# NOTES ON RATNER'S MEASURE CLASSIFICATION THEOREM

WENHAN DAI

We report on Ratner's Measure Classification Theorem for dynamic systems. The two principal references are [Mor05] together with [EK07]. The ultimate goal is to understand the simplified proof for our main theorem in [Mor05, Chapter 5]. Before we reach this, an interlude on two particular cases performs in detail with some preliminaries. Furthermore, an application of Ratner's theorem on the Oppenheim conjecture follows as the coda.

Contents



## 1. Introduction

We begin with a relatively elementary phenomenon. Let $f$ be the obvious covering map from Euclidean plane $\mathbb{R}^2$ to the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. It is well known that if $L$ is any straight line in

*Date*: December 14, 2021.

The author claims no originality of these notes. Some of the contents have been copied from references listed in the end directly.

$\mathbb{R}^2$, then the closure of $f(L)$ is a very nice submanifold of the torus. Furthermore, if the slope of $L$ is irrational, it is classical that $f(L)$ is dense; otherwise, it should be compact.

More generally, we may replace the Euclidean space with any Lie group $G$ and allows $L$ to be a subgroup satisfying some simple conditions. To make it precise, let $G$ be a connected Lie group, and let $\Gamma$ be a lattice in $G$. Look at the action of a closed subgroup $H$ of $G$ on $\Gamma\backslash G$. In general, the orbits of $H$ are chaotic (if $H$ acts ergodically, then almost all of them are dense). There have already been some examples where the closures of all orbits were homogeneous spaces themselves. This is, for some well-known instances, the case in Kronecker's theorem about translations on a torus, or in Hedlund's Minimality Theorem. One may wonder how general this phenomenon is. In the 1980s, some conjectures were formulated in this respect.

In 1990, Marina Ratner proved her result about this observation in full generality, with any Lie group $G$, which allows $\Gamma$ to be any subgroup of $G$ generated by *unipotent* elements. Ratner's theorems assert that, in general, the closure of an orbit will be a very nice topological set.

The following discussion is about the history of Ratner's theorem that has been summarized in [Mor05]. In the 1930s, G. Hedlund proved that if $G = \mathrm{SL}_2(\mathbb{R})$ and $\Gamma\backslash G$ is compact, then unipotent flows on $\Gamma\backslash G$ are ergodic and minimal, see [Hed08].

It was not until 1970 that H. Furstenberg proved these flows are uniquely ergodic, thus establishing the Measure Classification Theorem for this case. At about the same time, W. Parry proved an Orbit Closure Theorem, Measure Classification Theorem, and Equidistribution Theorem for the case where $G$ is nilpotent [Par71], and G. A. Margulis used the polynomial speed of unipotent flows to prove the fact that unipotent orbits cannot go off to infinity [Mar71].

Inspired by these and other early results, M. S. Raghunathan conjectured a version of the Orbit Closure Theorem and showed it would imply the Oppenheim Conjecture. He did not publish this conjecture, but it appeared in a paper of S. G. Dani [Dan81] in 1981. In this paper, Dani conjectured a version of the Measure Classification Theorem.

Dani also generalized Furstenberg's Theorem to the case where $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ is not compact [Dan78]. Publications of R. Bowen [Bow76], S. G. Dani, R. Ellis, and W. Perrizo [EP78], and W. Veech [Vee77] proved further generalizations for the case where the unipotent subgroup $U$ is horospherical. Results for horosphericals also follow from a method in the thesis of G. A. Margulis, see Lemma 5.2 of [Mar04].

M. Ratner began her work on the subject at about this time, proving her Rigidity Theorem, Quotients Theorem, Joinings Theorem, and other fundamental results in the early 1980's. In a series of her four papers, M. Ratner has just obtained a very satisfactory answer by solving a conjecture of M.S. Raghunathan [Rat82b] [Rat82a] [Rat83]. Using Ratner's methods, D. Witte generalized her rigidity theorem to all $G$ in [Wit87].

S. G. Dani and J. Smillie proved the Equidistribution Theorem when $G = \mathrm{SL}_2(\mathbb{R})$ in [DS84]. S. G. Dani [Dan86] showed that unipotent orbits spend only a negligible fraction of their life near infinity. A. Starkov [Sta89] proved an orbit closure theorem for the case where $G$ is solvable.

Using unipotent flows, G. A. Margulis' developed methods that allowed him to prove in 1987 the Oppenheim Conjecture, Theorem 1.1 in the following, on values of quadratic forms.

**Theorem 1.1** (Margulis, 1987)**.** *Let $Q$ be an indefinite non-degenerate quadratic form on $\mathbb{R}^n (n \geqslant 3)$. It is assumed that $Q$ is not a multiple of a rational form. Then, for any $\varepsilon > 0$, there exists a vector $v$ of $\mathbb{Z}^n$ such that $0 < Q(v) < \varepsilon$.*

He and S. G. Dani then proved several results, including the first example of an orbit closure theorem for actions of nonhorospherical unipotent subgroups of a semisimple Lie group [DM90] – namely, for so-called generic one-parameter unipotent subgroups of $\mathrm{SL}_3(\mathbb{R})$. G. A Margulis has pointed out that the methods could yield proof of the general case of the Orbit Closure Theorem.

Then M. Ratner proved her amazing theorems (largely independently of the work of others) [Rat91], by expanding the ideas from her earlier study of horocycle flows.

To state it, let us agree that an element $g$ of a Lie group $G$ is *unipotent* if the adjoint automorphism $\mathrm{Ad}_g$ of the Lie algebra $\mathfrak{g}$ is unipotent, i.e. has only 1 as its eigenvalue. A subgroup $H$ of $G$ is *unipotent* if all its elements are unipotent.

**Theorem 1.2** (Ratner, 1990)**.** *If $G$ is a closed, connected subgroup of $\mathrm{SL}_\ell(\mathbb{R})$ for some $\ell$, $\Gamma$ is a discrete subgroup of $G$, $u^t$ is a unipotent one-parameter subgroup of $G$, and $\mu$ is an ergodic $u^t$-invariant probability measure on $\Gamma\backslash G$, then $\mu$ is homogeneous. More precisely, there exist a closed, connected subgroup $S$ of $G$, and a point $x$ in $\Gamma\backslash G$, such that*

(1) *$\mu$ is $S$-invariant, and*
(2) *$\mu$ is supported on the orbit $xS$.*

This is not the best possible statement; we will give other versions later. Not surprisingly, this result has consequences for Diophantine approximations.

In the meantime, N. Shah [Sha91] showed that the Measure Classification Theorem implied an Equidistribution Theorem for many cases when $G = \mathrm{SL}_3(\mathbb{R})$.

Ratner's Theorems were soon generalized to $p$-adic groups, by M. Ratner [Rat95] and, independently, by G. A. Margulis and G. Tomanov [MT96]. N. Shah [Sha98] generalized the results to subgroups generated by unipotent elements. For *connected* subgroups generated by unipotent elements, this has been proved in Ratner's original papers.

## 2. Basics of ergodic theory

This section simply gathers some necessary background results, mostly without proof. Firstly, a result of the ergodic decomposition of measures (see Theorem 2.5) is given. After this, we state the pointwise ergodicity (see Theorem 2.7) (cf. Birkhoff's ergodic theorem) and a consequence of it (see Corollary 2.9), which reveals that distinct ergodic measures are mutually singular. Furthermore, the uniqueness theorem for ergodic measures has been discussed as well. A general reference for ergodic theory is Petersen's book [Pet83], but we only focus on Chapter 1.3 and 3.3 of [Mor05], and Chapter 2 of Einsiedler–Ward [Ein10].

2.1. **Ergodic decomposition of measures.** Let $X$ be a topological space, and $T\colon X \to X$ a map. We assume that there is a finite measure $\mu$ on $X$ which is preserved by $T$. One usually normalizes $\mu$ so that $\mu(X) = 1$, in which case $\mu$ is called a probability measure.

Sometimes, instead of a transformation $T$ one considers a flow $\varphi_t, t \in \mathbb{R}$. For a fixed $t, \varphi_t$ is a map from $X$ to $X$. We will only use definitions and theorems in this part for flows such as $\varphi_t$ later.

**Definition 2.1.** A measure-preserving flow $\varphi_t$ on a probability space $(X, \mu)$ is *ergodic* if, for each $\varphi_t$-invariant subset $A$ of $X$, we have either $\mu(A) = 0$ or $\mu(A) = 1$.

**Example 2.2.** For $G = \mathrm{SL}_2(\mathbb{R})$ and $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, the horocycle flow $\eta_t$ and the geodesic flow $\gamma_t$ are ergodic on $\Gamma\backslash G$ with respect to the Haar measure on $\Gamma\backslash G$.

Ergodic decomposition says that every measure-preserving flow can decompose into a union of ergodic flows. We begin with an example.

**Example 2.3.** Let $v = (\alpha, 1, 0) \in \mathbb{R}^3$ for some irrational $\alpha$, $\varphi_t$ be the corresponding flow on $\mathbb{T}^3 = \mathbb{R}^3/\mathbb{Z}^3$, and $\mu$ be the Lebesgue measure on $\mathbb{T}^3$. Then $\varphi_t$ is not ergodic, because sets of the form $\mathbb{T}^2 \times A$ are invariant.

However, the flow decomposes into a union of ergodic flows: for each $z \in \mathbb{T}$, let $T_z = \{z\} \times \mathbb{T}^2$ and $\mu_z$ be the Lebesgue measure on the torus $T_z$. Then

- $\mathbb{T}^3$ is the disjoint union $\bigcup_z T_z$,
- the restriction of $\varphi_t$ to each subtorus $T_z$ is ergodic (with respect to $\mu_z$), and
- the measure $\mu$ is the integral of the measures $\mu_z$ by Fubini's Theorem.

The following proposition shows that every measure $\mu$ can be decomposed into ergodic measures. Each ergodic measure $\mu_z$ is called an *ergodic component* of $\mu$.

**Proposition 2.4.** *If $\mu$ is any $\varphi_t$-invariant probability measure on $X$, then there exist a measure $v$ on a space $Z$, and a measurable family $\{\mu_z\}_{z\in Z}$ of ergodic measures on $X$, such that*

$$\mu = \int_Z \mu_z dv,$$

*that is, for every $f \in L^1(X,\mu)$,*

$$\int_X f d\mu = \int_Z \int_X f d\mu_z dv(z).$$

The proof of Proposition 2.4 requires some functional analysis and is not entirely in line with the main thrust of the argument, we do choose to omit it.

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

Proposition 2.4 indeed yields a decomposition of the measure $\mu$, but, unlike Example 2.3, it does not provide a decomposition of the space $X$. However, any two ergodic measures must be mutually singular, so a little more work yields the following geometric version of the ergodic decomposition. This often allows one to reduce a general question to the case where the flow is ergodic.

We also point out that ergodic measures always exist, whenever we obtain the existence of invariant measures. The probability measures form a convex set, and the ergodic probability measures are the extreme points of this set, which can be compared with the Krein–Milman theorem.

**Theorem 2.5** (Ergodic Decomposition)**.** *If $\mu$ is a $\varphi_t$-invariant probability measure on $X$, then there exists a measurable family $\{\mu_z\}_{z\in Z}$ of ergodic measures on $X$, a measure $v$ on $Z$, and a measurable function $\psi\colon X \to Z$, such that*

1. *$\mu = \int_Z \mu_z dv$, and*
2. *$\mu_z$ is supported on $\psi^{-1}(z)$, for almost every $z \in Z$.*

The proof of ergodic decomposition, which is also omitted, morally relies on the following very useful generalization of Fubini's Theorem.

**Proposition 2.6.** *Let $X$ and $Y$ be complete and separable metric spaces; $\mu$ and $v$ be probability measures on $X$ and $Y$, respectively; and $\Psi\colon X \to Y$ be a measure-preserving Borel map. Then there is a Borel map $\lambda\colon Y \to \mathrm{Prob}(X)$, such that*

1. *$\mu = \int_Y \lambda_y d\nu(y)$, and*
2. *$\lambda_y(\psi^{-1}(y)) = 1$, for all $y \in Y$.*

*Furthermore, $\lambda$ is unique up to measure zero.*

2.2. **Pointwise ergodicity.** In the proof of Ratner's Theorem and many other situations, one wants to know that the orbits of a flow are uniformly distributed. It is rarely the case that every orbit is uniformly distributed, that is, what it means, to say the flow is uniquely ergodic; but the Pointwise Ergodic Theorem 2.7 shows that if the flow is ergodic, a much weaker condition, then almost every orbit is uniformly distributed.

**Theorem 2.7** (Pointwise Ergodicity)**.** *Suppose $\mu$ is a probability measure on a locally compact and separable metric space $X$, $\varphi_t$ is an ergodic and measure-preserving flow on $X$, and $f \in L^1(X,\mu)$. Then*

$$\frac{1}{T}\int_0^T f(\varphi_t(x))dt \to \int_X f d\mu,$$

*for almost every $x \in X$.*

The integral 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. Theorem 2.7 is amazing in its generality: the only assumption is the ergodicity of the measure $\mu$. We note that this is some sort of irreducibility assumption.

**Definition 2.8.** A point $x \in X$ is *generic* for $\mu$ if pointwise ergodicity holds for every uniformly continuous, bounded function on $X$. In other words, a point is generic for $\mu$ if its orbit is uniformly distributed in $X$.

**Corollary 2.9.** *If $\varphi_t$ is ergodic, then almost every point of $X$ is generic for $\mu$.*

The converse of the corollary above is true as well. The last corollary will be used in the proof of Ratner's measure classification for $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ (see Theorem 4.1).

**Corollary 2.10.** *Let $T_t\colon X \to X$ be a flow preserving an ergodic probability measure $\mu$, and let $f \in L^1(\mu)$. For any $\varepsilon > 0$ and $\delta > 0$ there exists $\tau_0 > 0$ and a set $E$ with $\mu(E) < \varepsilon$ such that for any $x \notin E$ and any $\tau > \tau_0$ we have*

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

*In other words, the average of $f$ over the orbit of $x$ converges to the average of $f$ over $X$ uniformly outside of a set of small measure.*

*Proof.* Let $E_n$ be the set of $x \in X$ such that for some $\tau > n$,

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

By Theorem 2.7, $\mu(\bigcap_n E_n) = 0$; thus for some $n$ we have $\mu(E_n) < \varepsilon$. Let $\tau_0 = n$ and $E = E_n$. □

2.3. **Uniquely ergodic systems.** In some applications (in particular to number theory) we need some analogue of Theorem 2.7 for *all* points $x \in X$, 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 pointwise ergodicity is not helpful. However, there is one situation where we can show that it is still valid for all $x \in X$.

**Definition 2.11.** An ergodic measure-preserving flow $\varphi_t$ on a probability space $(X, \mu)$ is *uniquely ergodic* if there exists a unique invariant probability measure $\mu$ for it.

Intuitively, the meaning of unique ergodicity is as follows: the properties that hold *almost everywhere* for ergodic transformations will hold *everywhere* for uniquely ergodic ones. For another example, suppose we have an ergodic flow $\varphi_t$ and a measurable set $A \subset X$. For almost every $x \in X$ it makes sense to speak of the proportion of time the orbit of $x$ under $\varphi_t$ spends in $A$, and this proportion will be $\mu(A)$, as long as we assume $\mu(\partial A) = 0$. If $\varphi_t$ were uniquely ergodic, this statement would be true for all $x \in X$.

**Proposition 2.12.** *Suppose $X$ is compact, $\varphi_t$ is uniquely ergodic, and let $\mu$ be the invariant probability measure. Suppose $f\colon X \to \mathbb{R}$ is continuous. Then for all $x \in X$, the conclusion of Theorem 2.7 holds.*

*Remark* 2.13. The main point of the proof is the construction of an invariant measure (denoted by $\delta_\infty$ in most references) 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 $\delta_\infty$ 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 2.12 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. In fact, 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.

## 3. Flows and homogeneous spaces

This section serves as a preparation for those sections to appear. We discuss some examples of homogeneous spaces and flows on them such as the space of lattices $\mathrm{SL}_n(\mathbb{Z})\backslash\mathrm{SL}_n(\mathbb{R})$, and non-ergodicity of geodesic flow for $\mathrm{SL}_n(\mathbb{R})/\mathrm{SO}(n)$.[1] It is also a well-known idea to regard upper-half plane $\mathbb{H}$ as a homogeneous space. As a relatively basic but indispensable part for preliminaries, it refers to Chapter II§4 and Chapter V§2 of Bekka–Mayer [BM00].

[1] Probably, we want to work with finite-volume quotients of this symmetric space.

3.1. **Some geometry on $\mathbb{H}$.** We will now discuss the geometry of the complex upper half-plane, $\mathbb{H}$. The material in the next several paragraphs is standard, and some of the details are omitted; they can be found in Chapter 5.4 of [KH95].

To give $\mathbb{H}$ the structure of a Riemannian manifold, we must specify a metric, that is, the inner product on the tangent bundle. For $z \in \mathbb{H}$ and $u_1 + iv_1, u_2 + iv_2 \in T_z\mathbb{H}$ we define the hyperbolic metric on $\mathbb{H}$ by

$$\langle u_1 + iv_1, u_2 + iv_2 \rangle_z = \operatorname{Re} \frac{(u_1 + iv_1)(u_2 - iv_2)}{(\operatorname{Im} z)^2} = \frac{u_1u_2 + v_1v_2}{(\operatorname{Im} z)^2}.$$

The verification that this is a metric is standard. First, a lemma that will let us establish what the geodesics on the upper half-plane is in the following.

