M. E.
T. W.

Homogeneous Dynamics and applications (draft title)

– Monograph –

December 15, 2016

Springer

Contents

Introduction

These notes will eventually become a volume on the interaction between number theory and dynamics via homogeneous spaces.

A big portion of this text (specifically, most of Part I and Part III) should be digestible for a reader who is familiar with measure theory and the basics of functional analysis. We will make efforts to avoid going beyond these prerequisites (except for in Part II — see below). There are additional prerequisites at various places of the text, and some of these can be avoided. For example, because the development of the theory is not strictly linear the reader could skip some sections of the text that need more powerful background without losing track of the main ideas. We have marked those sections by footnotes.

The basics of Lie theory, including for instance basic facts about the Lie algebra, the exponential map, and the adjoint representation, will be assumed after a brief review in Chapter 2. Unless the reader strives for maximal generality in her understanding, the more concrete case of linear groups (that is, subgroups of $\mathrm{SL}_d(\mathbb{R})$) is quite sufficient both in breadth of applications and in terms of issues arising. On the other hand, at certain places in Chapter 2 we will use the full force of Lie theory (including the Cartan decomposition, the Levi decomposition, and the Jacobson–Morozov theorem). However, that portion of Chapter 2 could also be skipped, and is not used later.

This text is part of a larger project that started with the book *'Ergodic theory with a view towards Number theory'* [**?**] and is (at modest but positive speed) being developed in parallel with *'Entropy in ergodic theory and homogeneous dynamics'* [**?**]. We will not repeat material from the other two volumes, and will refer to them as needed. Initially we need very few facts from [**?**], namely the Poincaré recurrence theorem and the pointwise ergodic theorem.

The text *'Entropy in ergodic theory and homogeneous dynamics'* only becomes relevant in Part II of the current notes (to be honest, some of the theorems in Chapter 6 will use the results of Part II to give the reader motivation for the latter).

In order to focus on the theory we plan to be quite brief about historical remarks throughout the text, and apologize in advance for any missing references that we should have included. We also try to develop the theory partly from a logical and partly from an instructional point of view. Historically many theorems in these notes were much harder to prove initially, and as we have no desire to suffer ourselves, nor to cause suffering, where it is not strictly necessary we take the logically simpler route (even at the price of ignoring some interesting connections to other topics). On the other hand, from a purely logical point of view we should start immediately with homogeneous spaces defined using algebraic groups over local fields of zero or positive characteristic and also develop the entropy theory much more generally so that the case of smooth maps on manifolds is included. However, as such a text would be quite hard to read for anyone who does not already know the field that we hope to introduce, we instead start with homogeneous spaces defined using linear groups and introduce the language of algebraic groups relatively slowly (mostly over local fields of zero characteristic), starting in Chapter 3. Moreover, we only develop the entropy theory for homogeneous dynamics (which in many ways is easier than the entropy theory for smooth maps).

We hope you will enjoy these notes and the theory that they introduce.

Notation

The symbols $\mathbb{N}=\{1,2,\dots\}$, $\mathbb{N}_0=\mathbb{N}\cup\{0\}$, and $\mathbb{Z}$ denote the natural numbers, non-negative integers and integers; $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ denote the rational numbers, real numbers and complex numbers; the multiplicative and additive circle are denoted $\mathbb{S}^1=\{z\in\mathbb{C}\mid |z|=1\}$ and $\mathbb{T}=\mathbb{R}/\mathbb{Z}$ respectively. The real and imaginary parts of a complex number are denoted $x=\Re(x+\mathrm{i}y)$ and $y=\Im(x+\mathrm{i}y)$. The order of growth of real- or complex-valued functions f,g defined on $\mathbb{N}$ or $\mathbb{R}$ with $g(x)\neq 0$ for large x is compared using Landau's notation:

$$f\sim g \text{ if } \left|\frac{f(x)}{g(x)}\right|\longrightarrow 1 \text{ as } x\to\infty;$$
$$f=\mathrm{o}(g) \text{ if } \left|\frac{f(x)}{g(x)}\right|\longrightarrow 0 \text{ as } x\to\infty.$$

For functions f,g defined on $\mathbb{N}$ or $\mathbb{R}$, and taking values in a normed space, we write $f=\mathrm{O}(g)$ if there is a constant $A>0$ with $\|f(x)\|\leqslant A\|g(x)\|$ for all x. In particular, $f=\mathrm{O}(1)$ means that f is bounded. Where the dependence of the implied constant A on some set of parameters $\mathscr{A}$ is important, we write $f=\mathrm{O}_{\mathscr{A}}(g)$. The relation $f=\mathrm{O}(g)$ will also be written $f\ll g$, particularly when it is being used to express the fact that two functions are commensurate, $f\ll g\ll f$. A sequence $a_1,a_2,\dots$ will be denoted (a_n). Unadorned norms $\|x\|$ will only be used when x lives in a Hilbert space (usually L^2) and always refer to the Hilbert space norm. For a topological space X, $C(X)$, $C_{\mathbb{C}}(X)$, $C_c(X)$ denote the space of real-valued, complex-valued, compactly supported continuous functions on X respectively, with the supremum norm. For sets A, B, denote the set difference by

$$A \smallsetminus B = \{x \mid x \in A, x \notin B\}.$$

Additional specific notation is collected in an index of notation on page 457.

Acknowledgements

We thank[†] Anthony Flatters, Anish Ghosh, Elon Lindenstrauss, Manuel Lüthi, Alex Maier, John Parker, René Rühr, Uri Shapira, Barak Weiss, and Tamar Ziegler for their comments and suggestions. We thank Elon Lindenstrauss especially for suggesting Theorem 5.11 and its proof.

The second named author is grateful for hospitality at various moments from the Universiteit Utrecht, Imke Rust & Gunther Cornellisen in Utrecht, Grete Einsiedler in Zürich, the Pleasaunce Hotel in Overstrand, and the Erwin Schrödinger Institute in Vienna

December 15, 2016 ME, Zürich
TW, Norwich & Durham

[†] Your name may still appear here if you send us useful remarks on the text.

Part I
Orbits and Dynamics on Locally Homogeneous Spaces

Chapter 1
Lattices and the Space of Lattices

We recall that an *action* of a group G on a space X is a map $G \times X \to X$, written $(g, x) \mapsto g{\bullet}x$, with the property that $g{\bullet}(h{\bullet}x) = (gh){\bullet}x$ and $I{\bullet}x = x$ for all $g, h \in G$ and $x \in X$, where I is the identity element of G. Furthermore, for any $x \in X$ the set $G{\bullet}x = \{g{\bullet}x : g \in G\}$ is called the G-*orbit* of x.

One of our interests in this volume is to study the relationship between orbits, orbit closures and arithmetic properties of groups.

In this chapter we discuss discrete subgroups Γ of a locally compact σ-compact metric group G, the quotient space $X = \Gamma\backslash G$, which we will refer to as a locally homogeneous space, and the question of whether or not there is a G-invariant Borel probability measure on X. We finish by studying the central example $d = \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})$. In other words, we define the spaces (and the canonical measures) on which (or with respect to which) we will later discuss dynamical and arithmetic properties.

1.1 Discrete Subgroups and Lattices

1.1.1 Metric, Topological, and Measurable Structure

In this section, let G be a locally compact σ-compact metric group endowed with a left-invariant metric d_G giving rise to the topology of G. For example, d_G could be the metric derived from a Riemannian metric on a connected Lie group G, but in fact any topological group with a countable basis for the topology has such a metric (see Lemma A.2). We note that the left-invariance of the metric implies that

$$\mathsf{d}_G(g, I) = \mathsf{d}_G(g^{-1}g, g^{-1}) = \mathsf{d}_G(g^{-1}, I)$$

for any $g \in G$. Write $B_r^G = B_r^G(I)$ for the metric open ball of radius r around the identity $I \in G$. If Γ is a *discrete* subgroup (which means that I

is an isolated point of Γ), then there is an induced metric on the quotient space[†] $X = \Gamma\backslash G$ defined by

$$\mathsf{d}_X(\Gamma g_1, \Gamma g_2) = \inf_{\gamma_1,\gamma_2\in\Gamma} \mathsf{d}_G(\gamma_1 g_1, \gamma_2 g_2) = \inf_{\gamma\in\Gamma} \mathsf{d}_G(\gamma g_1, g_2) \tag{1.1}$$

for any $\Gamma g_1, \Gamma g_2 \in X$, where both infima are minima if the metric is proper[‡]. We note that $\mathsf{d}_X(\cdot,\cdot)$ indeed defines a metric on X and that we will always use the topology induced by this metric. In particular, a sequence $\Gamma g_n \in X$ converges to Γg as $n \to \infty$ if and only if there exists a sequence $\gamma_n \in \Gamma$ such that $\gamma_n g_n \to g$ as $n \to \infty$. Another consequence of the definition of this metric is that X and G are *locally isometric* in the following sense.

Lemma 1.1 (Injectivity radius). *Let Γ be a discrete subgroup in a group equipped with a left-invariant metric. For any compact subset $K \subseteq X = \Gamma\backslash G$ there exists some $r = r(K) > 0$, called the* injectivity radius *on K, with the property that for any $x_0 \in K$ the map*

$$B_r^G \ni g \longmapsto x_0 g \in B_r^X(x_0)$$

is an isometry between B_r^G and $B_r^X(x_0)$. If $K = \{x_0\}$ where $x_0 = \Gamma h$ for some $h \in G$, then

$$r = \tfrac{1}{4} \inf\nolimits_{\gamma\in\Gamma\smallsetminus\{I\}} \mathsf{d}_G(h^{-1}\gamma h, I) \tag{1.2}$$

has this property.

Proof. We first show this locally, for $K = \{x_0\}$ where $x_0 = \Gamma h$. Let r be as in (1.2), which is positive since $h^{-1}\Gamma h$ is also a discrete subgroup. Then, for $g_1, g_2 \in B_r^G$,

$$\mathsf{d}_X(\Gamma h g_1, \Gamma h g_2) = \inf_{\gamma\in\Gamma} \mathsf{d}_G(h g_1, \gamma h g_2) = \inf_{\gamma\in\Gamma} \mathsf{d}_G(g_1, h^{-1}\gamma h g_2).$$

We wish to show that the infimum is achieved for $\gamma = e$. Suppose that $\gamma \in \Gamma$ has

$$\mathsf{d}_G(g_1, h^{-1}\gamma h g_2) \leqslant \mathsf{d}_G(g_1, g_2) < 2r$$

then

$$\mathsf{d}_G(h^{-1}\gamma h g_2, I) \leqslant \mathsf{d}_G(h^{-1}\gamma h g_2, g_1) + \mathsf{d}_G(g_1, I) < 3r$$

since $g_1 \in B_r^G$, and similarly

$$\begin{aligned}\mathsf{d}_G(h^{-1}\gamma h, I) &= \mathsf{d}_G(e, h^{-1}\gamma^{-1} h)\\ &\leqslant \mathsf{d}_G(e, g_2) + \mathsf{d}_G(g_2, h^{-1}\gamma^{-1} h)\\ &\leqslant r + \mathsf{d}_G(h^{-1}\gamma h g_2, I) < 4r.\end{aligned}$$

[†] As usual in geometry and number theory, we consider $\Gamma\backslash G$ instead of G/Γ; the latter is also often considered in dynamics. The two set-ups are equivalent via the bijection sending $\Gamma g \in \Gamma\backslash G$ to $g^{-1}\Gamma \in G/\Gamma$.

[‡] A metric is *proper* if any ball of finite radius has a compact closure.

This implies that $\gamma = I$.

The lemma now follows by compactness of K. For x_0 and r as above it is easily checked that any $y \in B^X_{r/2}(x_0)$ satisfies the first claim of the proposition with r replaced by $r/2$. Hence K can be covered by balls so that on each ball there is a uniform injectivity radius. Now take a finite subcover and the minimum of the associated injectivity radii. □

Notice that given an injectivity radius, any smaller number will also be an injectivity radius. We define *the maximal injectivity radius* r_{x_0} at $x_0 \in X$ as the supremum of the possible injectivity radii for the set $K = \{x_0\}$ as in the lemma (see also Exercise 1.1.3). If $x_0 = \Gamma h$ then

$$\frac{1}{4} \inf_{\gamma \in \Gamma} \mathsf{d}_G(h^{-1}\gamma h, I) \leqslant r_{x_0} \leqslant \inf_{\gamma \in \Gamma} \mathsf{d}_G(h^{-1}\gamma h, I) \tag{1.3}$$

by Lemma 1.1.

We also define the natural quotient map

$$\begin{aligned} \pi_X : G \ &\to \ X = \Gamma\backslash G \\ g &\longmapsto \Gamma g, \end{aligned}$$

and note that π_X is locally an isometry by left invariance of the metric and Lemma 1.1. Clearly $X = \Gamma\backslash G$ is a homogeneous space in the sense of algebra, but due to this local isometric property we will call X a *locally homogeneous space*.

One (rather abstract) way to understand the quotient space $X = \Gamma\backslash G$ may be to consider a subset $F \subseteq G$ for which the projection π_X, when restricted to F, is a bijection. This motivates the following definition.

Definition 1.2 (Fundamental domain). Let $\Gamma \leqslant G$ be a discrete subgroup. A *fundamental domain* $F \subseteq G$ is a measurable[†] set with the property that

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

(where $\bigsqcup$ denotes a disjoint union). Equivalently, $\pi_X|_F : F \to \Gamma\backslash G$ is a bijection. A measurable set $B \subseteq G$ will be called *injective (for Γ)* if $\pi_X|_B$ is an injective map, and *surjective (for Γ)* if $\pi_X(B) = \Gamma\backslash G$.

Example 1.3. The set $[0,1)^d \subseteq \mathbb{R}^d$ is a fundamental domain for the discrete subgroup $\Gamma = \mathbb{Z}^d \leqslant \mathbb{R}^d = G$.

We will see more examples later, but the existence of a fundamental domain is a general property.

Lemma 1.4 (Existence of fundamental domains). *Let G be a locally compact σ-compact group equipped with a left-invariant metric $\mathsf{d}_G(\cdot,\cdot)$. If Γ*

[†] Unless indicated otherwise, measurable always means Borel-measurable.

is a discrete subgroup of G and $B_{inj} \subseteq B_{surj} \subseteq G$ are injective (resp. surjective) sets, then there exists a fundamental domain F with $B_{inj} \subseteq F \subseteq B_{surj}$. Moreover, $\pi_X|_F : F \to X = \Gamma\backslash G$ is a bi-measurable[†] bijection for any fundamental domain $F \subseteq G$.

PROOF. Notice first that $\mathsf{d}_X(\pi_X(g_1), \pi_X(g_2)) \leqslant \mathsf{d}_G(g_1, g_2)$ for all $g_1, g_2 \in G$. Therefore, π_X is continuous (and hence measurable). Using the assumption that G is σ-compact and Lemma 1.1, we can find a sequence of sets (B_n) with $B_n = g_n B_{r_n}^G$ for $n \geqslant 1$ such that $\pi_X|_{B_n}$ is an isometry, and $G = \bigcup_{n=1}^{\infty} B_n$. It follows that for any Borel set $B \subseteq G$ the image $\pi_X(B \cap B_n)$ is measurable for all $n \geqslant 1$, and so $\pi_X(B)$ is measurable. This implies the final claim of the lemma.

Now let $B_{\text{inj}} \subseteq B_{\text{surj}} \subseteq G$ be as in the lemma. Define inductively the following measurable subsets of G:

$$\begin{aligned}
F_0 &= B_{\text{inj}}, \\
F_1 &= B_{\text{surj}} \cap B_1 \smallsetminus \pi_X^{-1}\left(\pi_X(F_0)\right), \\
F_2 &= B_{\text{surj}} \cap B_2 \smallsetminus \pi_X^{-1}\left(\pi_X(F_0 \cup F_1)\right),
\end{aligned}$$

and so on. Then $F = \bigsqcup_{n=0}^{\infty} F_n$ satisfies all the claims of the lemma. Clearly F is measurable and $B_{\text{inj}} \subseteq F \subseteq B_{\text{surj}}$. If now $g \in G$ is arbitrary we need to show that $(\Gamma g) \cap F$ consists of a single element. If Γg intersects B_{inj} nontrivially, then the intersection is a singleton by assumption and F_n will be disjoint to Γg for all $n \geqslant 1$ by construction. If Γg intersects B_{inj} trivially, then we choose $n \geqslant 1$ minimal such that Γg intersects $B_{\text{surj}} \cap B_n$. By the properties of B_n this intersection is again a singleton, by minimality of n the point in the intersection also belongs to F_n, and Γg will intersect F_k trivially for $k > n$. Hence in all cases we conclude that $(\Gamma g) \cap F$ is a singleton, or equivalently F is a fundamental domain. $\square$

In some special cases, for example $\mathbb{Z}^d < \mathbb{R}^d$, we will be able to give very concrete fundamental domains with better properties, where in particular the boundary of the fundamental domain consists of lower-dimensional objects. In those situations one could and should also ask about how the various pieces of the boundary are glued together under Γ. For instance, in the case of $\mathbb{Z}^d$ we know that opposite sides of $[0,1)^d$ are to be identified. Another such situation will arise in the discussion in Section 1.2. As our goal is more general quotients where this is typically not so easily done, we will not pursue this further.

[†] That is, both $\pi_X|_F$ and its inverse are measurable maps.

1.1.2 Haar Measure and the Natural Action on the Quotient

Recall (see [**?**, Sec. 8.3] for an outline and the monograph of Folland [**?**, Sec. 2.2] for a full proof) that any metric, σ-compact, locally compact group G has a (left) Haar measure m_G which is characterized (up to proportionality) by the properties

- $m_G(K) < \infty$ for any compact $K \subseteq G$;
- $m_G(O) > 0$ for any non-empty open set $O \subseteq G$;
- $m_G(gB) = m_G(B)$ for any $g \in G$ and measurable $B \subseteq G$.

Similarly there also exists a right Haar measure $m_G^{(r)}$ with the first two properties and invariance under right translation instead of left translation as above. For concrete examples it is often not so difficult to give a concrete description of the Haar measure, see Exercise 1.1.5 and Exercise 1.1.6.

Lemma 1.5 (Independence of choice of fundamental domain). *Let Γ be a discrete subgroup of G. Any two fundamental domains for Γ in G have the same Haar measure. In fact, if $B_1, B_2 \subseteq G$ are injective sets for Γ with $\pi_X(B_1) = \pi_X(B_2)$ then*[†] *$m_G(B_1) = m_G(B_2)$.*

Alternatively we may phrase this lemma as follows. For any discrete subgroup $\Gamma < G$, the left Haar measure m_G induces a natural measure m_X on $X = \Gamma\backslash G$ such that

$$m_X(B) = m_G(\pi_X^{-1}(B) \cap F)$$

where $F \subseteq G$ is any fundamental domain for Γ in G.

Proof of Lemma 1.5. Suppose B_1 and B_2 are injective sets with

$$\pi_X(B_1) = \pi_X(B_2).$$

Then

$$B_1 = \bigsqcup_{\gamma \in \Gamma} B_1 \cap (\gamma B_2)$$

and

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

Note that the discrete subgroup $\Gamma < G$ must be countable as G is σ-compact. Therefore, we see that

$$m_G(B_1) = \sum_{\gamma \in \Gamma} m_G(B_1 \cap \gamma B_2) = \sum_{\gamma \in \Gamma} m_G \left(\gamma^{-1} B_1 \cap B_2\right) = m_G(B_2)$$

[†] As the proof will show, we only need left-invariance of the measure under Γ. We will use this strengthening later.

as required. □

Note that G acts naturally on $X = \Gamma\backslash G$ via right multiplication

$$g\raisebox{0.1ex}{\tiny$\bullet$}x = R_g(x) = xg^{-1}$$

for $x \in X$ and $g \in G$, and that this action satisfies

$$\pi_X(g_1 g_2^{-1}) = \pi_X(g_1)g_2^{-1} = g_2\raisebox{0.1ex}{\tiny$\bullet$}\pi_X(g_1)$$

for all $g_1, g_2 \in G$. Also note that $g_2\raisebox{0.1ex}{\tiny$\bullet$}g_1 = g_1 g_2^{-1}$ for $g_1 \in G$ is the natural action of $g_2 \in G$ on G on the right so that π_X satisfies the equivariance property $\pi_X(g_2\raisebox{0.1ex}{\tiny$\bullet$}g_1) = g_2\raisebox{0.1ex}{\tiny$\bullet$}\pi_X(g_1)$. We are interested in whether X supports a G-invariant probability measure, a property discussed in the next proposition and definition.

