Yuri Lima

Lectures on Ratner's Theory

Dynamics Beyond Uniform Hyperbolicity

May 6, 2016

Contents

1 Dynamics on hyperbolic surfaces ... 7
1.1 The hyperbolic plane $\mathbb{H}$ ... 7
1.1.1 Inner product ... 8
1.1.2 Möbius transformations ... 8
1.1.3 Geodesics of $\mathbb{H}$ ... 9
1.1.4 Hyperbolic surfaces ... 9
1.1.5 The geodesic flow ... 10
1.1.6 The horocycle flows ... 11
1.2 Algebraic description of g, h^+, h^- ... 12
1.2.1 The group $\mathrm{SL}(2, \mathbb{R})$ ... 13
1.2.2 The linear action of $\mathrm{SL}(2, \mathbb{R})$ on $\mathbb{R}^2$ ... 13
1.2.3 The hyperbolic action of $\mathrm{SL}(2, \mathbb{R})$ on $\mathbb{H}$ ... 14
1.2.4 The action of $\mathrm{PSL}(2, \mathbb{R})$ on $T^1\mathbb{H}$ ... 14
1.2.5 The geodesic flow in $\mathrm{PSL}(2, \mathbb{R})$-coordinates ... 14
1.2.6 The horocycle flow in $\mathrm{PSL}(2, \mathbb{R})$-coordinates ... 15
1.2.7 Comparison of $\mathrm{SL}(2, \mathbb{R})$ and $\mathrm{PSL}(2, \mathbb{R})$ ... 16
1.2.8 Ergodic theory of Anosov flows ... 16

2 Ratner's theorems ... 19
2.1 Homogeneous spaces ... 19
2.1.1 Group actions ... 19
2.1.2 Homogeneous spaces: general definition ... 20
2.1.3 Lie groups ... 20
2.1.4 Invariant distances ... 20
2.1.5 Haar measures ... 21
2.1.6 Quotient spaces and lattices ... 22
2.1.7 Homogeneous spaces as quotient spaces ... 25
2.2 Unipotent actions ... 25
2.2.1 Lie algebras ... 25
2.2.2 Adjoint transformations and unipotent elements ... 26
2.3 Ratner's theorems ... 27

2.3.1 Ratner's orbit closure theorem 27
2.3.2 Ratner's measure classification theorem 28
2.3.3 Ratner's equidistribution theorem 28
2.3.4 A simple example: toral translations 29

3 Horocycle flows of hyperbolic surfaces 31
3.1 More on the group $\mathrm{SL}(2,\mathbb{R})$ 32
3.1.1 KAN decomposition 32
3.1.2 Cartan decomposition 32
3.1.3 $\mathrm{SL}(2,\mathbb{R})$ is generated by unipotent elements 33
3.1.4 Decomposition of Haar measures 33
3.1.5 Haar measure of $\mathrm{SL}(2,\mathbb{R})$ 34
3.2 Hedlund's theorem 34
3.3 Mautner phenomenon 35
3.3.1 Representations 35
3.3.2 Mautner's lemma 36
3.3.3 Mautner phenomenon for $\mathrm{SL}(2,\mathbb{R})$ 36
3.4 Howe-Moore's theorem 37
3.5 Furstenberg's theorem 38
3.5.1 Hedlund and Ratner imply Furstenberg 40
3.6 Dani-Smillie's theorem 41
3.6.1 Hedlund and Ratner imply Dani-Smillie 41

4 Oppenheim's conjecture 43
4.1 From Oppenheim to Raghunathan 43
4.1.1 The assumptions on Q are necessary 44
4.2 The moduli space $\mathrm{SL}(n,\mathbb{R})/\mathrm{SL}(n,\mathbb{Z})$ 45
4.2.1 Mahler compactness criterium 45
4.2.2 X_n has finite volume 47
4.3 Proof of Oppenheim's conjecture 47
4.3.1 Raghunathan's observation 47
4.3.2 Proof of Oppenheim's conjecture 48

5 Appendix 51
5.1 $\mathrm{SO}(2,1)^\circ$ and $\mathrm{SL}(2,\mathbb{R})$ are locally isomorphic 51
5.2 The universal elliptic curve 52
5.2.1 Semidirect products 52
5.2.2 Description of X 53

Preface

The sole purpose of these lecture notes is to help the author in the organization of the mini-course "Introduction to Ratner theory", to be held at IMPA on August 2013, as part of the activities of the thematic semester "Dynamics Beyond Uniform Hyperbolicity".

The aim of the mini-course is to introduce Ph.D. students to some beautiful topics of dynamics on homogeneous spaces, with emphasis on Ratner's theorems on the classification of orbit closures, invariant measures and equidistribution of actions of subgroups generated by unipotent elements on homogeneous spaces.

We tried to give as much as possible details in two situations: when the topic is not of common knowledge for IMPA students, and when the details provide a better understanding of the underlying idea. In all other cases we skipped the details and gave references for the interested reader. Some recommended references for beginners: [1, 3, 7, 19].

Again: these notes were intended for the author's understanding, thus be aware of the many mistakes contained in the text.

1

Dynamics on hyperbolic surfaces

This first chapter deals with the simplest example of dynamics on homogeneous spaces: the geodesic and horocycle flows on surfaces of constant negative curvature. It is the “toy model” to understand the role of dynamics in homogeneous spaces, both because of its simple description and richness of dynamical phenomena.

1.1 The hyperbolic plane $\mathbb{H}$

Let $\mathbb{H} = \{z = x + iy : y > 0\}$ denote the upper-half plane of $\mathbb{C}$. $\mathbb{H}$ is a manifold of dimension two. Each tangent space $T_z\mathbb{H}$ is naturally identified with $\mathbb{R}^2 \cong \mathbb{C}$. Also, the identity $z \in \mathbb{H} \mapsto z \in \mathbb{R}^2$ is a global chart, thus $T\mathbb{H} \cong \mathbb{H} \times \mathbb{R}^2 \cong \mathbb{H} \times \mathbb{C}$.

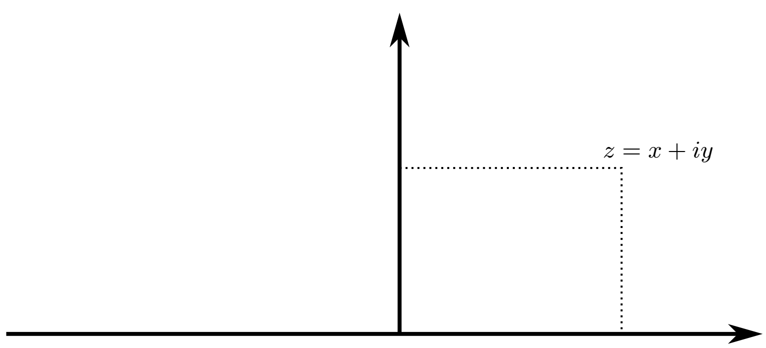

Fig. 1.1. The hyperbolic plane $\mathbb{H}$.

1.1.1 Inner product

We consider a metric in $\mathbb{H}$, called the *hyperbolic metric*, as follows: given $v, w \in T_z\mathbb{H} \cong \mathbb{R}^2$ with $z = x + iy$, we let

$$(v, w)_z = \frac{1}{y^2}(v, w),$$

where (v, w) represents the usual euclidean inner product in $\mathbb{R}^2$. The hyperbolic metric is conformal to the euclidean metric, thus any isometry of $\mathbb{H}$ preserves euclidean angles[1]. If $y = 1$, then $(v, w)_z = (v, w)$ and so the hyperbolic distance of points on the line $y = 1$ is the euclidean distance. As y grows, the distances decrease.

The hyperbolic plane together with the hyperbolic metric is called the *hyperbolic space* or the *Lobachevsky space*. It is a complete manifold. We will usually call it just hyperbolic plane and denote it simply by $\mathbb{H}$, the hyperbolic metric being implicit.

$\mathbb{H}$ is homogeneous: given any two points $z, w \in \mathbb{H}$, there is an isometry that sends z to w. Thus $\mathbb{H}$ is highly symmetric, all parts of the space looking the same.

1.1.2 Möbius transformations

A way of understanding $\mathbb{H}$ is by first identifying its isometries. Let

$$\mathrm{SL}(2, \mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R}, ad - bc = 1 \right\}.$$

Given $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$, consider the linear fractional transformation $g : \mathbb{C} \to \mathbb{C}$ defined by $g.z = \frac{az+b}{cz+d}$. Observe that $g.\mathbb{H} = \mathbb{H}$: if $z = x + iy \in \mathbb{H}$, then

$$g.z = \frac{az + b}{cz + d} = \frac{ac(x^2 + y^2) + (ad + bc)x + bd}{|cz + d|^2} + i\frac{y}{|cz + d|^2} \in \mathbb{H}. \tag{1.1}$$

We call both $g : \mathbb{C} \to \mathbb{C}$ and its restriction $g : \mathbb{H} \to \mathbb{H}$ a *Möbius transformation*. Let Mob denote the set of all Möbius transformations. Mob is a group under the composition operation, and it acts on $\mathbb{H}$ in the canonical way.

Given $g, h \in \mathrm{SL}(2, \mathbb{R})$, the Möbius transformation induced by the product gh is the composition of the Möbius transformations induced by g and h. Thus the map $\mathrm{SL}(2, \mathbb{R}) \mapsto$ Mob described above is a group homomorphism. In particular, $\mathrm{SL}(2, \mathbb{R})$ acts on $\mathbb{H}$.

The Möbius transformations represent a large source of symmetries of $\mathbb{H}$.

[1] Thinking of lines as circles passing through the point at infinity, an isometry of $\mathbb{H}$ sends circles to circles.

Lemma 1.1. *Every $g \in \mathrm{Mob}$ is an isometry of $\mathbb{H}$.*

Proof. If $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$, then $dg(z) = \frac{a(cz+d)-(az+b)c}{(cz+d)^2} = \frac{1}{(cz+d)^2}$. This and equation (1.1) imply that if $v, w \in T_z\mathbb{H}$, then

$$\begin{aligned}
(D(g)v, D(g)w)_{g.z} &= \frac{1}{\frac{y^2}{|cz+d|^4}}(D(g)v, D(g)w) \\
&= \frac{|cz+d|^4}{y^2}\left(\frac{v}{|cz+d|^2}, \frac{w}{|cz+d|^2}\right) \\
&= \frac{1}{y^2}(v, w) \\
&= (v, w)_z.
\end{aligned}$$

Indeed, Mob is the set of all orientation-preserving isometries of $\mathbb{H}$.

1.1.3 Geodesics of $\mathbb{H}$

Given $v \in T\mathbb{H}$, there is exactly one geodesic with velocity v. If $v = (1, 0) \in T_i\mathbb{H}$, then this geodesic is the vertical line $x = 0$. Because the Möbius transformations preserve angles, they send this vertical line to a semicircle orthogonal to the real axis. Because they are isometries of $\mathbb{H}$, any such semicircle is again a geodesic. Running over all elements of $\mathrm{SL}(2, \mathbb{R})$, we obtain all geodesics of $\mathbb{H}$: they are semicircles orthogonal to the real axis and vertical lines.

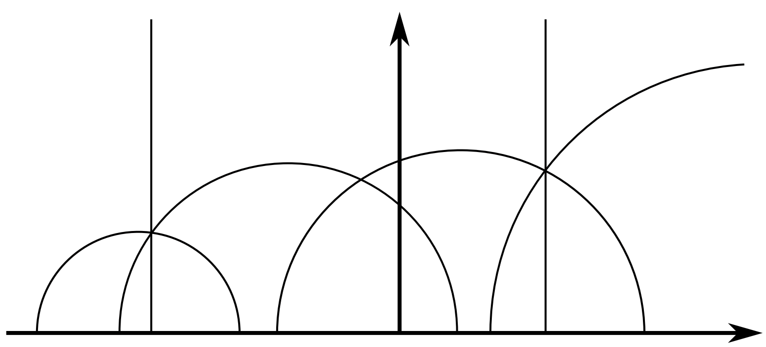

Fig. 1.2. The geodesics of $\mathbb{H}$.

1.1.4 Hyperbolic surfaces

A complete, orientable smooth surface X is called a *hyperbolic surface* if it has constant curvature equal to -1. The hyperbolic plane $\mathbb{H}$ gives rise to all hyperbolic surfaces: by the Riemann's uniformization theorem, if X is a hyperbolic surface, then its universal cover is $\mathbb{H}$. In other words, $X = \Gamma\backslash\mathbb{H}$, where Γ is a discrete subgroup of orientation-preserving isometries of $\mathbb{H}$, i.e. Γ is a discrete subgroup of Mob. In this case, $TX = \Gamma\backslash T\mathbb{H}$.

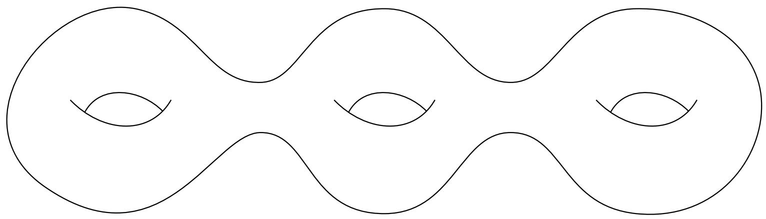

Fig. 1.3. An example of a hyperbolic surface.

1.1.5 The geodesic flow

The geodesic flow on $\mathbb{H}$ is the flow defined on the unit tangent bundle $T^1\mathbb{H}$ whose orbits are the geodesics of $\mathbb{H}$, parameterized by arc length. Formally, let $v \in T^1\mathbb{H}$, and let γ be the unique geodesic such that $\gamma'(0) = v$. The geodesic flow $g = \{g_t\}_{t\in\mathbb{R}}$ is given by $g_t v = \gamma'(t)$. It is well defined because $\mathbb{H}$ is complete.

The path that a geodesic defines in $\mathbb{H}$ is a curve which goes to infinity for positive and for negative times. Thus the geodesic flow has no recurrent properties. The situation changes if $X = \Gamma\backslash\mathbb{H}$ is e.g. a compact hyperbolic surface. The geodesic flow of X is the projection of $g : T^1\mathbb{H} \to T^1\mathbb{H}$ to $TX = \Gamma\backslash T^1\mathbb{H}$.

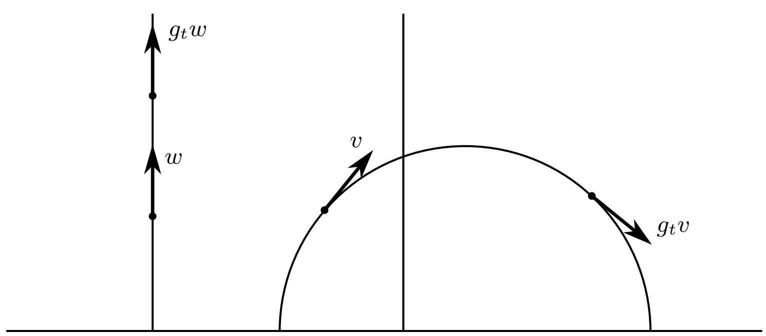

Fig. 1.4. The geodesic flow of $\mathbb{H}$.

1.1.6 The horocycle flows

Even not being interesting in the ergodic theoretical point of view, the geodesic flow of $\mathbb{H}$ is very special dynamically: it has stable and unstable foliations.

Given $v \in T\mathbb{H}$, we define the *stable manifold* or *stable leaf* of v as the set

$$W^s(v) = \left\{ w \in T\mathbb{H} : \lim_{t \to +\infty} d(g_t v, g_t w) = 0 \right\}$$

of all vectors with the same future as v, and the *unstable manifold* or *unstable leaf* of v as the set

$$W^u(v) = \left\{ w \in T\mathbb{H} : \lim_{t \to -\infty} d(g_t v, g_t w) = 0 \right\}$$

of all vectors with the same past as v.

For example, if $\mathbf{i}$ denotes the vector $i \in T_i^1\mathbb{H}$, then $W^s(\mathbf{i})$ is the set of unit vertical vectors with base point in the line $y = 1$ and pointing upward, and $W^u(\mathbf{i})$ is the set of unit vectors pointing outside and orthogonal to the circle that is orthogonal to $\mathbf{i}$ and tangent to the real axis (see Figure 1.5).