**Lemma 3.1.** *Let $M$ be a Riemannian manifold, and let $\Gamma$ be a group of isometries of $M$ that is transitive on unit vectors: that is, for all $v, \widetilde{v}$ in the unit tangent bundle on $M$, there exists $\phi_{v,\widetilde{v}} \in \Gamma$ with $\phi_{v,\widetilde{v}}(v) = \widetilde{v}$. Let $\mathcal{C}$ be a nonempty family of unit-speed curves satisfying the following properties:*

1. *$\mathcal{C}$ is closed under the action of $\Gamma$: that is, for all $c \in \mathcal{C}$ and $\phi \in \Gamma$, the composition $\phi \circ c \in \mathcal{C}$;*
2. *$\Gamma$ is transitive on $\mathcal{C}$: that is, for any $c, \widetilde{c} \in \mathcal{C}$, there exists $\phi_{c,\widetilde{c}} \in \Gamma$ with $\phi_{c,\widetilde{c}} \circ c = \widetilde{c}$; and*
3. *$\mathcal{C}$ consists of the axes of $\Gamma$; that is, for all $c \in \mathcal{C}$, there exists $\phi_c \in \Gamma$ such that $c$ is the set of fixed points of $\phi_c$.*

*Then $\mathcal{C}$ is the family of (all) unit-speed geodesics on $M$.*

*Proof.* First, we show that $\mathcal{C}$ contains all the geodesics. Let $v \in T_pM$ be a unit tangent vector to $M$ at $p$. It determines a unique geodesic $\gamma_v$ with $\dot{\gamma}_v(0) = v$. Now take some $c \in \mathcal{C}$, and let $\widetilde{v} = \dot{c}(0)$. Consider the action of $\phi_{\widetilde{v},v}$: it maps $c$ to a curve $\widetilde{c}$ that is tangent to $\gamma_v$ at $p$. Now we look at $\phi_{\widetilde{c}}$, the isometry that fixes $\widetilde{c}$: it must map $\gamma_v$ to a geodesic, but since it fixes $\widetilde{c}$ it must also fix $v$, and consequently $\phi_{\widetilde{c}} \circ \gamma_v$ and $\gamma_v$ are tangent to each other at $p$. Two tangent geodesics must coincide, so $\gamma_v$ is fixed by $\phi_{\widetilde{c}}$, and consequently $\gamma_v = \widetilde{c} \in \mathcal{C}$.

Finally, since $\mathcal{C}$ contains geodesics and $\Gamma$ is transitive on it, every curve in $\mathcal{C}$ is the isometric image of a geodesic, and thus itself a geodesic. □

As a simple application, we can now characterize the geodesics on the standard 2-sphere: let $\mathcal{C}$ be the family of great circles parametrized with unit speed, and let $\Gamma$ be the group generated by rotations and reflections in great circles. Checking the conditions of the lemma is easy; we conclude that $\mathcal{C}$ is exactly the group of unit-speed geodesics on the 2-sphere. We now consider the isometries of $\mathbb{H}$. The projective special linear group $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ acts on $\mathbb{H}$ by fractional linear transformations:

$$z \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \frac{az + b}{cz + d}$$

For later discussion, it is more convenient to let this be a right action, and let later actions be left ones. The group $\mathrm{PSL}_2(\mathbb{R})$ is generated by translations $z \mapsto z + b$ corresponding to $\left(\begin{smallmatrix} 1 & 0 \\ b & 1 \end{smallmatrix}\right)$; inversions $z \mapsto -1/z$ corresponding to $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$; and scaling $z \mapsto a^2 z$ corresponding to $\left(\begin{smallmatrix} a & 0 \\ 0 & 1 \end{smallmatrix}\right)$. All three of these clearly map points in $\mathbb{H}$ to $\mathbb{H}$. Also, for $T \in \mathrm{PSL}_2(\mathbb{R})$, we have

$$\operatorname{Im} T(z) = |T'(z)| \operatorname{Im}(z).$$

We only need to check this for the two generating transformations, since this formula respects composition. For $z \mapsto z + b$ and $z \mapsto a^2 z$ this is clear, and for $z \mapsto -1/z$ we have $\operatorname{Im}(-1/z) = \operatorname{Im}(z)/|z|^2$ as required. We can now check that the action of $PGL_2(\mathbb{R})$ on $\mathbb{H}$ is isometric:

$$\begin{aligned}
\langle T'(z)(u_1 + iv_1), T'(z)(u_2 + iv_2) \rangle_{T(z)} &= \operatorname{Re} \frac{T'(z)(u_1 + iv_1)\overline{T'(z)(u_2 + iv_2)}}{(\operatorname{Im} T(z))^2} \\
&= \operatorname{Re} \frac{(u_1 + iv_1)(u_2 - iv_2)}{(\operatorname{Im} z)^2} \\
&= \langle u_1 + iv_1, u_2 + iv_2 \rangle.
\end{aligned}$$

Note that the action of $\mathrm{PSL}_2(\mathbb{R})$ on the unit tangent bundle of $\mathbb{H}$ is transitive. Indeed, any $z \in \mathbb{H}$ may be translated onto the positive imaginary axis $i\mathbb{R}_+$, and then scaled onto $i$. It remains to check that $\mathrm{PSL}_2(\mathbb{R})$ is transitive on $T_i(\mathbb{H})$; the transformation

$$z \mapsto \frac{\cos(\theta/2)z + \sin(\theta/2)}{-\sin(\theta/2)z + \cos(\theta/2)},$$

corresponding to the matrix

$$\begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix},$$

sends $v \in T_i\mathbb{H}$ to $v(\cos\theta + i\sin\theta)$, i.e. rotates $v$ by $\theta$.

We are now in a position to classify the geodesics on $\mathbb{H}$.

**Theorem 3.2.** *The geodesics on $\mathbb{H}$ are vertical lines or semicircles with center on the real axis.*

*Proof.* To prove this theorem, we enlarge the group of isometries of $\mathbb{H}$ by reflection in the $i\mathbb{R}$ axis, $z \mapsto -\bar{z}$. The resulting group is, of course, still transitive on the unit tangent bundle. To show property (1) of Lemma 3.1, namely that our family of curves $\mathcal{C}$ is closed under the action of our isometry group $\Gamma$, we note that this is clear for $z \mapsto -\bar{z}$. For the Möbius transformations we will show transitivity first, and then observe that any Möbius transformation acting on $i\mathbb{R}_+$ sends it to an element of $\mathcal{C}$. Finally, transitivity together with the fact that $i\mathbb{R}_+$ is the set of fixed points of $z \mapsto -\bar{z}$ will show property (3) of Lemma 3.1.

We now show that any vertical line or semicircle with center on the real axis can be mapped to $i\mathbb{R}_+$. For a vertical line through $b \in \mathbb{R}$, the transformation $z \mapsto z - b$ works. For a semicircle through $b, b + a^2 \in \mathbb{R}$ we translate left by $b$ and scale by $a^{-2}$ to get a semicircle through 0 and 1. Now consider the map $z \mapsto z/(1-z)$. Its inverse is $z \mapsto z/(z+1)$, which sends $i\mathbb{R}_+$ onto the semicircle with endpoints 0 and 1, since

$$\left|\frac{it}{1+it} - \frac{1}{2}\right| = \left|\frac{2it - (1+it)}{2(1+it)}\right| = \frac{1}{2}$$

Therefore, we mapped our semicircle onto $i\mathbb{R}_+$ as required. Finally, we must show that Möbius transformations map $i\mathbb{R}_+$ into $\mathcal{C}$. It suffices to check this for the generators of the group of Möbius transformations: $z \mapsto z + b$ sends $i\mathbb{R}_+$ to the vertical line through $b \in \mathbb{R}$; $z \mapsto az$ sends $i\mathbb{R}_+$ to itself, reparametrizing it along the way; and $z \mapsto -1/z$ sends $i\mathbb{R}_+$ to itself but with the opposite parametrization.

Thus, we are in a position to apply Lemma 3.1 and conclude that the geodesics of $\mathbb{H}$ are the vertical lines and semicircles with center on the real axis. ☐

Not only is $\mathrm{PSL}_2(\mathbb{R})$ transitive on $T^1\mathbb{H}$, but the action is free: that is, the transformation $g$ mapping $v \in T_p\mathbb{H}$ to $v' \in T'_p\mathbb{H}$ is unique. Indeed, $g$ must map the unique geodesic tangent to $v$ at $p$ to the unique geodesic tangent to $v'$ at $v$. On the other hand, a Möbius transformation is uniquely determined by where it maps any three points; therefore, $g$ is unique. We may therefore identify the unit tangent bundle of $\mathbb{H}$ with $\mathrm{PSL}_2(\mathbb{R})$: we identify $v \in T_z\mathbb{H}$ with the (unique) transformation that sends the upwards unit vector at $i$ to $v$. The upwards unit vector at $i$ is therefore identified with the identity matrix.

3.2. **Geodesic and horocycle flows.** We now consider the geodesic flow on the unit tangent bundle of $\mathbb{H}$, described as follows: for $v \in T_z\mathbb{H}$ we define $\phi_t(v) = \dot{\gamma}(t) \in T_{\gamma(z)}\mathbb{H}$, where $\gamma$ is the unique unit-speed geodesic with $\dot{\gamma}(0) = v$. Under the identification of the unit tangent bundle with $\mathrm{PSL}_2(\mathbb{R})$, we have the following

**Lemma 3.3.** *The geodesic flow on the unit tangent bundle of $\mathbb{H}$ corresponds to the flow on the group $\mathrm{PSL}_2(\mathbb{R})$ given by left translation $g \mapsto h_t g$ with*

$$h_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/e} \end{pmatrix}$$

*for $t \in \mathbb{R}$.*

*Proof.* Let $v$ be the unit upward vector at $i$: then, by definition, $\phi_t(v)$ is the unit (in the $\mathbb{H}$-metric) upward vector at the point $z$ on the geodesic $i\mathbb{R}_+$ a distance $t$ away from $i$. Now, claim that $d(i, e^t i) = t$, indeed,

$$d(i, e^t i) = \int_1^{e^t} \frac{dy}{y} = t$$

The unit upward vector at $e^t i$ has the form $e^t v$; therefore, $\phi_t(v) = e^t v \in T_{e^t i}$. On the other hand, $h_t(i, v) = (e^t i, e^t v)$. Consequently, the actions agree on this unit tangent vector.

Now, let $\zeta \in T_z\mathbb{H}$ be arbitrary, and let $g_{z\zeta} \in \mathrm{PSL}_2(\mathbb{R})$ be such that $\zeta = v g_{z\zeta}$, where $v$ is still the unit upward vector at $i$. Since $g_{z\zeta}$ is an isometry on $\mathbb{H}$, we have

$$\phi_t(\zeta) = \phi_t(v g_{z\zeta}) = \phi_t(v) g_{z\zeta} = h_t v g_{z\zeta} = h_t \zeta$$

which is the desired result. □

Now, let $v \in T_p\mathbb{H}$ and $w \in T_q\mathbb{H}$ be two vectors in the unit tangent bundle of $\mathbb{H}$. We would like to define the distance between them. Consider the unique (unit-speed) geodesic $\gamma$ with $\gamma(0) = p$ and passing through $q$, and the vector field along $\gamma$ having the same angle with $\gamma'$ as the angle between $v$ and $\gamma'(0)$. Define the angle between $v$ and $w$ to be the angle between the vector field at $q$ and $w$. We just described the process of parallel-translating $v$ along $\gamma$ to a tangent vector at $q$. Now define

$$d(v, w) = \sqrt{(\angle(v, w))^2 + d(p, q)^2}.$$

This is the standard definition of distance on the unit tangent bundle (see, for example, [KH95, §A.4]), and in particular does define a distance function.

Now consider the unit upward vector at $i$ and at $x + i$ for some $x \in \mathbb{R}$. Their orbits under the geodesic flow will be $t \mapsto ie^t$ and $t \mapsto x + ie^t$. The hyperbolic distance between these two points is easily seen to be bounded by $xe^{-t}$ by considering the horizontal line segment joining $ie^t$ and $x + ie^t$; moreover, the angle between $v$ and $w$ is readily seen to be $2\tan^{-1}(x/(2e^t))$, and in particular is also bounded by $xe^{-t}$. Therefore, the distance between the tangent vectors to these geodesics is bounded by $\sqrt{2}xe^{-t}$: the orbits of the upward vertical unit vectors at all points $x \in \mathbb{R} + i$ are positively asymptotic to that of $i$ (and to each other).

Applying $z \mapsto -1/z$, we see that the orbits of the outward unit normals to the circle of (Euclidean) radius $1/2$ centered at $i/2$ are negatively asymptotic to that of $i$ (and to each other). This brings us to the concept of horocycles.

**Definition 3.4.** A *horocycle* on $\mathbb{H}$ is either a circle tangent to $\mathbb{R}$ at $x$ or a horizontal line $\mathbb{R} + ir = \{t + ir : t \in \mathbb{R}\}$. In the first case, we say that the horocycle is centered at $x$; in the second case, we say that it is centered at $\infty$.

All horocycles are isometric to the line $\mathbb{R} + i$. Indeed, for horocycles $H = \mathbb{R} + ri$ the isometry $T(z) = z/r$ suffices; for a horocycle centered at $x \in \mathbb{R}$ of Euclidean diameter $r$, take $T_1(z) = z - x, T_2(z) = z/r$ (applying $T_2 \circ T_1$ gets us a horocycle centered at 0 of Euclidean diameter 1), and $T_3(z) = -1/z$. Then $T = T_3 \circ T_2 \circ T_1$ maps our horocycle isometrically onto $\mathbb{R} + i$.

The horocycle flow on the unit tangent bundle is described as follows: for $v \in T_z\mathbb{H}$ there exists a unique horocycle passing through $z$ whose inward normal is $v$. (We define "inward" to be "up" for the horocycle at $\infty$; with this definition, each horocycle rests at the $+\infty$ end of the geodesic tangent to $v$.) The action of $\psi_s$ is to move $z$ to a point $s$ units away on the horocycle, and parallel-transport the unit tangent vector to an inward normal at that point. Equivalently, for $v$ an upward normal to a point $z \in \mathbb{R} + i$, we define $\psi_s(v)$ to be the upward normal to the point $z + s \in \mathbb{R} + i$, noting that on the line $\mathbb{R} + i$ the hyperbolic metric coincides with the Euclidean one.

From this description, we conclude that under the identification of the unit tangent bundle of $\mathbb{H}$ with $\mathrm{PSL}_2(\mathbb{R})$ the horocycle flow is given by the left action of

$$u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \quad s \in \mathbb{R}.$$

3.3. **Ergodicity of the flows.** Consider flows on the upper half-plane to be actions on $\mathrm{PSL}_2(\mathbb{R})/\Gamma$. The idea is given in the following.

Let $\Gamma$ be a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$ with finite covolume, i.e. a lattice. Then $\Gamma$ acts freely and discontinuously on $\mathbb{H}$, and therefore is the group of deck transformations for $\mathbb{H}$ regarded as a covering space of $\mathbb{H}/\Gamma$. The identification of the unit tangent space of $\mathbb{H}$ with $\mathrm{PSL}_2(\mathbb{R})$ induces an identification of the unit tangent space of $\mathbb{H}/\Gamma$ with $\mathrm{PSL}_2(\mathbb{R})/\Gamma$; in this identification, for $t \in \mathbb{R}$, the geodesic and horocycle flows respectively correspond to the action of

$$h_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, \quad u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

We use the following notation: let

$$N^+ = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}, \quad N^- = \left\{ \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} : x \in \mathbb{R} \right\},$$

and

$$A = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} : a > 0 \right\}.$$

Then $N^+, N^-$, and $A$ together generate all of $\mathrm{PSL}_2(\mathbb{R})$. Recall that $A$ is the set of matrices in the geodesic flow.

We will show that the geodesic flow on $\mathbb{H}/\Gamma$ is ergodic for every lattice $\Gamma$ in $\mathbb{H}$. To do this, we will show that if a function $f$ is invariant under the action of $A$, then it must be invariant under the action of all of $\mathrm{PSL}_2(\mathbb{R})$, and then if $f$ is in $L^2$, it must be constant. This will establish ergodicity. We observe the lemma as follows.

**Lemma 3.5.** *For $g = g_a \in A$ and $h \in N^+$ if $a < 1$, or $h \in N^-$ if $a > 1$, we have*

$$\lim_{n\to\infty} g^n h g^{-n} = e.$$

*That is, conjugation by $g$ contracts the relevant $h$.*

*Proof.* By direct computation:

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & a^2 x \\ 0 & 1 \end{pmatrix},$$

so

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^{n} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^{-n} = \begin{pmatrix} 1 & a^{2n} x \\ 0 & 1 \end{pmatrix}.$$

Also,

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ a^{-2n} x & 1 \end{pmatrix}.$$

This completes the proof. □

Now, we can use the following result due to Mautner.

**Lemma 3.6** (Mautner lemma)**.** *Consider a unitary representation on* $\mathrm{PSL}_2(\mathbb{R})/\Gamma$ *as above, and suppose $g, h \in \mathrm{SL}_2(\mathbb{R})$ satisfying $\lim_{n\to\infty} g^n h g^{-n} = 1$. Then all $f \in L^2(\mathrm{PSL}_2(\mathbb{R})/\Gamma)$ that are invariant under the action of $g$ are also invariant under the action of $h$.*

*Proof.* We use the associated operators of these actions instead. Note that

$$\|T_h f - f\| = \|T_h T_{g^{-n}} f - T_{g^{-n}} f\| = \|T_{g^n} T_h T_{g^{-n}} f - f\|.$$

We may let $n \to \infty$ inside the norm, concluding that $\|T_h f - f\| = 0$ and $f$ is invariant under $h$.