Proposition 1.6 (Finite volume quotients). *Let G be a locally compact σ-compact group with a left-invariant metric d_G, and let $\Gamma \leqslant G$ be a discrete subgroup. Then the following properties are equivalent:*

(a) *$X = \Gamma\backslash G$ supports a G-invariant probability measure, that is a probability measure m_X which satisfies $m_X(g\raisebox{0.1ex}{\tiny$\bullet$}B) = m_X(B)$ for all measurable $B \subseteq X$ and all g in G;*
(b) *There is a fundamental domain F for $\Gamma \leqslant G$ with $m_G(F) < \infty$;*
(c) *There is a fundamental domain $F \subseteq G$ which has finite right Haar measure $m_G^{(r)}(F) < \infty$ and $m_G^{(r)}$ is left Γ-invariant.*

If any (and hence all) of these conditions hold, then G is unimodular *(that is, the left-invariant Haar measure is also right-invariant).*

Definition 1.7 (Lattices). A discrete subgroup $\Gamma \leqslant G$ is called a *lattice* if $X = \Gamma\backslash G$ supports a G-invariant probability measure. In this case we also say that X has finite volume[†].

In the proof we will use the modular character and the pigeonhole principle for ergodic theory.

In any metric, locally compact σ-compact group G right multiplication may not preserve the left Haar measure m_G. However, there is a continuous homomorphism, the *modular character*, $\mathrm{mod} : G \to \mathbb{R}_{>0}$ with the property that $m_G(Bg^{-1}) = \mathrm{mod}(g)m_G(B)$ for all measurable $B \subseteq G$ and $g \in G$ (see [**?**, Sec. 8.3] for the details and references).

The modular character may also be defined using a right Haar measure $m_G^{(r)}$ via $m_G^{(r)}(gB) = \mathrm{mod}(g)m_G^{(r)}(B)$ for all measurable $B \subseteq G$ and $g \in G$, and the left and right Haar measures may be normalized to have $m_G^{(r)}(B) = m_G(B^{-1})$ for any Borel set $B \subseteq G$, where

[†] Given a fixed left Haar measure m_G on G, we can define the volume of X as $m_G(F)$ for any fundamental domain $F \subseteq G$ for Γ. Somewhat perversely we will often normalize the Haar measure m_G to have $m_X(X) = 1$.

$$B^{-1} = \{g^{-1} \mid g \in B\}.$$

The pigeonhole principle for ergodic theory is the *Poincaré recurrence theorem*, which may be formulated as follows in the metric setting. We refer to [**?**, Th. 2.21] and Exercise 1.1.7 for the proof.

Theorem 1.8 (Poincaré recurrence). *Let X be a locally compact metric space, and let μ be a Borel probability measure preserved by a continuous map $T : X \to X$. Then for μ-almost every $x \in X$ there is a sequence $n_k \to \infty$ with $T^{n_k} x \to x$ as $k \to \infty$.*

Proof of Proposition 1.6. We will start by proving that (a) $\Longrightarrow$ (c). Suppose therefore that m_X is a probability measure on $X = \Gamma\backslash G$ invariant under the action of G on the right. Then we can define a measure μ on G via the Riesz representation theorem by letting

$$\int f \,\mathrm{d}\mu = \int \sum_{\pi(g)=x} f(g) \,\mathrm{d}m_X(x) \tag{1.4}$$

for any $f \in C_c(G)$. Here the function defined by the sum

$$F : x = \Gamma g \mapsto \sum_{\gamma \in \Gamma} f(\gamma g),$$

on the right hand side belongs to $C_c(X)$ — indeed the sum vanishes if $x \notin \pi(\operatorname{Supp} f)$ and for every given $g \in G$ (and also on any compact neighborhood of g) the sum can be identified with a sum over a finite subset of Γ which implies continuity.

By invariance of μ under the action of G, we see that $\mu = m_G^{(r)}$ is a right Haar measure on G (the reader may check all the characterizing properties of Haar measures from page 11, or rather their analogues for right Haar measures). By the construction above, $m_G^{(r)}$ is left-invariant under Γ. Finally, (1.4) extends using dominated and monotone convergence to any measurable non-negative function f on G. Applying this to $f = \mathbb{1}_F$ for a fundamental domain $F \subseteq G$ shows that $m_G^{(r)}(F) = 1$, hence (c).

Now suppose that (c) holds, and let F be the fundamental domain. We define a measure m_X on X by

$$m_X(B) = \frac{1}{m_G^{(r)}(F)} m_G^{(r)}\left(F \cap \pi_X^{-1}(B)\right).$$

By Lemma 1.5 (and its footnote), this definition is independent of the particular fundamental domain used. Thus for $g \in G$ and $B \subseteq X$ we have

$$
\begin{aligned}
m_X(Bg) &= \frac{1}{m_G^{(r)}(F)} m_G^{(r)}\left(F \cap \pi_X^{-1}(Bg)\right) \\
&= \frac{1}{m_G^{(r)}(F)} m_G^{(r)}\left(F \cap \pi_X^{-1}(B)g\right) \\
&= \frac{1}{m_G^{(r)}(Fg^{-1})} m_G^{(r)}\left(Fg^{-1} \cap \pi_X^{-1}(B)\right) = m_X(B),
\end{aligned}
$$

since $Fg^{-1} \subseteq G$ is also a fundamental domain. This shows (a). It follows that (a) and (c) are equivalent.

We also note that (b) $\Longrightarrow$ (c) rather quickly: If F is a fundamental domain with $m_G(F) < \infty$ and $g \in G$, then Fg is another fundamental domain. Therefore, by Lemma 1.5, $m_G(F) = m_G(Fg) = m_G(F)\operatorname{mod}(g^{-1})$, so G is unimodular and (c) follows.

In the proof that (a) (or, equivalently, (c)) implies (b), we will again show first that G is unimodular. Note that this implies that (b) and (c) are the same statement. Also note that by the equivalence of (a) and (c) above and the uniqueness of Haar measures we know that the measure m_X on X is derived (up to a scalar) from the right Haar measure $m_G^{(r)}$ on G restricted to a fundamental domain $F \subseteq G$. Let $B = B_r^G \subseteq G$ be a compact neighborhood of the identity I in G so that $r > 0$ is an injectivity radius at $\Gamma I \in X$ as in Lemma 1.1. Then $m_X(\pi_X(B)) = m_G^{(r)}(B)$ by Lemma 1.1 and (1.4) (for $\mu = m_G^{(r)}$ and the characteristic function of B). By the properties of the Haar measure we have also $m_X(\pi_X(B)) = m_G^{(r)}(B) > 0$.

Let now g be an element of G; we wish to show that $\operatorname{mod}(g) = 1$, and only know that g preserves a finite measure m_X on X (which we may assume without loss of generality to be a probability measure). By Poincaré recurrence there exists some $b \in B$ and sequences $(n_k), (\gamma_k), (b_k)$ with

$$n_k \nearrow \infty, \gamma_k \in \Gamma, b_k \in B$$

such that

$$bg^{-n_k} = \gamma_k b_k$$

for all $k \geqslant 1$. Applying the modular character, and noticing that

$$\operatorname{mod}(\Gamma) = \{1\}$$

by (c), we see that

$$\operatorname{mod}(g)^{n_k} = \frac{\operatorname{mod}(b)}{\operatorname{mod}(b_k)}$$

belongs to a compact neighborhood of $1 \in (0, \infty)$ for all $k \geqslant 1$. It follows that $\operatorname{mod}(g) = 1$, as required. □

Proposition 1.9 (Haar measure on $X = \Gamma \backslash G$). *Let G and Γ be as in Proposition* 1.6*, and suppose in addition that G is unimodular. Then the*

Haar measure m_G on G induces a locally finite G-invariant measure m_X, also called the Haar measure *on $X = \Gamma\backslash G$, such that*†

$$\int_G f \,\mathrm{d}m_G = \int_X \sum_{\gamma\in\Gamma} f(\gamma g)\,\mathrm{d}m_X(\Gamma g) \tag{1.5}$$

for all $f \in L^1_{m_G}(G)$.

Proof. Since we assume that G is unimodular, the argument that (c) implies (a) in the proof of Proposition 1.6 can be used to define the measure m_X. Once again Lemma 1.5 shows that m_X is independent of the choice of fundamental domain $F \subseteq G$ used in the definition, and shows that m_X is G-invariant. By definition (1.5) holds for $f = \mathbb{1}_B$ if $B \subseteq F$ or $B \subseteq \gamma F$ for some $\gamma \in \Gamma$. By linearity (1.5) also holds for any measurable $B \subseteq G$ and hence for any simple function. In particular, the sum on the right hand side of (1.5) is a measurable function on X (or equivalently on F). The measurability of the sum and the equality of the integrals now extend by monotone convergence to show that (1.5) holds for any measurable non-negative function. □

Notice that Lemma 1.1 implies that any compact set $K_X \subseteq X$ is the image $K_X = \pi_X(K_G)$ of a compact set $K_G \subseteq G$. In particular, this implies that a compact quotient $\Gamma\backslash G$ is of finite volume in the sense of Definition 1.7.

Definition 1.10 (Uniform lattice). A discrete subgroup $\Gamma \leqslant G$ is called a (*co-compact* or) *uniform lattice* if the quotient space $X = \Gamma\backslash G$ is compact‡.

Roughly speaking, $\Gamma \leqslant G$ is a uniform lattice if the quotient space $\Gamma\backslash G$ is small topologically (compact) as well as measurably (of finite volume). At first sight, motivated by the abelian paradigm from $\mathbb{Z}^d \leqslant \mathbb{R}^d$, it seems reasonable to require that $\Gamma\backslash G$ should always be compact in defining a lattice. However, as we will soon see, this would exclude some of the most natural lattices and their quotient spaces.

† This formula is sometimes referred to as method of summation, *folding* (if used from the left hand side to the right hand side), or *unfolding* (if used in the other direction).

‡ A consequence is that there is a choice of injectivity radius that is *uniform* across all of $\Gamma\backslash G$.

1.1.3 Divergence in the Quotient by a Lattice

† In allowing non-compact quotients, it is natural to ask how compact subsets of $X=\Gamma\backslash G$ can be described or, equivalently, to characterize sequences (x_n) in X that *go to infinity* (that is, leave any compact subset of X).

Proposition 1.11 (Abstract divergence criterion). *Let G be a locally compact σ-compact group and let $\Gamma<G$ be a lattice. Then the following properties of a sequence (x_n) in $X=\Gamma\backslash G$ are equivalent:*

(1) *$x_n\to\infty$ as $n\to\infty$, meaning that for any compact set $K\subseteq X$ there is some $N=N(K)\geqslant 1$ such that $n\geqslant N$ implies that $x_n\notin K$.*
(2) *The maximal injectivity radius at $x_n=\Gamma g_n$ goes to zero as $n\to\infty$. That is, there exists a sequence (γ_n) in $\Gamma\smallsetminus\{I\}$ such that $g_n^{-1}\gamma_n g_n\to I\in G$ as $n\to\infty$.*

Proof. We note that the two statements in (2) are equivalent due to (1.3).

Suppose that (1) holds, so that $x_n\to\infty$ as $n\to\infty$. We need to show that the maximal injectivity radius r_{x_n} at x_n goes to zero. So suppose the opposite, then we would have $r_{x_n}\geqslant\varepsilon>0$ for some $\varepsilon>0$ and infinitely many n, and by choosing this subsequnce we may assume without loss of generality that $r_{x_n}\geqslant\varepsilon>0$ for all $n\geqslant 1$.

Decreasing ε if necessary, we may assume that $\overline{B_\varepsilon^G}$ is compact (since G is locally compact). Therefore there is some N_1 with

$$x_n\notin x_1\overline{B_\varepsilon^G}$$

for $n\geqslant N_1$. Now remove the terms $x_2,\dots,x_{N_1-1}$ from the sequence. Similarly, there is an $N_2\geqslant 1$ with

$$x_n\notin x_1\overline{B_\varepsilon^G}\cup x_{N_1}\overline{B_\varepsilon^G}$$

for $n\geqslant N_2$. Repeating this process infinitely often, and renaming the thinned-out sequence remaining (x_n) again, we may assume without loss of generality that $\mathsf{d}(x_n,x_m)\geqslant\varepsilon$ for all $m\neq n$. This now gives a contradiction to the assumption that X has finite volume: if $x_n=\pi_X(g_n)$ then

$$X\supseteq\bigsqcup_{n=1}^\infty x_nB_{\varepsilon/2}^G=\Gamma\left(\bigsqcup_{n=1}^\infty g_nB_{\varepsilon/2}^G\right),$$

and

$$\bigsqcup_{n=1}^\infty g_nB_{\varepsilon/2}^G$$

† In the remainder of the section we collect more fundamental results about locally homogeneous orbits, but the reader in a hurry could also move on to Section 1.2 and return to the material here later as needed.

is a disjoint union of infinite measure, and is an injective set.

Suppose now that (1) does not hold, so there exists some compact $K \subseteq X$ with $x_n \in K$ for infinitely many n. By Lemma 1.1 there exists an injectivity radius $r > 0$ on K and we see that $r_{x_n} \geqslant r$ for infinitely many n, so that (2) does not hold either. □

1.1.4 Orbits of Subgroups

In the following we will also be interested in orbits of subgroups $H \leqslant G$. Given an action of G on a space X, which we will write $(x, g) \mapsto g \cdot x$, the *H-orbit* of $x \in X$ is the set

$$H \cdot x = \{h \cdot x \mid h \in H\} \cong H / \operatorname{Stab}_H(x) \cong \operatorname{Stab}_H(x) \backslash H,$$

where

$$\operatorname{Stab}_H(x) = \{h \in H \mid h \cdot x = x\}$$

is the *stabilizer subgroup* of $x \in X$ and the isomorphisms are sending $h \cdot x$ to $h \operatorname{Stab}_H(x)$ resp. to $\operatorname{Stab}_H(x) h^{-1}$. Note that if $X = \Gamma \backslash G$ and $x = \Gamma g$, then

$$\operatorname{Stab}_H(x) = H \cap g^{-1} \Gamma g$$

is a discrete subgroup of H. Fixing a Haar measure m_H on H we define the *volume of the H-orbit*, $\operatorname{volume}(H \cdot x)$ to be $m_H(F_H)$ where $F_H \subseteq H$ is a fundamental domain for $\operatorname{Stab}_H(x)$ in H.

Clearly if an H-orbit $xH \subseteq X = \Gamma \backslash G$ is compact, it is also closed. In fact the same conclusion can be reached for finite volume orbits.

Proposition 1.12 (Finite volume orbits are closed). *Let G be a locally compact σ-compact group equipped with a left-invariant metric d, let $\Gamma \leqslant G$ be a discrete subgroup, and let $H \leqslant G$ be a closed subgroup. Suppose that the point $x \in X = \Gamma \backslash G$ has a finite volume H-orbit. Then $xH \subseteq X$ is closed.*

We note that Proposition 1.12 can also be deduced from Proposition 1.11 (see Exercise 1.1.9 and compare the two approaches), but we will also give an independent proof.

Proof of Proposition 1.12. Suppose that $y \in \overline{xH}$. By Lemma 1.1 there exists a neighborhood B_r^G of $I \in G$ such that the map $g \mapsto yg$ is injective on B_r^G. Let $V \subseteq H \cap B_{r/2}^G$ be a compact neighborhood of I in H. By assumption, there is a sequence (z_n) with $z_n = xh_n = yg_n \in (xH) \cap (yB_r^G)$ for some $h_n \in H$, $g_n \in B_{r/2}^G$ for each $n \geqslant 1$, and with $g_n \to e$ as $n \to \infty$. If $z_n V \cap yV \neq \varnothing$ for some n, then $y \in z_n V V^{-1} \subseteq xH$ as desired. Assume therefore that $z_n V \cap yV = \varnothing$ for all $n \geqslant 1$. Geometrically (and roughly speaking), we may interpret this situation by saying that $z_n V$ approaches yV from a direction transverse to H, as illustrated in Figure 1.1.

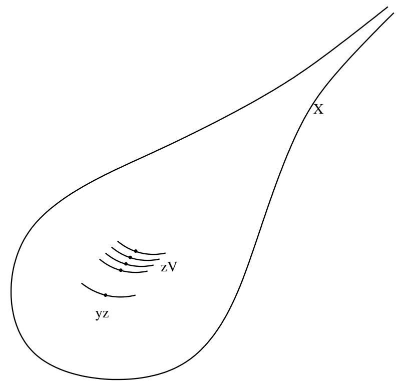

Fig. 1.1 We assume indirectly that the sets $z_n V$ approach yV transverse to the orbit direction.

Compactness of V implies that for any fixed n the set $z_m V$ (which approaches yV) must also be disjoint from $z_n V$ (which has positive distance from yV) for large enough m. Thus we may choose a subsequence and assume that

$$z_n V \cap z_m V = \varnothing$$

for any $n < m$. However, since $z_n = xh_n = yg_n$ as above, each set $z_n V$ is the injective image of the map

$$V \ni h \longmapsto z_n h = yg_n h$$

since $g_n V \subseteq B_{r/2}^G B_{r/2}^G \subseteq B_r^G$. In other words

$$\bigsqcup_{n=1}^{\infty} h_n V \subseteq H$$

is injective for $\mathrm{Stab}_H(x)$. However, this gives

$$m_H \left(\bigsqcup_{n=1}^{\infty} h_n V \right) = \infty,$$

which contradicts the assumption that the orbit xH has finite volume. $\square$

Clearly if we are interested in finding finite volume H-orbits (that will carry finite H-invariant measures), then we need to restrict to unimodular subgroups $H \leqslant G$ (by Proposition 1.6). If H is unimodular (and, as before, we have fixed some Haar measure m_H) then the *volume measure* volume_{xH} on the H-orbit is defined by

$$\mathrm{volume}_{xH}(B) = m_H \left(\{ h \in F \mid xh \in B \} \right)$$

where $F \subseteq H$ is a fundamental domain for $\mathrm{Stab}_H(x)$ in H. This measure may be finite or infinite (and in the latter case it may be locally finite considered on X or not), but is always invariant under the right action of H due to Proposition 1.9 applied to $\mathrm{Stab}_H(x)\backslash H \cong xH$.

Proposition 1.13 (Closed orbits are embedded). *Let G be a locally compact σ-compact group equipped with a left-invariant metric d, let $\Gamma \leqslant G$ be a discrete subgroup, and let $H \leqslant G$ be a closed subgroup. Suppose that the point $x \in X = \Gamma\backslash G$ has a closed H-orbit. Then $xH \subseteq X$ is* embedded*, meaning that the map $h \in \mathrm{Stab}_H(x)\backslash H \to xh \in xH$ is a homeomorphism. In particular,* volume_{xH} *is a locally finite measure on X.*

PROOF. Clearly the map $\mathrm{Stab}_H(x)\backslash H \longrightarrow xH \subseteq X$ is continuous, and we wish to show that its inverse is also continuous.