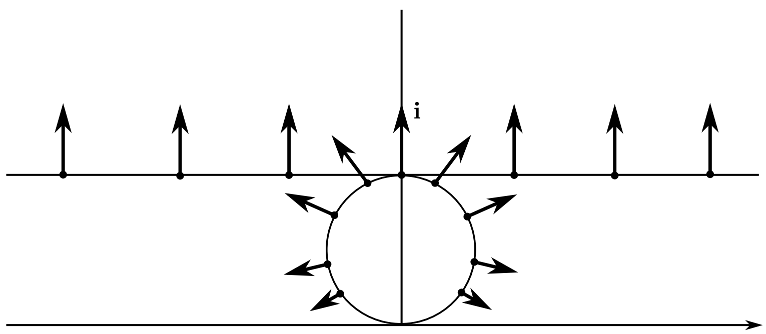

Fig. 1.5. The stable and unstable leafs of $\mathbf{i}$: vectors on the horizontal line form $W^s(\mathbf{i})$, and vectors orthogonal to the circle form $W^u(\mathbf{i})$.

The images of $W^s(\mathbf{i})$ and $W^u(\mathbf{i})$ under elements of Mob provide all other stable and unstable leaves. The stable leaves are of two types:

(i) If v is a unit vertical vector pointing upward, then $W^s(v)$ is the set of unit vertical vectors with base point in the same horizontal line as v, and pointing upward.

(ii) If v is a unit vector and it is not a vertical vector pointing upward, then $W^s(v)$ is the set of unit vectors pointing inside and orthogonal to the circle that is orthogonal to v and tangent to the real axis.

A similar description holds for the unstable leaves:

(i) If v is a unit vertical vector pointing downward, then $W^u(v)$ is the set of unit vertical vectors with base point in the same horizontal line as v, and pointing downward.
(ii) If v is a unit vector and it is not a vertical vector pointing downward, then $W^u(v)$ is the set of unit vectors pointing outside and orthogonal to the circle that is orthogonal to v and tangent to the real axis.

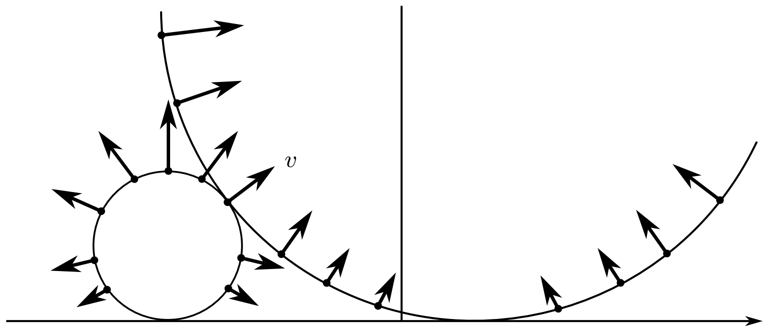

Fig. 1.6. The stable and unstable leafs of a generic $v \in T\mathbb{H}$: vectors orthogonal to the circle on the right form $W^s(v)$, and vectors orthogonal to the circle on the left form $W^u(v)$.

The family of stable leaves defines a one dimensional foliation of $T^1\mathbb{H}$, called the *stable foliation*, and the family of unstable leaves defines the *unstable foliation*. If we parameterize each stable leaf by arc length, then the stable foliation gives rise to a flow h^+, called the *stable horocycle flow* of $\mathbb{H}$. Doing the same for the unstable leafs, we define a flow h^-, called the *unstable horocycle flow* of $\mathbb{H}$. Whenever there is no confusion, we will suppress the words stable/unstable and symbols $\pm$, and refer to these flows simply as horocycle flows and denote them by h.

Like in the geodesic flow, the horocycle flows of $\mathbb{H}$ have no recurrent properties, but the horocycle flows of hyperbolic surfaces $X = \Gamma\backslash\mathbb{H}$ do if X is e.g. compact. The horocycle flow of X is the projection of $h : T^1\mathbb{H} \to T^1\mathbb{H}$ to $TX = \Gamma\backslash T^1\mathbb{H}$.

1.2 Algebraic description of g, h^+, h^-

In this section we will obtain an algebraic description of the flows g, h^+, h^-. We already started doing this in the description of a group homomorphism between $\mathrm{SL}(2,\mathbb{R})$ and Mob. Let Φ denote this homomorphism. Φ is "almost" an isomorphism, because its kernel is $\{\pm\mathrm{Id}\}$: $\frac{az+b}{cz+d} = z$ iff $cz^2 + (d-a)z - b = 0$ iff $(a,b,c,d) = \pm(1,0,0,1)$. Thus Φ induces an isomorphism between $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm\mathrm{Id}\}$ and Mob. We will see in Section 1.2.7 that the difference between $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{PSL}(2,\mathbb{R})$ is not relevant for us, because any information on $\mathrm{SL}(2,\mathbb{R})$ and its subgroups translates to information on their projective versions. Nevertheless, in the algebraic description of g, h^+, h^- we will use $\mathrm{PSL}(2,\mathbb{R})$.

1.2.1 The group $\mathrm{SL}(2,\mathbb{R})$

$\mathrm{SL}(2,\mathbb{R})$ has many important subgroups. Here are the important ones for us:

$$\begin{aligned}
K &= \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\} \\
A &= \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} : a > 0 \right\} \\
A^+ &= \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} : a \geq 1 \right\} \\
U^+ &= \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{R} \right\} \\
U^- &= \left\{ \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} : b \in \mathbb{R} \right\}.
\end{aligned} \tag{1.2}$$

Each of them is one dimensional. K is called the *rotation group*, and it is isomorphic to $\mathbb{S}^1$. A is called the *diagonal group*, and it is isomorphic to $\mathbb{R}^*$. U^+, U^- are called the *unipotent groups*, and they are isomorphic to $\mathbb{R}$. An element of K is called *elliptic*, an element of A is called *hyperbolic*, and an element of U^+ or U^- is called *parabolic*. This nomenclature comes from the way each of them acts on $\mathbb{R}^2$, as we'll now see.

1.2.2 The linear action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{R}^2$

Below we list some useful properties of the linear action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{R}^2$. All of them are easily proved.

- K acts by rotation.
- A acts by hyperbolic matrices.
- U^+ fixes each point in the x-axis.
- U^+ translates horizontal lines $x = c$, for $c \neq 0$.
- U^- fixes each point in the y-axis.
- U^- translates vertical lines $y = c$, for $c \neq 0$.

1.2.3 The hyperbolic action of $\mathbf{SL(2,\mathbb{R})}$ on $\mathbb{H}$

Let's see what are the Möbius transformations associated to A and U^+. If $a = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \in A$, then $a.z = \lambda^2 z$ is a dilation. If $n = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \in U^+$, then $n.z = z + b$ is a horizontal translation.

In particular, $\mathrm{SL}(2,\mathbb{R})$ acts transitively on $\mathbb{H}$: for any $z, w \in \mathbb{H}$, there is $g \in \mathrm{SL}(2,\mathbb{R})$ such that $g.z = w$. To see this, fix $z = i$ and let $w = x + iy$. Thus $\begin{pmatrix} \sqrt{y} & 0 \\ 0 & \sqrt{y}^{-1} \end{pmatrix}$ sends i to iy, and $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$ sends iy to $x + iy$.

Here are some other (easily proved) properties of the hyperbolic action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H}$.

- K is the stabilizer of i, and it acts on $T_i\mathbb{H}$ by rotation.
- K acts transitively on any circle centered at i.
- A^+ is a dilation with factor > 1 that preserves the y-axis.

1.2.4 The action of $\mathbf{PSL(2,\mathbb{R})}$ on $\boldsymbol{T^1}\mathbb{H}$

Each element of $\mathrm{PSL}(2,\mathbb{R})$ defines an isometry of $\mathbb{H}$, thus $\mathrm{PSL}(2,\mathbb{R})$ acts on $T^1\mathbb{H}$ as well. Let us show that this action is also transitive: given $v, w \in T^1\mathbb{H}$, there is $g \in \mathrm{PSL}(2,\mathbb{R})$ such that $Dg(v) = w$.

Fix $w = \mathbf{i}$ and $v \in T^1_z\mathbb{H}$. By the previous section, there is $h \in \mathrm{PSL}(2,\mathbb{R})$ such that $h.z = i$. The group K acts transitively on $T^1_i\mathbb{H}$. Thus there is $k \in K$ that rotates the vector $D(h)v$ to $\mathbf{i}$. By the chain rule, $g = kh$ satisfies $D(g)v = D(k)D(h)v = \mathbf{i}$.

The existence of g above is unique: if $h \in \mathrm{PSL}(2,\mathbb{R})$ also satisfies $D(h)v = w$, then $h = g$. To see this, consider g, h as elements of $\mathrm{SL}(2,\mathbb{R})$ and note that gh^{-1} fixes $\mathbf{i}$. Because it fixes i, we have $gh^{-1} \in K$. Because it also fixes the direction $\mathbf{i}$, we have $g = \pm h$. Coming back to $\mathrm{PSL}(2,\mathbb{R})$, this means that $h = g$. Thus the map $g \in \mathrm{PSL}(2,\mathbb{R}) \mapsto D(g)\mathbf{i} \in T^1\mathbb{H}$ is a continuous bijection, i.e. $T^1\mathbb{H}$ can be viewed as a Lie group.

In the sequel we represent the geodesic and horocycle flows via this algebraic perspective.

1.2.5 The geodesic flow in $\mathbf{PSL(2,\mathbb{R})}$-coordinates

We start by defining the geodesic through $\mathbf{i}$. This geodesic is the imaginary axis, and its parameterization by arc length is $\gamma(t) = e^t i$, $t \in \mathbb{R}$. Indeed,

$$(\gamma'(t), \gamma'(t))_{\gamma(t)} = \frac{1}{e^{2t}}(e^t i, e^t i) = 1.$$

Thus $g_t(\mathbf{i}) = e^t i \in T^1_{e^t i}\mathbb{H}$. The matrix of $\mathrm{PSL}(2,\mathbb{R})$ that sends $\mathbf{i}$ to $g_t(\mathbf{i})$ is the diagonal matrix $\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}$. To see this, note that $\dfrac{e^{\frac{t}{2}} \cdot i + 0}{0 \cdot i + e^{-\frac{t}{2}}} = e^t i$, and that $D\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}\mathbf{i}$ is a vertical unit vector.

Now let $v \in T^1\mathbb{H}$ be arbitrary, and let $g \in \mathrm{PSL}(2,\mathbb{R})$ such that $D(g)\mathbf{i} = v$. Because g is an isometry, it sends the geodesic tangent to $\mathbf{i}$ to the geodesic tangent to v. In particular, it sends $\gamma'(t)$ to $g_t(v)$.

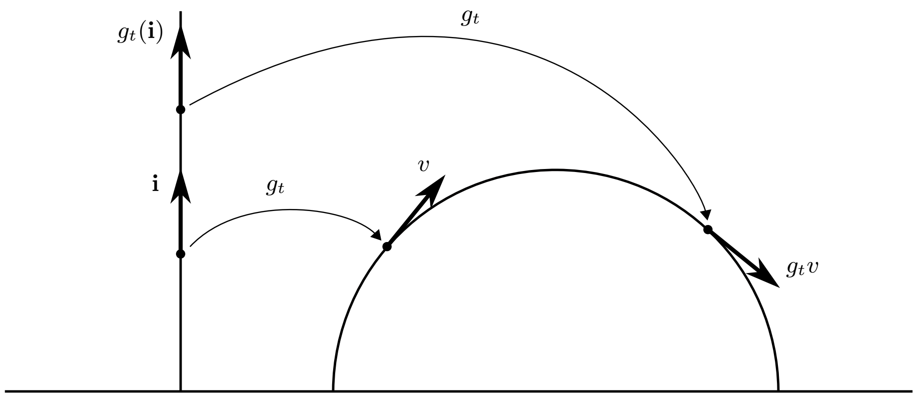

Fig. 1.7. The geodesic flow in $\mathrm{SL}(2,\mathbb{R})$-coordinates.

By composition, the element of $\mathrm{PSL}(2,\mathbb{R})$ that sends $\mathbf{i}$ to $g_t(v)$ is the product $g\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}$. Thus in $\mathrm{PSL}(2,\mathbb{R})$-coordinates g_t is a right multiplication:

$$\begin{aligned} g_t : \mathrm{PSL}(2,\mathbb{R}) &\to \mathrm{PSL}(2,\mathbb{R}) \\ g &\mapsto g\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}. \end{aligned} \tag{1.3}$$

Now let $X = \Gamma\backslash\mathbb{H}$ be an arbitrary hyperbolic surface, and let g_t also represent the geodesic flow of X. Because $\mathbb{H}$ is the universal cover of X, $T^1X = \Gamma\backslash T^1\mathbb{H}$ and g_t is the projection of the geodesic flow of $\mathbb{H}$ under the quotient map $\mathbb{T}^1\mathbb{H} \mapsto \Gamma\backslash T^1\mathbb{H}$. Thus g_t can also be represented in $\mathrm{PSL}(2,\mathbb{R})$-coordinates as a right multiplication:

$$\begin{aligned} g_t : \Gamma\backslash\mathrm{PSL}(2,\mathbb{R}) &\to \Gamma\backslash\mathrm{PSL}(2,\mathbb{R}) \\ \Gamma g &\mapsto \Gamma g\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}. \end{aligned} \tag{1.4}$$

1.2.6 The horocycle flow in $\mathbf{PSL(2,\mathbb{R})}$-coordinates

We proceed as above to obtain an algebraic description of the horocycle flows h^+, h^-. We will do the calculations for h^+, and leave the calculations for h^- as an exercise to the reader.

Again, fix $\mathbf{i}$. Its stable horocycle is $W^s(\mathbf{i}) = \{i \in T^1_z\mathbb{H} : z \in \mathbb{R} + i\}$ (see Figure 1.5). On $W^s(\mathbf{i})$, the hyperbolic metric coincides with the euclidean metric, thus $\gamma(t) = t + i$, $t \in \mathbb{R}$, parameterizes the horizontal line $\mathbb{R} + i$ by arc length, and $h_t^+(\mathbf{i}) = \gamma'(t) = i \in T^1_{t+i}\mathbb{H}$. The element of $\mathrm{PSL}(2,\mathbb{R})$ that sends $\mathbf{i}$ to $\gamma'(t)$ is the element $n \in U^+$ that sends i to $\gamma(t)$, i.e. it is $n = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$. Consequently, if $X = \Gamma\backslash\mathbb{H}$ is an arbitrary hyperbolic surface, then in $\mathrm{PSL}(2,\mathbb{R})$-coordinates h_t^+ is a right multiplication:

$$\begin{aligned} h_t^+ : \Gamma\backslash\mathrm{PSL}(2,\mathbb{R}) &\to \Gamma\backslash\mathrm{PSL}(2,\mathbb{R}) \\ \Gamma g &\mapsto \Gamma g \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{1.5}$$

Analogously,

$$\begin{aligned} h_t^- : \Gamma\backslash\mathrm{PSL}(2,\mathbb{R}) &\to \Gamma\backslash\mathrm{PSL}(2,\mathbb{R}) \\ \Gamma g &\mapsto \Gamma g \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} \end{aligned}$$

is the horocycle flow of X on $\mathrm{PSL}(2,\mathbb{R})$-coordinates.

1.2.7 Comparison of $\mathbf{SL(2,\mathbb{R})}$ and $\mathbf{PSL(2,\mathbb{R})}$

Our main object of interest is the unit tangent bundle of hyperbolic surfaces. As we saw in Section 1.2.4, they take the form $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$, for some discrete subgroup Γ of $\mathrm{PSL}(2,\mathbb{R})$. We want to convince the reader that we can work with $\mathrm{SL}(2,\mathbb{R})$ instead of $\mathrm{PSL}(2,\mathbb{R})$. The reason is that unit tangent bundles of hyperbolic surfaces can also be represented as $\Gamma\backslash\mathrm{SL}(2,\mathbb{R})$, where Γ is a discrete subgroup of $\mathrm{SL}(2,\mathbb{R})$.

To see this, let Γ be a symmetric subgroup of $\mathrm{SL}(2,\mathbb{R})$, i.e. $-g \in \Gamma$ whenever $g \in \Gamma$. Let $P : \mathrm{SL}(2,\mathbb{R}) \to \mathrm{PSL}(2,\mathbb{R})$ and $\pi : \mathrm{PSL}(2,\mathbb{R}) \to P\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$ be the canonical projections. Thus $\mathrm{Ker}(\pi P) = \Gamma$, and so $\Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ is isomorphic to $P\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$. As Γ runs over the symmetric subgroups of $\mathrm{SL}(2,\mathbb{R})$, $P\Gamma$ runs over the subgroups of $\mathrm{PSL}(2,\mathbb{R})$. Most of the time we will be working with $\mathrm{SL}(2,\mathbb{R})$.

1.2.8 Ergodic theory of Anosov flows

Let X be a compact hyperbolic surface. Its geodesic flow represents the basic example of Anosov flows.