Ergodicity of the geodesic flow on $\mathrm{PSL}_2(\mathbb{R})/\Gamma$. If $T_a f = f$ for every $a \in A$, then combining Lemma 3.5 with the Mautner lemma we derive that $T_g f = f$ for every $g \in \mathrm{PSL}_2(\mathbb{R})$. For $f \in L^2$ this means that $f$ is essentially constant, and transformation by $A$ is ergodic as required. We can also show that the horocycle flow is ergodic: Ergodicity of the horocycle flow on $\mathrm{PSL}_2(\mathbb{R})/\Gamma$. Let $f \in L^2(\mathrm{PSL}_2(\mathbb{R})/\Gamma)$ be invariant under $N^+$. We will show that $f$ must be invariant under $A$ as well, and then proceed as in the proof of the ergodicity of the horocycle flow.

For each $g \in \mathrm{SL}_2(\mathbb{R})$ define $\phi(g) = \langle T_g f, f\rangle$. Since $f$ is invariant under all of $N^+$, the operator $\phi$ is constant on every double coset $N^+ g N^+$.

Now let $\lambda_n \to 0, \lambda_n \neq 0$ for every $n$, and take

$$g_n = \begin{pmatrix} 0 & \lambda_n^{-1} \\ \lambda_n & 0 \end{pmatrix}, \quad a = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \in A$$

and then

$$\begin{pmatrix} 1 & \alpha\lambda_n^{-1} \\ 0 & 1 \end{pmatrix} g_n \begin{pmatrix} 1 & \alpha^{-1}\lambda_n^{-1} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ \lambda_n & \alpha^{-1} \end{pmatrix}.$$

That is, $\phi(a) = \lim_{n\to\infty} \phi(g_n)$. Since $g_n$ does not depend on $a$, we conclude that $\phi$ is constant on $A$; therefore, $\langle T_a f, f\rangle = \langle f, f\rangle$, and by Cauchy–Schwarz it must be that $a$ acts on $f$ by multiplication by some constant $\chi(a)$. However, it is easy to see that this constant must be 1: that is, $f$ is invariant under $A$.

We can now follow the same logic as for the geodesic flow. □

This proof relied essentially on the interaction between the geodesic and the horocycle flow: the geodesic flow contracts one direction of the horocycle flow and expands the other direction.

3.4. **The space of lattices.** We aim to give an interpretation that why it is admissible to regard $\mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$ as a space of unimodular lattices. The first step is to give a definition of unimodular lattices. The following definitions comes from [EK07, §2.1].

**Definition 3.7.** A subset $L$ of $\mathbb{R}^n$ is called a *lattice* if it is a module of rank $n$ whose basis is a basis for $\mathbb{R}^n$. Equivalently, $L$ is a discrete subgroup of $\mathbb{R}^n$ that spans $\mathbb{R}^n$ as a subset of an $\mathbb{R}$-vector space.

Every lattice thus has a basis $v_1, \dots, v_n$ of linearly independent vectors. We derive from this two corollaries. First, $\mathrm{GL}_n(\mathbb{R})$ acts on the space of lattices as follows: $g$ acts on $L$ by applying $g$ to each element of $L$, in particular applying $g$ to the $\mathbb{Z}$-basis of $L$. Second, every lattice $L$ is an image of the standard lattice $\mathbb{Z}^n$ under some transformation in $\mathrm{GL}_n(\mathbb{R})$ (the columns of $g$ are given by the basis vectors of $L$).

**Definition 3.8.** A lattice $L$ in $\mathbb{R}^n$ is *unimodular* if the covolume of $L$, that is, the volume of the compact set $\mathbb{R}^n/L$, is 1.

Let $\mathcal{L}_n$ denote the space of unimodular lattices in $\mathbb{R}^n$. $G = \mathrm{SL}_n(\mathbb{R})$ acts on $\mathcal{L}_n$, because the covolume of $gL$ is the covolume of $L$ multiplied by the determinant of $g$. The action is transitive: every unimodular lattice is the mage of $\mathbb{Z}^n$ under some transformation in $\mathrm{SL}_n(\mathbb{R})$. The element of $\mathrm{GL}_n(\mathbb{R})$ that effects the transformation must have determinant 1, since both lattices are unimodular.

Any endomorphism of a lattice must send the basis vectors to some $\mathbb{Z}$-linear combinations of them; that is, a lattice endomorphism may be represented by a matrix with integer entries. The endomorphism is invertible if the matrix is invertible. Since the matrix determinant is an integer, this can happen only if the determinant is $\pm 1$. Conversely, the expansion by minors formula shows that if $M$ is a matrix with integer coefficients and $\det M = \pm 1$, then $M$ is invertible over $\mathbb{Z}$, and thus represents an automorphism of a lattice.

We conclude that the space of all unimodular lattices in $\mathbb{R}^n$ can be identified with $\mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$. This geometric observation will be useful for us in the discussion of Ratner's theorems.

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

**Theorem 3.9** (Mahler compactness)**.** *For any $\varepsilon > 0$ the set $\mathcal{L}_n(\varepsilon)$ is compact.*

The detailed proof of this deep result can be found in Chapter V§3 of [BM00]. While the proof of the theorem is not entirely in line with the main thrust of the argument, we do rely heavily on Mahler compactness in our treatment of lattices.

We will now consider $n = 2$. Given a pair of vectors $v_1, v_2$ such that $\begin{pmatrix} v_1 & v_2 \end{pmatrix} \in G$ (that is, $\det \begin{pmatrix} v_1 & v_2 \end{pmatrix} = 1$), we can find a unique rotation matrix $k \in K = \mathrm{SO}_2(\mathbb{R})$ so that $kv_1$ is pointing along the positive $x$-axis and $kv_2$ is in the upper half-plane. The map $\begin{pmatrix} v_1 & v_2 \end{pmatrix} \mapsto kv_2$ gives an identification of $K\backslash G$ with the complex upper half-plane. $G$ (and in particular $\Gamma \subset G$) acts on $K\backslash G$ by multiplication on the right; this is a variant of the usual action by fractional linear transformations. Subsection 3.1 and a few following ones have more details on the geometry of this construction; we take the quotient $K\backslash G$ instead of $G$ here because we are interested in $\mathbb{H}$ rather than its unit tangent bundle. We recall a few things about the geodesic and horocycle flows. Let

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

and 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 $U$ on $G = \mathrm{SL}_2(\mathbb{R})$ by left multiplication is the horocycle flow, and the action of $A$ on $\mathrm{SL}_2(\mathbb{R})$ the geodesic flow. It is worth repeating that that $U, A, V$ are subsets of $G = \mathrm{SL}_2(\mathbb{R})$, and therefore act on $\mathcal{L}_2$. The basic commutation relations are

$$a_t u_s a_t^{-1} = u_{e^{2t}s} \quad a_t v_s a_t^{-1} = v_{e^{-2t}};$$

that is, conjugation by $a_t$ for $t > 0$ contracts $V$ and expands $U$. Projecting the orbits of the geodesic flow to $K\backslash G$ gives vertical lines or semicircles orthogonal to the $x$-axis; projecting the orbits of the horocycle flow gives horizontal lines or circles tangent to the $x$-axis.

Finally, we define *flowboxes*, which will be our basic open sets of interest.

**Definition 3.10.** Let $W_+ \subset U$, $W_- \subset V, W_0 \subset A$ be images of open intervals containing 0, i.e. the identity matrix. A *flowbox* is a subset of $G$ of the form $W_+W_0W_-g$ for some $g \in G$; it is an open set containing $g$.

Recall that right multiplication by $g$ is an isometry, so the flowbox is isometric to $W_+W_0W_-$.

3.5. **Non-divergence of unipotent flows.** The last part of this section includes an elementary non-divergence result given by Lemma 3.11 and 3.12 below. The emphasis lies in that this result holds for $\mathrm{SL}_2(\mathbb{R})$ only.

**Lemma 3.11.** *There exists an absolute constant $\varepsilon > 0$ such that the following holds: let $L \in \mathcal{L}_2$ be a unimodular lattice, then $L$ does not contain two linearly independent vectors each of length less than $\varepsilon$.*

*Proof.* Let $v_1$ be the shortest vector in $L$, and let $v_2$ be the shortest vector linearly independent of $v_1$. Then the covolume of $L$ is $\leqslant \|v_1\|\|v_2\|$, which therefore must be at least 1. Thus, we may choose $\varepsilon = 1$. ☐

**Lemma 3.12.** *Let $L \subset \mathcal{L}_2$ be a unimodular lattice. If $L$ does not contain a horizontal vector, then there exists $t \geqslant 0$ such that $a_t^{-1}L \subset \mathcal{L}_2(\varepsilon)$. Consequently, there exists a sequence of $t_n \to \infty$ such that $a_{t_n}^{-1}L \subset \mathcal{L}_2(\varepsilon)$.*

*Proof.* Suppose $L$ does not contain a horizontal vector, and $L \notin \mathcal{L}_2(\varepsilon)$. Then $L$ contains a vector $v$ of norm less than $\varepsilon$, which is not horizontal. Note that $a_t^{-1}$ stretches the second coordinate of $v$, so in particular there exists a smallest $t_0 > 0$ such that $\|a_{t_0}^{-1}v\| = \varepsilon$. Since for $t \in [0, t_0)$, the lattice $a_t^{-1}L$ contains no vectors shorter than $\varepsilon$ except $a_t^{-1}v$ and possibly multiples of it, we derive that $a_{t_0}^{-1}L \in \mathcal{L}_2(\varepsilon)$. ☐

*Remark* 3.13. We note that Lemma 3.11 and thus Lemma 3.12 are specific to dimension 2.

## 4. Interlude I: The particular case of $G = \mathrm{SL}_2(\mathbb{R})$

Now we wanna discuss Ratner's theorem in the case of homogeneous spaces under $\mathrm{SL}_2(\mathbb{R})$. We have recalled the geometry of the hyperbolic plane and defined geodesic and horocyclic flows on the unit tangent bundle as well as the identification with $\mathrm{PSL}_2(\mathbb{R})$ in Section 3. Our interest especially lies in the ergodicity of the geodesic and the horocyclic flow on a finite volume hyperbolic surface. By the way, we show the invariance of Haar measure under the flows before that ergodicity.

There are two main results of this section: one is the classification theorem of ergodic measures which are invariant under horocyclic flow; the other is Ratner's orbit closure theorem for $G = \mathrm{SL}_2(\mathbb{R})$. This section refers to Chapter II§1, II§3, IV of Bekka–Mayer's book [BM00].

4.1. **The classification of $U$-invariant measures.** Note that for $L \in \mathcal{L}_2$, the $U$-orbit of $L$ is closed if and only if $L$ 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 $\nu$ denote the Haar measure on $\mathcal{L}_2 = G/\Gamma = \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$. The measure $\nu$ is normalized so that $\nu(\mathcal{L}_2) = 1$. Recall that $\nu$ is ergodic for both the horocycle and the geodesic flows, and this follows from the Moore ergodicity theorem. Our main goal is the following.

**Theorem 4.1.** *Suppose $\mu$ is an ergodic $U$-invariant probability measure on $\mathcal{L}_2$. Then either $\mu$ is supported on a closed orbit, or $\mu$ is the Haar measure $\nu$.*

*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 $\mu$ is an ergodic $U$-invariant probability measure on $\mathcal{L}_2$. By ergodicity of $\mu$, either $\mu(\mathcal{L}_2') = 0$ or $\mu(\mathcal{L}_2') = 1$. If the latter holds, it is easy to show that $\mu$ 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 $f\colon \mathcal{L}_2 \to \mathbb{R}$ and $\varepsilon > 0$ such that

$$\left| \int_{\mathcal{L}_2} f d\mu - \int_{\mathcal{L}_2} f d\nu \right| > \varepsilon. \tag{1}$$

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 $L'' \in \mathcal{L}_2$,

$$|f(vaL'') - f(L'')| < \frac{\varepsilon}{3}. \tag{2}$$

Recall that $\pi\colon G \to G/\Gamma \cong \mathcal{L}_2$ denotes the natural projection. Since $\mathcal{L}_2(\varepsilon_0)$ is compact the injectivity radius on $\mathcal{L}_2(\varepsilon_0)$ 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 $\pi$ to the flowbox $W_-W_0W_+g$ is injective. We may also assume that $W_- \subset W_-'$ and $W_0 \subset W_0'$. Let $\delta = \nu(W_-W_0W_+)$ denote the Lebesque measure of the flowbox.

By Corollary 2.10 applied to the Lebesque measure $\nu$, there exists a set $E \subset \mathcal{L}_2$ with $\nu(E) < \delta$ and $T_1 > 0$ such that for any interval $I$ with $|I| \geqslant T_1$ and any $L' \notin E$,

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

Now let $L$ 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 $T_2$,

$$\left| \frac{1}{|I|} \int_I f(u_t L) dt - \int_{\mathcal{L}_2} f d\mu \right| < \frac{\varepsilon}{3}. \tag{4}$$

Since $\mu(\mathcal{L}_2') = 0$, we may assume that $L$ does not contain any horizontal vectors. Then by repeatedly applying Lemma 3.12 we can construct arbitrarily large $t > 0$ such that

$$a_t^{-1} L \in \mathcal{L}_2(\varepsilon).$$

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

$$Q = (a_t W_- a_t^{-1}) W_0 (a_t W_+ a_t^{-1}) L, \tag{5}$$

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 $L' \in Q$, let $I(L')$ denote the connected component containing the origin of the set

$$\{t \in \mathbb{R} \colon u_t L' \in Q\}.$$

Note that the length of $I(L')$ is independent of $L'$ (it is just the length of $W_+$ multiplied by $e^{2t}$). By choosing $t$ sufficiently large, we may assume that $|I(L')| \geqslant \max(T_1, T_2)$. By commutation relations $a_t u_s a_t^{-1} = u_{e^{2t}s}$ and $a_t v_s a_t^{-1} = v_{e^{-2t}}$, we know $a_t W_- a_t^{-1} \subset W'_-$. Also, by construction, $W_0 \subset W'_0$. Thus, by (2), we have for any $L' \in Q$,

$$\left| \frac{1}{|I(L')|} \int_{I(L')} f(u_t L')dt - \frac{1}{|I(L)|} \int_{I(L)} f(u_t L)dt \right| < \frac{\varepsilon}{3}, \tag{6}$$

which 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 $L' \in Q \cap E^c$. Now (3) holds with $I = I(L')$, and (4) holds with $I = I(L)$. These estimates together with (5) contradict the initial inequality (1). □

*Remark* 4.2. The above proof works with minor modifications if $\Gamma$ is an arbitrary lattice in $\mathrm{SL}_2(\mathbb{R})$, and not just $\mathrm{SL}_2(\mathbb{Z})$. Furthermore, if $\Gamma$ is a uniform lattice in $\mathrm{SL}_2(\mathbb{R})$ then the horocycle flow on $G/\Gamma$ is uniquely ergodic. This is a theorem of Furstenberg.

The proof of Theorem 4.1 does not generalize to classification of measures invariant under a one-parameter unipotent subgroup on e.g. $\mathcal{L}_n, n \geqslant 3$. Completely different ideas are needed.

4.2. **Comment: Horospherica subgroups and a theorem of Dani.** The key property of $U$ in dimension 2 which is used in the proof of Theorem 4.1 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 $\mathrm{SL}_2(\mathbb{R})$. An argument similar in spirit to the proof of Theorem 4.1 can be used to classify the measures invariant under the action of a horospherical subgroup. This is a theorem of Dani [Dan81] which was proved before Ratner's measure classification theorem. However, the details, and in particular the non-divergence results needed are much more complicated. In fact, the horospherical case also allows for an analytic approach.

4.3. **Orbit closures.** We are now ready to prove Ratner's orbit closure theorem for $G = \mathrm{SL}_2(\mathbb{R})$. It is a general version of the phenomenon which we have stated at the very beginning of Section 1.

**Lemma 4.3.** *For $L \in \mathcal{L}_2$, the $U$-orbit of $L$ is closed if and only if $L$ contains a horizontal vector.*

*Proof.* Note that the action of $U$ preserves the $y$-components of vectors, and fixes horizontal vectors. Therefore, if $v \in L$ is a horizontal vector, then $v$ is contained in $u_t L$ for all $t$, and therefore is contained in $\overline{U(L)}$. Now, let a matrix for $L$ be $\left(\begin{smallmatrix} a & c \\ 0 & d \end{smallmatrix}\right)$ containing the fundamental horizontal vector $(a, 0)^{\mathrm{T}}$. Then all vectors in $U(L)$ will have $y$-components that are multiples of $d$, and in particular the horizontal vectors in $\overline{U(L)}$ will be the same as those in $L$. Consequently, the matrix of any lattice $L'$ in $\overline{U(L)}$ can be written as $\left(\begin{smallmatrix} a & c' \\ 0 & d' \end{smallmatrix}\right)$. Note that the covolume of the lattice is $|ad|$, and therefore $d' = \pm d$; without loss of generality, let $d' = d$. We finally observe that $c' = c + td$ for some $t$ since $d \neq 0$ (otherwise $L$ is not a lattice); therefore, $L' = u_t(L)$.