Replacing $x = \Gamma g$ and H simultaneously with Γ and gHg^{-1}, we may assume for simplicity that $x = \Gamma$ so that $\mathrm{Stab}_H(x) = \Gamma \cap H$.

By Exercise 1.1.2† the quotient G/H is a locally compact metric space. We claim that our assumption that ΓH is closed in $\Gamma\backslash G$ also shows‡ that ΓH is closed as a subset of G/H. Indeed, suppose that $(\gamma_n H)$ converges to gH in G/H. Then we can find a sequence (h_n) in H such that $\gamma_n h_n \to g \in G$ as $n \to \infty$, showing that $\Gamma h_n \to \Gamma g$. However, this implies by our assumption that $\Gamma g \in \Gamma H$, so that there is some $\gamma \in \Gamma$ and $h \in H$ with $g = \gamma h$. This shows that $gH = \gamma H \in \Gamma H$ as needed.

Next we claim that ΓH is a discrete subset of G/H. If not we may choose a sequence (η_n) so that $\eta_n H \to gH$ as $n \to \infty$ for some $g \in G$, but $\eta_n H \neq gH$ for $n \geqslant 1$. Then $gH = \eta H$ for some $\eta \in \Gamma$ as ΓH is closed. Multiplying the sequence on the left by $\gamma\eta^{-1}$ for an arbitrary $\gamma \in \Gamma$ gives a sequence in ΓH with limit γH such that the limit is not achieved in the sequence. This shows that any element of ΓH is an accumulation§ point of ΓH. As Γ is countable (since G is σ-compact) we can write $\Gamma H = \{\gamma_1 H, \gamma_2 H, \dots\}$. Now $O_n = \Gamma H \smallsetminus \{\gamma_n H\}$ is an open dense subset of ΓH, which implies by the Baire category theorem that $\bigcap_n O_n$ must be dense in ΓH, which gives a contradiction as the intersection is empty.

Now suppose that $\Gamma h_n \to \Gamma h$ as $n \to \infty$ in $\Gamma\backslash G$. Then there exists a sequence (γ_n) in Γ with $\gamma_n h_n \to h \in H$ as $n \to \infty$, which implies that

$$\gamma_n H \to H$$

as $n \to \infty$ in G/H. By the discreteness of $\Gamma H \subseteq G/H$, it follows that $\gamma_n \in H$ for large enough n, so that we also have

$$(\Gamma \cap H)\, h_n \longrightarrow (\Gamma \cap H)\, h$$

† The result in the exercise of course also holds for G/H instead of $H\backslash G$.

‡ The argument will show the equivalence of the two.

§ Thus ΓH is a closed *perfect set*(1).

as $n \to \infty$ in $\Gamma \cap H \backslash H$.

For the last claim of the proposition notice that every compact set $K \subseteq X$ intersects xH in a compact set which has finite measure with respect to volume$_{xH}$ (as $K \cap xH$ also corresponds to a compact set in $\mathrm{Stab}_H(x)\backslash H$). $\square$

Exercises for Section 1.1

Exercise 1.1.1. Let G be equipped with a left-invariant metric, and let Γ be a discrete subgroup of G. Show that

$$\mathsf{d}_X(x, xg) \leqslant \mathsf{d}_G(I, g)$$

for all $x \in X$ and $g \in G$, where as usual $X = \Gamma \backslash G$.

Exercise 1.1.2. Let G be a topological group, let d be a left-invariant metric, and let $H < G$ be a closed subgroup. Imitate the definition in (1.1) to define a metric on $H\backslash G$. Show that $H\backslash G$ is locally compact and σ-compact if G is. Assuming only that G is locally compact, show that both G and $H\backslash G$ are complete as metric spaces.

Exercise 1.1.3. Show that the maximal injectivity radius as defined after Lemma 1.1 is indeed an injectivity radius. Show the upper bound in (1.3).

Exercise 1.1.4. Show that the topology induced by the metric $d_X(\cdot,\cdot)$ on $X = \Gamma\backslash G$ is the quotient topology of the topology on G for the natural map $\pi_X : G \to X$ (i.e. finest topology on X for which π_X is continuous).

Exercise 1.1.5. Show that the bi-invariant Haar measure $m_{\mathrm{GL}_d(\mathbb{R})}$ on the locally compact group

$$\mathrm{GL}_d(\mathbb{R}) = \bigl\{g = (g_{ij})_{i,j} \in \mathrm{Mat}_d(\mathbb{R}) \mid \det(g) \neq 0\bigr\},$$

which is called the *general linear group*, can be defined by the formula

$$\mathrm{d}m_{\mathrm{GL}_d(\mathbb{R})}(g) = \frac{\prod_{i,j=1}^d \mathrm{d}g_{ij}}{(\det g)^d}.$$

Exercise 1.1.6. Let $d \geqslant 2$. Show that

$$m_{\mathrm{SL}_d(\mathbb{R})}(B) = m_{\mathbb{R}^{d^2}}(\{tb : t \in [0,1], b \in B\})$$

for any measurable $B \subseteq \mathrm{SL}_d(\mathbb{R})$ defines a (bi-invariant) Haar measure on the locally compact group

$$\mathrm{SL}_d(\mathbb{R}) = \bigl\{g \in \mathrm{Mat}_d(\mathbb{R}) \mid \det(g) = 1\bigr\},$$

which is called the *special linear group*, where $m_{\mathbb{R}^{d^2}}$ is the Lebesgue measure on the matrix algebra $\mathrm{Mat}_d(\mathbb{R})$ viewed as the vector space $\mathbb{R}^{d^2}$.

Exercise 1.1.7. Show that Theorem 1.8 follows from the conventional formulation of Poincaré recurrence: if $(X, \mathscr{B}, \mu, T)$ is a measure-preserving system and $\mu(A) > 0$ then there is some $n \geqslant 1$ for which $\mu(A \cap T^{-n}A) > 0$ (see [**?**, Sec. 2.1]).

Exercise 1.1.8. Rephrase Proposition 1.11 as a compactness criterion characterizing compact subsets of $X = \Gamma\backslash G$ in terms of the injectivity radius.

Exercise 1.1.9. Prove Proposition 1.12 using Proposition 1.11, also showing that the inclusion

$$\begin{aligned} \mathrm{Stab}_H(x)\backslash HX &= \Gamma\backslash G \\ \mathrm{Stab}_H(x)h &\longmapsto xh \end{aligned}$$

is a proper map.

Exercise 1.1.10. Let $G < \mathrm{SL}_d(\mathbb{R})$ be a closed linear group, and let

$$\Gamma = G \cap \mathrm{SL}_d(\mathbb{Z}) < G$$

be a non-uniform lattice in G. Show that Γ must contain a unipotent matrix† (that is, a matrix for which 1 is the only eigenvalue).

Exercise 1.1.11. Let $\Gamma < G$ be a uniform lattice in a connected σ-compact locally compact group G equipped with a proper left-invariant metric. Show that Γ is finitely generated‡.

Exercise 1.1.12. Let $\Gamma < G$ be a discrete subgroup, let $x \in X = \Gamma\backslash G$, and let $H_1, H_2 < G$ be two closed subgroups for which xH_1 and xH_2 are closed orbits. Prove that $x(H_1 \cap H_2) \subseteq (xH_1) \cap (xH_2)$ is a closed orbit.

1.2 A Brief Review of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$

1.2.1 The Space

We recall (see, for example, [**?**, Ch. 9]) that the upper half-plane

$$\mathbb{H} = \{z = x + \mathrm{i}y \in \mathbb{C} \mid y = \Im(z) > 0\}$$

equipped with the Riemannian metric

$$\langle (z,u),(z,v)\rangle_z = \frac{(u \cdot v)}{y^2}$$

for $(z,u),(z,v) \in \mathrm{T}_z\mathbb{H} = \{z\} \times \mathbb{C}$ is the *upper half-plane model* of the hyperbolic plane (where $u \cdot v$ denotes the inner product after identifying u and v with elements of $\mathbb{R}^2$). Moreover, the group $\mathrm{SL}_2(\mathbb{R})$ acts on $\mathbb{H}$ transitively and isometrically via the Möbius transformation

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto g \bullet z = \frac{az+b}{cz+d}. \tag{1.6}$$

† This is true in general, as conjectured by Selberg and proved by Každan and Margulis [**?**]; also see Raghunathan [**?**, Ch. XI]. However, the proof for subgroups of the form $\Gamma = G \cap \mathrm{SL}_d(\mathbb{Z})$ is significantly easier.

‡ This again holds more generally, but for connected groups and for compact quotients the proof is straightforward. We refer to Raghunathan [**?**, Remark 13.21] for the general case.

The stabilizer of $\mathrm{i} \in \mathbb{H}$ is $\mathrm{SO}(2)$ so that

$$\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}(2) \cong \mathbb{H}$$

under the map sending $g\,\mathrm{SO}(2)$ to $g\cdot\mathrm{i}$.

The action of $\mathrm{SL}_2(\mathbb{R})$ is differentiable, and so gives rise to a derived action on the tangent bundle $\mathrm{T}\mathbb{H} = \mathbb{H} \times \mathbb{C}$ by

$$\mathrm{D}\,g : (z, u) \longmapsto \left(g\cdot z, \frac{1}{(cz+d)^2} u \right)$$

where

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This action gives rise to the simply transitive action of

$$\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$$

on the unit tangent bundle

$$\mathrm{T}^1\mathbb{H} = \{(z, v) \in \mathrm{T}\mathbb{H} \mid \|(z, v)\|_z^2 = \langle (z, v), (z, v) \rangle_z = 1\},$$

so that

$$\mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{T}^1\mathbb{H}$$

by sending g to $\mathrm{D}\,g(\mathrm{i}, \mathrm{i})$.

The region E illustrated by shading in Figure 1.2 is a fundamental region for the action of the discrete subgroup $\mathrm{PSL}_2(\mathbb{Z})$ on $\mathbb{H}$ (strictly speaking we should describe carefully which parts of the boundary of the hyperbolic triangle shaded belong to the domain but as the boundary is a nullset one usually ignores that issue — we will comply with this tradition), see Exercise 1.2.1.

This shows that we can define a fundamental domain for $\mathrm{PSL}_2(\mathbb{Z})$ in

$$\mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{T}^1\mathbb{H}$$

by taking all vectors (z, u) whose base point z lies in E, giving the set

$$F = \{g \in \mathrm{PSL}_2(\mathbb{R}) \mid \mathrm{D}\,g(\mathrm{i}, \mathrm{i}) = (z, u) \text{ with } z \in E\}.$$

(Once again, strictly speaking we should describe more carefully which vectors attached to points $z \in \partial E$ are allowed in F.) Furthermore, we can lift the set $F \subseteq \mathrm{PSL}_2(\mathbb{R})$ to a surjective set $F \subseteq \mathrm{SL}_2(\mathbb{R})$ for $\mathrm{SL}_2(\mathbb{Z})$. We claim that this argument shows that

$$\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$$

has finite volume. In order to see this, we recall some basic facts from [**?**, Ch. 9] (which we will prove in greater generality for $\mathrm{SL}_d(\mathbb{R})$ in Section 1.3.4):

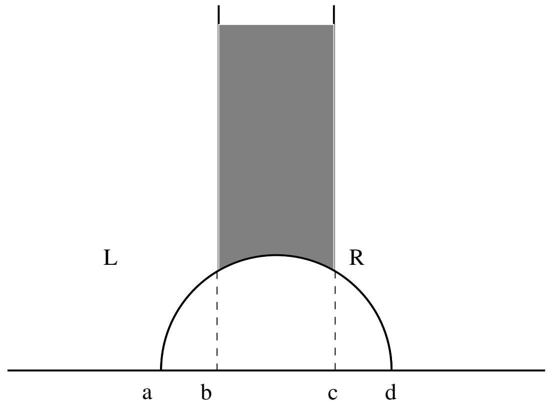

Fig. 1.2 A fundamental domain $E \subseteq \mathbb{H}$ for the action of $\mathrm{SL}_2(\mathbb{Z})$.

- $\mathrm{SL}_2(\mathbb{R})$ is unimodular (see Exercise 1.1.6).
- $\mathrm{SL}_2(\mathbb{R}) = NAK$ with†

$$N = \left\{\begin{pmatrix}1 & * \\ & 1\end{pmatrix}\right\}, A = \left\{\begin{pmatrix}a & \\ & a^{-1}\end{pmatrix} \mid a > 0\right\}$$

  and $K = \mathrm{SO}(2)$, in the sense that every $g \in \mathrm{SL}_2(\mathbb{R})$ can be written uniquely(2) as a product $g = nak$ with $n \in N$, $a \in A$ and $k \in K$.
- Let $B = NA = AN$ be the subgroup $B = \left\{\begin{pmatrix}a & t \\ & a^{-1}\end{pmatrix} \mid a > 0, t \in \mathbb{R}\right\}$. The Haar measure $m_{\mathrm{SL}_2(\mathbb{R})}$ decomposes in the coordinates $g = bk$, meaning that

$$m_{\mathrm{SL}_2(\mathbb{R})} \propto m_B \times m_K$$

  where $\propto$ denotes proportionality (with the constant of proportionality dependent only on the choices of Haar measures). Moreover, the left Haar measure m_B decomposes in the coordinate system

$$b(x, y) = \begin{pmatrix}1 & x \\ & 1\end{pmatrix}\begin{pmatrix}y^{1/2} & \\ & y^{-1/2}\end{pmatrix}$$

  with $x \in \mathbb{R}$, $y > 0$, as

$$\mathrm{d}m_B = \frac{1}{y^2}\,\mathrm{d}x\,\mathrm{d}y.$$

† We sometimes indicate by $*$ any entry of a matrix which is only restricted to be a real number, and do not write entries that are zero.

- We also note that $b(x,y)\centerdot\mathrm{i} = \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix}\centerdot(\mathrm{i}y) = x + \mathrm{i}y$, and that the Haar measure m_B on B is identical to the hyperbolic area measure on $\mathbb{H}$ under the map $b(x,y) \mapsto b(x,y)\centerdot\mathrm{i} = x + \mathrm{i}y$.

Combining these facts we get

$$m_{\mathrm{SL}_2(\mathbb{R})}(F) < \int_{-1/2}^{1/2}\int_{\sqrt{3}/2}^{\infty}\int_0^{2\pi} \frac{1}{y^2}\,\mathrm{d}\theta\,\mathrm{d}y\,\mathrm{d}x < \infty.$$

The argument above also helps us to understand the space

$$2 = \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$$

globally: it is, apart from some difficulties arising from the distinguished points $\mathrm{i}, \frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i} \in E$, the unit tangent bundle of the surface[†] $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. This surface may be thought of as being obtained by gluing the two vertical sides in Figure 1.2 together using the action of $\begin{pmatrix} 1 & \pm 1 \\ & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ and the third side to itself using the action of $\begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. In particular, 2 is non-compact.

1.2.2 The Geodesic Flow — the Subgroup A

We recall that

$$g_t : x \longmapsto x \begin{pmatrix} \mathrm{e}^{t/2} & \\ & \mathrm{e}^{-t/2} \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{-t/2} & \\ & \mathrm{e}^{t/2} \end{pmatrix} \centerdot x$$

defines the geodesic flow on 2, whose orbits may also be described in the fundamental region as in Figure 1.3.

The diagonal subgroup

$$A = \left\{ \begin{pmatrix} \mathrm{e}^{-t/2} & \\ & \mathrm{e}^{t/2} \end{pmatrix} \mid t \in \mathbb{R} \right\}$$

is also called the *torus* or *Cartan subgroup*. We recall that A acts ergodically on 2 with respect to the Haar measure m_2 (see [**?**, Sec. 9.5]; we will also discuss this from a more general point of view in Chapter 2). There are many different types of A-orbits, which include the following:

[†] Because of the distinguished points this surface is a good example of an *orbifold*, but not an example of a manifold.

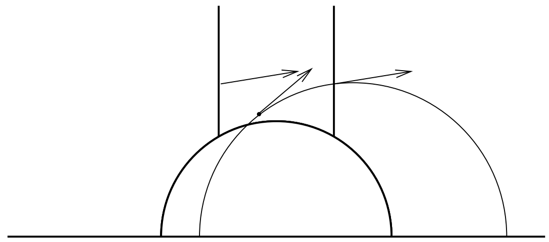

Fig. 1.3 The geodesic flow follows the circle determined by the arrow which intersects $\mathbb{R}\cup\{\infty\}=\partial\mathbb{H}$ normally, and is moved back to F via a Möbius transformation in $\mathrm{SL}_2(\mathbb{Z})$ once the orbit leaves F.

- Divergent trajectories, for example the orbit $\mathrm{SL}_2(\mathbb{Z})A$ which corresponds to the vertical geodesic through (i,i) in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{T}^1\mathbb{H}$.
- Compact trajectories, for example $\mathrm{SL}_2(\mathbb{Z})g_{\mathrm{golden}}A$ is compact, where the matrix $g_{\mathrm{golden}}\in K$ has the property† that

$$g_{\mathrm{golden}}^{-1}\begin{pmatrix}1&1\\1&2\end{pmatrix}g_{\mathrm{golden}}=\begin{pmatrix}\frac{3+\sqrt{5}}{2}&\\&\frac{3-\sqrt{5}}{2}\end{pmatrix}\in A.$$

Now notice that

$$\mathrm{SL}_2(\mathbb{Z})g_{\mathrm{golden}}\begin{pmatrix}\frac{3+\sqrt{5}}{2}&\\&\frac{3-\sqrt{5}}{2}\end{pmatrix}=\mathrm{SL}_2(\mathbb{Z})\begin{pmatrix}1&1\\1&2\end{pmatrix}g_{\mathrm{golden}}=\mathrm{SL}_2(\mathbb{Z})g_{\mathrm{golden}}$$

This identity shows that the orbit $\mathrm{SL}_2(\mathbb{Z})g_{\mathrm{golden}}A$ is compact (see also Figure 1.4 in which $\lambda=\frac{1+\sqrt{5}}{2}$).
- The set of dense trajectories, which includes (but is much larger than) the set of equidistributed trajectories of typical points in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$.
- Orbits that are neither dense nor closed.

Finally we would like to point out — in a sense to be made precise in Sections 3.1 and 3.5 — that there is a correspondence between rational (or arithmetic) objects and closed A-orbits as in the first two types of A-orbit considered above (see Exercise 1.2.3 and 1.2.4).

† The eigenvalues of $\begin{pmatrix}1&1\\1&2\end{pmatrix}$ are $\frac{3\pm\sqrt{5}}{2}$, and there is such a matrix $g_{\mathrm{golden}}\in K$ because $\begin{pmatrix}1&1\\1&2\end{pmatrix}$ is symmetric.

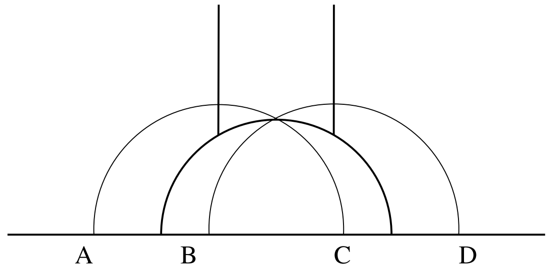

Fig. 1.4 The union of the two geodesics considered in 2 with both directions allowed is a periodic A-orbit, and comprises the orbit $\mathrm{SL}_2(\mathbb{Z})g_{\mathrm{golden}}A$.