Definition 1.2. *Let M be a smooth compact manifold, and let φ be a smooth flow on M. We say that φ is Anosov if there is a continuous decomposition $TM = E^s \oplus E^c \oplus E^u$ and numbers $C > 0$, $\lambda > 0$ such that:*

(i) E^c is tangent to the orbits of the flow,
(ii) E^s, E^c, E^u are $D\varphi$-invariant,
(iii) $\|D\varphi_t v^s\| \le Ce^{-\lambda t}$ for all $v^s \in T^1E^s$ and $t \ge 0$,
(iv) $\|D\varphi_{-t} v^u\| \le Ce^{-\lambda t}$ for all $v^s \in T^1E^u$ and $t \ge 0$.

Anosov proved that a flow satisfying (i)–(iv) exhibits chaotic behavior in many different perspectives. For example, it has infinitely many periodic orbits. Actually, Margulis proved that if X is a compact hyperbolic surface and $\pi(T)$ is the number of closed geodesics of X of length at most T, then

$$\lim_{T\to+\infty} 2Te^{-T}\pi(T) = 1.$$

In particular, it is impossible to classify all invariant measures for g_t: they are just too many. We will see that the horocycle flow belongs to the other spectrum, and that it is uniquely ergodic in most situations. The main reason for this is that, contrary to the geodesic flow, which is defined by hyperbolic matrices, the horocycle flow is defined by unipotent matrices, and their "polynomial behavior" does not allow chaotic behavior. This is the role of Ratner's theory.

Let $\mu =$ (Haar measure on X) $\times$ (Haar measure on S^1). μ is a measure supported on T^1X, invariant under g_t. Below we list some properties of geodesic flows of compact hyperbolic surfaces.

- g has positive topological entropy[2].
- g has Markov partitions. In general, Axiom A flows have Markov partitions (Bowen [2]).
- μ is mixing (see Chapter 3).
- (g, μ) is a K-flow (Anosov and Sinai).
- (g, μ) is a Bernoulli-flow: every pair (g_t, μ) is metrically conjugate to a full shift on finitely many symbols with a product measure (Ornstein and Weiss [20]).

We end this chapter proving that both h^+, h^- have zero topological entropy. Let us show this for h^+. By the matrix identity

$$\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-\frac{t}{2}} & 0 \\ 0 & e^{\frac{t}{2}} \end{pmatrix} = \begin{pmatrix} 1 & se^t \\ 0 & 1 \end{pmatrix} \tag{1.6}$$

we have $g_t h_s g_{-t} = h_{se^t}$, i.e. h_s and h_{se^t} are conjugate. In particular, $s = 1$ and $t = \log 2$ give that h_1 and h_2 are conjugate, thus h_1 has zero entropy.

[2] The topological entropy of a flow φ is defined as the topological entropy of its time-one map.

2

Ratner's theorems

In this chapter we will state Ratner's theorems, and provide the necessary background to understand their setup. In a few words, the idea is to use horocycle flows of hyperbolic surfaces as a basis model: hyperbolic surfaces will be changed to quotient spaces of Lie groups by lattices, and horocycle flows will be changed to actions of subgroups generated by unipotent elements.

2.1 Homogeneous spaces

2.1.1 Group actions

Let G be a locally compact Hausdorff topological group with identity e, and let X be a locally compact Hausdorff topological space. A *left action* of G on X is a continuous map

$$\begin{aligned} G \times X &\to X \\ (g, x) &\mapsto gx \end{aligned}$$

such that:

(i) e acts by identity: $ex = x$ for all $x \in X$,
(ii) associative: $(gh)x = g(hx)$ for all $g, h \in G$ and $x \in X$.

A *right action* of G on X is defined similarly: it is a continuous map

$$\begin{aligned} X \times G &\to X \\ (x, g) &\mapsto xg \end{aligned}$$

such that:

(i) e acts by identity: $xe = x$ for all $x \in X$,
(ii) associative: $x(gh) = (xg)h$ for all $g, h \in G$ and $x \in X$.

We will mostly consider right actions. We can always assume that this is the case, because every left action defines a right action and vice-versa: if $(g, x) \in G \times X \mapsto gx \in X$ is a left action, then $(x, g) \in X \times G \mapsto g^{-1}x \in X$ is a right action.

2.1.2 Homogeneous spaces: general definition

Assume that G acts on X on the right.

Definition 2.1. *We say that X is a homogeneous G-space or simply a homogeneous space if G acts transitively, i.e. if for each $x, y \in X$ there is $g \in G$ such that $xg = y$.*

If we think of the elements of G as isometries of X, then a homogeneous space is a space that locally looks the same at each point.

We will see in the next sections that a homogeneous space X is homeomorphic to a quotient space $\Gamma \backslash G$, where Γ is a discrete subgroup of G. If furthermore X has a G-invariant probability measure, then Γ is a lattice.

Classically, the notion of homogeneous spaces is restricted to the action of a Lie group on a manifold (see e.g. [14]). When G is a Lie group, the classical definition coincides with Definition 2.1.

2.1.3 Lie groups

A smooth manifold G is called a *Lie group* if it has group operations of multiplication and inversion such that the map

$$\begin{aligned} G \times G &\to G \\ (x, y) &\mapsto xy^{-1} \end{aligned} \tag{2.1}$$

is smooth. As being a group, G itself acts on G both on the right and on the left.

The smoothness of the map (2.1) implies that, for each $h \in G$, both the *left translation* $L_h : g \mapsto hg$ and the *right translation* $R_h : g \mapsto gh$ are smooth. To see this, note that L_h is the composition of the immersion $g \in G \mapsto (h, g^{-1}) \in G \times G$ with the map (2.1), and a similar statement holds for R_h.

Here are some examples of Lie groups:

- $(\mathbb{S}^1, \cdot)$ is a one dimensional commutative compact Lie group.
- $(\mathbb{R}^n, +)$ is a commutative Lie group of dimension n.
- K, A, U^+, U^- defined in Section 1.2.1 are one dimensional commutative Lie groups.
- $(\mathrm{GL}(n, \mathbb{R}), \cdot)$ is a noncommutative Lie group of dimension n^2.
- $(\mathrm{SL}(n, \mathbb{R}), \cdot)$ and $(\mathrm{PSL}(n, \mathbb{R}), \cdot)$ are noncommutative Lie groups of dimension $n^2 - 1$.
- $(\mathrm{SO}(m, n), \cdot)$ is a noncommutative Lie group of dimension $\frac{(m+n)(m+n-1)}{2} - 1$.

2.1.4 Invariant distances

The action of G on itself allows us to define distances with additional symmetries. Let $(\cdot, \cdot)_e$ be an inner product on T_eG. For each $g \in G$, define an inner product $(\cdot, \cdot)_g$ on T_gG by

$$(v,w)_g = ((DL_{g^{-1}})_g v, (DL_{g^{-1}})_g w)_e \ , \ \forall\, v,w \in T_gG. \tag{2.2}$$

This defines a smooth metric on G. It is smooth because the expression (2.2) is smooth on g, v and w. Let d be its associate distance. We claim that d is left-invariant, i.e. $d(hg_1, hg_2) = d(g_1, g_2)$ for all $g_1, g_2, h \in G$. In other words, we claim that every L_h is an isometry of (G, d). To see this, fix $g \in G$ and $v, w \in T_gG$. By the definition of $(\cdot,\cdot)_g$ and by the chain rule,

$$\begin{aligned}
((DL_h)_g v, (DL_h)_g w)_{hg} &= \big((DL_{(hg)^{-1}})_{hg}(DL_h)_g v, (DL_{(hg)^{-1}})_{hg}(DL_h)_g w\big)_e \\
&= \big((DL_{g^{-1}})_g v, (DL_{g^{-1}})_g w\big)_e \\
&= (v,w)_g.
\end{aligned}$$

In general, d is not right-invariant. If it is, we say that d is *bi-invariant*. In these notes we will only consider left-invariant distances.

2.1.5 Haar measures

A Borel measure μ on G is called *left-invariant* if $\mu(X) = \mu(gX)$ for all Borel sets $X \subset G$ and all $g \in G$. One such measure always exists, and moreover it is regular in many senses.

Theorem 2.2. *There is a countably additive Borel measure μ on G, unique up to multiplication by scalar, satisfying the conditions below.*

(i) μ is left-invariant.
(ii) $\mu(K) < \infty$ for all $K \subset G$ compact.
(iii) μ is outer regular: for any Borel set $X \subset G$,

$$\mu(X) = \inf\{\mu(A) : A \supset X \text{ is open}\}.$$

(iv) μ is inner regular: for any Borel set $X \subset G$,

$$\mu(X) = \sup\{\mu(K) : K \subset X \text{ is compact}\}.$$

See e.g. [10]. We call μ a *left Haar measure*. By symmetry, there is a unique (up to multiplication by scalar) *right Haar measure*: a countably additive Borel measure which is right-invariant and satisfies (ii)–(iv). For general groups, left and right Haar measures do not coincide. If they do, G is called *unimodular*. Here are some examples of unimodular Lie groups:

- Compact groups, e.g. $\mathbb{S}^1$ and $\mathrm{SO}(2)$.
- Semisimple Lie groups (see Section 2.2.1 for the definition), e.g. $\mathrm{GL}(n,\mathbb{R})$ and $\mathrm{SL}(n,\mathbb{R})$.

In Chapter 3 we will describe the Haar measure of $\mathrm{SL}(2,\mathbb{R})$, and we will show that $\mathrm{SL}(2,\mathbb{R})$ is unimodular.

In these notes we will only consider left Haar measures. Here are some examples:

- The Haar measure of $\mathbb{R}^n$ is the Lebesgue measure.
- Let G, H be groups, and let $\varphi : G \to H$ be a isomorphism. Then $\varphi_* \mu_G = \mu_H$.
- Let $S = AU^+$. In the parameterization

$$S = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} : a > 0, b \in \mathbb{R} \right\} \tag{2.3}$$

  we have $d\mu_S = \frac{1}{a} da db$, as we'll prove in Section 3.1.5.

2.1.6 Quotient spaces and lattices

Let G be a Lie group, and let Γ be a subgroup of G. Let $\sim$ be the equivalence relation defined by: $g \sim h$ iff $gh^{-1} \in \Gamma$. An equivalence class of $\sim$ is called a *right coset*, and denoted by Γg. The space of cosets is called the *quotient space*, and denoted by $\Gamma \backslash G$. $\Gamma \backslash G$ is a group iff Γ is a normal subgroup.

G acts on $\Gamma \backslash G$ on the right: if $x = \Gamma g \in \Gamma \backslash G$ and $h \in G$, then $xh = \Gamma(gh)$. Such action is well defined: if $g_1 \sim g_2$, then $g_1 h \sim g_2 h$.

Definition 2.3. *Γ is called discrete if it is discrete in the topology of G, i.e. if for every $g \in \Gamma$ there is a ball $B_\varepsilon(g) \subset G$ such that $B_\varepsilon(g) \cap \Gamma = \{g\}$.*

By the invariance of d, Γ is discrete iff there is a universal $\varepsilon > 0$ such that $B_\varepsilon(g) \cap \Gamma = \{g\}$ for every $g \in G$.

If Γ is discrete, then $\Gamma \backslash G$ is a manifold of same dimension as G. Furthermore, it is a Riemannian manifold, because we can consider a left-invariant metric d on G and induce it to a metric d_Γ on $\Gamma \backslash G$ by

$$d_\Gamma(\Gamma g, \Gamma h) = \inf\{d(g, \gamma h) : \gamma \in \Gamma\}.$$

Note that

- d_Γ is symmetric: this follows from the left-invariance of d.
- d_Γ is non-degenerate: let $\varepsilon > 0$ such that $B_\varepsilon(g) \cap \Gamma = \{g\}$ for all $g \in G$. If $d_\Gamma(\Gamma g, \Gamma h) = 0$, then there are $\gamma_n \in \Gamma$ such that $d(g, \gamma_n h) \to 0$. In particular, there is $r > 0$ such that $\gamma_n h \in B_r(g)$ for all n. Because the right translation $R_{h^{-1}}$ is smooth, there is $C > 0$ such that $R_{h^{-1}}|_{B_r(g)}$ is C-Lipschitz. Thus

$$d(\gamma_n, \gamma_m) = d(R_{h^{-1}}(\gamma_n h), R_{h^{-1}}(\gamma_m h)) \le C d(\gamma_n h, \gamma_m h) < \varepsilon$$

  if m, n are large enough. This contradicts the discreteness of Γ.
- d_Γ satisfies the triangular inequality: let $g, h, k \in G$. If $\gamma_1, \gamma_2 \in \Gamma$, then

$$d(g, \gamma_1\gamma_2 k) \le d(g, \gamma_1 h) + d(\gamma_1 h, \gamma_1\gamma_2 k) = d(g, \gamma_1 h) + d(h, \gamma_2 k).$$

  Running over γ_1, γ_2 gives that $d_\Gamma(\Gamma g, \Gamma k) \le d_\Gamma(\Gamma g, \Gamma h) + d_\Gamma(\Gamma h, \Gamma k)$.

Locally, $\Gamma\backslash G$ looks like G. Let $\pi : G \to \Gamma\backslash G$ be the projection. Each $g \in G$ has an *injectivity radius*, i.e. a positive real number $r = r(g)$ such that the restriction $\pi|_{B_r(g)}$ is a homeomorphism onto its image. Because π is continuous, we just need to guarantee injectivity. Choose r such that $d(hk^{-1}, e) < \varepsilon$ for all $h, k \in B_r(g)$. The choice is possible because the map $(h,k) \in G \times G \mapsto hk^{-1} \in G$ is continuous and sends (g,g) to e. Thus $\pi|_{B_r(g)}$ is injective: if $h, k \in B_r(g)$ with $\Gamma h = \Gamma k$, then $\gamma = hk^{-1} \in \Gamma$ and $d(\gamma, e) < \varepsilon$, i.e. $h = k$. When $\Gamma\backslash G$ is compact, there is a universal $r = r(\Gamma) > 0$ such that $\pi|_{B_r(g)}$ is injective for every $g \in G$.

$\Gamma\backslash G$ also has volume measures. For that we need the notion of fundamental domain. A *fundamental domain of* Γ is a Borel set $F \subset G$ such that

$$G = \bigsqcup_{\gamma \in \Gamma} \gamma F.$$

Fundamental domains always exist (see e.g. [12]). Let us give two examples.

(i) $\mathbb{Z}^n$ is a discrete subgroup of $\mathbb{R}^n$, and $F = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : 0 \leq x_1, \ldots, x_n < 1\}$ is a fundamental domain of $\mathbb{Z}^n$.

(ii) $\mathrm{PSL}(2,\mathbb{Z})$ is a discrete subgroup of $\mathrm{PSL}(2,\mathbb{R})$. Figure 2.1 depicts a fundamental domain of the action of $\mathrm{PSL}(2,\mathbb{Z})$ on $\mathbb{H}$: it is the region enclosed by the vertical lines $x = 0.5$ and $x = -0.5$ and the circle $C = \{|z| = 1\}$. Via the isomorphism $\mathrm{PSL}(2,\mathbb{R}) \cong T^1\mathbb{H}$, the unit tangent bundle of this region "represents"[1] a fundamental domain of $\mathrm{PSL}(2,\mathbb{Z})$.

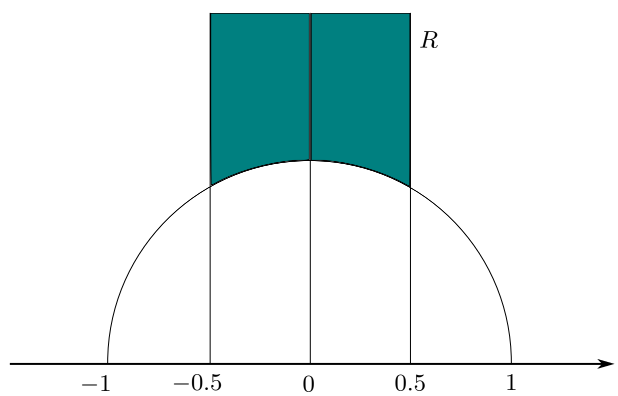

Fig. 2.1. The unit tangent bundle of the region R enclosed by the vertical lines $x = 0.5$ and $x = -0.5$ and the circle $\{|z| = 1\}$ "is" a fundamental domain of $\mathrm{PSL}(2,\mathbb{Z})$.

[1] We put quotes because there are some technicalities regarding the three points of intersection of C with the vertical lines $x = -0.5$, $x = 0$ and $x = 0.5$: on these points the fundamental domain folds into itself.

We will now show that fundamental domains are universal in the sense of the Haar measure μ.