On the other hand, suppose $L$ does not contain a horizontal vector; then it is generated by two vectors whose $y$-coordinates are incommensurable. In particular, $L$ contains vectors whose $y$-coordinates are arbitrarily close to 0. Let $v_n \in L$ be primitive vectors satisfying $0 < (v_n)_y < \frac{1}{n}$. Pick $t$ such that $u_n = u_t v_n = (1, (v_n)_y)^{\mathrm{T}} \in u_t(L)$; and find a second vector generating the lattice $u_t(L)$. It can be chosen so that its $x$-coordinate is in $[0, 1]$; the $y$ coordinate must be approximately 1 because the covolume of the lattice is 1. Letting $(v_2)_n$ be the sequence of such second vectors, we note that all the $(v_2)_n$ are contained in a compact set, and therefore have a converging sequence. Then the sequence of pairs $(u_{n_k}, (v_2)_{n_k})$ converges to some pair of generators for a lattice with the first vector horizontal: that is, $U(L) \neq \overline{U(L)}$ in this case.

Now, any closed $U$-orbit supports a $U$-invariant probability measure. Moreover, we have the Haar measure $\nu$ on $\mathcal{L}_2 = G/\Gamma$, normalized so that $\nu(\mathcal{L}_2) = 1$; this $\nu$ is ergodic for both the horocycle and the geodesic flows. Ratner's measure classification theorem asserts that these are the only $U$-invariant ergodic probability measures on $\mathcal{L}_2$. □

**Theorem 4.4** (Orbit Closures in Dimension 2)**.** *Let $L \in \mathcal{L}_2 = \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$. Then the $U$-orbit of $L$ is either closed or dense.*

*Proof.* Suppose $UL$ is not closed. By Lemma 4.3, this means that $L \notin \mathcal{L}'_2$. We wish to show that $UL$ passes though every open set $\widetilde{O} \subset \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$.

Find an open subset $O$ of a compact subset $C$ of $\widetilde{O}$ (we like working with functions of compact support, so all of $\widetilde{O}$ might be too large for us). Let $f$ be a uniformly continuous, nonnegative function supported on $C$ and equal to 1 on $O$; then $0 < \nu(O) \leqslant \int_{\mathcal{L}_2} f d\nu \leqslant \nu(\widetilde{O})$. That is, we approximate the characteristic function of $O$ by a uniformly continuous function of compact support. Let $\varepsilon < \nu(O)$.

Since our $U$-orbit is not closed, it is the orbit of some lattice $L \notin \mathcal{L}'_2$. Let the sequence $t_n \to \infty$, the flowbox $Q$, and the exceptional set $E$ be as in the proof of measure classification above. Since $\mu(Q) > \mu(E)$, for a large enough $t_n$ we can put together (4) and (6) to find an interval $I$ such that

$$\left| \frac{1}{|I|} \int_I f(u_t L) dt - \int_{\mathcal{L}_2} f d\nu \right| < \varepsilon.$$

However, this is only possible if $f(u_t L)$ actually visits $O \subset \widetilde{O}$; since $\widetilde{O}$ was arbitrary, we conclude that the $U$-orbit of $L$ is dense. □

## 5. Interlude II: The particular case of $H = \mathrm{SL}_2(\mathbb{R})$

We aim to discuss Ratner's Theorem 1.2 in the case of $\mathrm{SL}_2(\mathbb{R})$-invariant measures on general homogeneous spaces. This section contains some of the background needed in the proof: Lie groups and Lie algebras, finite-dimensional representations of $\mathrm{SL}_2(\mathbb{R})$, and Mautner phenomenon for $\mathrm{SL}_2(\mathbb{R})$, as well as explains and illustrates the proof for it.

The first three subsections in the following list the facts and notions needed for the proof of Theorem 1.2, of which, except for the Mautner's Phenomenon in Subsection 5.3, can be found in any introduction to Lie groups respect to ergodic theory. We strongly recommend Einsiedler's article [Ein06] that also includes most of the background needed to be the essential material for dealing with this case.

5.1. **Lie groups and Lie algebras.** At the identity element $e \in G$ for a Lie group $G$, we define the Lie algebra $\mathfrak{g}$ to be the tangent space there. The exponential map $\exp\colon \mathfrak{g} \to G$ and the locally defined inverse, the logarithm map, give local isomorphisms $\mathfrak{g} \cong G$. For any $g \in G$ the derivative of the conjugation map is the adjoint transformation $\mathrm{Ad}_g\colon \mathfrak{g} \to \mathfrak{g}$ and satisfies

$$\exp \mathrm{Ad}_g(v) = g \exp(v) g^{-1}, \quad \forall g \in G, v \in \mathfrak{g}.$$

For linear groups this could not be easier, the Lie algebra is a linear subspace of the space of matrices, $\exp(\cdot)$ and $\log(\cdot)$ are defined as usual by power series, and the adjoint transformation $\mathrm{Ad}_g$ is still conjugation by $g$.

Closed subgroups $L < G$ are almost completely described by their respective Lie algebras $\mathfrak{l}$ inside $\mathfrak{g}$ as follows. Let $L^\circ$ be the connected component of $L$ that contains the identity $e$. Then the Lie algebra $\mathfrak{l}$ of $L$ (and $L^\circ$) uniquely determines $L^\circ$, which is the subgroup generated by $\exp(\mathfrak{l})$. Moreover, any element $\ell \in L$ sufficiently close to $e$ is actually in $L^\circ$ and equals $\ell = \exp(v)$ for some small $v \in \mathfrak{l}$.

Using an inner product on $\mathfrak{g}$ we can define a left invariant Riemannian metric $d(\cdot,\cdot)$ on $G$. We will be using the restriction of $d(\cdot,\cdot)$ to subgroup $L < G$ and denote by $B^L_r$ the $r$-ball in $L$ around $e \in L$.

If $\Gamma < G$ is a discrete subgroup, then $X = \Gamma\backslash G$ has a natural topology and in fact a metric defined by

$$d(\Gamma g, \Gamma h) := \min_{\gamma \in \Gamma} d(g, \gamma h), \quad \forall g, h \in G,$$

which uses left invariance of $d(\cdot,\cdot)$. With this metric and topology $X$ can locally be described by $G$ as follows. For any $x \in X$ there is an $r > 0$ such that the map $\iota\colon g \mapsto xg$ is an homeomorphism between $B_r^G$ and a neighborhood of $x$. Moreover, if $r$ is small enough, $\iota\colon B_r^G \to X$ is in fact an isometric embedding. For a given $x$, a number $r > 0$ with these properties is called an *injective radius* at $x$.

5.2. **Finite-dimensional representation of** $\mathrm{SL}_2(\mathbb{R})$**.** The first property of $\mathrm{SL}_2(\mathbb{R})$ we will need is the following standard fact. Let $V$ be a finite dimensional real vector space and suppose $\mathrm{SL}_2(\mathbb{R})$ acts on $V$. Then any $\mathrm{SL}_2(\mathbb{R})$-invariant subspace $W < V$ has an $\mathrm{SL}_2(\mathbb{R})$-invariant complement $W' < V$ such that $V = W \oplus W'$. The above implies that all finite dimensional representations of $\mathrm{SL}_2(\mathbb{R})$ can be written as a direct sum of irreducible representations.

The second fact we need is the descrption of these irreducible representations. Let $A = (1,0)^{\mathrm{T}}$ and $B = (0,1)^{\mathrm{T}}$ denote the standard basis of $\mathbb{R}^2$ so that the unipotent multiplication

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} A = A, \quad \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} B = B + tA.$$

Any irreducible representation is obtained as a symmetric tensor product $\mathrm{Sym}^n(\mathbb{R}^2)$ of the standard representation on $\mathbb{R}^2$ for some $n$. $\mathrm{Sym}^n(\mathbb{R}^2)$ has $A^n, A^{n-1}B, \ldots, B^n$ as a basis, and every element we can view as a homogeneous polynomial $p(A,B)$ of degree $n$. The action of multiplication above can now be described by substitution, $p(A,B)$ is mapped to $p(A, B+tA)$. More concretely, $p(A,B) = c_0 A^n + c_1 A^{n-1} B + \cdots + c_n B^n$ is mapped to

$$p(A, B+tA) = (c_0 + \cdots + c_n t^n) A^n + (c_1 + \cdots + c_n n t^{n-1}) A^{n-1} B + \cdots + c_n B^n,$$

where the coefficients in front of the various powers of $t$ are the original components of the vector $p(A,B)$ multiplied by binomial coefficients. Notice that all components of $p(A,B)$ appear in the image vector in the component corresponding to $A^n$. Moreover, for any component of $p(A,B)$ the highest power of $t$ it gets multiplied by appears in the resulting component corresponding to $A^n$. For that reason, when $t$ grows (and say $p(A,B)$ is not just a multiple of $A^n$) the image of $p(A,B)$ under the unipotent multiplication will always grow fastest in the direction of $A^n$ when $t \to \infty$.

5.3. **Mautner's phenomenon for** $\mathrm{SL}_2(\mathbb{R})$**.** To be able to apply the ergodic theorem as stated in Subsection 5.4 in the proof of Theorem 1.2, we will need to know that the $\mathrm{SL}_2(\mathbb{R})$-invariant and ergodic probability measure is also ergodic under a one-parameter flow. The corresponding fact is best formulated in terms of unitary representations and is due to Moore [Moo80] and is known as the *Mautner phenomenon*. For completeness we prove the special case needed.

**Proposition 5.1.** *Let $\mathfrak{H}$ be a Hilbert space, and suppose $\phi\colon \mathrm{SL}_2(\mathbb{R}) \to U(\mathfrak{H})$ is a continuous representation on $\mathfrak{H}$. In other words, $\phi$ is a homomorphism into the group of unitary automorphisms $\mathbb{U}(\mathfrak{H})$ of $\mathfrak{H}$ such that for every $v \in \mathfrak{H}$ the vector $\phi(g)(v) \in \mathfrak{H}$ depends continuously on $g \in \mathrm{SL}_2(\mathbb{R})$. Then any vector $v \in \mathfrak{H}$ that is invariant under the upper unipotent matrix group*

$$U = \left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \right\}$$

*is in fact invariant under* $\mathrm{SL}_2(\mathbb{R})$.

Since any measure preserving action on $(X,\mu)$ gives rise to a continuous unitary representation on $\mathfrak{H} = L^2(X,\mu)$ the above gives immediately what we need (see also Proposition 5.2 in [Rat90] for another elementary treatment):

**Corollary 5.2.** *Let $\mu$ be an $H$-invariant and ergodic probability measure on $X = \Gamma\backslash G$ with $\Gamma < G$ discrete, and $H < G$ isomorphic to* $\mathrm{SL}_2(\mathbb{R})$*. Then $\mu$ is also ergodic with respect to the one-parameter unipotent subgroup $U$ of $H$ corresponding to the upper unipotent subgroup in* $\mathrm{SL}_2(\mathbb{R})$*.*

In fact, an invariant function $f \in L^2(X,\mu)$ that is invariant under $U$ must be invariant under $\mathrm{SL}_2(\mathbb{R})$ by Proposition 5.1. Since the latter group is assumed to be ergodic, the function must be constant as required.

*Proof of Proposition 5.1.* Following Margulis' idea, define the auxiliary function $p(g) = (\phi(g)v, v)$. Notice first that the function $p(\cdot)$ characterizes invariance in the sense that $p(g) = (v,v)$ implies $\phi(g)v = v$.

By continuity of the representation $p(\cdot)$ is also continuous. Moreover, by our assumption on $v$ the map $p(\cdot)$ is bi-$U$-invariant since

$$p(ugu') = (\phi(u)\phi(g)\phi(u')v, v) = (\phi(g)v, \phi(u^{-1})v) = p(g).$$

Let $\varepsilon, r, s \in \mathbb{R}$ and calculate

$$\begin{pmatrix}1 & r\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ \varepsilon & 1\end{pmatrix}\begin{pmatrix}1 & s\\ 0 & 1\end{pmatrix} = \begin{pmatrix}1+r\varepsilon & r+s+rs\varepsilon\\ \varepsilon & 1+s\varepsilon\end{pmatrix}.$$

Now fix some $t \in \mathbb{R}$, let $\varepsilon$ be close to zero but nonzero, choose $r = (e^t-1)/\varepsilon$ and $s = -r/(1+r\varepsilon)$. Then the above matrix simplifies to$\left(\begin{smallmatrix}e^t & 0\\ \varepsilon & e^{-t}\end{smallmatrix}\right)$. In particular, this shows that

$$p\begin{pmatrix}1 & 0\\ \varepsilon & 1\end{pmatrix} = p\begin{pmatrix}e^t & 0\\ \varepsilon & e^{-t}\end{pmatrix}$$

is both close to $p(e)$ and to $p\left(\begin{smallmatrix}e^t & 0\\ 0 & e^{-t}\end{smallmatrix}\right)$. Therefore, the latter equals $(v,v)$ which implies that $v$ is invariant under $\left(\begin{smallmatrix}e^t & 0\\ 0 & e^{-t}\end{smallmatrix}\right)$ as mentioned before.

The above implies now that $p(\cdot)$ is bi-invariant under the diagonal subgroup. Using this and the above argument once more, it follows that $v$ is also invariant under $\left(\begin{smallmatrix}1 & 0\\ s & 1\end{smallmatrix}\right)$ for all $s \in \mathbb{R}$. □

5.4. **Outline of the 4-steps proof for $H = \mathrm{SL}_2(\mathbb{R})$.** In this subsection we prove Theorem 1.2 for $H = \mathrm{SL}_2(\mathbb{R})$ using the prerequisites that we have discussed. Let us mention again that the general outline of the proof is very similar to the strategy Ratner used to prove her theorems. In this outline of the proof, many details are omitted for their is no space in the report to write them down.

From now on let $\mu$ be an $H$-invariant and ergodic probability measure on $X = \Gamma\backslash G$. Our goal is in the following.

**Theorem 5.3** (Ratner, $H = \mathrm{SL}_2(\mathbb{R})$)**.** *Let $G$ be a Lie group, $\Gamma < G$ a discrete subgroup, and $H < G$ a subgroup isomorphic to $\mathrm{SL}_2(\mathbb{R})$. Then any $H$-invariant and ergodic probability measure $\mu$ on $X = \Gamma\backslash G$ is homogeneous, i.e. there exists a closed connected subgroup $L < G$ containing $H$ such that $\mu$ is $L$-invariant and some $x_0 \in X$ such that the $L$-orbit $x_0L$ is closed and supports $\mu$. In other words $\mu$ is an $L$-invariant volume measure on $x_0L$.*

**Step 1.** It is easy to check that

$$\mathrm{Stab}_G(\mu) = \{g \in G\colon \text{right multiplication with } g \text{ on } X \text{ preserves } \mu\}$$

is a closed subgroup of $G$. Let $L = \mathrm{Stab}_G(\mu)^\circ$ be the connected component. Then as discussed any element of $\mathrm{Stab}_G(\mu)$ sufficiently close to $e$ belongs to $L$. Also since $\mathrm{SL}_2\,\mathbb{R})$ is connected we have $H < L$.

We will show that $\mu$ is concentrated on a single orbit of $L$, i.e. that there is some $L$-orbit $L.x_0$ of measure one $\mu(L.x_0) = 1$. Then by $L$-invariance of $\mu$ and uniqueness of Haar measure, $\mu$ would have to be the $L$-invariant volume form on this orbit $L.x_0$. However, since $\mu$ is assumed to be a probability measure this also implies that the orbit L. $x_0$ is closed as seen in the next lemma.

**Lemma 5.4.** *If $\mu$ is concentrated on a single $L$-orbit $L.x_0$ and is $L$ invariant, then $L.x_0$ is closed and $\mu$ is supported on $L.x_0$.*

The main argument will be to show that if $\mu$ is not concentrated on a single orbit of $L$, then there are other elements of $\mathrm{Stab}_G(\mu)$ close to $e$. This shows that we should have started with a bigger subgroup $L'$. If we repeat the argument with this bigger $L'$, we will either achieve our goal or make $L'$ even bigger. We start by giving a local condition for a measure $\mu$ to be concentrated on a single orbit.

**Lemma 5.5.** *Suppose $x_0 \in X$ has the property that $\mu(B_\delta^L \cdot x_0) > 0$ for some $\delta > 0$, then $\mu$ is concentrated on $L.x$. So either the conclusion of Theorem 5.3 holds for $L$ and $x_0$, or for every $x_0$ we have $\mu(B_\delta^L \cdot x_0) = 0$.*

We will be achieving the assumption to the last lemma by studying large sets $X' \subset X$ of points with good properties. Let $x_0 \in X'$ be such that all balls around $x_0$ have positive measure. Suppose $X'$ has the property that points $x'$ close to $x_0$ that also belong to $X'$ give rise to additional invariance of $\mu$ unless $x$ and $x'$ are locally on the same $L$-orbit (i.e. $x' = \ell.x$ for some $\ell \in L$ close to $e$). Then either $L$ can be made bigger or $B_\delta^L.x \cap X' \subset B_\delta^L.x_0$ for some $\delta > 0$ and therefore the latter has positive measure. However, to carry that argument through requires a lot more work. We start by a less ambitious statement where two close by points in a special position from each other give rise to invariance of $\mu$. Recall that

$$U = \left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \right\}$$

**Proposition 5.6.** *There is a set $X' \subset X$ of $\mu$-measure one such that if $x, x' \in X'$ and $x' = c.x$ with*

$$c \in C_G(U) = \{g \in G \colon gu = ug \text{ for all } u \in U\},$$

*then $c$ preserves $\mu$.*

The set $X'$ in the above proposition we define to be the set of $\mu$ generic points (for the one paramenter subgroup defined by $U$). A point $x \in X$ is *$\mu$-generic* if

$$\frac{1}{T}\int_0^T f(u_t.x)dt \to \int f d\mu, \quad T \to \infty$$

for all compactly supported, continuous functions $f \colon X \to \mathbb{R}$. Recall that by the Mautner phenomenon $\mu$ is $U$-ergodic. Now the ergodic theorem implies that the set $X'$ of all $\mu$-generic points has measure one. (Here one first applies the ergodic theorem for a countable dense set of compactly supported, continuous functions and then extends the statement to all such functions by approximation.)