1.2.3 The Horocycle Flow — the Subgroup $U^- = N$

We recall that the (stable) horocycle flow on 2 is defined by the action

$$h_s : x \longmapsto x\begin{pmatrix}1 & -s\\ & 1\end{pmatrix} = u(s)\bullet x$$

for $s \in \mathbb{R}$. Here the matrices

$$\begin{pmatrix}1 & s\\ & 1\end{pmatrix} = u(s)$$

are unipotent (that is, only have 1 as an eigenvalue) and the corresponding subgroup

$$U^- = \left\{\begin{pmatrix}1 & s\\ & 1\end{pmatrix} \mid s \in \mathbb{R}\right\}$$

is precisely the stable horospherical subgroup of the geodesic flow, in the sense that

$$U^- = \left\{g \in \mathrm{SL}_2(\mathbb{R}) \mid \begin{pmatrix}\mathrm{e}^{-t/2} & \\ & \mathrm{e}^{t/2}\end{pmatrix} g \begin{pmatrix}\mathrm{e}^{t/2} & \\ & \mathrm{e}^{-t/2}\end{pmatrix} \to I_2 \text{ as } t \to \infty\right\}.$$

This implies that

$$\mathsf{d}\left(g_t(x), g_t(u(s)\bullet x)\right) \to 0$$

as $t \to \infty$ for any $x \in 2$ and $s \in \mathbb{R}$, see Exercise 1.1.1.

Geometrically, the horocycle orbits $U^-\bullet x = xU^-$ can be described as circles touching the real axis with the arrows (that is, the tangent space component) normal to the circle pointing inwards or as horizontal lines with the arrows pointing upwards, as in Figure 1.5.

We recall that U^- also acts ergodically on 2 with respect to the Haar measure m_2 (see [**?**, Sec. 11.3] and Chapter 2). However, unlike the case of A-orbits, the classification of U^--orbits on 2 is shorter (we will discuss

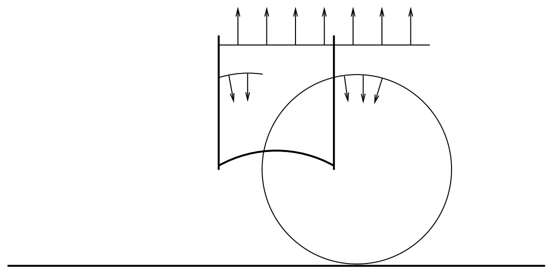

Fig. 1.5 The picture shows the two types of horocycle orbits; the orbits in 2 can again be understood by using the appropriate Möbius transformation whenever the orbit leaves the fundamental domain.

this phenomenon again, and in particular we will prove the facts below in Chapter 5 and more general results in Chapter 6). The possibilities are as follows:

- Compact trajectories, for example $\mathrm{SL}_2(\mathbb{Z})U^-$ is compact and corresponds to the horizontal orbit through $(\mathrm{i},\mathrm{i}) \in \mathrm{T}^1\mathbb{H}$.
- Dense trajectories, which are automatically also equidistributed with respect to m_2.

This gives the complete list of types of U^--orbits, and once more gives substance to the claim that there is a correspondence between rational objects and closed orbits (see Exercise 1.2.5).

1.2.4 The Subgroups K and B

For $\mathrm{SL}_2(\mathbb{R})$ there are two more connected subgroups of importance (and up to conjugation this completes the list of connected subgroups), namely

- $K = \mathrm{SO}(2) \subseteq \mathrm{SL}_2(\mathbb{R})$, and
- $B = U^-A = \left\{ \begin{pmatrix} a & s \\ & a^{-1} \end{pmatrix} \mid a > 0, s \in \mathbb{R} \right\}$

However, we note that for these two there is no correspondence between closed orbits and rational objects: for example, every K-orbit is compact since K itself is compact. On the other hand, every B-orbit is dense, independently of any rationality questions. In fact the latter follows from the properties of the horocycle flow. If xU^- is not periodic, then it is dense by the mentioned classification of U^--orbits in Section 1.2.3. If xU^- is periodic, then one can choose $a \in A$ so that xaU^- is a much longer periodic orbit. However, long periodic U^--orbits equidistribute in 2 (see Sarnak [**?**] and Section 5.3.1).

This shows that the phenomenon of a correspondence between closed orbits and rational objects is more subtle. It can only hold in certain situations, which we will discuss in Chapter 3, Chapter 4, and Chapter 7.

Exercises for Section 1.2

Exercise 1.2.1. Let E be as in Figure 1.2.

(1) Use $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ to show that $\mathrm{SL}_2(\mathbb{Z})\bullet E$ is 'uniformly open', meaning that there exists some $\delta > 0$ such that $z \in \mathrm{SL}_2(\mathbb{Z})\bullet E$ implies that

$$B_\delta(z) \subseteq \mathrm{SL}_2(\mathbb{Z})\bullet E.$$

Conclude that $\mathrm{SL}_2(\mathbb{Z})\bullet E = \mathbb{H}$.

(2) Suppose that both z and $\gamma\bullet z$ lie in E for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. Show that either $\gamma = \pm I$ or $z \in \partial E$.

(3) Conclude that E can be modified (by defining which parts of the boundary of E should be included) to become a fundamental domain.

Exercise 1.2.2. Show that $\mathrm{SL}_2(\mathbb{R})$ is generated by the unipotent subgroups

$$\begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & \\ * & 1 \end{pmatrix}.$$

Exercise 1.2.3. Show that $\mathrm{SL}_2(\mathbb{Z})gA$ is a divergent trajectory (that is, the map $A \ni a \mapsto \mathrm{SL}_2(\mathbb{Z})ga$ is a proper map) if and only if $ga \in \mathrm{SL}_2(\mathbb{Q})$ for some $a \in A$.

Exercise 1.2.4. Show that to any compact A-orbit in $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ one can attach a real quadratic number field K such that the length of the orbit is $\log|\xi|$, where ξ in O_K^* is a unit in the order O_K of K. Prove that there are only countably many such orbits.

Exercise 1.2.5. Show that $\mathrm{SL}_2(\mathbb{Z})gU^-$ is compact if and only if $g(\infty)$ lies in $\mathbb{Q} \cup \{\infty\}$. Show that if $\mathrm{SL}_2(\mathbb{Z})gU^-$ is compact, then any other compact orbit is of the form $\mathrm{SL}_2(\mathbb{Z})gaU^-$ for some $a \in A$.

Exercise 1.2.6. Show that $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R}) \cong \{\mathbb{Z}^2 g \mid g \in \mathrm{SL}_2(\mathbb{R})\}$ can be identified with lattices $\mathbb{Z}^2 g \subseteq \mathbb{R}^2$ of co-volume $\det g = 1$. Use the isomorphism with $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{T}^1\mathbb{H}$ discussed in this section to characterize compact subsets K of $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ in terms of elements of the lattices $\mathbb{Z}^2 g$ for $\mathrm{SL}_2(\mathbb{Z})g \in K$. More precisely, calculate the relationship between the shortest vector $ng \in \mathbb{Z}^2 g$ and the imaginary part of $g\mathrm{i} \in \mathbb{H}$ under the assumption that the representative $g \in \mathrm{SL}_2(\mathbb{R})$ has been chosen with $g\mathrm{i} \in E$.

1.3 The Space d of Lattices in $\mathbb{R}^d$

In this section we will introduce the most important locally homogeneous space for ergodic theory and its connections to number theory, namely

$$d = \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R}),$$

which gives rise to other arithmetical quotients by looking at orbits of subgroups of $\mathrm{SL}_d(\mathbb{R})$ on d. Such orbits will be discussed starting in Chapter 3.

1.3.1 Basic Definitions

A lattice in $\mathbb{R}^d$ in the sense of Definition 1.7 has the form $\Lambda = \mathbb{Z}^d g$ for some $g \in \mathrm{GL}_d(\mathbb{R})$ (see Exercise 1.3.1). A fundamental domain for Λ is given by the parallelepiped $[0,1)^d g$ which is spanned by the row vectors of g, and has Lebesgue measure $|\det g|$. This measure is also called the *covolume* (Λ) of Λ. A lattice $\Lambda \subseteq \mathbb{R}^d$ is called *unimodular* if the co-volume is 1. The space of all unimodular lattices in $\mathbb{R}^d$ — the *moduli space of lattices* — is therefore

$$d = \{\mathbb{Z}^d g \mid g \in \mathrm{SL}_d(\mathbb{R})\},$$

which is the orbit of $\mathbb{Z}^d$ under the right action of $\mathrm{SL}_d(\mathbb{R})$ on the subsets of $\mathbb{R}^d$: for $B \subseteq \mathbb{R}^d$ and $g \in \mathrm{SL}_d(\mathbb{R})$ the right action sends (g, B) to $Bg = \{vg : v \in B\}$. Notice that

$$\mathrm{Stab}_{\mathrm{SL}_d(\mathbb{R})}(\mathbb{Z}^d) = \mathrm{SL}_d(\mathbb{Z}),$$

so that†

$$d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$$

where $\mathrm{SL}_d(\mathbb{Z})g$ corresponds to the lattice $\mathbb{Z}^d g$. To understand d better, we need to develop a better understanding of lattices in $\mathbb{R}^d$.

1.3.2 Geometry of Numbers

The next result will be almost immediate from the abstract results in Section 1.3.1. It is a weak form of a classical result due to Minkowski in 1896 (see [**?**] for a modern reprinting).

Theorem 1.14 (Minkowski's first theorem). *If $\Lambda \subseteq \mathbb{R}^d$ is a lattice of co-volume V, then there exists a non-zero vector in Λ of length $\ll \sqrt[d]{V}$, with the implicit constant depending only on d.*

Recall that $f \ll g$ if there is a constant $C > 0$ with $f \leqslant Cg$, and $f \asymp g$ if $f \ll g$ and $g \ll f$; where the constant depends on other parameters these will appear as subscripts as, for example in the obvious bound

$$|\Lambda \cap B_1^{\mathbb{R}^d}(0)| \ll_\Lambda 1.$$

† We will think of this isomorphism in the following indeed always as an equality. In particular, the topology, the action of $G = \mathrm{SL}_d(\mathbb{R})$, and the Haar measure on d are as discussed in Section 1.1.

Since we will not be varying d throughout any of our discussions, we will not indicate dependencies on d in this way. We use this notation here as the particular value of the constants appearing in Theorems 1.14 and 1.15 will not be important for our purposes.

PROOF OF THEOREM 1.14. Choose $r_d > 0$ so that $B_{r_d}^{\mathbb{R}^d}(0)$ has Lebesgue measure 2 (any measure exceeding 1 will do). Then $\sqrt[d]{V} B_{r_d}^{\mathbb{R}^d}(0)$ has measure $2V$, and so cannot be an injective domain in the sense of Definition 1.2. It follows that there must exist $x_1 \neq x_2$ in $\sqrt[d]{V} B_{r_d}^{\mathbb{R}^d}(0)$ with $x_1 - x_2 = \lambda \in \Lambda \smallsetminus \{0\}$ of length $\|\lambda\| \leqslant 2 r_d \sqrt[d]{V}$. □

Barak says: we confuse the history here as Minkowski's minima are different — check Cassel's book and the bible on geometry of numbers by Gruber-Lek*** (difficult name)

Theorem 1.15 (Minkowski's successive minima). *Let $\Lambda \subseteq \mathbb{R}^d$ be a lattice. We define the successive minima $\lambda_1(\Lambda), \dots, \lambda_d(\Lambda)$ of Λ by*

$$\lambda_k(\Lambda) = \min\{r \mid \Lambda \text{ contains } k \text{ linearly independent vectors of norm} \leqslant r\}.$$

Then

$$\lambda_1(\Lambda) \cdots \lambda_d(\Lambda) \asymp (\Lambda).$$

Moreover, if†

$$\alpha_k(\Lambda) = \min\{(\Lambda \cap V) \mid V \subseteq \mathbb{R}^d \text{ is a subspace of rank } k\},$$

then

$$\alpha_k(\Lambda) \asymp \lambda_1(\Lambda) \cdots \lambda_k(\Lambda)$$

for $1 \leqslant k \leqslant d$.

For a subspace $V \subseteq \mathbb{R}^d$ there are two possibilities: either $V \cap \Lambda$ spans V or it does not. In the first case $\Lambda \cap V$ is a lattice in V, we say that V is Λ-*rational*, and the co-volume‡ $(\Lambda \cap V)$ of $\Lambda \cap V$ in V is finite. In the second case, we write $(\Lambda \cap V) = \infty$.

The proof of Theorem 1.15 is geometric, and relies on starting with a shortest vector (of size $\lambda_1(\Lambda)$) and then extending it with other vectors, chosen to be almost orthogonal to obtain a basis of $\mathbb{R}^d$.

PROOF OF THEOREM 1.15. We use induction on the dimension d. For $d = 1$ (and so also $k = 1$), it is clear that

$$\lambda_1(\Lambda) = \alpha_1(\Lambda) = (\Lambda).$$

Assume therefore that the theorem holds for $d - 1$, and let $\Lambda \subseteq \mathbb{R}^d$ be a lattice. It is clear by construction that

† See Exercise 1.3.2

‡ Strictly speaking we have to mention how we are normalizing the Haar measures of the different subspaces $V \subseteq \mathbb{R}^d$. However, we do this as one would expect: The Euclidean norm on $\mathbb{R}^d$ induces a Euclidean norm on V by restriction which in turn induces the Haar measure on V such that a unit cube in V has volume one.

$$\lambda_1(\Lambda) \leqslant \lambda_2(\Lambda) \leqslant \cdots \leqslant \lambda_d(\Lambda).$$

Pick a vector $v_1 \in \Lambda$ of length $\lambda_1(\Lambda)$, and define $W = (\mathbb{R}v_1)^\perp \subseteq \mathbb{R}^d$. Also let $\pi : \mathbb{R}^d \to W$ be the orthogonal projection along $\mathbb{R}v_1$ onto W. We claim that $\Lambda_W = \pi(\Lambda) \subseteq W$ is a discrete subgroup in W such that all of its nonzero vectors have length $\gg \lambda_1(\Lambda)$.

To see the claim, assume for the purpose of a contradiction that

$$w = \pi(v) \in \Lambda_W \smallsetminus \{0\}$$

has length less than $\frac{\sqrt{3}}{2}\|v_1\|$. Here $v = w + tv_1 \in \Lambda$ for some $t \in \mathbb{R}$, and we may assume (by replacing $v \in \Lambda$ with $v + nv_1 \in \Lambda$ for a suitable $n \in \mathbb{Z}$) that $t \in [-\frac{1}{2}, \frac{1}{2})$. However, since v_1 and w are orthogonal by construction, this implies that

$$\|v\|^2 = \|w\|^2 + t^2\|v_1\|^2 < \tfrac{3}{4}\|v_1\|^2 + \tfrac{1}{4}\|v_1\|^2 = \|v_1\|^2,$$

which contradicts the choice of v_1 as a non-zero vector in Λ of smallest length.

Next we claim that Λ_W is a lattice and that

$$\lambda_k(\Lambda_W) \asymp \lambda_{k+1}(\Lambda) \tag{1.7}$$

for $k = 1, \dots, d-1$. To see this, consider a fundamental domain F_W for Λ_W inside W. Then $F = [0,1)v_1 + F_W$ is a fundamental domain for Λ, and we get

$$(\Lambda) = \lambda_1(\Lambda)(\Lambda_W). \tag{1.8}$$

This shows that Λ_W is a lattice in W. Now assume that $v_1, v_2, \dots, v_{k+1} \in \Lambda$ are linearly independent and of length no more than $\lambda_{k+1}(\Lambda)$, so that

$$\pi(v_2), \dots, \pi(v_{k+1}) \in \Lambda_W$$

are linearly independent and also have length no more than $\lambda_{k+1}(\Lambda)$. Hence

$$\lambda_k(\Lambda_W) \leqslant \lambda_{k+1}(\Lambda)$$

for any $k = 1, \dots, d-1$. On the other hand, assume that

$$w_1 = \pi(v_2), \dots, w_k = \pi(v_{k+1}) \in \Lambda_W$$

are linearly independent of length no more than $\lambda_k(\Lambda_W)$. As above, we may assume $v_{j+1} = w_j + t_j v_1 \in \Lambda$ with $t_j \in [-\frac{1}{2}, \frac{1}{2})$, and so

$$\|v_{j+1}\| \ll \lambda_k(\Lambda_W) + \lambda_1(\Lambda) \ll \lambda_k(\Lambda_W),$$

since $\lambda_1(\Lambda) \ll \lambda_1(\Lambda_W) \leqslant \lambda_k(\Lambda_W)$.

By the inductive assumption and the statement above, we get that

$$\operatorname{vol}(\Lambda_W) \asymp \lambda_1(\Lambda_W) \cdots \lambda_{d-1}(\Lambda_W) \asymp \lambda_2(\Lambda) \cdots \lambda_d(\Lambda).$$

Together with (1.8) this gives $\operatorname{vol}(\Lambda) \asymp \lambda_1(\Lambda) \cdots \lambda_d(\Lambda)$ as claimed in the theorem.

To see the last statement in the theorem, we proceed similarly. If $v_j \in \Lambda$ has norm $\lambda_j(\Lambda)$ for $j = 1, \dots, k$, $v_1, \dots, v_k$ are linearly independent (over $\mathbb{R}$), and $V = \mathbb{R}v_1 + \cdots + \mathbb{R}v_k$ then†

$$\operatorname{vol}(\Lambda \cap V) \leqslant \operatorname{vol}(\mathbb{Z}v_1 + \cdots + \mathbb{Z}v_k) \leqslant \|v_1\| \cdots \|v_k\| = \lambda_1(\Lambda) \cdots \lambda_k(\Lambda),$$

and so $\alpha_k(\Lambda) \leqslant \lambda_1(\Lambda) \cdots \lambda_k(\Lambda)$. On the other hand, if $V \subseteq \mathbb{R}^n$ has dimension k and is Λ-rational, then we may apply the above to the lattice $\Lambda \cap V$ in V to get

$$\operatorname{vol}(\Lambda \cap V) \asymp \lambda_1(\Lambda \cap V) \cdots \lambda_k(\Lambda \cap V) \geqslant \lambda_1(\Lambda) \cdots \lambda_k(\Lambda),$$

which shows that $\alpha_k(\Lambda) \gg \lambda_1(\Lambda) \cdots \lambda_k(\Lambda)$ and proves the theorem. □

Using the same inductive argument (by projection to the orthogonal complement of the shortest vector) we also get the following.

Corollary 1.16 (Basis of a lattice). *Let $\Lambda \subseteq \mathbb{R}^d$ be a lattice. Then there is a $\mathbb{Z}$-basis $v_1, \dots, v_d \in \Lambda$ of Λ such that*

$$\|v_1\| = \lambda_1(\Lambda), \|v_2\| \asymp \lambda_2(\Lambda), \dots, \|v_d\| \asymp \lambda_d(\Lambda).$$

Moreover, the projection $\pi_k(v_k)$ of v_k onto the orthogonal complement of

$$\mathbb{R}v_1 + \cdots + \mathbb{R}v_{k-1}$$

has

$$\|\pi_k(v_k)\| \asymp \lambda_k(\Lambda) \asymp \|v_k\|$$

for $k = 2, \dots, d$

Corollary 1.16 may seem obvious, but our intuition about lattices does not extend to higher dimensions without some additional complexities. In particular, it is not true that there always exists a $\mathbb{Z}$-basis $v_1, \dots, v_d$ for a lattice with

$$\|v_1\| = \lambda_1(\Lambda), \|v_2\| = \lambda_2(\Lambda), \dots, \|v_d\| = \lambda_d(\Lambda),$$

see Exercise 1.3.3 for a simple counterexample.

† The first inequality holds as $\Lambda \cap V$ may have more lattice elements than

$$\mathbb{Z}v_1 + \cdots + \mathbb{Z}v_k \subseteq \Lambda \cap V,$$

and the second follows as the volume of a parallelepiped is less than the product of the lengths of its sides (or, more formally, from Proposition 1.19).