Lemma 2.4. *Let $B_1, B_2 \subset G$ be Borel sets such that $\pi|_{B_1}$, $\pi|_{B_2}$ are isomorphisms onto their images, and $\pi(B_1) = \pi(B_2)$. Then $\mu(B_1) = \mu(B_2)$.*

Proof. We claim that

$$B_1 = \bigsqcup_{\gamma \in \Gamma} (B_1 \cap \gamma B_2) \quad \text{and} \quad B_2 = \bigsqcup_{\gamma \in \Gamma} (\gamma B_1 \cap B_2). \tag{2.4}$$

Let us prove the first equality (the second is analogous). If $g \in B_1$, then $\pi(g) \in \pi(B_1) = \pi(B_2)$, thus there is $\gamma \in \Gamma$ such that $\gamma g \in B_2$. Because $\pi|_{B_2}$ is injective, γ is unique. Equality (2.4) and the left-invariance of μ give that

$$\mu(B_1) = \sum_{\gamma \in \Gamma} \mu(B_1 \cap \gamma B_2) = \sum_{\gamma \in \Gamma} \mu(\gamma^{-1} B_1 \cap B_2) = \sum_{\gamma \in \Gamma} \mu(\gamma B_1 \cap B_2) = \mu(B_2).$$

Let F be a fundamental domain of Γ. Define a measure $\mu_{\Gamma \backslash G}$ on $\Gamma \backslash G$ by $\mu_{\Gamma \backslash G}(A) = \mu(F \cap \pi^{-1}(A))$. By Lemma 2.4, $\mu_{\Gamma \backslash G}$ does not depend on the choice of F. In general, $\mu_{\Gamma \backslash G}$ is not finite nor G-invariant.

Definition 2.5. *Γ is called a lattice if $\mu_{\Gamma \backslash G}$ is finite. If $\Gamma \backslash G$ is also compact, we call Γ a uniform lattice or a cocompact lattice.*

Only unimodular groups have lattices, thus whenever we refer to a lattice Γ, it will be implicit that G is unimodular. Quotients by lattices are special because they carry G-invariant measures (remember that G acts on $\Gamma \backslash G$ on the right).

Proposition 2.6. *Let G be a Lie group, and let Γ be a discrete subgroup of G. The following are equivalent:*

(i) Γ is a lattice.
(ii) Γ has a fundamental domain F such that $\mu(F) < +\infty$.
(iii) $\Gamma \backslash G$ supports a G-invariant probability measure.

In this case, the G-invariant probability measure is unique and equal to a multiple of $\mu_{\Gamma \backslash G}$.

From now on, we normalize $\mu_{\Gamma \backslash G}$ to become a probability measure. Thus $\mu_{\Gamma \backslash G}$ is the unique G-invariant probability measure on $\Gamma \backslash G$. We call $\mu_{\Gamma \backslash G}$ the *Haar measure of* $\Gamma \backslash G$.

If $G = \mathbb{R}^n$ then the notion of lattice equals to the geometrical notion of lattices. For example, $\mathbb{Z}^n$ is a lattice of $\mathbb{R}^n$. It is even cocompact: $\mathbb{T}^n = \mathbb{Z}^n \backslash \mathbb{R}^n$ is compact.

Here is another example, not coming from euclidean spaces: $\mathrm{PSL}(2, \mathbb{Z})$ is a lattice of $\mathrm{PSL}(2, \mathbb{R})$. To prove that it is a lattice, we calculate the Haar measure of the fundamental domain described in Figure 2.1 (see Section 3.1.5 for the proper definition of the Haar measure of $\mathrm{PSL}(2,\mathbb{R})$):

$$\mu(T^1 R) = \int_{-\pi}^{\pi}\int_R \frac{1}{y^2} dxdyd\theta < 2\pi \int_{-0.5}^{0.5}\int_{0.5}^{\infty} \frac{1}{y^2} dxdy < +\infty.$$

2.1.7 Homogeneous spaces as quotient spaces

If Γ is a discrete subgroup of G, then $\Gamma\backslash G$ is a homogeneous space in the sense of Definition 2.1: if $x = \Gamma g, y = \Gamma h \in \Gamma\backslash G$, then $x(g^{-1}h) = y$.

Reciprocally, homogeneous spaces are homeomorphic to quotient spaces, perhaps by a non discrete subgroup. To see this, let X be a homogeneous space, and let $\mathrm{Stab}(x) = \{g \in G : xg = x\}$ denote the *stabilizer of* x. $\mathrm{Stab}(x)$ is a subgroup of G: if $g, h \in \mathrm{Stab}(x)$, then $x(gh) = (xg)h = xh = x$. Fix some $L = \mathrm{Stab}(x)$. Then the map $Lg \in L\backslash G \to xg \in X$ is an homeomorphism.

Assume that L is discrete. Then X carries a G-invariant probability measure μ_X iff L is a lattice of G. In this case, μ_X is unique and equal to the push-forward of the Haar measure $\mu_{\Gamma\backslash G}$ under the homeomorphism described above.

From now on, we restrict the term *homogeneous space* to refer to a quotient $X = \Gamma\backslash G$ such that Γ is a lattice. We call the unique G-invariant probability measure μ_X the *Haar measure of* X.

2.2 Unipotent actions

2.2.1 Lie algebras

A *Lie algebra* is a real vector space $\mathfrak{b}$ together with a bilinear map $[\cdot,\cdot] : \mathfrak{b} \times \mathfrak{b} \to \mathfrak{b}$ satisfying:

(i) antisymmetry: $[X, Y] = -[Y, X]$ for all $X, Y \in \mathfrak{b}$, and
(ii) Jacobi identity: $[X, [Y, Z]]+[Y, [Z, X]]+[Z, [X, Y]] = 0$ for all $X, Y, Z \in \mathfrak{b}$.

Given $X, Y \in \mathfrak{b}$, the vector $[X, Y] \in \mathfrak{b}$ is called the *bracket* of X and Y. The most important example of a Lie algebra is defined on the tangent space T_eG of a Lie group G, as we'll now see.

Let $\mathfrak{b}$ be the set of vector fields[2] on G. $\mathfrak{b}$ is an infinite dimensional real vector space. The usual bracket $[X, Y] = XY - YX$ defines a Lie algebra $(\mathfrak{b}, [\cdot,\cdot])$. We say that $X \in \mathfrak{b}$ is *left-invariant* if $L_g X = X$ for all $g \in G$. Let $\mathfrak{g} \subset \mathfrak{b}$ be the subspace of left-invariant vector fields. $\mathfrak{g}$ is closed under brackets: if $X, Y \in \mathfrak{g}$, then $L_g[X, Y] = [L_g X, L_g Y] = [X, Y]$.

[2] A vector field on G is a smooth map $X : G \to TG$ such that $X(g) \in T_gG$ for all $g \in G$. If $\phi : G \to G$ is a diffeomorphism, we denote by ϕX the vector field $(\phi X)(g) = D\phi_{\phi^{-1}(g)}(X(\phi^{-1}(g)))$.

Definition 2.7. $(\mathfrak{g}, [\cdot,\cdot])$ *is called the Lie algebra of* G.

Here is an infinitesimal interpretation of $(\mathfrak{g}, [\cdot,\cdot])$. As a vector space, $\mathfrak{g}$ is isomorphic to T_eG via the evaluation map $X \in \mathfrak{g} \mapsto X(e) \in T_eG$. Thus we can induce the bracket on T_eG: given $v, w \in T_eG$, define $[v,w] := [X,Y](e)$, where X and Y are the left-invariant vector fields with $X(e) = v$ and $Y(e) = w$.

Below we list some examples of Lie algebras, and refer to chapter 15 of [14] for details.

(i) $\mathbb{R}^n$: $\mathfrak{g} \cong \mathbb{R}^n$ and $[X,Y] = 0$.
(ii) $\mathbb{S}^1$: $\mathfrak{g} \cong \mathbb{R}$ and $[X,Y] = 0$.
(iii) $\mathrm{GL}(n,\mathbb{R})$: for each $X \in \mathrm{M}(n,\mathbb{R})$,

$$\alpha(t) = \exp(tX) = I + tX + \frac{t^2X^2}{2} + \frac{t^3X^3}{6} + \cdots \tag{2.5}$$

defines a curve in $\mathrm{GL}(n,\mathbb{R})$ passing through I with velocity X. Thus $\mathfrak{g} \cong \mathrm{M}(n,\mathbb{R})$. The bracket is the commutator of matrices: $[X,Y] = XY - YX$. We denote $(\mathfrak{g}, [\cdot,\cdot])$ by $\mathfrak{gl}(n,\mathbb{R})$.
(iv) $\mathrm{SL}(n,\mathbb{R})$: $T_I\mathrm{SL}(n,\mathbb{R})$ is isomorphic to the space of trace zero matrices of $\mathrm{GL}(n,\mathbb{R})$. The bracket is again the commutator of matrices. We denote $(\mathfrak{g}, [\cdot,\cdot])$ by $\mathfrak{sl}(n,\mathbb{R})$.

The Lie algebra gives information on the local algebraic structure of the Lie group (one cannot expect to get an isomorphism between Lie groups and their Lie algebras, because the former is an infinitesimal object). Here is an example of its role on this relation.

Theorem 2.8 (Fundamental theorem of Sophus Lie). *Two Lie groups are locally isomorphic (there is a local smooth isomorphism) iff their Lie algebras are isomorphic.*

Reference [14] contains this and other results on Lie groups.

2.2.2 Adjoint transformations and unipotent elements

Let $h \in G$, and let $C : g \in G \to hgh^{-1} \in G$ be the conjugation by h. C fixes the identity e, thus DC_e is a linear isomorphism of $\mathfrak{g}$. We call it an *adjoint transformation of* G, and denote it by Ad_g.

To exemplify, let's calculate the adjoint for linear groups, e.g. $\mathrm{GL}(n,\mathbb{R})$. Fix $g \in \mathrm{GL}(n,\mathbb{R})$ and $X \in \mathfrak{gl}(n,\mathbb{R})$, and let α as in (2.5). The conjugacy by g sends α to $g\exp(tX)g^{-1}$, thus $\mathrm{Ad}_g X = gXg^{-1}$ is again a conjugacy.

In linear algebra, a linear transformation A is *unipotent* if all of its eigenvalues are equal to 1, or equivalently if $A = I + N$ where N is nilpotent. The Lie algebra allows us to make a similar statement for elements of Lie groups.

Definition 2.9. *An element* $g \in G$ *is called unipotent if* Ad_g *is a unipotent linear transformation.*

This definition coincides with the classical one for linear groups. To see this, let $g = I + N \in \mathrm{GL}(n, \mathbb{R})$, where $N^{k+1} = 0$. Thus

$$\mathrm{Ad}_g X = (I + N)X(I - N + \cdots + (-1)^k N^k) = X + f(X),$$

where $f(X) = \sum_{i+j>0} a_{ij} N^i X N^j$. Iterating $2k$ times, we get that $f^{2k}(X) = \sum_{i+j>2k} b_{ij} N^i X N^j = 0$, because either $N^i = 0$ or $N^j = 0$. Thus $\mathrm{Ad}_g = I + f$, where f is nilpotent.

2.3 Ratner's theorems

We are finally able to state Ratner's theorems. The setup is:

- G is a connected Lie group.
- Γ is a lattice of G.
- X is the homogeneous space $\Gamma\backslash G$.
- U is a closed connected subgroup of G generated by unipotent elements.

Ratner's theorems are rigidity theorems about the U-action on X, at three different levels: orbit closure rigidity, measure rigidity and equidistribution. For the historical perspective that led to Ratner's theorems, we refer to [21].

2.3.1 Ratner's orbit closure theorem

Also known as *Raghunathan's conjecture*, the first theorem classifies orbit closures of U-actions. It was first proved by Ratner in [24].

Given $x \in X$, let $xU = \{xu : u \in U\}$ denote the U-orbit of x. In some cases xU is very simple, e.g. if $U = \{u_t : t \in \mathbb{R}\}$ is a one-parameter subgroup and $xu_t = x$ for some t, then xU is closed and isomorphic to $\mathbb{S}^1$.

In general, xU is not closed, e.g. the horocycle flow of hyperbolic surfaces has dense orbits. Ratner's orbit closure theorem says that, although xU may be complicated, its closure $\overline{xU}$ is nice: it is the L-orbit for some closed subgroup L containing U.

Definition 2.10. *A subset $A \subset X$ is called U-homogeneous if there is a closed subgroup $L \supset U$ and a point $x \in X$ such that $A = xL$ is a homogeneous L-space.*

Some remarks concerning the above definition:

- The L-action we consider in xL is the right action: $(xl_1)l_2 = x(l_1 l_2)$.
- This L-action is transitive: if $xl_1, xl_2 \in xL$, then $(xl_1)(l_1^{-1} l_2) = xl_2$.
- If $x = \Gamma g$, then xL is a homogeneous L-space iff $L \cap g^{-1}\Gamma g$ is a lattice of L: the map $l \in L \mapsto xl \in xL$ induces a homeomorphism between $\mathrm{Stab}(x)\backslash L$ and xL. Note that $\mathrm{Stab}(x) = L \cap g^{-1}\Gamma g$, because

$$xl = x \iff \Gamma gl = \Gamma g \iff glg^{-1} \in \Gamma \iff l \in L \cap g^{-1}\Gamma g.$$

By Proposition 2.6, xL is a homogeneous L-space iff $L \cap g^{-1}\Gamma g$ is a lattice of L.

Ratner's orbit closure theorem states that orbit closures of unipotent subgroups are homogeneous.

Theorem 2.11 (Ratner's orbit closure theorem). *Let $X = \Gamma\backslash G$ be a homogeneous space, where G is a connected Lie group, and let U be a closed connected subgroup of G generated by unipotent elements. If $x \in X$, then $\overline{xU}$ is U-homogeneous.*

This conclusion is far from being true if we drop the unipotent assumption, e.g. some orbit closures of geodesic flows are horseshoes.

Note that if $\overline{xU} = xL$, then L acts on $\overline{xU}$. Unless $L = U$, there is no obvious reason for this to be true. This means that the closure of xU is much richer and nicer that xU itself.

2.3.2 Ratner's measure classification theorem

The second theorem classifies what are the invariant Borel probability measures for U-actions. By the ergodic decomposition, any invariant probability measure is the convex combination of ergodic invariant probability measures. Thus it is enough to classify the ergodic invariant Borel probability measures. Dani conjectured that all such measures are algebraic in the sense to be defined below. Let $X = \Gamma\backslash G$ be a homogeneous space, and let μ be an ergodic U-invariant Borel probability measure on X.

Definition 2.12. *μ is called algebraic if there is a closed subgroup L of G and a point $x \in X$ such that xL is homogeneous and $\mu = \mu_{xL}$.*

In the series of works [22–24] Ratner established Dani's conjecture.

Theorem 2.13 (Ratner's measure classification theorem). *Let $X = \Gamma\backslash G$ be a homogeneous space, where G is a connected Lie group, and let U be a closed connected subgroup of G generated by unipotent elements. Then every ergodic U-invariant Borel probability measure on X is algebraic.*

2.3.3 Ratner's equidistribution theorem

The notion of equidistribution requires to consider averages along Fϕlner sequences. Thus we need to restrict ourselves to amenable subgroups. Let U be a *one-parameter unipotent subgroup*, i.e. $U = \{u_t : t \in \mathbb{R}\}$ where u_t is unipotent for all $t \in \mathbb{R}$. Given a bounded continuous functions $f : X \to \mathbb{R}$, we consider averages of the form $\frac{1}{T}\int_0^T f(xu_t)dt$, where $T > 0$.

Definition 2.14. *A point $x \in X$ is called U-generic if there is a closed subgroup $L \supset U$ such that xL is homogeneous and*

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T f(xu_t)dt = \int_X f d\mu_{xL}$$

for every bounded continuous $f : X \to \mathbb{R}$.

Theorem 2.15 (Ratner's equidistribution theorem). *Let $X = \Gamma \backslash G$ be a homogeneous space, where G is a connected Lie group, and let U be a one-parameter unipotent subgroup of G. Then every $x \in X$ is U-generic.*

In particular, the above theorem implies that every $x \in X$ is recurrent: there is a sequence $t_i \uparrow +\infty$ such that $xu_{t_i} \to x$. We would like to contrast this conclusion with Poincaré's recurrence theorem: while Poincaré's recurrence theorem implies that almost every $x \in X$ is recurrent, Ratner's equidistribution theorem asserts that **every** $x \in X$ is recurrent.

2.3.4 A simple example: toral translations

Here we discuss an application of Ratner's theorems for translations on tori. Toral translations are a classical object in ergodic theory, thus we use this example to clarify Ratner's theorems. The conclusions below were know even before Ratner's work, by means of tools such as harmonic analysis.

