometry of numbers.* Springer, 1989.

[Sha1] N. A. Shah. Uniformly distributed orbits of certain flows on homogeneous spaces. *Math. Ann.* **289** (1991), 315–334.

[Sha2] N. A. Shah. Limit distributions of polynomial trajectories on homogeneous spaces. *Duke Math. J.* **75**(1994), 711–732.

[Sha3] N. A. Shah. PhD thesis, Tata Institute for Fundamental Research.

[Sta1] A.N. Starkov. Solvable homogeneous flows. *Math. Sbornik* **176**(1987), 242–259. In Russian.

[Sta2] A.N. Starkov. The ergodic decomposition of flows on homogenous spaces of finite volume. *Math. Sbornik* **180**(1989), 1614–1633. In Russian.

[Sta3] A. Starkov, **Dynamical systems on homogeneous spaces**, Translations of Mathematical Monographs, Volume 190,Amer. Math. Soc., Providence, RI, 2000.

[Ter] A. Terras. *Harmonic analysis on Symmetric spaces and Applications II* Springer, 1988.

[V1] J. Vanderkam. Values at integers of homogeneous polynomials. *Duke Math. J.* **97** (1999), no. 2, 379–412.

[V2] J. Vanderkam. Pair correlation of four-dimensional flat tori. *Duke Math. J.* **97** (1999), no. 2, 413–438.

[V3] J. Vanderkam. Correlations of eigenvalues on multi-dimensional flat tori. *Comm. Math. Phys.* **210** (2000), no. 1, 203–223.

[Ve] W. Veech, *Siegel measures. Annals of Mathematics* **148** (1998), 895-944

[Vo1] Y. Vorobets. *Ergodicity of billiards in polygons.* Mat. Sb. 188 (3) (1997) pp. 65-112

[Vo2] Y. Vorobets. *Periodic geodesics on translation surfaces.* Eprint, arXiv:math.DS/0307249.

[Zim] R. J. Zimmer, **Ergodic theory and semisimple groups**, Birkhaüser, 1984.

[Zor] A. Zorich, *Flat surfaces*, in collection "Frontiers in Number Theory, Physics and Geometry. Volume 1: On random matrices, zeta functions and dynamical systems",