**Step 2.** In Proposition 5.6 we derived invariance of $\mu$ but only if we have two points $x, x' \in X'$ that are in a very special relationship to each other. On the other hand if $\mu$ is not supported on the single $L$-orbit, then we know that we can find many $y, y' \in X'$ that are close together but are not on the same $L$-leaf locally by Lemma 5.5. Without too much work we will see that we can assume

$$y' = \exp(v).y, \quad v \in \mathfrak{l}'$$

where $\mathfrak{l}'$ is an $\mathrm{SL}_2(\mathbb{R})$-invariant complement in $\mathfrak{g}$ of the Lie algebra $\mathfrak{l}$ of $L$, see Lemma 5.8. What we are going to describe is a version of the so-called H-principle as introduced by Ratner [Rat82a] [Rat83] and generalized in [Mor05].

By applying the same unipotent matrix $u \in U$ to $y$ and $y'$ we get

$$u.y' = (u\exp(v)u^{-1}).(u.y) = \exp(\mathrm{Ad}_u(v)).(u.y)$$

In other words, the divergence of the orbits through $y$ and $y'$ can be described by conjugation in $G^-$ or even by the adjoint representation on $\mathfrak{g}$. Since $H$ is assume to be isomorphic to $\mathrm{SL}_2(\mathbb{R})$ we will be able to use the theory on representations as in Section 5.2. In particular, recall that the fastest divergence is happening along a direction which is stabilized by $U$. Since all points on the orbit of a $\mu$-generic point are also $\mu$-generic, one could hope to flow along $U$ until the two points $x = u.y, x' = u.x'$ differ significantly but not yet to much. Then $y' = u.x' = h.(u.x) = h.y$ with $h$ *almost* in $C_G(U)$. To fix the *almost* in this statement we will consider points that are even closer to each other, flow along $U$ for a longer time, and get a sequence of pairs of $\mu$ generic points that differ more and more by some element of $C_G(U)$. In the limit we hope to get to points that differ precisely by some element of $C_G(U)$ which is not in $L$.

The main problem is that limits of $\mu$-generic points need not be $\mu$ generic (even for actions of unipotent groups). Therefore, we need to introduce quite early in the argument a compact subset $K \subset X'$ of almost full measure that consists entirely of $\mu$-generic points. When constructing $u.x', u.x$ we will make sure that they belong to $K$ – this way we will be able to go to the limit and get $\mu$-generic points that differ by some element of $C_G(U)$.

We are now ready to proceed more rigorously.

**Step 3.** Let $X'$ be the sets of $\mu$-generic points as above, and let $K \subset X'$ be compact with $\mu(K) > 0.9$. By the ergodic theorem

$$\frac{1}{T}\int_0^T 1_K(u_t.y)dt \to \mu(K)$$

for $\mu$-a.e. $y \in X$. In particular, we must have for a.e. $y \in X$

$$\frac{1}{T}\int_0^T 1_K(u_t.y)dt > 0.8, \quad T \gg 0.$$

Here $T$ may depend on $y$ but by choosing $T_0$ large enough we may assume that the set

$$X_1 = \left\{ y \in X \colon \frac{1}{T}\int_0^T 1_K(u_t.y)dt > 0.8 \text{ for all } T \geqslant T_0 \right\}$$

has measure $\mu(X_1) > 0.99$. By definition points in $X_1$ visit $K$ often enough so that we will be able to find for any $y, y' \in X_1$ many common values of $t$ with $u_t.y, u_t.y' \in K$.

The last preparation we need will allow us to find $y, y' \in X_1$ that differ by some $\exp(v)$ with $v \in \mathfrak{l}'$. For this we define

$$X_2 = \left\{ z \in X \colon \frac{1}{m_L(B_1^L)}\int_{B_1^L} 1_{X_1}(\ell.z)d(m_L(\ell)) > 0.9 \right\}$$

where $m_L$ is a Haar measure on $L$. Any other smooth measure would do here as well.

**Lemma 5.7.** $\mu(X_2) > 0.9$.

Let as before $\mathfrak{l} \subset \mathfrak{g}$ be the Lie algebra of $L < G$ and let $\mathfrak{l}' \subset \mathfrak{g}$ be an $\mathrm{SL}_2(\mathbb{R})$-invariant complement of $\mathfrak{l}$ in $\mathfrak{g}$. Then the map $\phi\colon \mathfrak{l}' \times \mathfrak{l} \to G$ defined by $\phi(v, w) = \exp(v)\exp(w)$ is $C^\infty$ and its derivative at $(0, 0)$ is the embedding of $\mathfrak{l}' \times \mathfrak{l}$ into $\mathfrak{g}$. Therefore, $\phi$ is locally invertible so that every $g \in G$ close to $e$ is a unique product $g = \exp(v)\ell$ for some $\ell \in L$ close to $e$ and some small $v \in \mathfrak{l}'$. We define $\pi_L(g) = \ell$. For simplicity of notation we assume that this map is defined on an open set containing $B_1^L$ (if necessary we rescale the metric).

**Lemma 5.8.** *For any $\varepsilon > 0$ there exists $\delta > 0$ such that for $g \in B_\delta^G$, and $z, z' = g.z \in X_2$ there are $\ell_2 \in B_1^L$ and $\ell_1 \in B_\varepsilon^L(\ell_2)$ with $\ell_1.z, \ell_2.z' \in X_1$ and $\ell_2 g \ell_1^{-1} = \exp(v)$ for some $v \in B_\varepsilon^{\mathfrak{l}'}(0)$.*

**Step 4.** Let $x_0 \in X_2 \cap \operatorname{supp}\mu|_{X_2}$ so that $\mu((B_\delta^G.x_0) \cap X_2) > 0$ for all $\delta > 0$. Now one of the following two statements must hold:

(1) there exists some $\delta > 0$ such that $B_\delta^G.x_0 \cap X_2 \subset B_\delta^L.x_0$, or

(2) for all $\delta > 0$ we have $B_\delta^G.x_0 \cap X_2 \not\subset B_\delta^L.x_0$.

We claim that actually only (1) above is possible if $L$ is really the connected component of $\mathrm{Stab}_G(\mu)$. Assuming this has been shown, then we have $\mu(B_\delta^L.x_0) > 0$ which was the assumption to Lemma 5.5. Therefore, $\mu(L.x_0) = 1$ and by Lemma 5.4, $L.x_0 \subset X$ is closed – Theorem 5.3 follows. So what we really have to show is that (2) implies that $\mu$ is invariant under a one parameter subgroup that does not belong to $L$.

**Lemma 5.9.** *Assuming (2) there are for every $\varepsilon > 0$ two points $y, y' \in X_1$ with $d(y, y') < \varepsilon$ and $y' = \exp(v).y$ for some nonzero $v \in B_\varepsilon^{\mathfrak{l}'}(0)$.*

Using $y, y' \in X_1$ and $v \in \mathfrak{l}'$ for all $\varepsilon > 0$ as in the above lemma we will show that $\mu$ is invariant under a one-parameter subgroup that does not belong to $L$. For this it is enough to show the following:

**Claim.** For any $\eta > 0$ there exists a nonzero $w \in B_\eta^{\mathfrak{l}'}(0)$ such that $\mu$ is invariant under $\exp(w)$.

To see that this is the remaining assertion, notice that we then also have invariance of $\mu$ under the subgroup $\exp(\mathbb{Z}w)$. While this subgroup could still be discrete, when $\eta \to 0$ we find by compactness of the unit ball in $\mathfrak{l}'$ a limiting one parameter subgroup $\exp(\mathbb{R}w)$ that leaves $\mu$ invariant.

We start proving the claim. Let $\eta > 0$ be fixed, and let $\varepsilon > 0$, $y, y' \in X_1$, and $v \in B_\varepsilon^{\mathfrak{l}'}(0)$ as above. We will think of $\varepsilon$ as much smaller than $\eta$ since we will below let $\varepsilon$ shrink to zero while not changing $\eta$. Let $\mathrm{Sym}^n(\mathbb{R}^2)$ be an irreducible representation as in Section 5.2, and let $p = p(A,B) \in \mathrm{Sym}^n(\mathbb{R}^2)$. Recall that $u_t = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$ applied to $p(A,B)$ gives $p(A, B+tA)$. We define

$$T_p = \frac{\eta}{\max(|c_1|, \ldots, |c_n|^{1/n})}$$

and set $T_p = \infty$ if the expression on the right is not defined. The significance of $T_p$ is that for $t = T_p$ at least one term in the sum $(c_0 + c_1 t + \cdots + c_n t^n)$ is of absolute value one while all others are less than that – recall that this sum is the coefficient of $A$ in $p(A, B+tA)$. To extend this definition to $\mathfrak{l}'$ which is not necessarily irreducible we split $\mathfrak{l}'$ into irreducible representations $\mathfrak{l}' = \bigoplus_{j=1}^k V_j$ and define for $v = (p_j)_{j=1,\ldots,k}$:

$$T_v = \min_j T_{p_j}.$$

**Lemma 5.10.** *There exists constants $n > 0$ and $C > 0$ that only depend on $\mathfrak{l}'$ such that for $v \in B_\varepsilon^{\mathfrak{l}'}(0)$ and $t \in [0, T_v]$ we have*

$$\mathrm{Ad}_{u_t}(v) = w + O(\varepsilon^{1/n})$$

*where $w \in B_{C\eta}^{\mathfrak{l}'}(0)$ is fixed under the subgroup $U = u_{\mathbb{R}}$. Here we write $O(\varepsilon^{1/n})$ to indicate a vector in $\mathfrak{l}'$ of norm less than $C\varepsilon^{1/n}$.*

If $v$ is already fixed by $U$ then $T_v = \infty$ (and other way around) and the above statement is rather trivial since $w = v$. Moreover, by definition of $X_1$ we have

$$\frac{1}{T}\int_0^T 1_K(u_t x_i)\,dt > 0.8$$

for $i = 1, 2$. From this it follows that there is some $t \in [0, T_0]$ with $u_t.x_1, u_t \cdot x_2 \in K$. Since $K \subset X'$ Lemma 5.6 proves (assuming $\varepsilon < \eta$) the claim in that case and we may from now on assume that $v$ is not fixed under the action of $U$ and so $T_v < \infty$.

**Lemma 5.11.** *There exists a constant $c > 0$ that only depends on $\mathfrak{l}'$ such that the decomposition $\mathrm{Ad}_{u_t}(v) = w + O(\varepsilon^{1/n})$ as above satisfies $\|w\| > c\eta$ for $t \in E_v$ where $E_v \subset [0, T_v]$ has Lebesgue measure at least $0.9T_v$.*

Recall that case (1) from the beginning of this section implies Theorem 5.3 and that we are assuming case (2). Moreover, recall that this implies for all $\varepsilon > 0$ the existence of $y, y' \in X_1$ with $y' = \exp(v).y$ for some nonzero $v \in B_\varepsilon^{\mathfrak{l}'}(0)$ by Lemma 5.9. By definition of $X_1$ the sets

$$E_T = \{t \in [0,T] \colon u_t.y \in K\},$$
$$E'_T = \{t \in [0,T] \colon u_t.y' \in K\}$$

have Lebesgue measure bigger than $0.8T$ whenever $T \geqslant T_0$. From the definition it is easy to see that $T_v \geqslant T_0$ once $\varepsilon$ and therefore $v$ are sufficiently small, so we can set to $T = T_v$. Moreover, let $E_v$ be as in Lemma 5.11. Then the union of the complements of these three sets in $[0, T_v]$ has Lebesgue measure less than $0.5T_v$. Therefore, there exists some $t \in E_{T_v} \cap E'_{T_v} \cap E_v$. We set $x = u_t.y$ and $x' = u_t \cdot y'$, both belonging to $K$ by definition of $E_{T_v}$ and $E'_{T_v}$. Moreover, $x' = \exp(w + O(\varepsilon^{1/n})).X$, where $w \in \mathfrak{l}'$ is stabilized by $U$ and satisfies $c\eta \leqslant \|w\| \leqslant C\eta$ by Lemma 5.10 and 5.11. We let $\varepsilon \to 0$ and choose converging subsequences for $x, x'$, and $w$. This shows the existence of $x$, $x' = \exp(w) \cdot x \in K$ and $w \in \mathfrak{l}'$ with $c\eta \leqslant \|w\| \leqslant C\eta$ which is stabilized by $U$. This is in effect our earlier claim which as we have shown implies that $\mu$ is invariant under a one-parameter subgroup not belong into $L$. This concludes the proof of Theorem 5.3.

## 6. Ratner's "tours de force" in full generality

This section discusses some of the ideas and problems of the proof of Ratner's Theorem in the general case. The phrase "tour de force" in French translates to be "masterpiece" in English. There will be no space in this report to give the technical details of the proof, so we only concentrate on the outline of the proof, comparing everything to the case of $\mathrm{SL}_2(\mathbb{R})$-invariant measures, and focus on the important ideas such as non-divergence of unipotent flows.

Chapter 5 of Morris' book [Mor05] is the main reference. We point out there are also useful remarks in Ghys' paper from Séminaire Bourbaki, see [Ghy92].

6.1. **Ratner's 9-steps proof.** This part cites from [Mor05], and we give out a fairly complete outline of proof, modulo some propositions and lemmas in need that are postponed to Subsection 6.2. Let us first state Theorem 1.2 again.

**Theorem 6.1** (Ratner, 1990)**.** *If $G$ is a closed, connected subgroup of $\mathrm{SL}_\ell(\mathbb{R})$ for some $\ell$, $\Gamma$ is a discrete subgroup of $G$, $u^t$ is a unipotent one-parameter subgroup of $G$, and $\mu$ is an ergodic $u^t$-invariant probability measure on $\Gamma\backslash G$, then $\mu$ is homogeneous. More precisely, there exist a closed, connected subgroup $S$ of $G$, and a point $x$ in $\Gamma\backslash G$, such that*

(1) *$\mu$ is $S$-invariant, and*
(2) *$\mu$ is supported on the orbit $xS$.*

*Remark* 6.2. If we write $x = \Gamma g$, for some $g \in G$, and let $\Gamma_S = (g^{-1}\Gamma g) \cap S$, then the conclusions imply that

- under the natural identification of the orbit $xS$ with the homogeneous space $\Gamma_S\backslash S$, the measure $\mu$ is the Haar measure on $\Gamma_S\backslash S$,
- $\Gamma_S$ is a lattice in $S$, and
- $xS$ is closed.

*Remark* 6.3. Note that $G$ is not assumed to be semisimple. Although the *semisimple* case is the most interesting, we allow ourselves more freedom, principally because the proof relies – at one point, in the proof of Theorem 6.21 – on induction on $\dim G$, and this induction is based on knowing the result for all connected subgroups, not only the semisimple ones.

**Step 1.** Let $S = \mathrm{Stab}_G(\mu)$. We wish to show that $\mu$ is supported on a single $S$-orbit. Let $\mathfrak{g}$ be the Lie algebra of $G$ and $\mathfrak{s}$ be the Lie algebra of $S$. The expanding and contracting subspaces of $a^s$ (for $s > 0$) provide decompositions

$$\mathfrak{g} = \mathfrak{g}_- + \mathfrak{g}_0 + \mathfrak{g}_+, \quad \mathfrak{s} = \mathfrak{s}_- + \mathfrak{s}_0 + \mathfrak{s}_+,$$

and we have corresponding subgroups $G_-, G_0, G_+, S_-$, $S_0$, and $S_+$ (see Notation 6.9). For convenience, let $U = S_+$. Note that $U$ is unipotent, and we may assume $\{u^t\} \subset U$, so $\mu$ is ergodic for $U$.

**Step 2.** We are interested in transverse divergence of nearby orbits. We ignore relative motion along the $U$-orbits, and project to $G \ominus U$. The shearing property of unipotent flows implies, for a.e. $x, y \in \Gamma\backslash G$, that if $x \approx y$, then the transverse divergence of the $U$-orbits through $x$ and $y$ is fastest along some direction in $S$ (see Proposition 6.4). Therefore, the direction belongs to $G_-G_0$ (see Corollary 6.10).

**Step 3.** We define a certain subgroup

$$\widetilde{S}_- = \{g \in G_- \mid \forall u \in U,\ u^{-1}gu \in G_-G_0U\}$$

of $G_-$ (cf. Definition 6.11). Note that $S_- \subset \widetilde{S}_-$. The motivation for this definition is that if $y \in x\widetilde{S}_-$, then all of the transverse divergence belongs to $G_-G_0-$ there is no $G_+$-component to any of the transverse divergence. For clarity, we emphasize that this restriction applies to all transverse divergence, not only the fastest transverse divergence.

**Step 4.** Combining Step 2 with the dilation provided by the translation $a^{-s}$ shows, for a.e. $x, y \in \Gamma\backslash G$, that if $y \in xG_-$, then $y \in x\widetilde{S}_-$ (see Corollary 6.15).