Let $\mathbb{T}^n = \mathbb{Z}^n \backslash \mathbb{R}^n$ be the n dimensional torus. Given $x \in \mathbb{R}^n$, denote $x \pmod{\mathbb{Z}^n}$ by $[x]$. Every $u \in \mathbb{R}^n$ is unipotent, because $\mathrm{Ad}_u(v) = u + v - u = v$. The flow generated by u is the translation on $\mathbb{T}^n$ by u, i.e. $[x]u_t = [x + tu]$. By Ratner's orbit closure theorem, the closure of $[x + \mathbb{R}u]$ is homogeneous: there is a closed subgroup L containing $\{tu : t \in \mathbb{R}\}$ such that $\overline{[x + \mathbb{R}u]} = [x + L]$ and $L \cap (x + \mathbb{Z}^n - x) = L \cap \mathbb{Z}^n$ is a lattice of L. Indeed, L is a subspace of $\mathbb{R}^n$ (exercise).

For example, let $n = 2$ and $u = (u_1, u_2)$.

- If u_1, u_2 are rationally dependent, then $[x + \mathbb{R}u]$ is periodic, i.e. L is a line with rational slope.
- If u_1, u_2 are rationally independent, then $[x + \mathbb{R}u]$ is dense in $\mathbb{T}^2$, i.e. $L = \mathbb{R}^2$. By Ratner's equidistribution theorem, every such x is generic wrt $\mu_{\mathbb{T}^2}$.

3

Horocycle flows of hyperbolic surfaces

Let $X = \Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ be a homogeneous space. In the sequel, we will fix the stable horocycle flow h_t^+. We will denote it simply by h_t and the subgroup U^+ by N. Firstly, we remind the description of h_t in $\mathrm{SL}(2,\mathbb{R})$-coordinates, given by equation (1.5):

$$\begin{aligned} h_t : X &\to X \\ x &\mapsto x \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{3.1}$$

The matrix $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ is unipotent. By Section 2.2.2, N is a one-parameter unipotent subgroup, thus h_t is an action on a homogeneous space by a one-parameter unipotent subgroup, which is the setup of Ratner's theorems. The goal of this chapter is to discuss some of the historical developments of orbit closure, measure rigidity and equidistribution for h_t that led to Ratner's theorems. They are:

- Hedlund: if X is compact, then h_t is minimal and μ_X is ergodic [11].
- Furstenberg: if X is compact, then h_t is uniquely ergodic [9].
- Dani-Smillie: if X is not compact, then the h_t-invariant probability measures are either μ_X or are supported on periodic orbits [6].

The idea is to discuss these results without the aid of Ratner's theorems. We will give a self-contained proof of Furstenberg's theorem, and also deduce it from Hedlund and Ratner's measure classification theorems.

Furstenberg explores the relation between g_t and h_t by means of two facts:

(i) g_t renormalizes h_t: this is the matrix identity (1.6). The orbits of h_t are expanded/contracted by g_t. This idea goes back to Hedlund [11].
(ii) (g_t, μ_X) is mixing.

We will give a self-contained proof of (ii) in Section 3.4. This is Howe-Moore's theorem.

3.1 More on the group $\mathrm{SL}(2,\mathbb{R})$

Firstly, we complement the discussion of Section 1.2.1 and collect more information about $\mathrm{SL}(2,\mathbb{R})$.

3.1.1 *KAN* decomposition

The subgroups K, A, N described in equation (1.2) give a coordinate system for $\mathrm{SL}(2,\mathbb{R})$, called the *KAN decomposition.*

Lemma 3.1. *$(k,a,n) \in K \times A \times N \mapsto kan \in \mathrm{SL}(2,\mathbb{R})$ is a diffeomorphism.*

Proof. The above map is differentiable, because multiplication on $\mathrm{SL}(2,\mathbb{R})$ is. Let $\{e_1, e_2\}$ be the canonical basis of $\mathbb{R}^2$. We consider the linear action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{R}^2$ (see Section 1.2.2) and identify $g \in \mathrm{SL}(2,\mathbb{R})$ with the positively oriented basis $\{ge_1, ge_2\}$ it defines. The parallelogram with sides ge_1, ge_2 has unit area. We prove surjectivity as follows: starting with an arbitrary such basis, we find elements $k \in K$, $a \in A$, $n \in N$ such that $n^{-1}a^{-1}k^{-1}$ sends $\{ge_1, ge_2\}$ back to $\{e_1, e_2\}$. Thus $n^{-1}a^{-1}k^{-1}g = e$, i.e. $g = kan$.

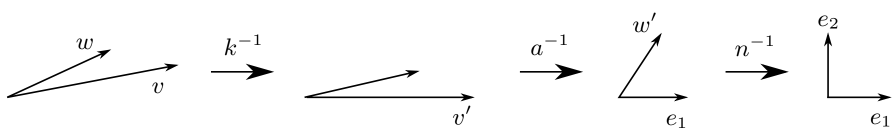

Fig. 3.1. The *KAN* decomposition of $\mathrm{SL}(2,\mathbb{R})$.

Let $\{ge_1, ge_2\} = \{v, w\}$. Choose $k \in K$ such that $k^{-1}v = v'$ is horizontal and has positive x-coordinate. Choose $a \in A$ such that $a^{-1}v' = e_1$. Thus $a^{-1}k^{-1}$ sends $\{v, w\}$ to $\{e_1, w'\}$. Because the parallelogram with sides e_1, w' has unit area, $w' = (x, 1)$ for some $x \in \mathbb{R}$. Choose $n \in N$ such that $n^{-1}w' = e_2$. Because n^{-1} fixes e_1, $n^{-1}a^{-1}k^{-1}$ sends $\{v, w\}$ to $\{e_1, e_2\}$.

Now we prove injectivity: assume that $k_1a_1n_1 = k_2a_2n_2$, i.e. $k_2^{-1}k_1 = a_2n_2n_1^{-1}a_1^{-1}$. Both A, N preserve the x-axis and its orientation. The only element of K that does this is e, thus $k_1 = k_2$, i.e. $a_1^{-1}a_2 = n_1n_2^{-1}$. Because $A \cap N = \{e\}$, it follows that $a_1 = a_2$ and $n_1 = n_2$.

3.1.2 Cartan decomposition

Let $A^+ = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} : a \geq 1 \right\}$. The *Cartan decomposition* of $\mathrm{SL}(2,\mathbb{R})$ is described in the lemma below.

Lemma 3.2. *$(k,a,l) \in K \times A^+ \times K \mapsto kal \in \mathrm{SL}(2,\mathbb{R})$ is a diffeomorphism.*

Proof. We consider the hyperbolic action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H}$ (see Section 1.2.3). Below, distances and balls are wrt the hyperbolic metric. We prove surjectivity as follows: given $v \in T_z^1\mathbb{H}$, there is $k, l \in K$ and $a \in A^+$ such that $D(kal)v = \mathbf{i}$. Let's first describe how K and A^+ act on $\mathbb{H}$. Let B denote the circle with center i and radius $d = d(z,i)$. B intersects the y-axis in two points, one below i and one above i. Let w be the below intersection. Choose $l \in K$ such that $l.z = w$, and choose $a \in A^+$ such that $a.w = i$. Thus $(al).z = w$ and $u = D(al)v \in T_i^1\mathbb{H}$. Choose $k \in K$ such that $D(k)u = \mathbf{i}$. By the chain rule, $D(kal)v = D(k)u = \mathbf{i}$.

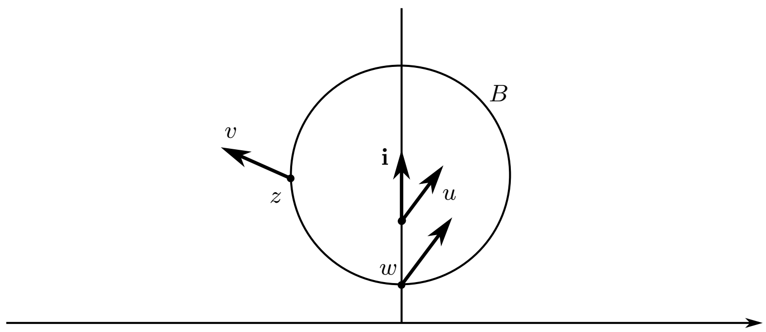

Fig. 3.2. The Cartan decomposition of $\mathrm{SL}(2,\mathbb{R})$.

To get injectivity, note that l is uniquely defined by the distance between z and w, and $a \in A^+$ is uniquely defined by the distance between w and i.

3.1.3 $\mathrm{SL}(2,\mathbb{R})$ is generated by unipotent elements

Remember U^+, U^- defined in equation (1.2).

Lemma 3.3. $\mathrm{SL}(2,\mathbb{R})$ *is generated by* U^+ *and* U^-.

Proof. We again consider the linear action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{R}^2$ (see Section 1.2.2). Let $\{v, w\} = \{ge_1, ge_2\}$. We have two cases:

- v does not belong to the x-axis: choose $u_3 \in U^+$ such that $u_3^{-1}v$ belongs to the line $x = 1$, and choose $u_2 \in U^-$ such that $u_2^{-1}u_3^{-1}v = e_1$. Because $w' = u_2^{-1}u_3^{-1}w$ belongs to the line $y = 1$, there is $u_1 \in U^+$ such that $u_1^{-1}w' = e_2$.
- v belongs to the x-axis: choose any $u \in U^-$ such that uv does not belong to the x-axis and apply the previous case to the pair $\{uv, uw\}$.

3.1.4 Decomposition of Haar measures

The decompositions of the previous sections allow us to describe $\mu_{\mathrm{SL}(2,\mathbb{R})}$ as a product of Haar measures of subgroups. We follow a more general setup, borrowed from part VIII of [13].

Let G be a unimodular Lie group, and let S, T be closed subgroups of G. Assume that $S \cap T = \{e\}$ and that ST contains a neighborhood V of some $g \in G$. The map $\varphi : (s,t) \in S \times T \mapsto st \in G$ is injective. It gives a coordinate system for its image $\mathrm{Im}(\varphi) = \varphi(S \times T)$, thus g has local charts in terms of S and T. Let $\nu = \mu_S \times \mu_T$. The lemma below says that the restriction of μ_G to V is, in the coordinate system φ, equal to ν.

Lemma 3.4. *After a normalization of ν, $\varphi_*^{-1}(\mu_G|_V) = \nu|_{\varphi^{-1}(V)}$.*

Haar measures of homogeneous spaces $\Gamma\backslash G$ can also be locally expressed as a product, because the projection $G \mapsto \Gamma\backslash G$ is a local diffeomorphism.

3.1.5 Haar measure of $\mathbf{SL(2,\mathbb{R})}$

Let $S = AN$, as in equation (2.3). By Lemma 3.1, the map $(a,n) \in A \times N \mapsto an \in S$ is a diffeomorphism. A, N and S are closed subgroups, thus by Lemma 3.4 we have $\mu_S = \mu_A \times \mu_N$. Because S and K are closed, we can repeat the argument above to get that $\mu_{\mathrm{SL}(2,\mathbb{R})} = \mu_K \times \mu_A \times \mu_N$.

Below we give explicit descriptions of $\mu_K, \mu_A, \mu_N, \mu_S, \mu_{\mathrm{SL}(2,\mathbb{R})}$ in terms of the parameterizations (1.2) and (2.3).

- The map $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in K \mapsto \theta \in \mathbb{S}^1$ is an isomorphism, thus $d\mu_K = d\theta$.
- The map $\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \in A \mapsto \log a \in \mathbb{R}$ is an isomorphism, thus $d\mu_A = \frac{1}{a} da$.
- The map $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \in N \mapsto b \in \mathbb{R}$ is an isomorphism, thus $d\mu_N = db$.
- $d\mu_S = d\mu_A d\mu_N = \frac{1}{a} da db$.
- $d\mu_{\mathrm{SL}(2,\mathbb{R})} = d\mu_K d\mu_A d\mu_N = \frac{1}{a} da db d\theta$.

3.2 Hedlund's theorem

Let $X = \Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ be a homogeneous space. Exploring the relation between h_t and g_t, Hedlund proved the following.

Theorem 3.5 (Hedlund [11]**).** *Every orbit of the horocycle flow is either periodic or dense in X, and μ_X is ergodic. In particular, if X is compact, then the horocycle flow is minimal.*

Note that the second conclusion indeed follows from the first one: because a horocycle orbit is a stable manifold of the geodesic flow, it can not be simultaneously closed and contained in a compact set.

Theorem 3.5 goes in the opposite direction of the geodesic flow, that has many periodic orbits (and thus is far from being minimal). In the next section we will prove part of Hedlund's theorem, namely that μ_X is ergodic for the horocycle flow. The proof uses the so called *Mautner phenomenon*.

3.3 Mautner phenomenon

We first need to discuss a little about representation theory.

3.3.1 Representations

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle\cdot,\cdot\rangle$, let $\mathrm{GL}(\mathcal{H})$ be the set of linear automorphisms of $\mathcal{H}$, and let $\mathrm{U}(\mathcal{H})$ be the set of unitary linear automorphisms of $\mathcal{H}$. I.e. $\varphi \in \mathrm{U}(\mathcal{H})$ iff $\varphi \in \mathrm{GL}(\mathcal{H})$ and $\langle\varphi(v),\varphi(w)\rangle = \langle v,w\rangle$ for all $v,w \in \mathcal{H}$. Both $\mathrm{GL}(\mathcal{H})$ and $\mathrm{U}(\mathcal{H})$ are groups with the composition operation.

Definition 3.6. *Let G be a group. A representation of G is a group homomorphism $\pi : G \to \mathrm{GL}(\mathcal{H})$. A unitary representation of G is a group homomorphism $\pi : G \to \mathrm{U}(\mathcal{H})$. π is called strongly continuos if the following holds: if $g_n \to g$ in G, then $\pi(g_n)v \to \pi(g)v$ in the strong topology of $\mathcal{H}$, for all $v \in \mathcal{H}$.*

In other words, for each $g \in G$ there is a linear automorphism $\pi(g) : \mathcal{H} \to \mathcal{H}$ and these maps satisfy $\pi(gh) = \pi(g)\pi(h)$. Given $v \in \mathcal{H}$, $\pi(g)v \in \mathcal{H}$ denotes the evaluation of the map $\pi(g)$ in the vector v. We already saw an example of a representation in Section 2.2.2. Let G be a Lie group, and let $\mathcal{H} = \mathfrak{g}$ be its Lie algebra. Define $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ by $\mathrm{Ad}(g) = \mathrm{Ad}_g$. Ad is called the *adjoint representation*.

There is another class of representations, coming from measure-preserving actions. It is called the *Koopman-von Neumann representation*. Consider a measure-preserving G-action on a measure space (X,μ). Let

$$\mathcal{H} = L^2_0(X,\mu) = \left\{ f \in L^2(X,\mu) : \int_X f d\mu = 0 \right\}$$

and let $\pi : G \to \mathrm{U}(L^2_0(X,\mu))$ be defined by $\pi(g)f = f \circ g^{-1}$, i.e. $\pi(g)f \in L^2_0(X,\mu)$ is the function defined by $[\pi(g)f](x) = f(xg^{-1})$. π is unitary, by the invariance of μ. If the G-action is continuous, then π is strongly continuous.

Koopman-von Neumann representations are important because they characterize many ergodic properties of the G-action. Remember that G acts on (X,μ) ergodically if every subset $A \subset X$ such that $Ag = A$ for all $g \in G$ is measure-theoretically trivial, i.e. $\mu(A) = 0$ or $\mu(X - A) = 0$. In the spectral perspective, call a function $f \in L^2_0(X,\mu)$ *invariant* if $\pi(g)f = f$ for all $g \in G$. Thus G acts ergodically iff 0 is the only invariant function.

Now assume that $G = \mathbb{Z}$. Mixing can also be translated to a spectral property of the Koopman-von Neumann representation. If 1 is a generator of $\mathbb{Z}$, then the $\mathbb{Z}$-action is mixing iff $\lim_{n\to+\infty}\langle\pi(n)f,g\rangle = 0$ for all $f,g \in L^2_0(X,\mu_X)$. This former expression defines a matrix coefficient of π.

Definition 3.7. *Let $\pi : G \to \mathrm{GL}(\mathcal{H})$ be a representation. Given $v,w \in \mathcal{H}$, the function $g \in G \mapsto f_{v,w}(g) = \langle\pi(g)v,w\rangle$ is called a matrix coefficient of π.*

Using the analogy of actions and their Koopman-von Neumann representations, we define notions of ergodicity and mixing for a representation. Given a representation $\pi : G \to \mathrm{GL}(\mathcal{H})$, we say that $v \in \mathcal{H}$ is invariant if $\pi(g)v = v$ for all $g \in G$.