Proof of Corollary 1.16. Assume the corollary for dimension $(d-1)$, and define $W = (\mathbb{R}v_1)^{\perp}$, $\pi = \pi_1$, and $\Lambda_W = \pi(\Lambda)$ as in the proof of Theorem 1.15. Recall that these assumptions lead to (1.7). By assumption, Λ_W has a $\mathbb{Z}$-basis $w_1 = \pi(v_2), \dots, w_{d-1} = \pi(v_d)$ satisfying all the claims. Once more we may assume that $v_k = w_{k-1} + t_k v_1 \in \Lambda$ with $t_k \in [-\frac{1}{2}, \frac{1}{2})$ so that $\|v_k\| \ll \lambda_k(\Lambda)$ as in the proof of Theorem 1.15. It follows that $v_1, \dots, v_d \in \Lambda$ is a $\mathbb{Z}$-basis of Λ with $\|v_1\| = \lambda_1(\Lambda)$, and $\|v_k\| \asymp \lambda_k(\Lambda)$ for $k = 2, \dots, d$.

For the last claim in the corollary, recall that we already showed that

$$\|v_2\| \asymp \|w_1\| \asymp \lambda_2(\Lambda),$$

which is the claim for $k = 2$. For $k > 2$, notice that $\pi_k \pi = \pi_k$ is (when restricted to W) also the orthogonal projection $\pi_{W,k-1}$ in W onto the orthogonal complement of $\mathbb{R}w_1 + \cdots + \mathbb{R}w_{k-2}$. Therefore, the inductive assumption applies to give

$$\|\pi_k(v_k)\| = \|\pi_{W,k-1}(w_{k-1})\| \asymp \lambda_{k-1}(\Lambda_W) \asymp \lambda_k(\Lambda) \asymp \|v_k\|,$$

which proves the corollary. □

1.3.3 Mahler's Compactness Criterion

The space $d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$ cannot be compact for $d \geqslant 2$, since d is the space of unimodular lattices, and it is possible to degenerate a sequence of lattices. For example, the sequence of unimodular lattices (Λ_n) defined by

$$\Lambda_n = (\tfrac{1}{n}\mathbb{Z}) \times (n\mathbb{Z}) \times \mathbb{Z}^{d-2}$$

has no subsequence converging to a unimodular lattice. Indeed, if we were to assign a limit to this sequence, then we could only have

$$\Lambda_n \to \mathbb{R} \times \{0\} \times \mathbb{Z}^{d-2}$$

as $n \to \infty$, so the putative 'limit' is not discrete and does not span $\mathbb{R}^d$.

More generally, any sequence (Λ_n) of unimodular lattices containing vectors with length converging to 0 (that is, with $\lambda_1(\Lambda_n) \to 0$ as $n \to \infty$) cannot converge in d. To see this concretely, suppose that $\Lambda_n = \mathbb{Z}^d g_n \to \mathbb{Z}^d g$. Then (after replacing g_n with $\gamma_n g_n$ for a suitable choice of $\gamma_n \in \mathrm{SL}_d(\mathbb{Z})$ if necessary) we can assume that $g_n \to g$ as $n \to \infty$ in the topology of $\mathrm{SL}_d(\mathbb{R})$ (cf. (1.1) on page 8). Thus we can write $g_n = g h_n$ with $h_n \to I_d$ as $n \to \infty$, which implies that $\lambda_1(\mathbb{Z}^d g_n) \to \lambda_1(\mathbb{Z}^d g) > 0$ (see Exercise 1.3.4).

A reasonable guess is that the argument above is the only way in which the non-compactness of d comes about (that is, a sequence (Λ_n) of lattices with no convergent subsequence has $\lambda_1(\Lambda_n) \to 0$ as $n \to \infty$; equivalently any closed subset of d on which λ_1 has a positive lower bound — a 'uniformly discrete' set of lattices — is pre-compact).

Theorem 1.17 (Mahler's compactness criterion). *A subset $B \subseteq d$ has compact closure if and only if there exists some $\delta > 0$ for which*

$$\Lambda \in B \implies \lambda_1(\Lambda) \geqslant \delta. \tag{1.9}$$

That is, B is compact if and only if it is closed and uniformly discrete.

Because of this result, it will be convenient to define the subset

$$d(\delta) = \{\Lambda \in d \mid \lambda_1(\Lambda) \geqslant \delta\}$$

for any $\delta > 0$. The condition in (1.9) will also be described by saying that elements of B do not contain any non-trivial δ-short vectors. An equivalent formulation of Theorem 1.17 is to say that a set $B \subseteq d$ of unimodular lattices is compact if and only if it is closed and the *height* function defined by

$$(\Lambda) = \frac{1}{\lambda_1(\Lambda)}$$

is bounded on B. Even though it is difficult to depict d on paper (for example, 3 is topologically an 8-dimensional space), it is conventionally depicted as in Figure 1.6, in part to express the meaning of Theorem 1.17.

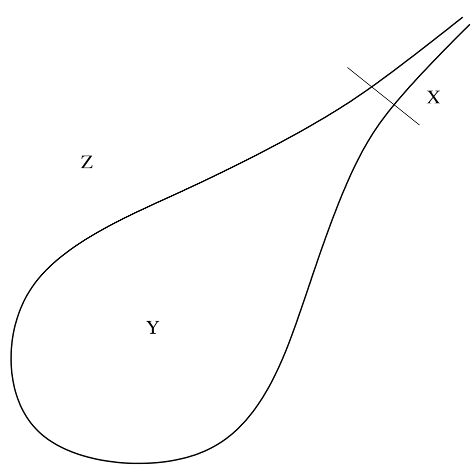

Fig. 1.6 A compact subset of d is contained in $d(\delta) = \{\Lambda \in d \mid \lambda_1(\Lambda) \geqslant \delta\}$ for some $\delta > 0$. The non-compact part $d \smallsetminus d(\delta)$, loosely referred to as a *cusp*, is depicted as a thin set to indicate the finite total volume. For $d > 2$ the geometry of the cusp is much more complicated than the cusp in the $d = 2$ case.

Proof of Theorem 1.17. We have already mentioned that λ_1 is a continuous function on d. Since λ_1 only achieves positive values, it follows that a compact subset of d must lie in $d(\delta)$ for some $\delta > 0$. It remains to prove that $d(\delta)$ is itself compact. Let $\left(\mathbb{Z}^d g_n\right)$ in $d(\delta)$ be any sequence. Then, by Corollary 1.16, the lattice $\mathbb{Z}^d g_n$ has a basis $v_1^{(n)}, \dots, v_d^{(n)}$ with

$$\delta \leqslant \lambda_1(\mathbb{Z}^d g_n) = \|v_1^{(n)}\| \ll \|v_2^{(n)}\| \ll \cdots \ll \|v_d^{(n)}\|$$

and

$$\|v_1^{(n)}\| \cdots \|v_d^{(n)}\| \ll 1,$$

which implies that

$$\|v_i^{(n)}\| \ll \delta^{-(d-1)}$$

for $i = 1, \dots, d$. This means that for some $\gamma_n \in \mathrm{SL}_d(\mathbb{Z})$, the entries of the matrix $\gamma_n g_n$ are all $\ll \delta^{-(d-1)}$. Thus there is a convergent subsequence

$$\gamma_{n_i} g_{n_i} \to g$$

as $i \to \infty$, so that $\mathrm{SL}_d(\mathbb{Z}) g_{n_i} \to \mathrm{SL}_d(\mathbb{Z}) g$ as required. □

1.3.4 d has Finite Volume

Write π for the canonical quotient map $\pi : \mathrm{SL}_d(\mathbb{R}) \to d$.

Theorem 1.18 (d has finite volume). $\mathrm{SL}_d(\mathbb{Z})$ *is a lattice in* $\mathrm{SL}_d(\mathbb{R})$.

We will prove the theorem by showing that Corollary 1.16 gives a surjective set of finite Haar measure — that is, a measurable set $F \subseteq \mathrm{SL}_d(\mathbb{R})$ (called a *Siegel domain*) with $\pi(F) = d$ and

$$m_{\mathrm{SL}_d(\mathbb{R})}(F) < \infty.$$

The fact that $m_{\mathrm{SL}_d(\mathbb{R})}(F)$ is finite is essentially a calculation, but is considerably helped by the Iwasawa decomposition†.

Proposition 1.19 (Iwasawa decomposition). *Let* $K = \mathrm{SO}(d)$ *and*‡

$$B = UA = \left\{ \begin{pmatrix} a_1 & & & \\ * & a_2 & & \\ \vdots & \vdots & \ddots & \\ * & * & \cdots & a_d \end{pmatrix} \mid a_1, \dots, a_d > 0, a_1 \cdots a_d = 1 \right\},$$

† This is also referred to as the NAK decomposition.

‡ We sometimes indicate by $*$ any entry of a matrix which is only restricted to be a real number, and do not write entries that are zero.

where

$$U = N = \left\{ \begin{pmatrix} 1 & & & \\ u_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ u_{d1} & u_{d2} & \cdots & 1 \end{pmatrix} \right\}$$

and

$$A = \left\{ \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix} \mid a_1, \dots, a_d > 0, a_1 \cdots a_d = 1 \right\}.$$

Then $\mathrm{SL}_d(\mathbb{R}) = BK = UAK$ *in the sense that for every* $g \in \mathrm{SL}_d(\mathbb{R})$ *there are unique matrices* $u \in U$, $a \in A$, $k \in K$ *with* $g = uak$.

Proof. This is the Gram–Schmidt procedure[(3)] in disguise. Let

$$g = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix},$$

where $w_1, \dots, w_d \in \mathbb{R}^d$ are the row vectors of g. We apply the Gram–Schmidt procedure to define

$$w_1' = \frac{1}{a_1} w_1$$

with $a_1 = \|w_1\| > 0$,

$$w_2^{(1)} = u_{21} w_1 + w_2$$

with $u_{21} \in \mathbb{R}$ such that $w_2^{(1)} \perp w_1$, and

$$w_2' = \frac{1}{a_2} w_2^{(1)}$$

with $a_2 = \|w_2^{(1)}\| > 0$ (by linear independence of w_1 and w_2). We continue this until

$$w_d^{(1)} = u_{d1} w_1 + u_{d2} w_2 + \cdots + w_d$$

with $u_{d1}, u_{d2}, \dots, u_{d(d-1)} \in \mathbb{R}$ such that

$$w_d^{(1)} \perp w_1, \dots, w_{d-1}$$

(or, equivalently, $w_d^{(1)} \perp w_1', \dots, w_{d-1}'$) and

$$w_d' = \frac{1}{a_d} w_d^{(1)}$$

with $a_d = \|w_d^{(1)}\| > 0$ (again by linear independence). This has the following effect. If

$$u = \begin{pmatrix} 1 & & & \\ u_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ u_{d1} & u_{d2} & \cdots & 1 \end{pmatrix}$$

and

$$a = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix}$$

then

$$ug = \begin{pmatrix} w_1 \\ w_2^{(1)} \\ \vdots \\ w_d^{(1)} \end{pmatrix}, a^{-1}ug = \begin{pmatrix} w_1' \\ \vdots \\ w_d' \end{pmatrix} = k.$$

By construction k has orthogonal rows, so that $\det(k) = \pm 1$. However,

$$\det(g) = 1 = \det(u)$$

and $\det(a) > 0$ which gives $\det(a) = 1 = \det(k)$. This shows the existence of the claimed $u \in U, a \in A$, and $k \in K$ with $g = u^{-1}ak$.

To see that this decomposition is unique, notice that B is a subgroup with $B \cap K = \{I_d\}$ so that $b_1k_1 = b_2k_2$ implies $b_2^{-1}b_1 = k_2k_1^{-1} = I_d$. Similarly, $A \cap U = \{I_d\}$, and the proposition follows. □

Our geometric arguments in Corollary 1.16 are closely related to the Gram–Schmidt procedure used in Proposition 1.19. Combining these gives the next result.

Definition 1.20 (Siegel domain for d). A set of the form

$$\Sigma_{s,t} = U_s A_t K$$

where $s > 0$, $t > 0$,

$$U_s = \left\{ \begin{pmatrix} 1 & & & \\ u_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ u_{d1} & u_{d2} & \cdots & 1 \end{pmatrix} \mid |u_{ij}| \leqslant s \right\},$$

and

$$A_t = \left\{ \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix} \mid \tfrac{a_{i+1}}{a_i} \geqslant t \text{ for } i = 1, \dots, d-1 \right\},$$

is called a *Siegel domain*.

We note that U_s is a compact subset of the lower unipotent subgroup but A_t is a non-compact subset of the diagonal subgroup.

Corollary 1.21 (Surjectivity of Siegel domains). *There exists some t_0*[†] *such that for $t \leqslant t_0$ and $s \geqslant \frac{1}{2}$ the Siegel domain $\Sigma_{s,t}$ is surjective (that is, $\pi(\Sigma_{s,t}) = d$).*

Proof. Let $\Lambda \in d$ be a unimodular lattice, and let $w_1, \dots, w_d$ be the $\mathbb{Z}$-basis as in Corollary 1.16. Replacing w_d by $-w_d$ if necessary, we may assume that $\det(g) = 1$, where

$$g = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}.$$

Now apply the Gram–Schmidt procedure as in the proof of Proposition 1.19 to g. By Corollary 1.16 we get

$$\begin{aligned} a_1 &= \|w_1\| = \lambda_1(\Lambda) \\ a_2 &= \|w_2^{(1)}\| \asymp \lambda_2(\Lambda) \\ &\vdots \\ a_d &= \|w_d^{(1)}\| \asymp \lambda_d(\Lambda) \end{aligned}$$

which satisfy

$$\frac{a_{i+1}}{a_i} \gg \frac{\lambda_{i+1}(\Lambda)}{\lambda_i(\Lambda)} \geqslant 1$$

for $i = 1, \dots, d-1$. Choosing t_0 and $t \leqslant t_0$ accordingly gives

$$a = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix} \in A_t.$$

Therefore $\Lambda = \mathbb{Z}^d g$ and $g = uak$ with $u \in U$ and $k \in K$. Notice that by replacing g by $u_{\mathbb{Z}} g$ with $u_{\mathbb{Z}} \in U(\mathbb{Z}) = U \cap \mathrm{Mat}_d(\mathbb{Z})$ we can easily ensure that $u_{(i+1)i} \in [-\frac{1}{2}, \frac{1}{2})$. Having achieved this we may use another $u_{\mathbb{Z}} \in U(\mathbb{Z})$ with $(u_{\mathbb{Z}})_{(i+1)i} = 0$ for $i = 1, \dots, d-1$, which makes it easy to calculate the next off-diagonal of $u_{\mathbb{Z}} u$ as follows:

$$\begin{aligned} (u_{\mathbb{Z}} u)_{(i+2)i} &= (u_{\mathbb{Z}})_{(i+2)i} + (u_{\mathbb{Z}})_{(i+2)(i+1)} u_{(i+1)i} + u_{(i+2)i} \\ &= (u_{\mathbb{Z}})_{(i+2)i} + 0 + u_{(i+2)i} \end{aligned}$$

[†] A more careful analysis of the proof shows that $t_0 = \frac{\sqrt{3}}{2}$ suffices in any dimension; see also Exercise 1.3.8 which can also be used to prove this claim.

for any $i = 1, \ldots, d-2$. Therefore, we can modify u by some $u_{\mathbb{Z}}$ as above to ensure that $u_{(i+2)i}$ lies in $[-\frac{1}{2}, \frac{1}{2})$ for $i = 1, \ldots, d-2$. Proceeding by induction gives

$$\Lambda = \mathbb{Z}^d g = \mathbb{Z}^d uak$$

for some $u \in U_{1/2}$, $a \in A_t$, and $k \in K$. $\square$

It remains to show that the Haar measure of the Siegel domains is finite. For this the Iwasawa decomposition also helps us to understand the Haar measure $m_{\mathrm{SL}_d(\mathbb{R})}$ as a result of the following general fact about locally compact groups.

Lemma 1.22 (Decomposition of Haar measure). *Let G be a unimodular, metric, σ-compact, locally compact group. Let $S, T \subseteq G$ be closed subgroups with $S \cap T = \{I\}$ and with the property that $m_G(ST) > 0$ (for example, because ST contains an open neighborhood of I). Then*

$$m_G|_{ST} \propto \phi_* \left(m_S \times m_T^{(r)} \right),$$

where $\phi : S \times T \to G$ is the product map $\phi : (s,t) \mapsto st$.

We refer to [**?**, Lemma 11.31] and Knapp [**?**] for the proof. The above lemma is useful for us because of the following.

Lemma 1.23. $\mathrm{SL}_d(\mathbb{R})$ *is unimodular.*

As an alternative to Exercise 1.1.6 (which is quite special but gives the above lemma) we start with a general lemma about the structure of $\mathrm{SL}_d(\mathbb{K})$ over any field $\mathbb{K}$, generalizing Exercise 1.2.2.

Lemma 1.24 (Unipotent Generation). *Over any field $\mathbb{K}$, the group $\mathrm{SL}_d(\mathbb{K})$ is generated by the elementary unipotent subgroups*

$$U_{ij}(\mathbb{K}) = \{u_{ij}(t) = I + tE_{ij} \mid t \in \mathbb{K}\}$$

with $i \neq j$ and E_{ij} being the elementary matrix with (i,j)th entry 1 and all other entries 0.

For $\mathbb{K} = \mathbb{R}$ (and for $\mathbb{K} = \mathbb{C}$), this implies that $\mathrm{SL}_d(\mathbb{R})$ (and $\mathrm{SL}_d(\mathbb{C})$) are connected as topological spaces, because each subgroup $U_{ij}(\mathbb{R})$ and $U_{ij}(\mathbb{C})$ is connected. In particular, this shows that $\mathrm{SL}_d(\mathbb{R})$ carries a left-invariant Riemannian metric, and by restriction of this metric to any closed subgroup of $\mathrm{SL}_d(\mathbb{R})$ (which may be connected or not) one has a left-invariant metric on the subgroup (which induces the locally compact, σ-compact, induced topology).

Outline proof of Lemma 1.24. Notice that the row (and column) operation of adding t times the jth row to the ith row (or t times the ith column to the jth column) corresponds to multiplication by the elements $u_{ij}(t) \in U_{ij}(\mathbb{K})$

on the left (resp. right) of a given matrix $g \in \mathrm{SL}_d(\mathbb{K})$. This restricted Gaussian elimination can be used to reduce the matrix g to the identity. To do this we may first ensure that $g_{12} \neq 0$ with a suitable row operation, then use another row operation to ensure that $g_{11} = 1$. Then suitable row and column operations can be used to obtain $g_{1i} = 0 = g_{i1}$ for $i > 1$, and we may then continue by induction. At the last step the fact that $\det(g) = 1$ is needed to ensure that the diagonal matrix produced is in fact the identity. This can be used to express g as a finite product of elementary unipotent matrices. □

Proof of Lemma 1.23. Recall the unipotent subgroups

$$U_{ij} = \{u_{ij}(t) = I + tE_{ij} \mid t \in \mathbb{R}\}$$

for $i \neq j$ from Lemma 1.24. Let $a \in A$ be any diagonal matrix, and notice that $au_{ij}(t)a^{-1} = u_{ij}(\frac{a_i}{a_j}t)$ for $t \in \mathbb{R}$. Therefore, the commutator satisfies

$$[a, u_{ij}(t)] = a^{-1}u_{ij}(-t)au_{ij}(t) = u_{ij}((1 - \tfrac{a_j}{a_i})t).$$

Choosing $a \in A$ correctly, it follows that the commutator group

$$[\mathrm{SL}_d(\mathbb{R}), \mathrm{SL}_d(\mathbb{R})]$$

contains U_{ij} for all $i \neq j$. By Lemma 1.24 it follows that

$$[\mathrm{SL}_d(\mathbb{R}), \mathrm{SL}_d(\mathbb{R})] = \mathrm{SL}_d(\mathbb{R}).$$

Since the modular character $\mathrm{mod} : \mathrm{SL}_d(\mathbb{R}) \to \mathbb{R}_{>0}$ is a homomorphism to an abelian group it follows that $\mathrm{mod}(\mathrm{SL}_d(\mathbb{R})) = \{1\}$, proving the lemma. □

To complete the proof of Theorem 1.18, it remains to show the following lemma.