**Step 5.** A Lie algebra calculation shows that if $y \approx x$, and $y = xg$, with $g \in (G_- \ominus \widetilde{S}_-)G_0G_+$, then the transverse divergence of the $U$-orbits through $x$ and $y$ is fastest along some direction in $G_+$ (see Lemma 6.16).

**Step 6.** Because the conclusions of Step 2 and Step 5 are contradictory, we see, for a.e. $x, y \in \Gamma\backslash G$, that if $x \approx y$, then $y \notin x(G_- \ominus \widetilde{S}_-)G_0G_+$ (cf. Corollary 6.17). Actually, a technical problem causes us obtain this result only for $x$ and $y$ in a set of measure $1 - \varepsilon$.

**Step 7.** The relation between stretching and entropy provides bounds on the entropy of $a^s$, in terms of the the Jacobian of $a^s$ on $U$ and (using Step 4) the Jacobian of $a^{-s}$ on $\widetilde{S}_-$:

$$J(a^s, U) \leqslant h_\mu(a^s) \leqslant J(a^{-s}, \widetilde{S}_-)$$

On the other hand, structure of $\mathfrak{sl}_2(\mathbb{R})$-modules implies that $J(a^s, U) \geqslant J(a^{-s}, \widetilde{S}_-)$. Thus, we conclude that $h_\mu(a^s) = J(a^{-s}, \widetilde{S}_-)$. This implies that $\widetilde{S}_- \subset \operatorname{Stab}_G(\mu)$, so we must have $\widetilde{S}_- = S_-$ (see Proposition 6.19).

**Step 8.** By combining the conclusions of Step 6 and Step 7, we show that $\mu(xS_-G_0G_+) > 0$, for some $x \in \Gamma\backslash G$ (see Proposition 6.20).

**Step 9.** By combining Step 8 with the (harmless) assumption that $\mu$ is not supported on an orbit of any closed, proper subgroup of $G$, we show that $S_- = G_-$(so $S_-$ is horospherical), and then there are a number of ways to show that $S = G$ (see Theorem 6.21).

### 6.2. **Key results of use.**

**Proposition 6.4.** *If $U$ is any connected, ergodic, unipotent subgroup of $S$, then there is a conull subset $\Omega$ of $\Gamma\backslash G$, such that, for all $x, y \in \Omega$, with $x \approx y$, the $U$-orbits through $x$ and $y$ diverge fastest along some direction that belongs to $S$.*

This immediately implies the following interesting special case of Ratner's Theorem.

**Corollary 6.5.** *If $U = \operatorname{Stab}_G(\mu)$ is unipotent (and connected), then $\mu$ is supported on a single $U$-orbit.*

Although Proposition 6.4 is true, it seems to be very difficult to prove from scratch, so we will be content with proving the following weaker version that does not yield a conull subset, and imposes a restriction on the relation between $x$ and $y$.

**Proposition 6.6.** *For any connected, ergodic, unipotent subgroup $U$ of $S$, and any $\varepsilon > 0$, there is a subset $\Omega_\varepsilon$ of $\Gamma\backslash G$, such that $\mu(\Omega_\varepsilon) > 1 - \varepsilon$, and for all $x, y \in \Omega_\varepsilon$, with $x \approx y$, and such that certain technical assumption, see Addendum 6.8, is satisfied, the fastest transverse divergence of the $U$-orbits through $x$ and $y$ is along some direction that belongs to $S$.*

*Proof.* Let us assume that no $N_G(U)$-orbit has positive measure, for otherwise it is easy to complete the proof. Then, for a.e. $x \in \Gamma\backslash G$, there is a point $y \approx x$, such that $y \notin xN_G(U)$, and $y$ is a generic point for $\mu$.

Because $y \notin xN_G(U)$, we know that the orbit $yU$ is not parallel to $xU$, so they diverge from each other. We know that the direction of fastest transverse divergence belongs to $N_G(U)$, so there exist $u, u' \in U$, and $c \in N_G(U) \ominus U$, such that $yu' \approx (xu)c$, and $\|c\| \asymp 1$ (i.e. $\|c\|$ is finite, but not infinitesimal).

Because $c \notin U = \operatorname{Stab}_G(\mu)$, we know that $c_*\mu \neq \mu$. Because $c \in N_G(U)$, this implies $c_*\mu \perp \mu$, so there is a compact subset $K$ with $\mu(K) > 1 - \varepsilon$ and $K \cap Kc = \varnothing$.

We would like to complete the proof by saying that there are values of $u$ for which both of the two points $xu$ and $yu'$ are arbitrarily close to $K$, which contradicts the fact that $d(K, Kc) > 0$. However, there are two technical problems:

(1) The set $K$ must be chosen before we know the value of $c$.

(2) The Pointwise Ergodic Theorem 2.7 implies (for a.e. $x$) that $xu$ is arbitrarily close to $K$ a huge proportion of the time. But this theorem does not apply directly to $yu'$, because $u'$ is a nontrivial function of $u$. To overcome this difficulty, we add an additional technical hypothesis on the element $g$ with $y = xg$. With this assumption, the result can be proved, by showing that the Jacobian of the change of variables $u \mapsto u'$ is bounded above and below on some set of reasonable size, and applying the uniform approximate version of the Pointwise Ergodic Theorem (see Theorem 2.7). The uniform estimate is what requires us to restrict to a set of measure $1 - \varepsilon$, rather than a conull set.

□

*Remark* 6.7. The fact that $\Omega_\varepsilon$ is not quite conull is not a serious problem, although it does make one part of the proof more complicated (cf. Proposition 6.20).

We will apply Proposition 6.4 only twice, in the proofs of Cororollary 6.15 and 6.17. In each case, it is not difficult to verify that the technical assumption is satisfied.

**Addendum 6.8.** The so-called *technical assumption* that appeared in Proposition 6.6 can be stated in the following explicit form if $g$ is infinitesimal: there are an (infinite) integer $n$, and a finite element $u_0$ of $U$, such that

- $a^{-n} u_0 a^n g \in G_- G_0 G_+$,
- $a^{-n} u_0 a^n g U$ is not infinitesimally close to $eU$ in $G/U$, and
- $a^n g a^{-n}$ is finite (or infinitesimal).

In non-infinitesimal terms, the assumption on $\{g_k\}$ is: there are a sequence $n_k \to \infty$, and a bounded sequence $\{u_k\}$ in $U$, such that

- $a^{-n_k} u_k a^{n_k} g_k \in G_- G_0 G_+$,
- no subsequence of $a^{-n_k} u_k a^{n_k} g_k U$ converges to $eU$ in $G/U$, and
- $a^{n_k} g_k a^{-n_k}$ is bounded.

**Notation 6.9.** The notations used in Step 1 are in the following.

- For a (small) element $g$ of $G$, we use $\underline{g}$ to denote the corresponding element $\log g$ of the Lie algebra $\mathfrak{g}$.
- Recall that $S = \mathrm{Stab}_G(\mu)^\circ$.
- By renormalizing, let us assume that $[u, a] = 2u$ (where $a = a^1$ and $u = u^1$).
- Let $\{v^r\}$ be the (unique) one-parameter unipotent subgroup of $L$, such that $[v, a] = -2v$ and $[v, u] = a$.
- Let $\bigoplus_{\lambda \in \mathbb{Z}} \mathfrak{g}_\lambda$ be the decomposition of $\mathfrak{g}$ into weight spaces of $a$: that is,
$$\mathfrak{g}_\lambda = \{\underline{g} \in \mathfrak{g} \mid [\underline{g}, a] = \lambda \underline{g}\}.$$
- Let $\mathfrak{g}_+ = \bigoplus_{\lambda > 0} \mathfrak{g}_\lambda$, $\mathfrak{g}_- = \bigoplus_{\lambda < 0} \mathfrak{g}_\lambda$, $\mathfrak{s}_+ = \mathfrak{s} \cap \mathfrak{g}_+$, $\mathfrak{s}_- = \mathfrak{s} \cap \mathfrak{g}_-$, and $\mathfrak{s}_0 = \mathfrak{s} \cap \mathfrak{g}_0$. Then
$$\mathfrak{g} = \mathfrak{g}_- + \mathfrak{g}_0 + \mathfrak{g}_+, \quad \mathfrak{s} = \mathfrak{s}_- + \mathfrak{s}_0 + \mathfrak{s}_+.$$
These are direct sums of vector spaces, although they are not direct sums of Lie algebras.
- Let $G_+, G_-, G_0, S_+, S_-, S_0$ be the connected subgroups of $G$ corresponding to the Lie subalgebras $\mathfrak{g}_+, \mathfrak{g}_-, \mathfrak{g}_0, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{s}_0$, respectively.
- Let $U = S_+$ (and let $\mathfrak{u}$ be the Lie algebra of $U$).

Because $S_- S_0 U = S_- S_0 S_+$ contains a neighborhood of $e$ in $S$, Proposition 6.6 states that the direction of fastest transverse divergence belongs to $S_- S_0$. The following corollary is a *priori* weaker (because $G_-$ and $G_0$ are presumably larger than $S_-$ and $S_0$), but it is the only consequence that we will need in our later arguments.

**Corollary 6.10.** *For any $\varepsilon > 0$, there is a subset $\Omega_\varepsilon$ of $\Gamma \backslash G$, such that $\mu(\Omega_\varepsilon) > 1 - \varepsilon$, and for all $x, y \in \Omega_\varepsilon$, with $x \approx y$, and such that a certain technical assumption in Addendum 6.8 is satisfied, the fastest transverse divergence of the $U$-orbits through $x$ and $y$ is along some direction that belongs to $G_- G_0$.*

**Definition 6.11.** Let

$$\widetilde{S} = \{ g \in G \mid u^{-1} g u \in \overline{G_- G_0 U}, \text{ for all } u \in U \}$$

and

$$\widetilde{S}_- = \widetilde{S} \cap G_-.$$

It is more or less obvious that $S \subset \widetilde{S}$. Although this is much less obvious, it should also be noted that $\widetilde{S}$ is a closed subgroup of $G$.

*Remark* 6.12. Here is an alternate approach to the definition of $\widetilde{S}$, or, at least, its identity component.

(1) Let

$$\widetilde{\mathfrak{s}} = \{ \underline{g} \in \mathfrak{g} \mid \underline{g} (\operatorname{ad} u)^k \in \mathfrak{g}_- + \mathfrak{g}_0 + \mathfrak{u}, \ \forall k \geqslant 0, \ \forall \underline{u} \in \mathfrak{u} \}.$$

Then $\mathfrak{s}$ is a Lie subalgebra of $\mathfrak{g}$, so we may let $\widetilde{S}^\circ$ be the corresponding connected Lie subgroup of $G$. (We will see in (3) below that this agrees with Definition 6.11)

(2) From the point of view in (1), it is not difficult to see that $\widetilde{S}^\circ$ is the unique maximal connected subgroup of $G$, such that $\widetilde{S}^\circ \cap G_+ = U$, and $\widetilde{S}^\circ$ is normalized by $a^t$. This makes it obvious that $S \subset \widetilde{S}^\circ$. It is also easy to verify directly that $\mathfrak{s} \subset \widetilde{\mathfrak{s}}$.

(3) It is not difficult to see that the identity component of the subgroup defined in Definition 6.11 is also the subgroup characterized in (2), so this alternate approach agrees with the original definition of $\widetilde{S}$.

**Example 6.13.** Remark 6.12 makes it easy to calculate $\widetilde{S}^\circ$.

(1) We have $\widetilde{S} = G$ if and only if $U = G_+$.

(2) If

$$G = \mathrm{SL}_3(\mathbb{R}), \quad a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \mathfrak{u} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ * & 0 & 0 \end{pmatrix},$$

then

$$\mathfrak{g}_+ = \begin{pmatrix} 0 & 0 & 0 \\ * & 0 & 0 \\ * & * & 0 \end{pmatrix}, \quad \widetilde{\mathfrak{s}} = \begin{pmatrix} * & 0 & * \\ 0 & * & 0 \\ * & 0 & * \end{pmatrix}.$$

(3) If

$$G = \mathrm{SL}_3(\mathbb{R}), \quad a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \mathfrak{u} = \begin{pmatrix} 0 & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{pmatrix},$$

then

$$\mathfrak{g}_+ = \begin{pmatrix} 0 & 0 & 0 \\ * & 0 & 0 \\ * & * & 0 \end{pmatrix}, \quad \widetilde{\mathfrak{s}} = \begin{pmatrix} * & 0 & * \\ * & * & * \\ * & 0 & * \end{pmatrix}.$$

(4) If

$$G = \mathrm{SL}_3(\mathbb{R}), \quad a = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad \mathfrak{u} = \mathbb{R} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

then

$$\mathfrak{g}_+ = \begin{pmatrix} 0 & 0 & 0 \\ * & 0 & 0 \\ * & * & 0 \end{pmatrix}, \quad \widetilde{\mathfrak{s}} = \mathbb{R} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{pmatrix} + \mathfrak{u}.$$

(5) If

$$G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}), \quad a = \left( \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right),$$

and

$$\mathfrak{u} = \mathbb{R} \left( \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right),$$

then

$$\mathfrak{g}_+ = \begin{pmatrix} 0 & 0 \\ * & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 0 \\ * & 0 \end{pmatrix}, \quad \mathfrak{s} = \mathbb{R}(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}) + \mathbb{R}\underline{a} + \mathfrak{u}.$$

Our ultimate goal is to find a conull subset $\Omega$ of $\Gamma\backslash G$, such that if $x, y \in \Omega$, then $y \in xS$. In this section, we establish two consequences of Corollary 6.10 that represent major progress toward this goal (see Corollary 6.15 and 6.17). These results deal with $\widetilde{S}$, rather than $S$, but that turns out not to be a very serious problem, because $\widetilde{S} \cap G_+ = S \cap G_+$ (see Remark 6.12(2)) and $\widetilde{S} \cap G_- = S \cap G_-$ (see Proposition 6.19).

**Notation 6.14.** Let

- $\mathfrak{g}_+ \ominus \mathfrak{u}$ be an $a^s$-invariant complement to $\mathfrak{u}$ in $\mathfrak{g}_+$,
- $\mathfrak{g}_- \ominus \widetilde{\mathfrak{s}}_-$ be an $a^s$-invariant complement to $\widetilde{\mathfrak{s}}_-$ in $\mathfrak{g}_-$,
- $G_+ \ominus U = \exp(\mathfrak{g}_- \ominus \mathfrak{u})$, and
- $G_- \ominus \widetilde{S}_- = \exp(\mathfrak{g}_- \ominus \widetilde{\mathfrak{s}}_-)$.

Note that the natural maps $(G_+ \ominus U) \times U \to G_+$ and $(G_- \ominus \widetilde{S}_-) \times \widetilde{S}_- \to G_-$ defined by $(g, h) \mapsto gh$ are diffeomorphisms.

**Corollary 6.15.** *There is a conull subset $\Omega$ of $\Gamma\backslash G$, such that if $x, y \in \Omega$, and $y \in xG_-$, then $y \in x\widetilde{S}_-$.*

*Proof.* Choose $\Omega_0$ as in the conclusion of Corollary 6.10 From the Pointwise Ergodic Theorem 2.7, we know that

$$\Omega = \{ x \in \Gamma\backslash G \mid \{ t \in \mathbb{R}^+ \mid xa^t \in \Omega_0 \} \text{ is bounded} \}$$

is conull.

We have $y = xg$, for some $g \in G_-$. Because $a^{-t} g a^t \to e$ as $t \to \infty$, we may assume, by replacing $x$ and $y$ with $xa^t$ and $ya^t$ for some infinitely large $t$, that $g$ is infinitesimal (and that $x, y \in \Omega_0$).

Suppose $g \notin \widetilde{S}_-$(this will lead to a contradiction). From the definition of $\widetilde{S}_-$, this means there is some $u \in U$, such that $u^{-1} g u \notin G_- G_0 U$: write $u^{-1} g u = hcu'$ with $h \in G_- G_0$, $c \in G_+ \ominus U$, and $u' \in U$. We may assume $h$ is infinitesimal (because we could choose $u$ to be finite, or even infinitesimal, if desired). Translating again by an (infinitely large) element of $\{a^t\}$, with $t \geqslant 0$, we may assume $c$ is infinitely large. Because $h$ is infinitesimal, this clearly implies that the orbits through $x$ and $y$ diverge fastest along a direction in $G_+$, not a direction in $G_- G_0$. This contradicts Corollary 6.10. ☐

An easy calculation (involving only algebra, not dynamics) establishes the following.

**Lemma 6.16.** *If $y = xg$ with $g \in (G_- \ominus \widetilde{S}_-) G_0 G_+$, and $g \approx e$, then the transverse divergence of the $U$-orbits through $x$ and $y$ is fastest along some direction in $G_+$.*

*Proof.* Choose $s > 0$ (infinitely large), such that $\widehat{g} = a^s g a^{-s}$ is finite, but not infinitesimal, and write $\widehat{g} = \widehat{g}_- \widehat{g}_0 \widehat{g}_+$, with $\widehat{g}_- \in G_-, \widehat{g}_0 \in G_0$, and $\widehat{g}_+ \in G_+$ Note that $\widehat{g}_0$ and $\widehat{g}_+$ are infinitesimal, but $\widehat{g}_-$ is not. Because $\widehat{g}_- \in G_- \ominus \widetilde{S}_-$, we know that $\widehat{g}$ is not infinitely close to $\widetilde{S}_-$, so there is some finite $u \in U$, such that $u^{-1}\widehat{g}$ is not infinitesimally close to $G_- G_0 U$.