Definition 3.8. *π is called ergodic if 0 is the only invariant vector. π is called mixing if its matrix coefficients vanish at infinity: for every $v, w \in \mathcal{H}$ we have $\lim_{g\to\infty} f_{v,w}(g) = 0$.*

By $\lim_{g\to\infty} f_{v,w}(g) = 0$ we mean the following: for every $\varepsilon > 0$ there is a compact $K \subset G$ such that $|f_{v,w}(g)| < \varepsilon$ for all $g \notin K$.

3.3.2 Mautner's lemma

Mautner's lemma is a functional analytical tool that provides extra invariance for unitary representations. It is a translation of the property that the geodesic flow renormalizes the horocycle flow.

Lemma 3.9 (Mautner's lemma). *Let $\pi : G \to \mathrm{U}(\mathcal{H})$ be a strongly continuous representation. Let $g_n, h \in G$ such that $g_n^{-1} h g_n \to e$, and let $v \in \mathcal{H}$. If $\pi(g_n)v = v$ for all n, then $\pi(h)v = v$.*

There is also a version of Mautner's lemma for the weak topology, which is more general and serves to our purposes. We denote convergence of v_n to v in the weak topology by $v_n \overset{\mathrm{w}}{\to} v$ or by $\mathrm{w}-\lim v_n = v$.

Lemma 3.10 (Mautner's lemma - weak topology version). *Let $\pi : G \to \mathrm{U}(\mathcal{H})$ be a strongly continuous representation. Let $g_n, h \in G$ such that $g_n^{-1} h g_n \to e$, and let $v, w \in \mathcal{H}$. If $\pi(g_n)v \overset{\mathrm{w}}{\to} w$, then $\pi(h)w = w$.*

Proof. We will prove that $\pi(hg_n)$ converges in the weak topology to both $\pi(h)w$ and w. By the uniqueness of w-limits, the lemma will be proved. Because π is unitary, we have $\pi(hg_n)v \overset{\mathrm{w}}{\to} \pi(h)w$. For the other convergence, note that $\|\pi(hg_n)v - \pi(g_n)v\| = \|\pi(g_n^{-1}hg_n)v - v\|$ converges to 0 in the strong topology, thus $\mathrm{w}-\lim \pi(hg_n)v = \mathrm{w}-\lim \pi(g_n)v = w$.

3.3.3 Mautner phenomenon for $\mathbf{SL(2,\mathbb{R})}$

Mautner phenomenon is a mechanism that gives full invariance of unitary representations of simple Lie groups. We focus on the case of $\mathrm{SL}(2,\mathbb{R})$. Firstly, we prove a lemma in the spirit of Mautner's lemma.

Lemma 3.11. *Let $\pi : G \to \mathrm{U}(\mathcal{H})$ be a strongly continuous representation. Let $g_n, k_n, l_n, h \in G$ such that $k_n \to e$ and $g_n^{-1} k_n l_n \to h$, and let $v \in \mathcal{H}$. If $\pi(g_n)v = \pi(l_n)v = v$ for all n, then $\pi(h)v = v$.*

Proof. $\|\pi(h)v-v\| = \lim_{n\to+\infty} \|\pi(g_n^{-1}k_nl_n)v-v\| = \lim_{n\to+\infty} \|\pi(k_n)v-v\| = 0$.

Theorem 3.12 (Mautner phenomenon for $\mathrm{SL}(2,\mathbb{R})$**).** *Let* $\pi : \mathrm{SL}(2,\mathbb{R}) \to \mathrm{U}(\mathcal{H})$ *be a strongly continuous representation. Let* $g \neq e$ *and* $v \in \mathcal{H}$ *such that* $\pi(g)v = v$. *If either* $g \in A$ *or* $g \in N$, *then* $\pi(h)v = v$ *for all* $h \in G$.

Proof. If $g \in A^+$, then $g^n u g^{-n} \to e$ for $u \in U^-$ and $g^{-n} u g^n \to e$ for $u \in U^+$ (see equation (1.6)). By Mautner's lemma (Lemma 3.9), $\pi(h)v = v$ for all $h \in U^+ \cup U^-$. Because U^+ and U^- generate $\mathrm{SL}(2,\mathbb{R})$, it follows that $\pi(h)v = v$ for all $h \in \mathrm{SL}(2,\mathbb{R})$.

Now assume that $g = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, $s \neq 1$. For every positive integer n it holds

$$g^{-2n} \begin{pmatrix} 1 & 0 \\ -\frac{1}{2ns} & 1 \end{pmatrix} g^n = \begin{pmatrix} 2 & 0 \\ -\frac{1}{2sn} & \frac{1}{2} \end{pmatrix}. \tag{3.2}$$

Because $\begin{pmatrix} 1 & 0 \\ -\frac{1}{2ns} & 1 \end{pmatrix} \to e$ and $\begin{pmatrix} 2 & 0 \\ \frac{1}{2sn} & \frac{1}{2} \end{pmatrix} \to h \in A$, Lemma 3.11 implies that $\pi(h)v = v$. Thus we are reduced to the first case.

Here is a dynamical consequence of Mautner phenomenon.

Theorem 3.13. *Let* $X = \Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ *be a homogeneous space, and let* $g \in A$ *or* $g \in N$. *Then* μ_X *is ergodic for the* $\mathbb{Z}$*-action defined by* g. *In particular, the geodesic and horocycle flows are ergodic.*

In the next section we will show that this approach via unitary representations can be sharpened to give mixing.

3.4 Howe-Moore's theorem

As explained in the beginning of this chapter, to prove that the horocycle flow is uniquely ergodic, Furstenberg used that μ_X is mixing for the geodesic flow, i.e. that the Koopman-von Neumann representation of the geodesic flow is mixing in the sense of Definition 3.8.

Howe-Moore's theorem states if G is a simple Lie group[1], then every ergodic, strongly continuous, unitary representation of G is mixing. We will concentrate in the case $G = \mathrm{SL}(2,\mathbb{R})$.

Theorem 3.14 (Howe-Moore for $\mathrm{SL}(2,\mathbb{R})$**).** *Every ergodic, strongly continuous, unitary representation of* $\mathrm{SL}(2,\mathbb{R})$ *is mixing.*

[1] Simple Lie groups have Cartan decomposition, e.g. $\mathrm{SL}(2,\mathbb{R}) = KA^+K$ (Lemma 3.2).

Proof. The proof is by contradiction. Fix $v, w \in \mathcal{H}$, and assume that there is a sequence $g_n \to \infty$ in G such that $|f_{v,w}(g_n)| > 2\varepsilon$ for some $\varepsilon > 0$. We will construct a non-zero vector $z \in \mathcal{H}$ such that $\pi(g)z = z$ for all $g \in \mathrm{SL}(2,\mathbb{R})$. By ergodicity, this is a contradiction.

Firstly, we claim that we can reduce the contradiction assumption to diagonal elements. For that we use the Cartan decomposition of $\mathrm{SL}(2,\mathbb{R})$ (Lemma 3.2). Write $g_n = k_n a_n l_n$, with $k_n, l_n \in K$ and $a_n \in A^+$. Because K is compact, we can assume that $k_n \to k$ and $l_n \to l$, with $k, l \in K$. Let $v' = \pi(l)v$ and $w' = \pi(k^{-1})w$. We claim that $|f_{v',w'}(a_n)| > \varepsilon$ for large n. Indeed,

$$\begin{aligned}|f_{v',w'}(a_n) - f_{v,w}(g_n)| &= |\langle \pi(a_n l)v, \pi(k^{-1})w\rangle - \langle \pi(a_n l_n)v, \pi(k_n^{-1})w\rangle| \\ &\leq |\langle \pi(a_n l)v, \pi(k^{-1})w - \pi(k_n^{-1})w\rangle| + \\ &\quad\ |\langle \pi(a_n)(\pi(l)v - \pi(l_n)v), \pi(k_n^{-1})w\rangle| \\ &\leq \|v\|\|\pi(k^{-1})w - \pi(k_n^{-1})w\| + \|w\|\|\pi(l)v - \pi(l_n)v\|\end{aligned}$$

converges to zero. Renaming the vectors, we assume that $|f_{v,w}(a_n)| > \varepsilon$. Because $g_n \to \infty$ in G and K is compact, $a_n \to \infty$ in G.

Look at the sequence of vectors $\{\pi(a_n)v\}$: it is contained in a ball of $\mathcal{H}$. By the Banach-Alaoglu theorem, we can assume that $\pi(a_n)v \overset{\mathrm{w}}{\to} z$, for some $z \in \mathcal{H}$. Observe that $z \neq 0$, because $|\langle z, w\rangle| = \lim_{n\to+\infty} |f_{v,w}(a_n)| \geq \varepsilon$.

For any $h \in N$, it holds $a_n^{-1} h a_n \to e$. By Mautner's Lemma (Lemma 3.10), $\pi(h)z = z$ for all $h \in N$. By Mautner phenomenon (Theorem 3.12), $\pi(h)z = z$ for all $h \in \mathrm{SL}(2,\mathbb{R})$.

Here is a dynamical consequence of Howe-Moore's theorem.

Theorem 3.15. *Let $X = \Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ be a homogeneous space, and let $g \in A$ or $g \in N$. Then μ_X is mixing for the $\mathbb{Z}$-action defined by g. In particular, the geodesic and horocycle flows are mixing.*

Even more is known: Marcus proved that the horocycle flow is mixing of all orders [15].

3.5 Furstenberg's theorem

We proceed to give a self-contained proof of Furstenberg's theorem. It says that horocycle flows on compact homogeneous spaces are measure-theoretically rigid.

Theorem 3.16 (Furstenberg [9]). *If $X = \Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ is a compact homogeneous space, then the horocycle flow is uniquely ergodic.*

Proof. Let $G = \mathrm{SL}(2,\mathbb{R})$, and let μ be a probability measure on X invariant under h_t. We want to show that $\mu = \mu_X$. The proof follows two steps:

Step 1. N has a Fϕlner sequence $\{B_n\}_{n\geq 1}$.

Step 2. For every $x \in X$ and every $f \in C_0(X)$ it holds

$$\lim_{n\to+\infty} \frac{1}{\mu_N(B_n)} \int_{B_n} f(xu)d\mu_N(u) = \int_X f d\mu_X.$$

By Step 2, every $x \in X$ is generic for μ_X, thus the horocycle flow is uniquely ergodic.

Fix $a = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix} \in A$. By direct calculation, we have

$$N = \{g \in G : a^n g a^{-n} \to e\} \tag{3.3}$$

and

$$AU^- = \left\{ \begin{pmatrix} \lambda & 0 \\ b & \lambda^{-1} \end{pmatrix} : \lambda > 0, b \in \mathbb{R} \right\} = \{g \in G : \{a^{-n} g a^n\}_{n\geq 0} \text{ bdd.}\}. \tag{3.4}$$

Let $P = AU^-$. Here are some properties of N and P.

- N and P are closed: clear.
- $N \cap P = \{e\}$: if $\begin{pmatrix} \lambda & 0 \\ b & \lambda^{-1} \end{pmatrix} = g = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, then $g = e$.
- NP contains a neighborhood of e: if $\begin{pmatrix} x & y \\ z & w \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R})$ with $x \neq 0$, then

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} x & 0 \\ z & x^{-1} \end{pmatrix} \begin{pmatrix} 1 & x^{-1}y \\ 0 & 1 \end{pmatrix}.$$

By Lemma 3.4, $\mu_G = \mu_N \times \mu_P$ on a neighborhood of e.

Let's prove Step 1. Let $B_n = a^{-n} B a^n$, where $B = B_\varepsilon(e) \subset N$. By (3.3), $N = \bigcup_{n\geq 1} B_n$ and the convolution $g \in N \mapsto a^{-1} g a \in N$ is well-defined. Such convolution preserves ratios of μ_N-measures: if $A \subset N$, then by equation (1.6) we have $\mu_N(a^{-1}Aa) = 4\mu(A)$. Thus for every compact $C \subset N$,

$$\begin{aligned} \frac{\mu_N(B_n \Delta B_n C)}{\mu_N(B_n)} &= \frac{\mu_N(a^{-n}Ba^n \Delta a^{-n}Ba^n C)}{\mu_N(a^{-n}Ba^n)} \\ &= \frac{\mu_N(a^{-n}(B\Delta B a^n C a^{-n})a^n)}{\mu_N(a^{-n}Ba^n)} \\ &= \frac{\mu_N(B\Delta B a^n C a^{-n})}{\mu_N(B)} \end{aligned}$$

converges to zero, because $a^n C a^{-n} \to e$ uniformly and $\mu_N(\partial B) = 0$.

Now we prove Step 2. Fix $f \in C_0(X)$. Let $V = B_\varepsilon(e) \subset P$. Given $\delta > 0$, choose ε small enough such that the conditions below are satisfied:

(i) $|f(xg) - f(x)| < \delta$ for all $x \in X$, $n \geq 0$ and $g \in a^{-n} V a^n$.
(ii) The diameter of BV is smaller than the injectivity radius of X.

The ratio-preserving property of μ_N, Lemma 3.4 and (i) above give that (integration on u is wrt μ_N and on v is wrt μ_P):

$$\begin{aligned}\frac{1}{\mu_N(B_n)}\int_{B_n} f(xu)d\mu_N(u) &= \frac{1}{\mu_N(B_n)\mu_P(V)}\int_V\int_{B_n} f(xua^{-n}va^n)+O(\delta)\\ &= \frac{1}{\mu_N(B)\mu_P(V)}\int_V\int_B f(xa^{-n}uva^n)+O(\delta)\\ &= \frac{1}{\mu_G(BV)}\int_{BV} f(xa^{-n}ga^n)d\mu_G(g)+O(\delta).\end{aligned}$$

By (ii), the map $g\in BV\mapsto xg\in X$ is injective, thus also is the map $g\in BV\mapsto xa^{-n}ga^n\in X$. This, together with Lemma 3.4 (more specifically, the comment after its proof) and the a-invariance of μ_X, give that

$$\begin{aligned}\int_{BV} f(xa^{-n}ga^n)d\mu_G(g) &= \int_{xa^{-n}BVa^n} f(y)d\mu_X(y)\\ &= \int_{xa^{-n}BV} f(ya^{-n})d\mu_X(y)\\ &= \int_X f(ya^{-n})\chi_{xa^{-n}BV}(y)d\mu_X(y).\end{aligned}$$

Because X is compact, we can assume that $xa^{-n}\to z$, thus $\|\chi_{xa^{-n}BV}-\chi_{zBV}\|_{L^2(X,\mu_X)}\to 0$. The Cauchy-Schwartz inequality and Theorem 3.15 imply that

$$\begin{aligned}\lim_{n\to+\infty}\int_X f(ya^{-n})\chi_{xa^{-n}BV}(y)d\mu_X(y) &= \lim_{n\to+\infty}\int_X f(ya^{-n})\chi_{zBV}(y)d\mu_X(y)\\ &= \mu_X(zBV)\int_X fd\mu_X\\ &= \mu_G(BV)\int_X fd\mu_X\end{aligned}$$

and so

$$\lim_{n\to+\infty}\frac{1}{\mu_N(B_n)}\int_{B_n} f(xu)d\mu_N(u) = \int_X fd\mu_X+O(\delta).$$

Because $\delta>0$ is arbitrary, the proof of Step 2 is complete.

3.5.1 Hedlund and Ratner imply Furstenberg

To give an example of the power of Ratner's theorems, we describe a proof of Furstenberg's theorem using Theorem 2.13: is μ is N-invariant, then there is $x\in X$ and a closed subgroup $L\supset U$ such that $\overline{xN}=xL$ and $\mu=\mu_{xL}$. By Hedlund's theorem (Theorem 3.5), $\overline{xN}=X$ and so $\mu=\mu_X$.

3.6 Dani-Smillie's theorem

If $X = \Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ is not compact, then the situation changes. The horocycle flow is not even minimal, because it has periodic orbits. For example, if $\Gamma = \mathrm{SL}(2,\mathbb{Z})$, then any vertical vector pointing up, whose base point has y-coordinate bigger than one, defines a periodic orbit (see Figure 3.3). Nevertheless, by Hedlund's theorem (Theorem 3.5), this is the only case when the orbit is not dense.

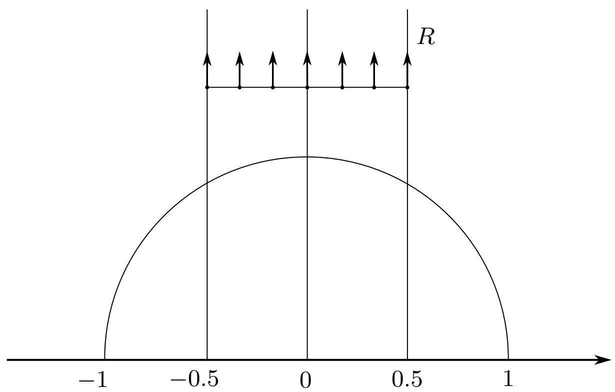

Fig. 3.3. A periodic orbit for the horocycle flow on $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$.