Lemma 1.25. *For any $s > 0$ and $t > 0$, we have $m_{\mathrm{SL}_d(\mathbb{R})}(\Sigma_{s,t}) < \infty$.*

Proof. Using Lemma 1.22 for $G = \mathrm{SL}_d(\mathbb{R})$, $S = B$, and $T = K$ we see that K can be ignored and we have to calculate $m_B(U_s A_t)$ (where as usual m_B denotes the left Haar measure on B). Note that $B = UA$ is not unimodular so that we cannot apply Lemma 1.22 again (indeed, applying it erroneously would not give the desired result). On the other hand, U and A are unimodular (see Exercise 1.3.6). Furthermore, the left Haar measure on B is given by a density function $\rho(a)$ with respect to $m_U \times m_A$ (using the coordinate system arising from $B = UA$). In fact

$$\mathrm{d}m_B \propto \rho(a)\,\mathrm{d}m_U \times \mathrm{d}m_A, \tag{1.10}$$

where

$$\rho\left(\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix}\right) = \prod_{i>j}\left(\frac{a_j}{a_i}\right).$$

Using the fact that the Haar measure on U is simply the Lebesgue measure (in the coordinate system implied by the way we write down these matrices) and that A normalizes U, the relation in (1.10) can be checked directly (see Exercise 1.3.7).

Using this, we get

$$m_B(U_s A_t) \ll \underbrace{m_U(U_s)}_{<\infty} \int_{A_t} \rho(a)\,\mathrm{d}m_A(a),$$

and so the problem is reduced to the integral over A_t.

Using the relations

$$\frac{a_j}{a_i} = \frac{a_j}{a_{j+1}} \cdots \frac{a_{i-1}}{a_i} = \prod_{k=j}^{i-1} \frac{a_k}{a_{k+1}}$$

for $i > j$, we also obtain the formula

$$\rho\left(\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix}\right) = \prod_{k=1}^{d-1}\left(\frac{a_k}{a_{k+1}}\right)^{r_k} = \prod_{k=1}^{d-1}\left(\frac{a_{k+1}}{a_k}\right)^{-r_k}$$

for some integers $r_k > 0$ (here $r_k = (d-k)k$ equals the number of tuples of indices (i,j) with $j \leqslant k < i$, but the exact form of r_k does not matter).

Next notice that

$$A \ni a = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix} \longmapsto (y_1, \dots, y_{d-1}) = \left(\log \tfrac{a_2}{a_1}, \dots, \log \tfrac{a_d}{a_{d-1}}\right) \in \mathbb{R}^{d-1}$$

is an isomorphism of topological groups which maps A_t to $[\log t, \infty)^{d-1}$, so that[†]

$$\int_{A_t} \rho(a)\,\mathrm{d}m_A(a) \propto \prod_{k=1}^{d-1} \int_{\log t}^{\infty} \mathrm{e}^{-r_k y_k}\,\mathrm{d}y_k < \infty$$

as claimed. □

The proof presented above is usually referred to as the *reduction theory* of SL_d, and this generalizes to other algebraic groups by a theorem of Borel

[†] The symbol $\propto$ denotes proportionality, and here the constant of proportionality depends on the choices of the Haar measures on A and on $\mathbb{R}^{d-1}$.

and Harish–Chandra [?] (see Siegel [?]). In Chapter 4 we will give a second proof which will also lead to the general result for other groups in Chapter 7.

Exercises for Section 1.3

Exercise 1.3.1. Check that any lattice in $\mathbb{R}^d$ (in the sense of Definition 1.7) is indeed of the form $\mathbb{Z}^d g$ for some $g \in \mathrm{GL}_d(\mathbb{R})$.

Exercise 1.3.2. Show that the minimum in the definition of $\alpha_k(\Lambda)$ in Theorem 1.15 is indeed achieved.

Exercise 1.3.3. Let $d \geqslant 5$. Let $\Lambda = \mathbb{Z}^{d-1} \times \{0\} + \mathbb{Z}v$ where $v = (\frac{1}{2}, \dots, \frac{1}{2})$. Show that $\lambda_1 = \dots = \lambda_d = 1$, $(\Lambda) = \frac{1}{2}$, and that there does not exist a basis of Λ consisting of vectors of length 1.

Exercise 1.3.4. (1) Show that $\lambda_1(\mathbb{Z}^d gh) \leqslant \lambda_1(\mathbb{Z}^d g)\|h\|$ for $g, h \in \mathrm{GL}_d(\mathbb{R})$, where $\|\cdot\|$ denotes the operator norm.
(2) Conclude that $\lambda_1 : d \to (0, \infty)$ is continuous.
(3) Generalize (2) to λ_k for $1 \leqslant k < d$.

Exercise 1.3.5. Can Mahler's compactness criterion also be phrased in terms of λ_d, or in terms of λ_j for $2 \leqslant j < d$?

Exercise 1.3.6. Prove that U and A are unimodular (and describe their Haar measures).

Exercise 1.3.7. Let $B = UA$, m_B, m_U, m_A, and ρ be as in the proof of Lemma 1.25. Let $f \geqslant 0$ be any measurable function on B, and fix some $b \in B$. Using Fubini's theorem and substitution prove that

$$\int_B f(bua)\rho(a)\,\mathrm{d}m_U(u)\,\mathrm{d}m_A(a) = \int_B f(ua)\rho(a)\,\mathrm{d}m_U(u)\,\mathrm{d}m_A(a),$$

first for $b = u_0 \in U$ and then for $b = \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_d \end{pmatrix} \in A$. Deduce that (1.10) holds.

Exercise 1.3.8. In this exercise a different proof of Corollary 1.21 will be given (which will not use Minkowski's theorem on successive minimas)[(4)]. For this let $v_1, \dots, v_d$ be an ordered basis of a unimodular lattice $\Lambda < \mathbb{R}^d$. For every $i = 1, \dots, d$ define v_i^* to be the projection of v_i onto the orthogonal complement of the linear span of $v_1, \dots, v_{i-1}$. Recall that $\|v_i^*\|$ is the ith diagonal entry of the A-component of the NAK-decomposition of the matrix g whose rows consist of $v_1, \dots, v_d$. We may assume that $\det g = 1$.

The basis is called *semi-reduced* if all linear coefficients of $v_i - v_i^*$, when expressed as a linear combination of $v_1, \dots, v_{i-1}$, are in $[-\frac{1}{2}, \frac{1}{2})$ (that is, the N-part of g in the NAK-decomposition belongs to $U_{\frac{1}{2}}$).

The basis is called *t-reduced* (for some fixed $t > 0$) if it is semi-reduced and if $\frac{\|v_{i+1}^*\|}{\|v_i^*\|} \geqslant t$ for $i = 1, \dots, d-1$ (that is, the A-part of g in the NAK-decomposition belongs to A_t).

Prove that the following algorithm terminates for every fixed $t < \frac{\sqrt{3}}{2}$ with a t-reduced ordered basis of Λ.

(1) Check if the ordered basis is semi-reduced. If not perform a simple change of basis (using only a change of basis in $N \cap \mathrm{SL}_d(\mathbb{Z})$) and produce a new ordered basis which is semi-reduced.
(2) Check if the basis is t-reduced. If so, the algorithm terminates.
(3) So assume that the ordered basis is not t-reduced but is semi-reduced. Then there exists a smallest i for which $\frac{\|v_{i+1}^*\|}{\|v_i^*\|} > t$. Now replace the basis with the new basis where the order of v_i and v_{i+1} is reversed (but all other basis elements retain their place), and start the algorithm from the beginning.

For the proof you may find useful the function θ of the ordered basis defined by

$$\theta(v_1, \dots, v_d) = \prod_{i=1}^{d} (\mathbb{Z}v_1 + \cdots \mathbb{Z}v_i).$$

Exercise 1.3.9. For any $f \in C_c(\mathbb{R}^d)$ we define the Siegel transform at $x \in d$ by

$$S_f(x) = \sum_{v \in \Lambda_x \smallsetminus \{0\}} f(v),$$

where $\Lambda_x = \mathbb{Z}^d g$ denotes the lattice corresponding to $x = \mathrm{SL}_d(\mathbb{R})g$. In this exercise we wish to show that there exists some $c > 0$ (depending on the choice of Haar measures) such that $\int_d S_f \,\mathrm{d}m_d = c \int_{\mathbb{R}^d} f(t)\,\mathrm{d}t$ for all $f \in C_c(\mathbb{R}^d)$.

(1) Show that $\int_d S_f \,\mathrm{d}m_d < \infty$.
(2) Show that the positive measure μ on $\mathbb{R}^d$ defined by Riesz representation and the functional $f \mapsto \int_d S_f \,\mathrm{d}m_d$ satisfies $\mu(\{0\}) = 0$.
(3) Show that μ is $\mathrm{SL}_d(\mathbb{R})$-invariant and conclude the claim.

Notes to Chapter 1

(1)(Page 19) In fact any perfect Polish space allows an embedding of the middle-third Cantor set into it, so in particular such a space has the cardinality of the continuum. We refer to Kechris [**?**, Sec. 6.A].

(2)(Page 22) This is a simple instance of the more general Iwasawa decomposition of a connected real semi-simple Lie group [**?**] (see also [**?**]).

(3)(Page 36) This method was presented by E. Schmidt [**?**, Sec. 3, p. 442], and he pointed out that essentially the same method was used earlier by Gram [**?**]; the modern view is that the methods differ, and that the Gram form was used earlier by Laplace [**?**, p. 497ff.] in a different setting.

(4)(Page 42) This is based on the so-called LLL algorithm of A. K. Lenstra, H. W. Lenstra, Jr., and Lovász [**?**].

Chapter 2

Ergodicity and Mixing on Locally Homogeneous Spaces

Throughout, we will assume that an acting group G is σ-compact, locally compact, and metrizable. Moreover, we will assume that X, the space G acts on, is a σ-compact locally compact metric space, and that the action is jointly continuous (see [**?**, Def. 8.1]). Such an action is said to be

- *measure-preserving* with respect to a probability measure μ on X if
$$\mu(g^{-1}\cdot B) = \mu(B)$$
for any $g \in G$ and measurable set $B \subseteq X$, in which case we say that μ is invariant;
- *ergodic* with respect to a probability measure μ if any measurable $B \subseteq X$ with the property $\mu(g^{-1}\cdot B \triangle B) = 0$ for all $g \in G$ has $\mu(B) \in \{0, 1\}$; and
- *mixing* with respect to a probability measure μ if
$$\mu(g^{-1}\cdot A \cap B) \longrightarrow \mu(A)\mu(B)$$
as $g \to \infty$ in G for any measurable sets $A, B \subseteq X$.

Here the notation $g \to \infty$ is shorthand for elements g of G running through a sequence $(g_n)_{n\geqslant 1}$ with the property that for any compact set $K \subseteq G$ there is an $N = N(K)$ such that $n \geqslant N(K)$ implies $g_n \notin K$. Notice that the property of mixing (of non-compact groups) is much stronger than ergodicity in the following sense. Mixing for the action implies that each element $g \in G$ with $g^n \to \infty$ as $n \to \infty$ is itself a mixing (and ergodic) transformation in the usual sense (where the acting group is a copy of $\mathbb{Z}$), while ergodicity *a priori* does not tell us anything at all about properties of the action of individual elements of G (see Exercise 2.2.1).

We will now recall also that ergodicity and mixing are *spectral properties* in the sense that they can be phrased in terms of the associated unitary representation or unitary action π of G defined by $\pi(g)f = f \circ g^{-1}$ for $f \in L^2(X, \mu)$ and $g \in G$. We note that this unitary representation has the following natural continuity property (which we will assume for all unitary representations discussed): given a function $f \in L^2(X,\mu)$ the map $g \in G \mapsto \pi(g)f \in L^2(X,\mu)$ is continuous (with respect to the given topology on G and the norm topology on $L^2(X,\mu)$), see [**?**, Def. 11.16 and Lemma 11.17].

Assuming the action is measure-preserving, then:

- the G-action is *ergodic* if and only if the constant function $\mathbb{1}$ is the only eigenfunction for the representation (up to multiplication by scalars);
- the G-action is *mixing* if and only if
$$\langle \pi(g)f_1, f_2\rangle \longrightarrow \int f_1 \,\mathrm{d}\mu \int \overline{f_2}\,\mathrm{d}\mu = \langle f_1, \mathbb{1}\rangle \langle \mathbb{1}, f_2\rangle$$
as $g \to \infty$ for any $f_1, f_2 \in L^2(X,\mu)$.

As a motivation for the study of ergodicity in this chapter we recall the pointwise ergodic theorem. The pointwise ergodic theorem holds quite generally for actions of amenable groups[(5)], but here we wish to only discuss the case of $\mathbb{R}^d$-flows (measure-preserving actions of $\mathbb{R}^d$).

Theorem 2.1. *Let $(t,x) \mapsto t\bullet x$ be a jointly continuous action of $\mathbb{R}^d$ on a σ-compact locally compact metric space X preserving a Borel probability measure μ. Then, for any $f \in L^1_\mu(X)$,*
$$\frac{1}{m_{\mathbb{R}^d}(B_r)} \int_{B_r} f(t\bullet x)\,\mathrm{d}t \longrightarrow \mathbb{E}(f \big| \mathscr{E})(x) \tag{2.1}$$
for μ-almost every $x \in X$, where $\mathscr{E} = \{B \subseteq X \mid \mu(B \triangle g\bullet B) = 0$ for all $g \in G\}$ denotes the σ-algebra of invariant sets under the action, and
$$B_r = \{t = (t_1, \dots, t_d) \in \mathbb{R}^d \mid 0 \leqslant t_i \leqslant r \text{ for } i = 1, \dots, d\}.$$

Remark 2.2. (1) This is a special case of [**?**, Th. 8.19], and the use of d-dimensional cubes as the averaging sequence is not necessary. As may be seen from conditions (P), (D), and (F) in [**?**, Sec. 8.6.2] any reasonable choice of metric balls containing the origin of $\mathbb{R}^d$ will suffice to achieve the almost everywhere convergence in (2.1).

(2) Notice that ergodicity for the action is equivalent to the invariant σ-algebra $\mathscr{E}$ being equivalent modulo μ-null sets to the trivial algebra $\{\varnothing, X\}$, so in this case the ergodic averages in (2.1) converge to $\int_X f\,\mathrm{d}\mu$.

(3) A consequence of Theorem 2.1 is that μ-almost every point in X has an orbit under the action that is not only dense but is equidistributed with respect to μ (see [**?**, Ch. 4.4.2] for the details in the case of a single transformation, and Section 6.3.1).

(4) The natural G-action on the quotient $X = \Gamma\backslash G$ by a lattice $\Gamma < G$ is ergodic with respect to the measure m_X inherited from Haar measure on G. However, as the group G is uncountable, it is not immediately obvious

that the absence of nontrivial invariant sets (which is obvious for the G-action on X) implies the triviality of the measure of sets that are invariant modulo m_X (as is required for ergodicity). For the fact that this is indeed the case we refer to [**?**, Sec. 8.1].

(5) As mentioned above, mixing is of course a stronger property than ergodicity in many different ways. More significantly for our purposes, we will see in Chapter 5 situations in which mixing allows us to prove even stronger results on the behavior of all orbits rather than just almost all orbits.

2.1 Real Lie Algebras and Lie Groups

†In this section we will set up the language concerning real Lie algebras and Lie groups that we need. For brevity we assume the basic definitions and properties of Lie groups are known. For proofs, background, and more details we refer to Knapp [**?**]. Not all of the theorems that we mention here will be used in an essential way, but for the most general theorem in this chapter we will use both the Levi decomposition and the Jacobson–Morozov theorem (Theorem 2.14).

2.1.1 Basic Notions

Recall that for any real Lie group G there is an associated real Lie algebra $\mathfrak{g}$ that *describes G near the identity*. There is a smooth map $\exp : \mathfrak{g} \to G$ with a local inverse $\log : B_\delta^G(I) \to \mathfrak{g}$ defined on some neighborhood $B_\delta^G(I)$ of the identity $I \in G$ with $\delta > 0$.

There is a linear representation of G on $\mathfrak{g}$, the *adjoint* representation

$$\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$$

for $g \in G$, satisfying

$$\exp(\mathrm{Ad}_g(v)) = g \exp(v) g^{-1}$$

for $g \in G$ and $v \in \mathfrak{g}$. Furthermore, there is a bilinear anti-symmetric Lie bracket

† This section can be skipped if the reader is familiar with the theory. Also, most of the section can be skipped if the reader is only interested in some main examples of the theory, for example, the important cases of the simple Lie group $G = \mathrm{SL}_d(\mathbb{R})$ or the semi-simple Lie groups

$$G = \mathrm{SL}_d(\mathbb{R}) \times \cdots \times \mathrm{SL}_d(\mathbb{R}).$$

In the latter case, the reader will need to familiarize herself with the notions used in Section 2.1.1, the notion of simple Lie ideals and Lie groups, and should also do Exercise 2.1.1.

$$[\cdot,\cdot]:\mathfrak{g}\times\mathfrak{g}\to\mathfrak{g}$$

and a related map $\mathrm{ad}_u:\mathfrak{g}\to\mathfrak{g}$ defined by

$$\mathrm{ad}_u(v)=[u,v]$$

for $u,v\in\mathfrak{g}$, which satisfies

$$\mathrm{Ad}_g([u,v])=[\mathrm{Ad}_g(u),\mathrm{Ad}_g(v)] \tag{2.2}$$

and

$$\exp(\mathrm{ad}_u)=\mathrm{Ad}_{\exp(u)} \tag{2.3}$$

for all $u,v\in\mathfrak{g}$ and all $g\in G$. Here $\mathrm{ad}_u:\mathfrak{g}\to\mathfrak{g}$ is an element of the algebra of linear maps $(\mathfrak{g})$,

$$\exp:(\mathfrak{g})\longrightarrow\mathrm{GL}(\mathfrak{g})$$

is the exponential map from $(\mathfrak{g})$ to the group $\mathrm{GL}(\mathfrak{g})$ of linear automorphisms of the vector space $\mathfrak{g}$, and $\mathrm{Ad}_{\exp(u)}$ is the adjoint representation defined by the element $\exp(u)\in G$.

Finally, the Lie bracket satisfies the *Jacobi identity*

$$[u,[v,w]]+[v,[w,u]]+[w,[u,v]]=0$$

for all $u,v,w\in\mathfrak{g}$. In the special case where G is a closed linear subgroup of $\mathrm{SL}_d(\mathbb{R})$ for some $d\geqslant 2$ (which is sufficient for all of our applications) the claims above are easy to verify, and

$$\mathfrak{g}\subseteq\mathfrak{sl}_d(\mathbb{R})=\{u\in\mathrm{Mat}_d(\mathbb{R})\mid\mathrm{tr}(u)=0\},$$

$$\mathrm{Ad}_g(u)=gug^{-1},$$

and

$$[u,v]=uv-vu.$$

2.1.2 An Aside on Complex Lie Algebras

The local relationship between a Lie group and its Lie algebra mentioned in Section 2.1.1 in fact goes much further. If G is connected and simply connected then its Lie algebra uniquely determines G. That is, any two connected and simply connected Lie groups with isomorphic Lie algebras are themselves isomorphic. Even without the assumption that the Lie groups G_1,G_2 are simply connected, one obtains a diffeomorphism ϕ between neighborhoods U_1 and U_2 of the identities in G_1 and G_2 if they have the same Lie algebra, such that products are mapped to products $\phi(gh)=\phi(g)\phi(h)$ as long as all the terms $g,h,gh\in U_1$ stay in the domain of the map ϕ. In this case we say that G_1 and G_2 are *locally isomorphic*. For this reason, one usually starts with a classification of Lie algebras, and this classification is easier in the case of complex Lie algebras, making this the conventional first case to consider.