Let $\widehat{u} = a^{-s} u a^s$, and consider $\widehat{u}^{-1} g \widehat{u} = a^{-s}(u^{-1}\widehat{g}u)a^s$. Because $u^{-1}\widehat{g}u$ is finite (since $u$ and $\widehat{g}$ are finite), we know that each of $(u^{-1}\widehat{g}u)_-$ and $(u^{-1}\widehat{g}u)_0$ is finite. Therefore $(\widehat{u}^{-1} g \widehat{u})_-$ and $(\widehat{u}^{-1} g \widehat{u})_0$ are finite, because conjugation by $a^s$ does not expand $G_-$ or $G_0$.

On the other hand, we know that $(\widehat{u}^{-1} g \widehat{u})_+$ is infinitely far from $U$, because the distance between $u^{-1}\widehat{g}u$ and $U$ is not infinitesimal, and conjugation by $a^s$ expands $G_+$ by an infinite factor.

Therefore, the fastest divergence is clearly along a direction in $G_+$. ☐

The conclusion of the above lemma contradicts the conclusion of Corollary 6.10 (2) (and the technical assumption in Addendum 6.8 is automatically satisfied in this situation), so we have the following conclusion.

**Corollary 6.17.** *For any $\varepsilon > 0$, there is a subset $\Omega_\varepsilon$ of $\Gamma\backslash G$, such that $\mu(\Omega_\varepsilon) > 1 - \varepsilon$, and for all $x, y \in \Omega_\varepsilon$, with $x \approx y$, we have $y \notin x(G_- \ominus \widetilde{S}_-)G_0G_+$.*

This can be restated in the following non-infinitesimal terms.

**Corollary 6.18.** *For any $\varepsilon > 0$, there is a subset $\Omega_\varepsilon$ of $\Gamma\backslash G$, and some $\delta > 0$, such that $\mu(\Omega_\varepsilon) > 1 - \varepsilon$, and for all $x, y \in \Omega_\varepsilon$, with $d(x, y) < \delta$, we have $y \notin x(G_- \ominus \widetilde{S}_-)G_0G_+$.*

**Proposition 6.19.** *We have $\widetilde{S}_- = S_-$.*

*Proof.* We already know that $\widetilde{S}_- \supset S_-$. Thus, because $\widetilde{S}_- \subset G_-$, it suffices to show that $\widetilde{S}_- \subset S$. That is, it suffices to show that $\mu$ is $\widetilde{S}_-$ invariant.

We have

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

From Corollary 6.15, we have

$$h_\mu(a) \leqslant \log J(a^{-1}, \widetilde{S}_-).$$

Combining these two inequalities with the fact that $h_\mu(a) = h_\mu(a^{-1})$, we have

$$\log J(a, U) \leqslant h_\mu(a^{-1}) = h_\mu(a) \leqslant \log J(a^{-1}, \widetilde{S}_-).$$

Thus, if we show that

$$\log J(a^{-1}, \widetilde{S}_-) \leqslant \log J(a, U),$$

then we must have equality throughout, and the desired conclusion will follow.

Because $u$ belongs to the Lie algebra $l$ of $L$ (see Notation 6.9), the structure of $\mathfrak{sl}_2(\mathbb{R})$-modules implies, for each $\lambda \in \mathbb{Z}^+$, that the restriction $(\operatorname{ad}_{\mathfrak{g}} u)^\lambda|_{\mathfrak{g}_{-\lambda}}$ is a bijection from the weight space $\mathfrak{g}_{-\lambda}$ onto the weight space $\mathfrak{g}_\lambda$. If $\mathfrak{g} \in \widetilde{\mathfrak{s}}_- \cap \mathfrak{g}_{-\lambda}$, then (1) of Remark 6.12 implies

$$\mathfrak{g}(\operatorname{ad}_{\mathfrak{g}} u)^\lambda \in (\mathfrak{g}_- + \mathfrak{g}_0 + \mathfrak{u}) \cap \mathfrak{g}_\lambda = \mathfrak{u} \cap \mathfrak{g}_\lambda,$$

so we conclude that $(\operatorname{ad}_{\mathfrak{g}} u)^\lambda|_{\mathfrak{s}_- \cap \mathfrak{g}_{-\lambda}}$ is an embedding of $\widetilde{\mathfrak{s}}_- \cap \mathfrak{g}_{-\lambda}$ into $\mathfrak{u} \cap \mathfrak{g}_\lambda$. So

$$\dim(\widetilde{\mathfrak{s}}_- \cap \mathfrak{g}_{-\lambda}) \leqslant \dim(\mathfrak{u} \cap \mathfrak{g}_\lambda).$$

The eigenvalue of $\operatorname{Ad}_G a = \exp(\operatorname{ad}_{\mathfrak{g}} a)$ on $\mathfrak{g}_\lambda$ is $e^\lambda$, and the eigenvalue of $\operatorname{Ad}_G a^{-1}$ on $\mathfrak{g}_{-\lambda}$ is also $e^\lambda$. Hence,

$$\begin{aligned}
\log J(a^{-1}, \widetilde{S}_-) &= \log \det(\operatorname{Ad}_G a^{-1})|_{\widetilde{\mathfrak{s}}_-} \\
&= \log \prod_{\lambda \in \mathbb{Z}^+} (e^\lambda) \dim(\widetilde{\mathfrak{s}}_- \cap \mathfrak{g}_{-\lambda}) \\
&= \sum_{\lambda \in \mathbb{Z}^+} (\dim(\widetilde{\mathfrak{s}}_- \cap \mathfrak{g}_{-\lambda})) \cdot \log e^\lambda \\
&\leqslant \sum_{\lambda \in \mathbb{Z}^+} (\dim \mathfrak{u} \cap \mathfrak{g}_\lambda) \cdot \log e^\lambda \\
&= \log J(a, U)
\end{aligned}$$

as desired. □

We wish to show, for some $x \in \Gamma\backslash G$, that $\mu(xS) > 0$. In other words, that $\mu(xS_-S_0S_+) > 0$. The following weaker result is a crucial step in this direction.

**Proposition 6.20.** *For some $x \in \Gamma\backslash G$, we have $\mu(xS_-G_0G_+) > 0$.*

*Proof.* Assume that the desired conclusion fails. This will lead to a contradiction. Let $\Omega_\varepsilon$ be as in Corollary 6.17, with $\varepsilon$ sufficiently small.

Because the conclusion of the proposition is assumed to fail, there exist $x, y \in \Omega_\varepsilon$, with $x \approx y$ and $y = xg$, such that $g \notin S_-G_0G_+$. Thus, we may write $g = vwh$ with $v \in S_-, w \in (G_- \ominus S_-)\backslash\{e\}$, and $h \in G_0G_+$.

For simplicity, let us pretend that $\Omega_\varepsilon$ is $S_-$-invariant. This is not so far from the truth, because $\mu$ is $S_-$-invariant and $\mu(\Omega_\varepsilon)$ is very close to 1, so the actual proof is only a little more complicated. Then we may replace $x$ with $xv$, so that $g = wh \in (G_- \ominus S_-)G_0G_+$. This contradicts the definition of $\Omega_\varepsilon$. □

We can now complete the proof using some of the theory of algebraic groups.

**Theorem 6.21.** *$\mu$ is supported on a single $S$-orbit.*

*Proof.* There is no harm in assuming that $G$ is almost Zariski closed. By induction on $\dim G$, we may assume that there does not exist a subgroup $H$ of $G$, such that $H$ is almost Zariski closed, $U \subset H$, and some $H$-orbit has full measure.

Then a short argument implies, for all $x \in \Gamma\backslash G$, that if $V$ is any subset of $G$, such that $\mu(xV) > 0$, then $G \subset \overline{\overline{V}}$, the Zariski closure of $V$. This hypothesis will allow us to show that $S = G$.

**Claim.** We obtain $S_- = G_-$.

Proposition 6.20 states that $\mu(xS_-G_0G_+) > 0$, so, from our hypothesis above, $G \subset \overline{\overline{S_-G_0G_+}}$. This implies that $S_-G_0G_+$ must contain an open subset of $G$. Therefore

$$\dim S_- \geqslant \dim G - \dim(G_0G_+) = \dim G_-.$$

Because $S_- \subset G_-$, and $G_-$ is connected, this implies that $S_- = G_-$, as desired.

The subgroup $G_-$ is a horospherical subgroup of $G$, so we have shown that $\mu$ is invariant under a horospherical subgroup of $G$. The fact below is used to complete the proof.

**Fact.** If $N$ is a unimodular, normal subgroup of $G$, $N$ is contained in $\mathrm{Stab}_G(\mu)$, and $N$ is ergodic on $\Gamma\backslash G$, then $\mu$ is homogeneous.

There are now at least three ways to complete the argument.

(a) We showed that $\mu$ is $G_-$-invariant. By going through the same argument, but with $v^r$ in the place of $u^t$, we could show that $\mu$ is $G_+$-invariant. So $S$ contains $\langle L, G_+, G_-\rangle$, which is easily seen to be a (unimodular) normal subgroup of $G$. Then the fact applies.

(b) By using considerations of entropy, one can show that $G_+ \subset S$, and then the fact applies, once again.

(c) If we assume that $\Gamma\backslash G$ is compact (and in some other cases), then a completely separate proof of the theorem is known for measures that are invariant under a horospherical subgroup. Such special cases were known several years before the general theorem.

More rigorous details are omitted here. □

## 7. Application to Oppenheim conjecture

In this section, we present a schematic proof of the Oppenheim conjecture using Ratner's theorems. Our main concern is the connection between the number theory of the Oppenheim conjecture and the theory of dynamical systems in Ratner's theorems. The connection comes via the group $\mathrm{SO}(Q) \subset \mathrm{SL}_n(\mathbb{R})$, the group of transformations that leaves the quadratic form $Q$ invariant. The theory of dynamical systems will let us show that $\overline{\mathrm{SO}(Q)}$ must be "nice", and we'll see that this leaves only two options for it — corresponding to $Q$ rational or $Q(\mathbb{Z}^n)$ dense in $\mathbb{R}$. The references are the original paper by Dani–Margulis [DM89] together with [Ghy92, §4.2].

### 7.1. Oppenheim conjecture.

Let

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

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

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

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

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

**Theorem 7.2** (Margulis, 1986)**.** *The Oppenheim Conjecture is true as long as $n \geqslant 3$. Thus, if $n \geqslant 3$ 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 $n \geqslant 21$, and for diagonal forms with $n \geqslant 5$.

*Remark* 7.3 (The necessity of the conditions)*.* If $Q$ is definite, the image of $\mathbb{Z}^n$ is confined to $\mathbb{R}_{\geqslant 0}$, and in fact is a lattice. The requirement of $n \geqslant 3$ variables is also necessary. Indeed, let $\alpha$ be a real algebraic number of degree 2; then it is well-known that $|\alpha - p/q| \geqslant C/q^2$ for some constant $C$ and all rationals $p/q$. Consequently, the quadratic form $Q(x, y) = y^2 - \alpha^2 x^2$ has the property that

$$|Q(x, y)| = |x^2(y/x - \alpha)(y/x + \alpha)| \geqslant C|\alpha|$$

and 0 is an isolated point in the image of $Q$. Finally, if $Q$ is degenerate, then after a suitable change of coordinates it is isomorphic to an $(n-1)$-form; and since the requirement of $n \geqslant 3$ variables is necessary, so is nondegeneracy.

Theorem 7.2 is true if we replace $\mathbb{Z}^n$ by the set of primitive vectors (a vector $p = (p_1, \ldots, p_n)$ is primitive if $\gcd(p_1, \ldots, p_n) = 1$). The general case of Oppenheim's conjecture can be reduced to the case of $n = 3$ variables. The argument is somewhat tedious, but straightforward.

7.2. **Margulis' original approach.** We begin with some definition.

**Definition 7.4.** If $Q$ is a quadratic form in $n$ variables, the *special orthogonal group* of $Q$ is

$$\mathrm{SO}(Q) = \{h \in \mathrm{SL}_n(\mathbb{R}) \mid Q(vh) = Q(v), \ \forall v \in \mathbb{R}^n\}.$$

We will let $\mathrm{SO}(Q)^\circ$ be the connected component of the identity in $\mathrm{SO}(Q)$.

Since every indefinite quadratic form has signature $(2, 1)$ or $(1, 2)$, and the two cases differ from each other only by an overall sign, we will let $Q_0$ denote the standard quadratic form of signature $(2, 1)$: that is, $Q_0(x_1, x_2, x_3) = x_1^2 + x_2^2 - x_3^2$. Then our arbitrary quadratic form $Q$ is conjugate to $\pm Q_0$. We will let $H = \mathrm{SO}(Q_0)^\circ$ stand for the connected component of the identity in the special orthogonal group of $Q_0$.

*Remark* 7.5*.* We do not lose much generality by working with $H$ rather than the entire special orthogonal group. $\mathrm{SO}(Q_0)$ has only two connected components: thus, $H$ has index 2 in $\mathrm{SO}(Q_0)$. This is a classical result.

This proof is closer to the original approach used by Margulis in his 1987 proof of the Oppenheim Conjecture, and relies on deep statements about the behavior of unipotent flows. The estimates derived by Margulis are weaker than Ratner's general estimates (especially the more quantitative ones); some of the spirit of the original argument is given in this section, although we avoid presenting the proof in full generality.

We will exploit the fact that $\mathrm{SO}(Q)$ is large, and a priori $\mathrm{SO}(Q)\mathbb{Z}^n$ is much larger than $\mathbb{Z}^n$. Ratner's theorem will let us quantify this: either $\mathrm{SO}(Q)\mathbb{Z}^3$ is dense in $\mathbb{R}^3$ or (after a few more arguments) $Q$ must be rational. The precise statement of Ratner's theorem (cf. Theorem 4.4) is in the following.

**Theorem 7.6** (Ratner's orbit closures theorem)**.** *Let $G$ be a connected Lie group, and let $\Gamma$ be a lattice in $G$. Let $H$ be a connected Lie subgroup of $G$ generated by unipotent one-parameter groups. Then for any $x \in G/\Gamma$ there exists a closed connected subgroup $P \subset G$ containing $H$ such that $\overline{Hx} = Px$ and $Px$ admits a $P$-invariant probability measure.*

The dichotomy in the statement of the Oppenheim conjecture results from the fact that the $H$ in question is a maximal connected subgroup of $G$, and therefore there are only two possible choices for $P$; namely, $Hx$ is either dense or closed, corresponding to the cases of $Q(\mathbb{Z}^n)$ dense in $\mathbb{R}$ and $Q$ proportional to an integer form respectively.

Let $g_Q \in \mathrm{SL}_3(\mathbb{R})$ and $\lambda \in \mathbb{R}^\times$ be such that $Q = \lambda Q_0 \circ g_Q$. In that case,

$$\mathrm{SO}(Q)^\circ = g_Q H g_Q^{-1}.$$

Now, $H = \mathrm{SO}(Q_0)^\circ \cong \mathrm{SL}_2(\mathbb{R})$ is generated by unipotent elements, and $\mathrm{SL}_3(\mathbb{Z})$ is a lattice in $\mathrm{SL}_3(\mathbb{R})$, so we can apply Ratner's Orbit Closure Theorem to obtain the following.

There exists a closed, connected subgroup $P \subset \mathrm{SL}_3(\mathbb{R})$ such that $H \subset P$, $\overline{Hg_Q} = \overline{Pg_Q}$, and there is an $P$-invariant probability measure on $Pg_Q$.

There are only two possibilities for a closed, connected subgroup of $\mathrm{SL}_3(\mathbb{R})$ containing $H = \mathrm{SO}(Q_0)$: namely, $S = H$ or $S = \mathrm{SL}_3(\mathbb{R})$. We consider these two cases separately.

**Case I.** Assume $S = \mathrm{SL}_3(\mathbb{R})$. In that case, $\mathrm{SL}_3(\mathbb{Z})gH$ is dense in $\mathrm{SL}_3(\mathbb{R})$. So

$$\begin{aligned} Q(\mathbb{Z}^3) &= Q_0(\mathbb{Z}^3 g_Q) && \text{(by definition of } g_Q) \\ &= Q_0(\mathbb{Z}^3 \,\mathrm{SL}_3(\mathbb{Z}) g_Q) && (\mathbb{Z}^3 = \mathbb{Z}^3 \,\mathrm{SL}_3(\mathbb{Z})) \\ &= Q_0(\mathbb{Z}^3 \,\mathrm{SL}_3(\mathbb{Z}) g_Q H) && (H = \mathrm{SO}(Q_0)^\circ) \end{aligned}$$

is dense in $Q(\mathbb{Z}^3 G)$ since $Q$ is continuous. On the other hand,

$$Q(\mathbb{Z}^3 G) = Q_0(\mathbb{R}^3 - \{0\}) = \mathbb{R}$$

since $vG = \mathbb{R}^3 - \{0\}$ for nonzero $v$. That is, since $Q$ is indefinite, it must map $\mathbb{R}^3$ onto $\mathbb{R}$, and we concluded that $\mathrm{SO}(Q)\mathbb{Z}^3$ is dense in $\mathbb{R}^3$ – so its image is dense in $\mathbb{R}$.

**Case II.** Assume $S = H = \mathrm{SO}(Q_0)$. This is the degenerate case, where $Q$ is a scalar multiple of a form with integer coefficients. We present two proofs of this: the first one relying on Margulis' lemma 7.7 in analysis, and the second on the theory of algebraic groups. The algebraic approach is more concise, but uses fairly deep results from the theory of algebraic groups; the analytic approach is closer to the argument used by Margulis in his original 1987 proof of the Oppenheim conjecture.