In the measure-theoretical level, the existence of periodic orbits forbids unique ergodicity. When $\Gamma = \mathrm{SL}(2,\mathbb{Z})$, Dani proved that the periodic orbits are the only obstruction for unique ergodicity.

Theorem 3.17 (Dani [5]). *Let μ be an horocycle-invariant ergodic probability measure on $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$. Then either μ is the Haar measure or μ is supported on a periodic orbit.*

Later on, Dani and Smillie generalized this result to general nonuniform lattices.

Theorem 3.18 (Dani-Smillie [6]). *Let $X = \Gamma\backslash\mathrm{SL}(2,\mathbb{R})$ be a noncompact homogeneous space, and let μ be a horocycle-invariant ergodic probability measure on X. Then either $\mu = \mu_X$ or μ is supported on a periodic orbit. If $x \in X$ is not periodic, then its orbit equidistributes wrt μ_X.*

3.6.1 Hedlund and Ratner imply Dani-Smillie

Let μ be an ergodic N-invariant probability measure. By Theorem 2.13, there is $x \in X$ and a closed subgroup $L \supset U$ such that $\overline{xN} = xL$ and $\mu = \mu_{xL}$. By Hedlund's theorem (Theorem 3.5), either xL is a periodic orbit or $xL = X$, thus either μ is supported on a periodic orbit or $\mu = \mu_X$.

4

Oppenheim's conjecture

A *quadratic form* is a homogeneous polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$ of degree two:

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

Oppenheim conjectured in 1929 that, under some natural conditions, $Q(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.

Conjecture (Oppenheim). Let Q be a non-degenerate, indefinite quadratic form in $n \geq 3$ variables that is not the multiple of a quadratic form with rational coefficients. Then $Q(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.

Around 1987, Margulis proved Oppenheim's conjecture using unipotent actions on homoegeneous spaces [16, 17]. In this chapter we will give a proof of Oppenheim's conjecture, modulo some (deep) results on algebraic groups. We mostly follow [19], and also borrow some discussions from [1, 8, 18].

Oppenheim's conjecture is a conjecture on diophantine approximations: small values of Q give diophantine properties of the coefficients a_{ij}. For example, let $Q(x, y) = x - \alpha y$, where $\alpha \notin \mathbb{Q}$. By Dirichlet's theorem, there exists infinitely many $\frac{x}{y} \in \mathbb{Q}$ such that $|\frac{x}{y} - \alpha| < \frac{1}{y^2}$, thus $0 < |Q(x, y)| < \frac{1}{y}$. Reciprocally, if $0 < |Q(x, y)| < \frac{1}{y}$ then $\frac{x}{y}$ is a good rational approximation of α. Another example we'll see below is $Q(x, y) = x^2 - \alpha^2 y^2$ (see Section 4.1.1).

4.1 From Oppenheim to Raghunathan

Firstly, we make some definitions. Let Q be a quadratic form.

Definition 4.1. *Q is called rational if it is a multiple of a quadratic form with rational coefficients.*

For example, $Q(x,y,z) = \sqrt{2}x^2 + \sqrt{8}y^2 - \sqrt{2}xz$ is rational, while $Q(x,y,z) = x^2 + \sqrt{2}y^2 - \sqrt{3}z^2$ is not.

Definition 4.2. *Q is called definite if either $Q(x) \geq 0$ for all $x \in \mathbb{R}^n$ or $Q(x) \leq 0$ for all $x \in \mathbb{R}^n$.*

For example, $Q(x,y) = x^2 - 2xy + y^2$ is definite, while $Q(x,y) = x^2 - 3xy + y^2$ is indefinite. Other examples of indefinite quadratic forms are $Q(x_1,\dots,x_n) = x_1^2 + \dots + x_i^2 - x_{i+1}^2 - \dots - x_n^2$, where $0 \leq i < n$.

A theorem of Meyer in number theory states that if Q is an indefinite, rational quadratic form in $n \geq 5$ variables, then there is $x \in \mathbb{Z}^n - \{0\}$ such that $Q(x) = 0$. In 1929, Oppenheim conjectured an analogue of Meyer's theorem for non-rational quadratic forms.

Conjecture (Oppenheim - weak form). Let Q be an non-degenerate, indefinite, non-rational quadratic form in $n \geq 5$ variables. Then for every $\varepsilon > 0$, there exists a vector $x \in \mathbb{Z}^n - \{0\}$ such that $|Q(x)| < \varepsilon$.

In 1946, Davenport extended this conjecture to $n \geq 3$ variables. Finally, in 1953 Oppenheim stated the conjecture in its final form, as stated in the beginning of this chapter. Because Q is homogeneous, we have that $Q(tv) = t^2Q(v)$ for $t \in \mathbb{R}$. Thus Oppenheim's conjecture is equivalent to the following statement: for every $\varepsilon > 0$, there is $x \in \mathbb{Z}^n - \{0\}$ such that $0 < |Q(x)| < \varepsilon$.

Before Margulis, all approaches used to prove the conjecture were number theorical. For example, Birch, Davenport and Ridout proved the conjecture for $n \geq 21$, and Davenport and Heilbronn proved it for diagonal quadratic forms in five variables. But the number theoretical arguments were not enough for the case $n = 3$ (which is the hardest case, according to Lemma 4.6).

The first one to spread (although not the first one to observe) the relation of Oppenheim's conjecture to dynamics on homogeneous spaces was Raghunathan. He noted that Oppenheim's conjecture is related to the relative compactness of orbits of a subgroup generated by unipotent elements on the moduli space of three dimensional lattices. This is Lemma 4.5.

4.1.1 The assumptions on Q are necessary

Let us explain why we need to assume that $n \geq 3$ and that Q is non-degenerate, indefinite and non-rational.

- Indefinite: if Q is definite, then there exists $C > 0$ such that $Q(x) \geq C\|x\|$ for all $x \in \mathbb{R}^n$. In particular, $Q(\mathbb{Z}^n)$ is discrete.
- Non-rational: if $a_{ij} \in \mathbb{Q}$, then $Q(\mathbb{Z}^n)$ is discrete. The same happens if Q is a multiple of a quadratic form in $\mathbb{Q}[x_1,\dots,x_n]$.
- $n \geq 3$: there are quadratic forms $Q \in \mathbb{R}[x,y]$ that do not satisfy Oppenheim's conjecture. For example, let $\alpha > 0$ be a quadratic irrational such

that α^2 is irrational[1]. These numbers are badly approximable by rationals: there exists $C > 0$ such that $\left|\frac{x}{y} - \alpha\right| > \frac{C}{y^2}$ for all $\frac{x}{y} \in \mathbb{Q}$. Thus $Q(x,y) = x^2 - \alpha^2 y^2$ does not satisfy Oppenheim's conjecture: if $x, y > 0$, then $|Q(x,y)| = |x - \alpha y||x + \alpha y| \geq C\alpha$.

- Non-degeneracy: $Q \in \mathbb{R}[x_1, \ldots, x_n]$ is *degenerate* if there is a change of variables that makes $Q \circ g$ a quadratic form with less than n variables. For example, $Q(a,b,c,d) = (a+b)^2 - \alpha^2(c+d)^2$ is degenerate, because the change of variables $x = a+b$ and $y = c+d$ takes Q to $(x,y) \mapsto x^2 - \alpha^2 y^2$. In particular, Q does not satisfy Oppenheim's conjecture.

4.2 The moduli space $\mathrm{SL}(n,\mathbb{R})/\mathrm{SL}(n,\mathbb{Z})$

The case $\Gamma = \mathrm{SL}(n,\mathbb{Z})$ is important not only for the proof of Oppenheim's conjecture, but also for other areas of mathematics, such as algebra and geometry. The reason is that the homogeneous space $X_n = \mathrm{SL}(n,\mathbb{R})/\mathrm{SL}(n,\mathbb{Z})$ is the moduli space of unimodular (covolume one) lattices of $\mathbb{R}^n$. Let's prove this. For each $g \in \mathrm{SL}(n,\mathbb{R})$ we can define a unimodular lattice of $\mathbb{R}^n$ by $\Gamma = ge_1\mathbb{Z} + \cdots + ge_n\mathbb{Z}$. Two elements $g, h \in \mathrm{SL}(n,\mathbb{R})$ define the same lattice iff $g\gamma = h$ for some $\gamma \in \mathrm{SL}(n,\mathbb{Z})$: $ge_1\mathbb{Z} + \cdots + ge_n\mathbb{Z} = he_1\mathbb{Z} + \cdots + he_n\mathbb{Z}$ iff there are $\gamma_{ij} \in \mathbb{Z}$ such that

$$\begin{cases} he_1 = \gamma_{11} ge_1 + \cdots + \gamma_{n1} ge_n \\ \qquad \vdots \\ he_n = \gamma_{1n} ge_1 + \cdots + \gamma_{nn} ge_n \end{cases} \iff h = g \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} \end{pmatrix}.$$

The action of $\mathrm{SL}(n,\mathbb{R})$ and its subgroups on X_n can be translated to number theoretical information. This is the starting point of Raghunathan's observation (Lemma 4.5).

Another description of X_2 as a moduli space comes from geometry. If M is a two dimensional flat torus with unit area, then $\mathbb{R}^2$ is its universal cover. If $\pi : \mathbb{R}^2 \to M$ is the covering map, then $\pi^{-1}(\pi(0))$ is a unimodular lattice of $\mathbb{R}^2$. If M' is another two dimensional flat torus with unit area, then the lattices are equal iff M and M' are isometric. Thus X_2 is the moduli space of two dimensional flat tori with area one.

4.2.1 Mahler compactness criterium

We follow Section 1.3.3 of [7]. Because $\mathrm{SL}(n,\mathbb{R})$ is a discrete subgroup, X_n is a topological space (with the quotient topology). Let us give an alternative description of this topology, in terms of lattices. A sequence of lattices $\gamma_k \in X_n$

[1] An irrational number α is called *quadratic irrational* if it solves a quadratic equation $x^2 + bx + c = 0$, where $b, c \in \mathbb{R}$. In this case, α^2 is irrational iff $b^2 - 4c \neq 0$.

converges to a lattice $\gamma \in X_n$ iff each γ_k has a basis $\{v_1^k, \dots, v_n^k\}$ and γ has a basis $\{v_1, \dots, v_n\}$ such that $v_i^k \to v_i$ as $k \to +\infty$, for all $i = 1, \dots, n$.

We define the *shortest length function* $\lambda : X_n \to \mathbb{R}$ by

$$\lambda(\gamma) = \min\{\|v\| : v \in \gamma - \{0\}\}.$$

λ is everywhere positive. It is also continuous, by the alternative description of the topology of X_n.

Let us show that X_n is not compact. Consider a sequence $\gamma_k \in X_n$ that degenerates λ:

$$\gamma_k = \frac{1}{k} e_1\mathbb{Z} + k e_2\mathbb{Z} + e_3\mathbb{Z} + \cdots + e_n\mathbb{Z}.$$

Thus $\lambda(\gamma_k) = \frac{1}{k}$. If $\gamma_k \to \gamma$, then $\lambda(\gamma) = \lim \lambda(\gamma_k) = 0$, which contradicts the positiveness of λ.

Actually, the degeneracy of λ characterizes the non-compactness of X_n: the only way for a sequence γ_k to diverge is if $\liminf \lambda(\gamma_k) = 0$. This is *Mahler compactness criterium.*

Theorem 4.3 (Mahler compactness criterium). *$A \subset X_n$ is relatively compact iff* $\inf_{\gamma\in A} \lambda(\gamma) > 0$.

In other words, $\frac{1}{\lambda}$ can be viewed as a height function on X_n. That is why we depict X_n as Figure 4.1.

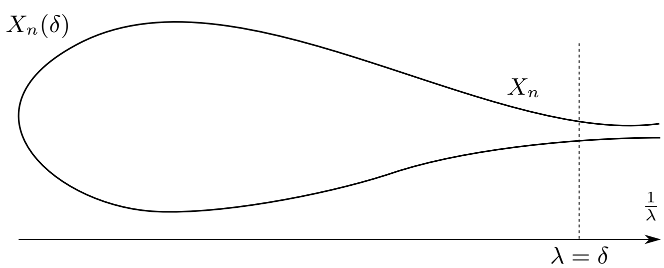

Fig. 4.1. The moduli space of unimodular lattices X_n. By Mahler compactness criterium, $X_n(\delta)$ is compact, i.e. a sequence converges to infinity iff λ converges to zero.

Proof. If A is relatively compact, then $\inf_{\gamma\in A} \lambda(\gamma) > 0$. Otherwise, the same argument used to show the non-compactness of X_n would imply that A has a sequence without a converging subsequence.

Now assume that $\delta_0 = \inf_{\gamma\in A} \lambda(\gamma) > 0$. For each $\delta > 0$, let $X_n(\delta) = \{\gamma \in X_n : \lambda(\gamma) \geq \delta\}$. It is enough to prove that $X_n(\delta)$ is compact: because $A \subset X_n(\delta_0)$, it follows that A is relatively compact.

To prove that $X_n(\delta)$ is compact, we apply Minkowski's successive minima theorem (see Theorem 1.14 of [7]). This theorem says that, for each n, there is a constant $C > 0$ with the following property: if $\gamma \in X_n$, then there exists a basis $\{v_1, \dots, v_n\}$ of γ such that $\|v_1\| = \lambda(\gamma)$ and $C^{-1} \leq \|v_1\| \cdots \|v_n\| \leq C$. In particular, if $\gamma \in X_n(\delta)$ then $\delta \leq \|v_1\|, \dots, \|v_n\| \leq C\delta^{-n+1}$. Thus every sequence on $X_n(\delta)$ has a converging subsequence.

4.2.2 X_n has finite volume

In Section 2.1.6 we proved that X_2 is a homogeneous space. The same happens for any X_n, and the procedure to prove is the same: we explicit a subset $R \subset \mathrm{SL}(n, \mathbb{R})$ that projects onto X_n and $\mu_{\mathrm{SL}(n,\mathbb{R})}(R) < +\infty$. This is a simple but tedious calculation. We refer to Section 1.3.4 of [7].

4.3 Proof of Oppenheim's conjecture

Margulis did not use Ratner's theorems (Ratner proved them a few years after Margulis' proof). He was able to prove a special case of Ratner's orbit closure theorem that was enough to establish the conjecture. Nowadays we can prove Oppenheim's conjecture with the combination of Ratner's theorems and some theory of real algebraic groups. This is what we'll do in this section.

4.3.1 Raghunathan's observation

Raghunathan observed that Oppenheim's conjecture can be rephrased to a statement about the orbits of the action of a subgroup of $\mathrm{SL}(n, \mathbb{R})$ on $\mathrm{SL}(n, \mathbb{R})/\mathrm{SL}(n, \mathbb{Z})$. Given $g \in \mathrm{GL}(n, \mathbb{R})$, let Qg denote the composition $(Qg)(x) = Q(gx)$.

Definition 4.4. *Let Q be a quadratic form in n variables. The set* $\mathrm{SO}(Q) = \{g \in \mathrm{SL}(n, \mathbb{R}) : Qg = Q\}$ *is called the special orthogonal group of Q.*

$\mathrm{SO}(Q)$ is a subgroup of $\mathrm{SL}(n, \mathbb{R})$. Here is an example: if $Q(x, y, z) = x^2 + y^2 + z^2$, then $\mathrm{SO}(Q) = \mathrm{SO}(3)$ is the classical special orthogonal group of $\mathbb{R}^3$. More generally, if $Q(x_1, \dots, x_n) = x_1^2 + \cdots + x_i^2 - x_{i+1}^2 - \cdots - x_n^2$, then $\mathrm{SO}(Q) = \mathrm{SO}(i, n-i)$ is the special orthogonal group with signature $(i, n-i)$. Let $\mathbb{Z}^n \in X_n$ denote the canonical lattice of $\mathbb{R}^n$.

Lemma 4.5 (Raghunathan's observation). *If* $\mathrm{SO}(Q)\mathbb{Z}^n$ *is not relatively compact on X_n, then Oppenheim's conjecture holds for Q.*

Proof. By Mahler compactness criterium (Theorem 4.3), there exists a sequence $u_k \in \mathrm{SO}(Q)$ such that $\lambda(u_k\mathbb{Z}^n) \to 0$, i.e. there are vectors $x_k \in \mathbb{Z}^n - \{0\}$ such that $\|u_k x_k\| \to 0$. In particular, $Q(u_k x_k) \to 0$. Because $u_k \in \mathrm{SO}(Q)$, we have $Q(u_k x_k) = Q(x_k)$, thus $Q(x_k) \to 0$ as well.