2.1.3 The Structure of Lie Algebras

A Lie ideal $\mathfrak{f} \lhd \mathfrak{g}$ is a subspace of $\mathfrak{g}$ with $[\mathfrak{f}, \mathfrak{g}] \subseteq \mathfrak{f}$. Lie ideals of Lie algebras of real Lie groups correspond to normal subgroups in the following sense. If $F \lhd G$ is a closed normal subgroup, then its Lie algebra $\mathfrak{f} \subseteq \mathfrak{g}$ is a Lie ideal (see Exercise 2.1.3). On the other hand, if $\mathfrak{f} \lhd \mathfrak{g}$ is a Lie ideal, then there is an immersed normal subgroup $F \lhd G$ with Lie algebra $\mathfrak{f}$. Here the term *immersed* allows for the possibility that the subgroup $F = \langle \exp(\mathfrak{f}) \rangle$ generated by $\mathfrak{f}$ is not closed in G (this arises, for example, for the abelian Lie algebras $\mathfrak{f} = \mathbb{R}v$ and $\mathfrak{g} = \mathbb{R}^2$ for the group $G = \mathbb{R}^2/\mathbb{Z}^2$ for most choices of v). In the situation where $F \lhd G$ is not closed, we note that $\overline{F} \lhd G$ would then correspond to another Lie ideal $\bar{\mathfrak{f}} \lhd \mathfrak{g}$ (which is determined by $\mathfrak{f}$ and G, but in general not by $\mathfrak{f}$ and $\mathfrak{g}$ alone).

In group theory the notion of the commutator subgroup

$$[G, G] = \langle [g, h] \mid g, h \in G \rangle \lhd G$$

(where $[g, h] = g^{-1}h^{-1}gh$) is an important measure of the extent to which G fails to be abelian. Recall that a group G is said to be *nilpotent* if the *lower central series* (G_i) defined by

$$\begin{aligned} G_0 &= G, \\ G_{i+1} &= [G, G_i] = \langle [g, h] \mid g \in G, h \in G_i \rangle \lhd G \end{aligned}$$

for $i \geqslant 1$ reaches the trivial group $G_r = \{I\}$ for some $r \geqslant 1$ (which is called the nilpotency degree). Similarly G is called *solvable* if the *commutator series* (G^i) defined by

$$\begin{aligned} G^0 &= G, \\ G^1 &= [G, G] \lhd G, \\ G^{i+1} &= [G^i, G^i] \lhd G \end{aligned}$$

for $i \geqslant 1$ reaches the trivial group $G_s = \{I\}$ for some $s \geqslant 1$. Every nilpotent group is solvable, while the group

$$G = B = \left\{ \begin{pmatrix} a & b \\ & 1 \end{pmatrix} \mid a > 0, b \in \mathbb{R} \right\}$$

is solvable but not nilpotent.

These fundamental notions in group theory have been translated into the theory of Lie algebras in a natural way. A Lie algebra $\mathfrak{g}$ is *nilpotent* if the lower central series

$$\mathfrak{g}_0 = \mathfrak{g} \rhd \mathfrak{g}_1 = [\mathfrak{g}, \mathfrak{g}_0] \rhd \cdots \rhd \mathfrak{g}_{i+1} = [\mathfrak{g}, \mathfrak{g}_i] \rhd \cdots$$

ends with the trivial subalgebra $\mathfrak{g}_r = \{0\}$ for some $r \geqslant 1$, and $\mathfrak{g}$ is *solvable* if the commutator series

$$\mathfrak{g}^0 = \mathfrak{g} \rhd \mathfrak{g}^1 = [\mathfrak{g}^0, \mathfrak{g}^0] \rhd \cdots \rhd \mathfrak{g}^{i+1} = [\mathfrak{g}^i, \mathfrak{g}^i] \rhd \cdots$$

ends with the trivial subalgebra $\mathfrak{g}^s = \{0\}$ for some $s \geqslant 1$.

By Ado's theorem [**?**, Th. B.8], every real (or complex) Lie algebra $\mathfrak{g}$ can be realized as a linear Lie algebra, meaning that $\mathfrak{g}$ can be embedded into $\mathfrak{gl}_d(\mathbb{R}) = \mathrm{Mat}_d(\mathbb{R})$ (or into $\mathfrak{gl}_d(\mathbb{C}) = \mathrm{Mat}_d(\mathbb{C})$) for some $d \geqslant 1$. By Lie's theorem [**?**, Th. 1.25], a complex Lie algebra $\mathfrak{g}$ is solvable if and only if it can be embedded into

$$\mathfrak{b}(\mathbb{C}) = \left\{ \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1d} \\ & a_{22} & a_{23} & \cdots & a_{2d} \\ & & \ddots & & \\ & & & a_{d-1,d-1} & a_{d-1,d} \\ & & & & a_{dd} \end{pmatrix} 816 a_{ij} \in \mathbb{C} \text{ for } i \leqslant j \right\}.$$

Since every real Lie algebra $\mathfrak{g}$ has a complexification $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} + \mathrm{i}\mathfrak{g}$ (see below) it also follows that every real Lie algebra can be embedded into $\mathfrak{b}(\mathbb{C})$ (but maybe not into the analogous real Lie algebra $\mathfrak{b}(\mathbb{R})$.)

By Engel's theorem [**?**, Th. 1.35], a real Lie algebra $\mathfrak{g}$ is nilpotent if and only if it can be embedded into

$$\mathfrak{n} = \left\{ \begin{pmatrix} 0 & a_{12} & \cdots & \cdots & a_{1d} \\ & 0 & a_{23} & \cdots & a_{2d} \\ & & \ddots & & \\ & & & 0 & a_{d-1,d} \\ & & & & 0 \end{pmatrix} 816 a_{ij} \in \mathbb{R} \text{ for } i < j \right\}.$$

It is interesting to note that the commutator $\mathfrak{g}^1 = [\mathfrak{g}, \mathfrak{g}]$ of a solvable Lie algebra is nilpotent (since $[\mathfrak{b}(\mathbb{C}), \mathfrak{b}(\mathbb{C})] \subseteq \mathfrak{n}(\mathbb{C})$)– there is no analog of this fact for abstract groups.

For a general Lie algebra $\mathfrak{g}$, the *radical* $\mathfrak{g}$ of $\mathfrak{g}$ is defined to be the subspace generated by all solvable Lie ideals $\mathfrak{f} \lhd \mathfrak{g}$, and this is a solvable Lie ideal of $\mathfrak{g}$.

A (real or complex) Lie algebra $\mathfrak{g}$ is said to be *semi-simple* if $\mathfrak{g} = \{0\}$. A (real or complex) Lie algebra is called *simple* if $\mathfrak{g}$ is non-abelian (that is, if $[\mathfrak{g}, \mathfrak{g}] \neq \{0\}$) and $\mathfrak{g}$ has no Lie ideals other than $\mathfrak{g}$ and $\{0\}$. We note that a real simple Lie algebra always has a semi-simple complexification

$$\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} + \mathrm{i}\mathfrak{g},$$

with the complexified Lie bracket defined by

$$[u + \mathrm{i}v, w + \mathrm{i}z] = [u, w] - [v, z] + \mathrm{i}\left([v, w] + [u, z]\right),$$

(but not a simple complexification automatically; see Exercise 2.1.2).

Every (real or complex) semi-simple Lie algebra $\mathfrak{g}$ is a direct sum of (real or complex) simple Lie subalgebras, each of which is a Lie ideal in $\mathfrak{g}$.

Finally, we note that solvable Lie algebras and semi-simple Lie algebras complement each other, and any Lie algebra can be described using Lie algebras of these two types in the following sense. The Levi decomposition

$$\mathfrak{g} = \mathfrak{g}_s + \mathfrak{g}$$

of a (real or complex) Lie algebra consists of a semi-simple Lie subalgebra $\mathfrak{g}_s$ of $\mathfrak{g}$ and the radical $\mathfrak{g} \lhd \mathfrak{g}$. In this decomposition $\mathfrak{g}$ is unique, but in general $\mathfrak{g}_s$ is not.

2.1.4 Almost Direct Simple Factors

A connected real (or complex) Lie group G is called *simple* or *semi-simple* if its Lie algebra $\mathfrak{g}$ is simple or semi-simple respectively.

If $\mathfrak{g}$ is a real (or complex) semi-simple Lie algebra then, as mentioned above, we have a decomposition

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_r$$

with simple Lie ideals $\mathfrak{g}_i \lhd \mathfrak{g}$ for $i = 1, \dots, r$. If G is a real (or complex) connected simply connected semi-simple Lie group then the stronger property

$$G \cong G_1 \times \cdots \times G_r, \tag{2.4}$$

holds, where each $G_i \lhd G$ is a connected simply connected Lie group with Lie algebra $\mathfrak{g}_i$.

The product decomposition in (2.4) does not hold for general semi-simple Lie groups without the assumption that the group is simply connected. However, the reason why the product decomposition fails is easy to understand.

Example 2.3. Let

$$G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})/\{(I, I), (-I, -I)\}$$

be the quotient by the normal subgroup N generated by $(-I, -I)$ in

$$\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}).$$

Notice that the Lie algebra of G is isomorphic to $\mathfrak{sl}_2(\mathbb{R}) \times \mathfrak{sl}_2(\mathbb{R})$ and that G is not simply connected. Furthermore,

$$G_1 = \mathrm{SL}_2(\mathbb{R}) \times \{I\}N/N$$

and

$$G_2 = \{I\} \times \mathrm{SL}_2(\mathbb{R})N/N$$

are both normal subgroups of G, are both isomorphic to $\mathrm{SL}_2(\mathbb{R})$, but

$$G \ncong G_1 \times G_2$$

unlike the simply connected case discussed above. Also note that $G_1 \cap G_2$ is generated by $(-I, I)N = (I, -I)N$ which is contained in the center of G.

Allowing for such phenomena along the center, one does get an almost direct product decomposition into almost direct factors of a real semi-simple Lie group as follows. Let G be a real semi-simple Lie group, and suppose that

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \cdots \oplus \mathfrak{g}_r$$

is the decomposition of its Lie algebra into real simple Lie subalgebras. Then for each $i = 1, \ldots, r$ there is a normal closed connected simple Lie subgroup G_i, which we will refer to as an *almost direct factor*, with Lie algebra $\mathfrak{g}_i$. These almost direct factors have the following properties.

- G_i commutes with G_j for $i \neq j$;
- $G = G_1 \cdots G_r$; and
- the kernel of the homomorphism

$$\begin{aligned} G_1 \times \cdots \times G_r &\longrightarrow G_1 \cdots G_r = G \\ (g_1, \ldots, g_r) &\longmapsto g_1 \cdots g_r \end{aligned}$$

  is contained in the center of $G_1 \times \cdots \times G_r$.

We define $G^+ \subseteq G$ to be the almost direct product of (i.e. the normal subgroup of G generated by) those almost direct factors G_i of G that are non-compact.

From now on, unless explicitly identified to be complex, we will always consider real Lie groups and Lie algebras.

Exercises for Section 2.1

Exercise 2.1.1. Show that $\mathfrak{sl}_d(\mathbb{R})$ (or $\mathfrak{sl}_d(\mathbb{C})$) is a real (resp. complex) simple Lie algebra for $d \geqslant 2$. Show that $\mathrm{SL}_d(\mathbb{R})$ and $\mathrm{SL}_d(\mathbb{C})$ are connected simple Lie groups.

Exercise 2.1.2. Show that $\mathfrak{sl}_d(\mathbb{C})$ for $d \geqslant 2$, when viewed as a real Lie algebra, is simple but its complexification is not.

Exercise 2.1.3. Show that if $F \lhd G$ is a closed normal subgroup of a Lie group G, then its Lie algebra $\mathfrak{f} \subseteq \mathfrak{g}$ is a Lie ideal.

Exercise 2.1.4. Let G be a real simple connected Lie group. Show that any proper normal subgroup of G for $d \geqslant 2$ is contained in the center of G.

Exercise 2.1.5. Show that the connected component of

$$\mathrm{SO}(2,2)(\mathbb{R}) = \{g \in \mathrm{SL}_4(\mathbb{R}) \mid g \text{ preserves the quadratic form } ad - bc\}$$

is isomorphic to the almost direct product discussed in Example 2.3.

2.2 Howe–Moore Theorem

Our first goal in relating the algebraic properties of G to properties of its measure-preserving actions is to show that for certain Lie groups ergodicity forces mixing (in contrast to the abelian case, where an ergodic action of $\mathbb{Z}^2$ could have no ergodic elements).

Theorem 2.4 (Howe–Moore, automatic mixing). *A measure-preserving and ergodic action on a probability space by a simple connected Lie group G with finite center is mixing.*

The assumption that the center be finite is necessary. If $G = \widetilde{\mathrm{SL}_2(\mathbb{R})}$ is the universal cover of $\mathrm{SL}_2(\mathbb{R})$, then there are ergodic actions of G on non-trivial probability spaces in which the infinite center (which is isomorphic to $\mathbb{Z}$) acts trivially (as for example the action of $\widetilde{\mathrm{SL}_2(\mathbb{R})}$ induced by the natural action of $\mathrm{SL}_2(\mathbb{R})$ on 2).

A more general formulation expresses this result in terms of *vanishing of matrix coefficients at infinity* in the associated unitary representations. Here a *unitary representation* is an action $\pi : G \times \mathscr{H} \to \mathscr{H}$ by unitary maps $\pi(g)$ for $g \in G$ such that for any given $v \in \mathscr{H}$ the map $G \ni g \mapsto \pi(g)v$ is continuous (with respect to the given topology on G and the norm topology on $\mathscr{H}$). Given a continuous action of a metric locally compact group G on a locally compact metric space X and a locally finite measure μ on X that is preserved by the action, the associated unitary representation

$$\pi(g)(f) = f \circ g^{-1}$$

for $f \in \mathscr{H} = L^2(X, \mu)$ indeed satisfies this continuity property (this may be seen, for example, in [**?**, Lemma 8.7]).

Theorem 2.5 (Howe–Moore, vanishing of matrix coefficients). *If a simple connected Lie group G with finite center acts unitarily on a Hilbert space $\mathscr{H}$, and the action has no non-trivial fixed vectors, then the associated matrix coefficients vanish at infinity in the sense that*

$$\langle \pi(g)v, w\rangle \longrightarrow 0$$

as $g \to \infty$ in G for any $v, w \in \mathscr{H}$.

One of the most important ingredients in the proof of the Howe–Moore theorem is the following weaker statement, which says that ergodicity of a G-action is inherited by unbounded subgroups of simple groups[(6)]. As mentioned earlier, this is far from true in the setting of abelian groups (see Exercise 2.2.1).

Theorem 2.6 (Mautner phenomenon for simple groups). *Let G be a simple connected Lie group with finite center acting unitarily on a Hilbert space $\mathscr{H}$. If $g \in G$ does not belong*[†] *to a compact subgroup of G, and $v \in \mathscr{H}$ is fixed under the action of g, then v is fixed under the action of G.*

2.2.1 Proof of the Howe–Moore Theorem

Assuming the Mautner phenomenon in Theorem 2.6 for simple groups with finite center, we will deduce the following generalization of the Howe–Moore theorem on vanishing of matrix coefficients. In order to state the theorem, we will use[‡] the terminology and results from Section 2.1.4.

Theorem 2.7 (Howe–Moore for semi-simple groups). *Let G be a semi-simple Lie group with finite center, and let $\pi : G \times \mathscr{H} \to \mathscr{H}$ be a unitary representation on a Hilbert space $\mathscr{H}$. For v_1, v_2 in $\mathscr{H}$ we have*

$$\langle \pi(g_n)v_1, v_2\rangle \longrightarrow 0 \tag{2.5}$$

as $n \to \infty$ in either of the following two situations:

(1) *For any of the simple non-compact factors G_i of G, there are no non-trivial G_i-fixed vectors in $\mathscr{H}$ and $g_n \to \infty$ as $n \to \infty$.*
(2) *$\mathscr{H}$ has no non-trivial G^+-fixed vectors, $g_n = g_n^{(1)} \cdots g_n^{(r)}$ with $g_n^{(i)} \in G_i$, and $g_n^{(i)} \to \infty$ as $n \to \infty$ for each simple non-compact factor*[§] *$G_i \subseteq G^+$ of G.*

In the proof of Theorem 2.7 we will make use of the general Cartan decomposition for semi-simple Lie groups with finite center, also known as

[†] Equivalently, if $g^n \to \infty$ in G as $n \to \infty$.

[‡] This is only needed because we state the theorem in greater generality. At its core the argument only needs basic functional analysis, see Exercise 2.2.3.

[§] Even though the decomposition of g_n into $g_n^{(1)} \cdots g_n^{(r)}$ with $g_n^{(i)} \in G_i$ is not unique, the requirement that $g_n^{(i)} \to \infty$ as $n \to \infty$ does make sense as the ambiguity in the decomposition is only up to the finite center of G.

the KAK decomposition (the existence of this decomposition with K compact is where the essential hypothesis that G have finite center enters the argument). Here $K < G$ is a maximal compact subgroup and $A < G$ is a Cartan subgroup[†]. For the case $G = \mathrm{SL}_d(\mathbb{R})$ this decomposition is easy to exhibit, as in this case $K = \mathrm{SO}(d)$, A is the subgroup of diagonal matrices with positive entries down the diagonal, and every matrix $g \in \mathrm{SL}_d(\mathbb{R})$ can be written in the form $g = ka\ell$ with $k, \ell \in K$ and $a \in A$ (see Exercise 2.2.2). We refer to Knapp [**?**, Sec. VII.3] or [**?**] for the proof in the general case.

Proof of Theorem 2.7 (Theorems 2.4–2.5) assuming Theorem 2.6. Assume that $g_n \to \infty$ in G as $n \to \infty$. We will show (2.5) by showing that there always exists a subsequence for which (2.5) holds.

This suffices by a simple indirect argument. Assume (2.5) does not hold, then there exists some $\varepsilon > 0$ and some subsequence n_k with $|\langle \pi(g_{n_k})v_1, v_2\rangle| \geqslant \varepsilon$. However, applying the above claim to this subsequence we find a subsequence of n_k for which (2.5) holds — a contradiction to the choice of n_k.

Using the Cartan decomposition of G, write

$$g_n = k_n a_n \ell_n$$

with $a_n \to \infty$ as $n \to \infty$ in $A < G$. We claim that in order to prove the theorem, it is enough to consider the case $g_n = a_n \to \infty$. Since K is compact and the representation is continuous the study of

$$\langle \pi\left(k_n a_n \ell_n\right) v_1, v_2 \rangle$$

can be reduced — by choosing a subsequence with $k_{n_j} \to k$ and $\ell_{n_j} \to \ell$ as $j \to \infty$, applying continuity of the representation and the Cauchy-Schwartz inequality — to the study of $\left\langle \pi(a_{n_j})\pi(\ell)v_1, \pi(k)^{-1}v_2 \right\rangle$ for some fixed $k, \ell \in K$ and $j \to \infty$. We define $v = \pi(\ell)v_1$.

By passing to a further subsequence if necessary (and dropping the resulting double subscript for convenience), we may also assume that

$$v^* = \lim_{n\to\infty} \pi(a_n)v \in \mathscr{H}$$

exists in the weak*-topology by the Tychonoff–Alaoglu theorem and since

$$\|\pi(a_n)v\| = \|v\|$$

by unitarity.