If $S = H$ then $g_Q H$ is closed, and therefore so is the $H$-orbit orbit of $g_Q \,\mathrm{SL}_3(\mathbb{Z})$ in $G/\Gamma = \mathrm{SL}_3(\mathbb{R})/\mathrm{SL}_3(\mathbb{Z})$. Let $x = g_Q \,\mathrm{SO}_3(\mathbb{Z}) \in G/\Gamma$ and $x_0 = \mathrm{SO}_3(\mathbb{Z}) \in G/\Gamma$. Then $\mathrm{SO}(Q)x_0 = g_Q^{-1} Hx$ is also closed. Let $\Delta = \mathrm{SO}_3(\mathbb{Z}) \cap \mathrm{SO}(Q)$.

Our strategy will be to show that there exist real symmetric $3 \times 3$ matrices $S$ satisfying $\gamma^t S \gamma = S$ for all $\gamma \in \Delta$, and that all such matrices correspond to quadratic forms that are proportional to $Q$. Since this system of equations for $S$ is defined over the integers, if it has some solution, it will have a rational solution – yielding a rational quadratic form proportional to $Q$.

In terms of quadratic forms, $\gamma^t S \gamma = S$ means $\Delta \subset \mathrm{SO}(Q')$ for a quadratic form $Q'$. Existence of such a $Q'$, therefore, is trivial: $\Delta \subset \mathrm{SO}(Q)$. The difficult part will be to show that if $\Delta \subset \mathrm{SO}(Q')$ then $Q$ and $Q'$ are proportional.

Now, $\mathrm{SO}(Q)^\circ$ is similar to $\mathrm{SO}(Q)$ (as follows from Remark 7.5, it is an index-2 subgroup), but $H = \mathrm{SO}(2,1)^\circ$ is isomorphic to $\mathrm{SL}_2(\mathbb{R})$, hence it is generated by unipotent one-parameter subgroups. We will show that $\Delta \subset \mathrm{SO}(Q')$ implies that all unipotent 1-parameter subgroups of $\mathrm{SO}(Q)$ are contained in $\mathrm{SO}(Q')$, and hence $\mathrm{SO}(Q)^\circ \subset \mathrm{SO}(Q')$.

Fix a point $p \in \mathbb{R}^3$, and consider $f_p\colon \mathrm{SO}(Q) \to \mathbb{R}, g \mapsto Q'(g^{-1}p)$. If $\Delta \subset \mathrm{SO}(Q')$, then $f_p$ factors through $\Delta$ to a continuous function

$$\widetilde{f_p}\colon \mathrm{SO}(Q)/\Delta \to \mathbb{R}$$

Now, let $\{u(t)\}_{t\in\mathbb{R}} \subset \mathrm{SO}(Q)$ be a unipotent one-parameter subgroup. The function $q\colon \mathbb{R} \to \mathbb{R}, t \mapsto f_p(u(t))$ is polynomial in $t$, since the entries of $u(t)$ are polynomial in $t$.

We now invoke Margulis' lemma without proof to produce $K \subset G/\Gamma$ compact such that the set $\{t \geqslant 0\colon u(t)x_0 \in K\}$ is unbounded: that is, a compact set to which the $u(t)$-orbit of $x_0$ returns infinitely often.

**Lemma 7.7** (Margulis' lemma). *Let $n \geqslant 2$. Let $\{u_t\}_{t\in\mathbb{R}}$ be a unipotent one-parameter subgroup of $\mathrm{SL}_n(\mathbb{R})$, and let $x \in \mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$. Then there exists a compact set $K \subset \mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$ such that $\{t \geqslant 0 \mid u_t x \in K\}$ is unbounded: that is, $u_t x$ does not tend to infinity as $t \to \infty$.*

Since $\mathrm{SO}(Q)x_0$ is closed, the map $\phi\colon \mathrm{SO}(Q)/\Delta \to \mathrm{SO}(Q)x_0$ via $g\Delta \mapsto gx_0$ is a homeomorphism, and $K' = \phi^{-1}(K)$ is compact. Therefore, $\widetilde{f}_p(K')$ is a compact subset of $\mathbb{R}$. On the other hand, $K$ was chosen so that $\{t \in \mathbb{R}\colon q(t) \in f_p(K')\}$ is unbounded, implying that $q$ is the constant polynomial.

That is, $f_p(u(t))$ is constant, and $Q'(u(t)p) = Q'(p)$ for all $t \in \mathbb{R}$. Since $p$ was arbitrary, this holds at every $p \in \mathbb{R}^3$, implying that $\{u(t)\}_{t\in\mathbb{R}} \subset \mathrm{SO}(Q')$. We can therefore conclude that if $\Delta \subset \mathrm{SO}(Q')$ then $\mathrm{SO}(Q)^\circ \subset \mathrm{SO}(Q')$. Now, let $\sigma$ and $\sigma'$ be the symmetric matrices corresponding to $Q$ and $Q'$ respectively. We have for all $h \in \mathrm{SO}(Q)^\circ$,

$$h\sigma'\sigma^{-1}h^{-1} = (h\sigma'h^t)((h^{-1})^t\sigma^{-1}h^{-1}) = \sigma'\sigma^{-1}.$$

Now, $H = \mathrm{SO}(Q_0)^\circ$ is centralized only by scalars, and the same holds for $\mathrm{SO}(Q)^\circ$ since it is conjugate to $H$. Therefore, $\sigma\sigma^{-1}$ is a scalar, i.e. the two matrices are proportional. This concludes the proof that $Q$ is proportional to a rational matrix.

7.3. **Simplified algebraic approach.** There's another proof in an algebraic way for Case II. The supplementary lemmas follow the outline of the argument.

**Outline.** If $S = H$, then the orbit $g_Q H = g_Q S$ has a finite $H$-invariant measure. Therefore, $\Gamma_{g_Q} = \Gamma \cap (g_Q H g_Q^{-1}) = \mathrm{SL}_3(\mathbb{Z}) \cap (g_Q H g_Q^{-1})$ is a lattice in $g_Q H g_Q^{-1} = \mathrm{SO}(Q)^\circ$. Since $H = \mathrm{SO}(2,1)^\circ \cong \mathrm{SL}_2(\mathbb{R})$ is generated by unipotents, Borel density theorem (Theorem 7.9) implies that $\mathrm{SO}(Q)^\circ$ is contained in the Zariski closure of $\Gamma_{g_Q}$; and since $\Gamma_{g_Q} \subset \Gamma = \mathrm{SL}_3(\mathbb{Z})$, we conclude that $\mathrm{SO}(Q)^\circ$ is defined over $\mathbb{Q}$ (Lemma 7.11). Consequently, up to a scalar multiple, $Q$ has integer coefficients (Lemma 7.12).

**Definition 7.8.** A subset $H \subset \mathrm{SL}_l(\mathbb{R})$ is *Zariski closed* if there exists a subset $S \subset \mathbb{R}[x_{1,1}, \ldots, x_{l,l}]$ such that $H = \{g \in \mathrm{SL}_l(\mathbb{R}) \mid Q(g) = 0,\ \forall Q \in S\}$, where we understand $Q(g)$ to denote the value obtained by substituting the matrix entries $g_{i,j}$ into the variables $x_{i,j}$. That is, $H$ is Zariski closed if the matrix entries are characterized by polynomials.

For $H \subset \mathrm{SL}_l(\mathbb{R})$, let $\overline{\overline{H}}$ denote the Zariski closure of $H$, that is, the unique smallest Zariski closed set containing $H$.

**Theorem 7.9** (Borel density theorem). *Let $G \subset \mathrm{SL}_l(\mathbb{R})$ be a closed subgroup, and let $\Gamma$ be a lattice in $G$. Then the Zariski closure $\overline{\overline{\Gamma}}$ of $\Gamma$ contains every unipotent element of $G$.*

Before we prove this, we introduce another lemma.

**Lemma 7.10.** *Let $g \in \mathrm{SL}_m(\mathbb{R})$ be unipotent, and let $\mu$ be a $g$-invariant probability measure on $P\mathbb{R}^{m-1} = (\mathbb{R}^m)^\times/\mathbb{R}^\times$. Then $\mu$ is supported on the set of fixed points of $g$.*

*Proof.* Let $T = g - I$ (then $T$ is nilpotent), and let $v \in (\mathbb{R}^m)^\times$. Let $r$ be such that $vT^r \neq 0$ but $vT^{r+1} = 0$. Then $gT^r v = T^r v$, so $[T^r v]$ is a fixed point of $g$.

On the other hand, it is easy to see $g^n[v] \to [T^r v]$ as $n \to \infty$. By Poincaré recurrence, for $\mu$-every $[v] \in P\mathbb{R}^{m-1}$ there exists a sequence $n_k \to \infty$ such that $g^{n_k}[v] \to [v]$. Since we know that $g^n[v]$ converges to a fixed point of $g$ as $n \to \infty$, we conclude that $\mu$-every point is a fixed point, i.e. $\mu$ is supported on the set of fixed points of $g$. □

*Proof of Borel density theorem.* By Chevalley's theorem (which will not be proved here), there exists a polynomial homomorphism $\rho\colon \mathrm{SL}(l,\mathbb{R}) \to \mathrm{SL}(m,\mathbb{R})$ for some $m$, and a vector $[v] \in P\mathbb{R}^{m-1}$, such that

$$\overline{\overline{\Gamma}} = \{g \in \mathrm{SL}(l,\mathbb{R}) \mid [v]\rho(g) = [v]\}$$

Therefore, $\rho$ induces a well-defined map on $G/\Gamma \to P\mathbb{R}^{m-1}$:

$$\bar{\rho}(g\Gamma) = \rho(g)[v]$$

Let $g \in G$ be unipotent, and let $\mu_0$ be a $G$-invariant probability measure on $G/\Gamma$. This is pushed to a $\rho(G)$-invariant measure on $P\mathbb{R}^{m-1}$ defined by $\mu(A) = \mu_0(\bar{\rho}^{-1}(A))$; and since $\rho(g)$ is unipotent, by the preceding lemma, $\mu$ is supported on the set of fixed points of $\rho(g)$. However, it is not hard to show that $[v]$ lies in the support of $\mu$; and therefore $\rho(g)$ must fix $[v]$, from which $g \in \overline{\overline{\Gamma}}$. □

**Lemma 7.11.** *Let $C$ be a subset of $\mathrm{SL}_l(\mathbb{Q})$; then $\overline{\overline{C}}$ is defined over $\mathbb{Q}$.*

*Proof.* Suppose $\overline{\overline{C}}$ is defined by $S \subset P^d$, where $P^d$ is the set of all polynomials of degree $\leqslant d$. Now, the subspace $\{Q \subset P^d : Q(C) = 0\}$ is defined by linear equations with rational coefficients; and therefore it is spanned by some rational vectors, which therefore determine the set $S$. □

**Lemma 7.12.** *For a nondegenerate quadratic form $Q$, $\mathrm{SO}(Q)$ is defined over $\mathbb{Q}$ if and only if $Q$ is proportional to a form with rational coefficients.*

*Proof.* If $Q$ is a rational form, then $\mathrm{SO}(Q)$ is quite apparently defined over $\mathbb{Q}$; note that $\mathrm{SO}(Q)$ does not depend on the scaling of $Q$.

Conversely, given $\mathrm{SO}(Q)$ defined over $\mathbb{Q}$, our $Q$ is uniquely determined up to scalar multiplication. Consider an automorphism $\phi$ of $\mathbb{R}/\mathbb{Q}$, and notice that $\mathrm{SO}(\phi Q) = \phi\,\mathrm{SO}(Q) = \mathrm{SO}(Q)$; that is, $\phi$ must send $Q$ to a scalar multiple of itself. Now scale $Q$ so that it has one rational coordinate; that coordinate will be fixed by $\phi$, and therefore the scalar multiple must in fact be 1. That is, the scaled $Q$ is invariant under all the automorphisms of $\mathbb{R}/\mathbb{Q}$, and consequently $Q$ is proportional to a form with rational coefficients. □

## References


[BM00] M. B. Bekka and Matthias Mayer. *Ergodic theory and topological dynamics of group actions on homogeneous spaces*, volume 269. Cambridge University Press, Cambridge, 2000.

[Bow76] Rufus Bowen. Weak mixing and unique ergodicity on homogeneous spaces. *Israel journal of mathematics*, 23(3):267–273, 1976.

[Dan78] S. G. Dani. Invariant measures of horospherical flows on noncompact homogeneous spaces. *Inventiones mathematicae*, 47(2):101–138, 1978.

[Dan81] S. G. Dani. Invariant measures and minimal sets of horospherical flows. *Inventiones mathematicae*, 64(2):357–385, 1981.

[Dan86] S. G. Dani. On orbits of unipotent flows on homogeneous spaces, ii. *Ergodic theory and dynamical systems*, 6(2):167–182, 1986.

[DM89] S. G. Dani and G. A. Margulis. Values of quadratic forms at primitive integral points. *Inventiones mathematicae*, 98(2):405–424, 1989.

[DM90] S. G. Dani and G. A. Margulis. Orbit closures of generic unipotent flows on homogeneous spaces of $\mathrm{SL}(3,\mathbb{R})$. *Mathematische annalen*, 286(1-3):101–128, 1990.

[DS84] S. G. Dani and John Smillie. Uniform distribution of horocycle orbits for Fuchsian groups. *Duke mathematical journal*, 51(1):185–194, 1984.

[Ein06] Manfred Einsiedler. Ratner's theorem on $\mathrm{SL}(2,\mathbb{R})$-invariant measures. *Jahresber*, 108(3):143–164, 2006.

[Ein10] Manfred Einsiedler. *Ergodic theory: with a view towards number theory*, volume 259. Springer, New York, 2010.

[EK07] A. Eskin and D. Kleinbock. Unipotent flows and applications. Lecture Notes for Clay Institute Summer School, 2007.

[EP78] Robert Ellis and William Perrizo. Unique ergodicity of flows on homogeneous spaces. *Israel journal of mathematics*, 29(2-3):276–284, 1978.

[Ghy92] Etienne Ghys. Dynamique des flots unipotents sur les espaces homogènes. *Astérisque*, 3(206):93–136, 1992. Séminaire Bourbaki, Vol. 1991/92.

[Hed08] Gustav A. Hedlund. The dynamics of geodesic flows. *Bulletin (new series) of the American Mathematical Society*, 45(4):241–260, 1939; 2008.

[KH95] A. B. Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54. Cambridge University Press, Cambridge; New York, NY, USA, 1995.

[Mar71] G. A. Margulis. On the action of unipotent groups in a lattice space. *Mathematics of the USSR. Sbornik*, 15(4):549–554, 1971.

[Mar04] Grigoriy A. Margulis. *On Some Aspects of the Theory of Anosov Systems: With a Survey by Richard Sharp: Periodic Orbits of Hyperbolic Flows*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[Moo80] Calvin C. Moore. The Mautner phenomenon for general unitary representations. *Pacific J. Math.*, 86(1):155–169, 1980.

[Mor05] Dave W. Morris. *Ratner's Theorems on Unipotent Flows*. University of Chicago Press, 2005.

[MT96] G. A. Margulis and G. M. Tomanov. Measure rigidity for almost linear groups and its applications. *Journal d'analyse mathématique (Jerusalem)*, 69(1):25–54, 1996.

[Par71] William Parry. Metric classification of ergodic nilflows and unipotent affines. *American journal of mathematics*, 93(3):819–828, 1971.

[Pet83] Karl E. Petersen. *Ergodic theory*, volume 2. Cambridge University Press, New York; Cambridge [Cambridgeshire], 1983.

[Rat82a] Marina Ratner. Factors of horocycle flows. *Ergodic theory and dynamical systems*, 2(3-4):465–489, 1982.

[Rat82b] Marina Ratner. Rigidity of horocycle flows. *Annals of mathematics*, 115(3):597–614, 1982.

[Rat83] Marina Ratner. Horocycle flows, joinings and rigidity of products. *Annals of mathematics*, 118(2):277–313, 1983.

[Rat90] Marina Ratner. On measure rigidity of unipotent subgroups of semisimple groups. *Acta mathematica*, 165(1):229–309, 1990.

[Rat91] Marina Ratner. On Raghunathan's measure conjecture. *Annals of mathematics*, 134(3):545–607, 1991.

[Rat95] Marina Ratner. Raghunathan's conjectures for Cartesian products of real and $\mathfrak{p}$-adic lie groups. *Duke mathematical journal*, 77(2):275–382, 1995.

[Sha91] Nimish A. Shah. Uniformly distributed orbits of certain flows on homogeneous spaces. *Mathematische annalen*, 289(1):315–334, 1991.

[Sha98] Nimish A. Shah. Invariant measures and orbit closures on homogeneous spaces for actions of subgroups generated by unipotent elements. *Lie Groups and Ergodic Theory*, pages 229–271, 1998.

[Sta89] A. N. Starkov. Solvable homogeneous flows. *Mathematics of the USSR. Sbornik*, 62(1):243–260, 1989.

[Vee77] William A. Veech. Unique ergodicity of horospherical flows. *American journal of mathematics*, 99(4):827–859, 1977.

[Wit87] Dave Witte. Zero-entropy affine maps on homogeneous spaces. *American journal of mathematics*, 109(5):927–961, 1987.

School of Mathematical Sciences, Peking University, 100871, Beijing, China
*Email address*: daiwenhan@pku.edu.cn