It seems that this implication already appeared implicitly in a paper of Cassels and Swinnerton-Dyer [4], as Margulis remarked in [18].

4.3.2 Proof of Oppenheim's conjecture

Firstly, we start with a reduction to the case $n = 3$.

Lemma 4.6. *Let Q be a non-degenerate, indefinite, non-rational quadratic form in $n \geq 3$ variables. Then there is a rational hyperplane Π such that the restriction $Q|_\Pi$ is a non-degenerate, indefinite, non-rational quadratic form.*

The proof is in Chapter VI of [1]. A consequence of the above lemma is that if Oppenheim's conjecture holds for $Q|_\Pi$, then it also holds for Q. Thus we can assume that $Q \in \mathbb{R}[x, y, z]$.

The subgroup $\mathrm{SO}(Q)$ is neither connected nor generated by unipotent elements. In order to apply Ratner's theorems, we restrict the action to the connected component of $\mathrm{SO}(Q)$ containing e. Denote this subgroup by $\mathrm{SO}(Q)^\circ$. Of course, if $\mathrm{SO}(Q)^\circ\mathbb{Z}^n$ is not relatively compact, neither is $\mathrm{SO}(Q)\mathbb{Z}^n$.

By definition, $\mathrm{SO}(Q)^\circ$ is connected. Let us show that it is also generated by unipotent elements. This follows from two observations:

- $\mathrm{SO}(Q)^\circ$ is isomorphic to $\mathrm{SO}(2,1)^\circ$.
- $\mathrm{SO}(2,1)^\circ$ is locally isomorphic to $\mathrm{SL}(2,\mathbb{R})$.

By Lemma 3.3, these will give that $\mathrm{SO}(Q)^\circ$ is generated by unipotent elements.

The second observation is proved in the appendix (Section 5.1). Here is the proof of the first observation: Q is indefinite, thus its signature is either $(2,1)$ or $(1,2)$. Assuming without loss of generality that it is $(2,1)$, there are $h \in \mathrm{SL}(3,\mathbb{R})$ and $\lambda \neq 0$ such that $\lambda Q h = Q_0$, where $Q_0(x,y,z) = x^2 + y^2 - z^2$. Thus $\mathrm{SO}(Q) = h\mathrm{SO}(Q_0)h^{-1} = h\mathrm{SO}(2,1)h^{-1}$. Because the conjugation by h fixes the identity, we have that

$$\mathrm{SO}(Q)^\circ = h\mathrm{SO}(2,1)^\circ h^{-1}. \tag{4.1}$$

Thus $\mathrm{SO}(Q)^\circ$ and $\mathrm{SO}(2,1)^\circ$ are isomorphic.

By Ratner's orbit closure theorem (Theorem 2.11), there is a closed subgroup L satisfying the following:

(i) $\overline{\mathrm{SO}(Q)^\circ\mathbb{Z}^n} = L\mathbb{Z}^n$.
(ii) $\mathrm{SO}(Q)^\circ \subset L \subset \mathrm{SL}(3,\mathbb{R})$.
(iii) $L \cap \mathrm{SL}(3,\mathbb{Z})$ is a lattice of L.

We claim that (ii) implies that either $L = \mathrm{SO}(Q)^\circ$ or $L = \mathrm{SL}(3,\mathbb{R})$. To prove this, let $\widetilde{L} = h^{-1}Lh$. Using equation (4.1), (ii) becomes:

(ii)' $\mathrm{SO}(2,1)^\circ \subset \widetilde{L} \subset \mathrm{SL}(3,\mathbb{R})$.

Because $\mathfrak{so}(2,1)$ is a maximal subalgebra of $\mathfrak{sl}(3,\mathbb{R})$ (see Exercise 1.2#17 of [19]), either $\widetilde{L} = \mathrm{SO}(2,1)^\circ$ or $\widetilde{L} = \mathrm{SL}(3,\mathbb{R})$. We analyze the two cases separately:

Case 1. $\widetilde{L} = \mathrm{SL}(3, \mathbb{R})$: Thus also $L = \mathrm{SL}(3, \mathbb{R})$, i.e. $\overline{\mathrm{SO}(Q)^\circ \mathbb{Z}^n} = X_n$. This implies that $\mathrm{SO}(Q)^\circ \mathbb{Z}^n$ is not relatively compact. By Raghunathans's observation (Lemma 4.5), we are done.

Case 2. $\widetilde{L} = \mathrm{SO}(2, 1)^\circ$: Some theory of real algebraic groups[2] gives that Q is rational, which contradicts our assumption.

This concludes the proof of Oppenheim's conjecture. Note that we proved more: because Case 2 does not happen, the orbit $\mathrm{SO}(Q)^\circ \mathbb{Z}^n$ is dense in X_n, i.e. we can use elements of $\mathrm{SO}(Q)^\circ$ to take $\mathbb{Z}^n$ arbitrarily close to any given unimodular lattice of $\mathbb{R}^n$.

[2] The properties we use are in Sections 4.7 and 4.8 of [19]. The first one is the Borel density theorem: lattices of semisimple real algebraic groups are Zariski dense. The second gives sufficient conditions for a closed subgroup of $\mathrm{SL}(n, \mathbb{R})$ to be defined over $\mathbb{Q}$ (i.e. the polynomials that define the subgroup can be taken to have rational coefficients). After some calculations, this last statement implies that Q is rational.

5

Appendix

5.1 SO$(2, 1)^\circ$ and SL$(2, \mathbb{R})$ are locally isomorphic

Here we prove that SO$(2,1)^\circ$ and SL$(2,\mathbb{R})$ are locally isomorphic. By the fundamental theorem of Sophus Lie (Theorem 2.8), we need to show that $\mathfrak{so}(2,1)$ and $\mathfrak{sl}(2,\mathbb{R})$ are isomorphic as Lie algebras.

We already claimed in Section 2.2.1 that

$$\mathfrak{sl}(2,\mathbb{R}) = \left\{ \begin{pmatrix} x & y \\ z & -x \end{pmatrix} : x, y, z \in \mathbb{R} \right\}.$$

This can be proved by standard calculations. $\mathfrak{sl}(2,\mathbb{R})$ has a basis such that the bracket product has a simple description. This basis is $\{E, F, H\}$ defined by:

$$E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We have $[H, E] = 2E$, $[H, F] = -2F$ and $[E, F] = H$. Thus it is enough to find a basis $\{X, Y, Z\}$ of $\mathfrak{so}(2,1)$ such that $[Z, X] = 2X$, $[Z, Y] = -2Y$ and $[X, Y] = Z$. By standard calculations, we have

$$\mathfrak{so}(2,1) = \left\{ \begin{pmatrix} 0 & x & y \\ -x & 0 & z \\ y & z & 0 \end{pmatrix} : x, y, z \in \mathbb{R} \right\}.$$

After some tentatives, we take

$$X = \begin{pmatrix} 0 & 1 & -\frac{1}{\sqrt{2}} \\ -1 & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 1 & \frac{1}{\sqrt{2}} \\ -1 & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & 0 & \sqrt{2} \\ 0 & 0 & -\sqrt{2} \\ \sqrt{2} & -\sqrt{2} & 0 \end{pmatrix}.$$

5.2 The universal elliptic curve

In this section we describe the quotient $X = G/\Gamma$, where $G = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$ and $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2$. X is called the *universal elliptic curve*. During his lectures, Eskin will sketch the proof of Ratner's theorems for the action of the unipotent subgroup $U^+ = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{R} \right\}$ on X. For further details, see Section 3 of [8].

5.2.1 Semidirect products

Let N be an additive group, and let H be a multiplicative group that acts on N.

Definition 5.1. *The semidirect product of H and N is the group $H \ltimes N$ of pairs (h, n) with the operation*

$$(h_1, n_1)(h_2, n_2) = (h_1 h_2, n_1 + h_1 n_2).$$

The operation is associative: if $(h_i, n_i) \in H \ltimes N$, $i = 1, 2, 3$, then

$$\begin{aligned} [(h_1, n_1)(h_2, n_2)](h_3, n_3) &= (h_1 h_2, n_1 + h_1 n_2)(h_3, n_3) \\ &= (h_1 h_2 h_3, n_1 + h_1 n_2 + h_1 h_2 n_3) \end{aligned}$$

and

$$\begin{aligned} (h_1, n_1)[(h_2, n_2)(h_3, n_3)] &= (h_1, n_1)(h_2 h_3, n_2 + h_2 n_3) \\ &= (h_1 h_2 h_3, n_1 + h_1(n_2 + h_2 n_3)) \\ &= (h_1 h_2 h_3, n_1 + h_1 n_2 + h_1 h_2 n_3). \end{aligned}$$

Also, $(h, n) = (h^{-1}, -h^{-1} n)$.

Topologically, $H \ltimes N$ equals the direct product $H \times N$, but they differ as groups. $H \ltimes N$ has copies of H and N:

- $H \times \{0\}$ is a subgroup of $H \ltimes N$ isomorphic to H. It acts on $H \ltimes N$ as
$$(h, 0)(h_1, n_1) = (h h_1, h n_1), \tag{5.1}$$
i.e. it acts coordinate by coordinate.
- $\{e\} \times N$ is a subgroup of $H \ltimes N$ isomorphic to N. It acts on $H \ltimes N$ as
$$(e, n)(h_1, n_1) = (h_1, n + n_1), \tag{5.2}$$
i.e. it acts by translation in the second coordinate. It is a normal subgroup:
$$\begin{aligned} (h_1, n_1)(e, n_2)(h_1, n_1)^{-1} &= (h_1, n_1)(e, n_2)(h_1^{-1}, -h_1^{-1} n_1) \\ &= (h_1, n_1)(h_1^{-1}, n_2 - h_1^{-1} n_1) \\ &= (h_1 h_1^{-1}, n_1 + h_1(n_2 - h_1^{-1} n_1)) \\ &= (e, h_1 n_2). \end{aligned}$$

We also denote these copies by H and N.

5.2.2 Description of X

$\mathrm{SL}(2,\mathbb{R})$ acts linearly on $\mathbb{R}^2$ (see Section 1.2.2). By the previous section, $G = \mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$ is a group. Analogously, $\Gamma = \mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2$ is a group. Indeed, it is a lattice of G, thus $X = G/\Gamma$ is a homogeneous space.

Let us describe what are the points of X. Let $(g,v),(h,w) \in G$ such that $(g,v)\Gamma = (h,w)\Gamma$. Because

$$(g,v)^{-1}(h,w) = (g^{-1}h, -g^{-1}v + g^{-1}w) = (g^{-1}h, g^{-1}(w-v)),$$

we get that $g\mathrm{SL}(2,\mathbb{Z}) = h\mathrm{SL}(2,\mathbb{Z})$ and $w - v \in g\mathbb{Z}^2$. As we saw in Section 4.2, $g\mathrm{SL}(2,\mathbb{Z})$ is the flat torus with fundamental domain defined by the lattice $g \in \mathbb{Z}^2$. Thus an element of X is a pair (Δ, v), where $\Delta \in X_2$ is a flat torus and v is a point of Δ. Geometrically, X is a fiber bundle with basis X_2 and fibers equal to flat tori (see Figure 5.1).

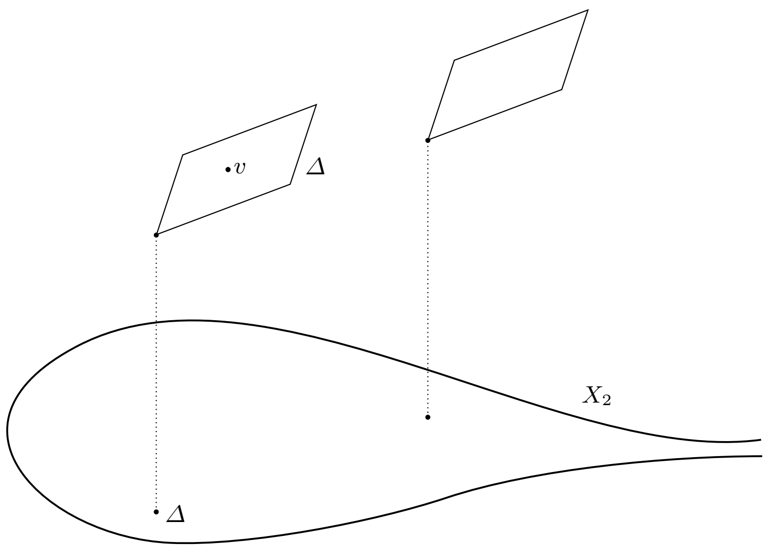

Fig. 5.1. The universal elliptic curve X: it is a fiber bundle with basis X_2 and fibers equal to flat tori.

Let us now describe the actions of $\mathrm{SL}(2,\mathbb{R})$ and $\mathbb{R}^2$ on X.

- Let $h \in \mathrm{SL}(2,\mathbb{R})$. By equation (5.1), h acts on each coordinate:

$$h(\Delta, v) = (h\Delta, hv).$$

- Let $w \in \mathbb{R}^2$. By equation (5.2), w acts by translation on the second coordinate:

$$w(\Delta, v) = (\Delta, v + w).$$

References

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

[2] Rufus Bowen, *Symbolic dynamics for hyperbolic flows*, American Journal of Mathematics **95** (1973), no. 2, 429–460.

[3] Pablo Carrasco, *Introduction to ratner's theory*. Available at `http://tinyurl.com/ratnersnotespc`.

[4] J. W. S. Cassels and H. P. F. Swinnerton-Dyer, *On the product of three homogeneous linear forms and indefinite ternary quadratic forms*, Philos. Trans. Roy. Soc. London. Ser. A. **248**.

[5] S.G. Dani, *Invariant measures of horospherical flows on noncompact homogeneous spaces*, Inventiones mathematicae **47** (1978), no. 2, 101–138.

[6] S.G. Dani and John Smillie, *Uniform distribution of horocycle orbits for fuchsian groups*, Duke Mathematical Journal **51** (1984), no. 1, 185–194.

[7] Manfred Einsiedler and Thomas Ward, *Homogeneous dynamics and applications*, 2013. Available at `http://maths.dur.ac.uk/~tpcc68/homogeneous/welcome.html`.

[8] Alex Eskin, *Unipotent flows and applications*, Homogeneous flows, moduli spaces and arithmetic **10** (2010), 71–129.

[9] Harry Furstenberg, *The unique ergodigity of the horocycle flow*, Recent advances in topological dynamics, 1973, pp. 95–115.

[10] Paul Halmos, *Measure theory*, Vol. 2, van Nostrand New York, 1950.

[11] Gustav A Hedlund, *Fuchsian groups and transitive horocycles*, Duke Mathematical Journal **2** (1936), no. 3, 530–542.

[12] Svetlana Katok, *Fuchsian groups*, University of Chicago Press, 1992.

[13] Anthony W. Knapp, *Lie groups beyond an introduction*, Vol. 140, Springer, 2002.

[14] John M. Lee, *Introduction to smooth manifolds*, Springer, 2002.

[15] Brian Marcus, *The horocycle flow is mixing of all degrees*, Inventiones mathematicae **46** (1978), no. 3, 201–209.

[16] G. A. Margulis, *Formes quadratiques indéfinies et flots unipotents sur les espaces homogènes*, CR Acad. Sci. Paris Sér. I Math **304** (1987), no. 10, 249–253.

[17] ———, *Lie groups and ergodic theory*, Algebra – some current trends, 1988, pp. 130–146.

[18] ———, *Oppenheim conjecture*, Fields Medallists lectures **5** (1997), 272–327.

[19] Dave Witte Morris, *Ratner's theorems on unipotent flows*, University of Chicago Press, 2005.

[20] Donald Ornstein and Benjamin Weiss, *Geodesic flows are Bernoullian*, Israel Journal of Mathematics **14** (1973), no. 2, 184–198.

[21] Marina Ratner, *Ratner theory.* Scholarpedia (`http://www.scholarpedia.org/w/index.php?title=Ratner_theory&oldid=36357`).

[22] ———, *On measure rigidity of unipotent subgroups of semisimple groups*, Acta Mathematica **165** (1990), no. 1, 229–309.

[23] ———, *Strict measure rigidity for unipotent subgroups of solvable groups*, Inventiones Mathematicae **101** (1990), no. 1, 449–482.

[24] ———, *On Raghunathan's measure conjecture*, The Annals of Mathematics **134** (1991), no. 3, 545–607.