Recall that $a_n \in A < G$ is the product $a_n = a_n^{(1)} \cdots a_n^{(r)}$ with $a_n^{(i)} \in G_i$ for $i = 1, \ldots, r$. We claim that v^* is fixed under a non-trivial unipotent[‡]

[†] Recall that a Cartan subgroup A is a maximal abelian connected subgroup of G for which Ad_a is $\mathbb{R}$-diagonalizable for all $a \in A$.

[‡] If $G \leqslant \mathrm{SL}_d(\mathbb{R})$ is a linear group, then $u \in G$ is unipotent if 1 is the only eigenvalue of u. In general we say that $u \in G$ is unipotent if $\mathrm{Ad}_u \in \mathrm{SL}(\mathfrak{g})$ is unipotent — this is often referred to as being Ad-unipotent.

element of every factor G_i of G (with respect to the action of π on $\mathscr{H}$), where $a_n^{(i)} \to \infty$ as $n \to \infty$. This claim implies the theorem via the Mautner phenomenon (Theorem 2.6): the vector v^* is fixed under all almost direct factors G_i of G for which $a_n^{(i)} \to \infty$. In both case (1) and case (2), this implies that $v^* = 0$, and hence the theorem.

To prove the claim, let $w \in \mathscr{H}$ be any element. Then since $a_n^{(i)} \to \infty$ as $n \to \infty$ by assumption on G_i, we may choose a subsequence so that there exists a non-trivial unipotent element $u \in G_i$ with

$$(a_n^{(i)})^{-1} u (a_n^{(i)}) \to e$$

as $n \to \infty$. This is easy to see for the case $G_i = \mathrm{SL}_d(\mathbb{R})$ (where u is an element of one of the elementary unipotent subgroups as in Lemma 1.24), and in general u is an element of one of the restricted root subgroups. Then

$$\begin{aligned}\langle \pi(u)v^*, w\rangle &= \left\langle v^*, \pi(u)^{-1} w\right\rangle \\ &= \lim_{n\to\infty} \left\langle \pi(a_n)v, \pi(u^{-1})w\right\rangle \\ &= \lim_{n\to\infty} \left\langle \pi(a_n^{-1} u a_n)v, \pi(a_n^{-1})w\right\rangle .\end{aligned}$$

However,

$$\lim_{n\to\infty} \|\pi(a_n^{-1} u a_n)v - v\| = 0,$$

so

$$\begin{aligned}\langle \pi(u)v^*, w\rangle &= \lim_{n\to\infty} \left\langle \pi(a_n^{-1} u a_n)v, \pi(a_n^{-1})w\right\rangle \\ &= \lim_{n\to\infty} \left\langle v, \pi(a_n^{-1})w\right\rangle \\ &= \lim_{n\to\infty} \langle \pi(a_n)v, w\rangle = \langle v^*, w\rangle .\end{aligned}$$

However, this implies that $\pi(u)v^* = v^*$, giving the claim and hence the theorem. □

Problems for Section 2.2

Exercise 2.2.1. (a) Let $G = \mathbb{Z}^d$ with $d \geqslant 2$. Find an ergodic action of G with the property that no subgroup of G with lower rank acts ergodically.
(b) Let $G = \mathbb{R}^d$ with $d \geqslant 1$. Prove that in any ergodic action of G almost every element of $\mathbb{R}^d$ acts ergodically. (This relies on the standing assumptions regarding X, which imply in particular that $L^2(X)$ is separable.)

Exercise 2.2.2. Prove that every element of $\mathrm{SL}_d(\mathbb{R})$ can be written in the form $ka\ell$ as claimed on page 55.

Exercise 2.2.3. Extract from the general proof of Theorem 2.7 above the special case of $\mathrm{SL}_2(\mathbb{R})$ (or $\mathrm{SL}_d(\mathbb{R})$ for $d \geqslant 2$), still assuming Theorem 2.6 for this (or these) groups.

2.3 The Mautner Phenomenon

The following key lemma(7) will be the main tool used for proving the inheritance property in Theorem 2.6 and its much more general version in Theorem 2.11.

Lemma 2.8 (The key lemma). *Let $\mathscr{H}$ be a Hilbert space carrying a unitary representation of a topological group G. Suppose that $v_0 \in \mathscr{H}$ is fixed by some subgroup $L \leqslant G$. Then v_0 is also fixed under every other element $h \in G$ with the property that there exists sequences $g_n \in G$, $\ell_n, \ell'_n \in L$ with $\lim_{n\to\infty} g_n = I$ and $h = \lim_{n\to\infty} \ell_n g_n \ell'_n$.*

Proof. By assumption, there exist three sequences (g_n) in G, (ℓ_n) in L, and (ℓ'_n) in L with $g_n \to e$ and $\ell_n g_n \ell'_n \to h$ as $n \to \infty$. This implies that

$$\|\pi(\ell_n g_n \ell'_n)v_0 - v_0\| = \|\pi(\ell_n)(\pi(g_n \ell'_n)v_0 - \pi(\ell_n^{-1})v_0)\| = \|\pi(g_n)v_0 - v_0\|$$

by invariance of v_0 under all elements of L and unitarity of $\pi(\ell_n)$. However, the left hand side converges to $\|\pi(h)v_0 - v_0\|$ by continuity of the representation and the right hand side converges to 0. □

2.3.1 The Case of $\mathrm{SL}_2(\mathbb{R})$

We now turn to the special (but important) case of $G = \mathrm{SL}_2(\mathbb{R})$. Any element $g \in \mathrm{SL}_2(\mathbb{R})$ is conjugate to one of the following three type of elements:

- an $\mathbb{R}$-diagonal matrix, that is one of the form $a = \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$ with $\lambda \in \mathbb{R}$;
- a unipotent matrix $u = \begin{pmatrix} 1 & \pm 1 \\ & 1 \end{pmatrix}$; or
- a matrix in the compact subgroup $\mathrm{SO}(2, \mathbb{R})$, that is one of the form

$$k = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

  for some $\phi \in \mathbb{R}$.

For the last case we can make no claim concerning ergodicity of the action of g. However, for the first two types we find the following phenomenon, where we write

$$C_G = \{g \in G \mid gh = hg \text{ for all } h \in G\}$$

for the center of G.

Proposition 2.9 (Mautner for $\mathrm{SL}_2(\mathbb{R})$). *Let $G = \mathrm{SL}_2(\mathbb{R})$ act unitarily on a Hilbert space $\mathscr{H}$, and suppose that $g \neq \pm I$ is unipotent or $\mathbb{R}$-diagonalizable and fixes a vector $v_0 \in \mathscr{H}$. Then all of G fixes v_0 also. The same holds for a connected Lie group G locally isomorphic*[†] *to $\mathrm{SL}_2(\mathbb{R})$ and $g \in G \smallsetminus C_G$ is such that Ad_g is unipotent or $\mathbb{R}$-diagonalizable with an eigenvalue λ with $|\lambda| \neq 1$.*

Suppose $g \in G$ satisfies the hypotheses of Proposition 2.9, and $h \in G$ has the property that hgh^{-1} fixes $v_0 \in \mathscr{H}$. Then g fixes $\pi^{-1}(h)v_0$ and so $v_0 = \pi^{-1}(h)v_0$ is fixed by G as needed. Thus it is sufficient to consider one representative of each conjugacy class for the proof of Proposition 2.9 and for the proof of similar statements that come later.

Proof of Proposition 2.9 for $\mathrm{SL}_2(\mathbb{R})$. For $a = \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$ with $\lambda \neq \pm 1$ a direct calculation shows that we can apply Lemma 2.8 with $L = a^{\mathbb{Z}}$ and any element of the unipotent subgroups $\begin{pmatrix} 1 & * \\ & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & \\ * & 1 \end{pmatrix}$ in $\mathrm{SL}_2(\mathbb{R})$. For example,

$$a^n \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} a^{-n} = \begin{pmatrix} 1 & \lambda^{2n} s \\ & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$$

if $\lambda^{2n} \to 0$ as $n \to \infty$. It follows that if a fixes some $v_0 \in \mathscr{H}$, then so do these two unipotent subgroups, and as they together generate $\mathrm{SL}_2(\mathbb{R})$ (see Exercise 1.2.2 and Lemma 1.24), we obtain Proposition 2.9 for this case.

If $u = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}$ then

$$\begin{aligned} u^n \begin{pmatrix} 1+\delta & \\ & \frac{1}{1+\delta} \end{pmatrix} u^{-n} &= \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix} \begin{pmatrix} 1+\delta & \\ & \frac{1}{1+\delta} \end{pmatrix} \begin{pmatrix} 1 & -n \\ & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1+\delta & \left(\frac{1}{1+\delta} - 1 - \delta\right) n \\ & \frac{1}{1+\delta} \end{pmatrix} \end{aligned}$$

can be made (since n can be chosen arbitrary) to converge to $\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}$ for $\delta \to 0$. It follows that if v_0 is fixed by $\begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}$ then it is also fixed by $\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}$ for any $s \in \mathbb{R}$ by Lemma 2.8 applied with

$$L = \left\{ \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix} \mid n \in \mathbb{Z} \right\}.$$

Applying Lemma 2.8 once more with

[†] This second case is not needed if one is only interested in closed linear subgroups G in $\mathrm{SL}_d(\mathbb{R})$. If G is a closed linear group linearly isomorphic to $\mathrm{SL}_2(\mathbb{R})$, then the theory of finite-dimensional representations of $\mathfrak{sl}_2(\mathbb{R})$ implies that $G \cong \mathrm{SL}_2(\mathbb{R})$ or $G \cong \mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm I\}$, and both of these cases are handled by the first part of the proposition.

$$L = \left\{ \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\}$$

to the matrix

$$\begin{pmatrix} 1 & s_1 \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & \\ \delta & 1 \end{pmatrix} \begin{pmatrix} 1 & s_2 \\ & 1 \end{pmatrix} = \begin{pmatrix} 1+\delta s_1 & s_2(1+\delta s_1)+s_1 \\ \delta & 1+\delta s_2 \end{pmatrix} = g_\delta \tag{2.6}$$

with s_1 chosen to have

$$1 + \delta s_1 = \mathrm{e}^{\alpha}$$

for some fixed $\alpha \in \mathbb{R}$, and with s_2 chosen to have

$$s_2(1+\delta s_1) + s_1 = 0$$

shows that v_0 is also fixed by

$$\begin{pmatrix} \mathrm{e}^{\alpha} & \\ & \mathrm{e}^{-\alpha} \end{pmatrix} = \lim_{\delta \to 0} g_\delta.$$

Applying the previous (diagonal) case, we see once again that v_0 is fixed by all of $\mathrm{SL}_2(\mathbb{R})$. This finishes the proof of the proposition for $\mathrm{SL}_2(\mathbb{R})$, and also the proof of the Howe–Moore theorem (Theorem 2.7) for $\mathrm{SL}_2(\mathbb{R})$ and for products of several copies of $\mathrm{SL}_2(\mathbb{R})$. □

For the second case of Proposition 2.9 where G is only assumed to be locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$ we are going to use the following more general lemma and also the calculations of the proof above. Here and in the following we will work more and more with elements $v \in \mathfrak{g}$ of the Lie algebra of G. If G acts unitarily on a Hilbert space $\mathscr{H}$, $w \in \mathscr{H}$, and $\pi(t\exp(v))w = w$ for all $t \in \mathbb{R}$ and some $v \in \mathfrak{g}$ then we say that v *fixes* w.

Elon asks: why is the case of little sl2 that different?

Lemma 2.10 (Key lemma for unipotent elements). *Let G be a connected Lie group with Lie algebra $\mathfrak{g}$. Let π be a unitary representation on a Hilbert space $\mathscr{H}$. Suppose that $g \in G$ fixes v_0. Then v_0 is also fixed by all elements of the subspace*

$$\mathrm{Im}\,(\mathrm{Ad}_g - I) \cap \ker\,(\mathrm{Ad}_g - I) \subseteq \mathfrak{g},$$

and all of these elements are nilpotent.

In particular, this applies to $g = \exp(u)$ if $u \in \mathfrak{g}$ is nilpotent and the subspace $\mathrm{Im}\,\mathrm{ad}_u \cap \ker \mathrm{ad}_u$.

Proof. Let $v \in \mathrm{Im}\,(\mathrm{Ad}_g - I) \cap \ker\,(\mathrm{Ad}_g - I)$. We wish to show that

$$\pi\left(\exp(v)\right) v_0 = v_0.$$

By assumption, there exists some $w \in \mathfrak{g}$ with

$$\left(\mathrm{Ad}_g - I\right)(w) = v$$

and
$$(\mathrm{Ad}_g - I)(v) = 0,$$
so that
$$\mathrm{Ad}_g(w) = w + v$$
and
$$\mathrm{Ad}_g(v) = v.$$
For $n \geqslant 1$ this gives
$$\mathrm{Ad}_g^n\left(\tfrac{1}{n}w\right) = \tfrac{1}{n}w + v,$$
and so
$$g^n \exp\left(\tfrac{1}{n}w\right) g^{-n} = \exp\left(\tfrac{1}{n}w + v\right). \tag{2.7}$$
The exponential in the left-hand side of (2.7) converges to I, but the right-hand side converges to $\exp(v)$ as $n \to \infty$. It follows by Lemma 2.8 that $\exp(v)$ fixes v_0.

For the last claim of the first part of the lemma we calculate
$$\mathrm{ad}_v = \lim_{n\to\infty} \mathrm{ad}_{\frac{1}{n}w+v} = \lim_{n\to\infty} \mathrm{ad}_{\mathrm{Ad}_g^n(\frac{1}{n}w)} = \lim_{n\to\infty} \mathrm{Ad}_g^n \circ \left(\frac{1}{n}\,\mathrm{ad}_w\right) \circ \mathrm{Ad}_g^{-n},$$
where we used (2.2). Since conjugation does not change the eigenvalues, it follows that ad_v is nilpotent.

Let now $u \in \mathfrak{g}$ be nilpotent as in the last part of the lemma, and let
$$g = \exp(u).$$
Then $\mathrm{Ad}_g = \exp(\mathrm{ad}_u) = I + \mathrm{ad}_u + \ldots + \frac{1}{n!}\,\mathrm{ad}_u^n$ for some n (see (2.3)). If now $v = \mathrm{ad}_u(w) \in \ker \mathrm{ad}_u$, then $\mathrm{Ad}_g(v) = v$ and $\mathrm{Ad}_g(w) = w + v$ and so the first part of the lemma applies. □

Proof of Proposition 2.9. Suppose now that G is only locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$. If Ad_g is $\mathbb{R}$-diagonalizable with an eigenvalue λ with $|\lambda| \neq 1$, then we may argue as above. Indeed suppose that $x \in \mathfrak{g}$ has $\mathrm{Ad}_g(x) = \lambda x$ with $|\lambda| < 1$. Then
$$g^n \exp(tx) g^{-n} = \exp\left(t\,\mathrm{Ad}_g^n(x)\right) \longrightarrow e$$
as $n \to \infty$, and Lemma 2.8 for $L = g^{\mathbb{Z}}$ shows that $\exp(\mathbb{R}x) \subseteq G$ fixes v_0. The same holds for $\exp(\mathbb{R}y)$ for any $y \in \mathfrak{g}$ with
$$\mathrm{Ad}_g(y) = \mu y$$
for some $|\mu| > 1$ by applying the same argument with $n \to -\infty$. Notice that the latter eigenvector must also exist, since otherwise $g \mapsto |\det \mathrm{Ad}_g|$ would be a non-trivial character from the simple group G to $\mathbb{R}^\times$. It follows that $[x, y]$ is an eigenvector for another eigenvalue since

$$\mathrm{Ad}_g([x,y]) = [\mathrm{Ad}_g(x), \mathrm{Ad}_g(y)] = \lambda\mu[x,y].$$

Hence $\exp(\mathbb{R}x)$ and $\exp(\mathbb{R}y)$ generate the 3-dimensional group G, which therefore fixes v_0.

So suppose now that we are in the second case where $\mathrm{Ad}_g \neq I$ is unipotent. Applying Lemma 2.10, we see that v_0 is fixed by all elements of

$$\mathrm{Im}\,(\mathrm{Ad}_g - I) \cap \ker\,(\mathrm{Ad}_g - I)\,.$$

By assumption ($g \notin C_G$ and Ad_g is unipotent) we know that this subspace is nontrivial. Therefore, there exists some $v \in \mathfrak{g} \smallsetminus \{0\}$ such that ad_v is nilpotent and v_0 is fixed by $\exp(\mathbb{R}v)$.

Now choose the isomorphism ϕ between $\mathfrak{g}$ and $\mathfrak{sl}_2(\mathbb{R})$ in such a way that v is mapped to

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{sl}_2(\mathbb{R}).$$

This is possible by the following simple observations. Since

$$w = \phi(v) \in \mathfrak{sl}_2(\mathbb{R})$$

has the property that ad_w is nilpotent, it follows that w also has to be nilpotent. Now recall that the Jordan normal form for matrices in $\mathbb{R}^2$ shows that there is only one conjugacy class $[u]$ of elements of $\mathfrak{sl}_2(\mathbb{R})$ for which u (and also ad_u) is nilpotent. Hence composing ϕ with an appropriate conjugation gives a new ϕ with $\phi(v) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

For α and $\delta > 0$ we define

$$s_1(\alpha) = \frac{\mathrm{e}^{\alpha} - 1}{\delta},$$

and

$$s_2(\alpha) = \frac{-s_1(\alpha)}{1 + \delta s_1(\alpha)}$$

so that

$$\begin{pmatrix} 1 & s_1(\alpha) \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & \\ \delta & 1 \end{pmatrix} \begin{pmatrix} 1 & s_2(\alpha) \\ & 1 \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{\alpha} & \\ \delta & \mathrm{e}^{-\alpha} \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{\alpha} & \\ & \mathrm{e}^{-\alpha} \end{pmatrix} \begin{pmatrix} 1 & \\ \mathrm{e}^{\alpha}\delta & 1 \end{pmatrix}$$

in $\mathrm{SL}_2(\mathbb{R})$ by (2.6). Clearly, if $\delta > 0$ and $\alpha > 0$ are chosen small enough and in that order, then the local isomorphism is defined on the matrices above. So let $g_\delta \in G$ be the element corresponding to

$$\begin{pmatrix} 1 & \\ \delta & 1 \end{pmatrix},$$

and let $h \in G$ be the element corresponding to

$$\begin{pmatrix} e^{\alpha} & \\ & e^{-\alpha} \end{pmatrix}.$$

We then have

$$\exp(s_1(\alpha)v)g_\delta \exp(s_2(\alpha)v) = hg_{e^\alpha \delta} \tag{2.8}$$

as an identity in G. We wish to conjugate both sides of this expression by h^n. Note that

$$h^n \exp(sv)h^{-n} = \exp(e^{2n\alpha}sv)$$

is already known to fix v_0 and that

$$h^n g_\delta h^{-n} = g_{e^{-2n\alpha}\delta}$$

converges to the identity as $n \to \infty$. Therefore, conjugating (2.8) by h^n gives

$$\exp(e^{2n\alpha}s_1(\alpha)v)g_{e^{-2n\alpha}\delta} \exp(e^{2n\alpha}s_2(\alpha)v) = hg_{e^{\alpha-2n\alpha}\delta}$$

It follows that h satisfies the assumptions of Lemma 2.8, and so fixes $v_0 \in \mathscr{H}$. We are therefore reduced to the first case of the proof. □

Exercises for Section 2.3.1

Exercise 2.3.1. Prove Proposition 2.9 (and hence Theorem 2.6) for the case of $\mathrm{SL}_d(\mathbb{R})$ for $d = 3$ or more generally for $d \geqslant 3$, either directly by a similar argument or using the case $\mathrm{SL}_2(\mathbb{R})$ considered above.

Exercise 2.3.2. Prove the analogue of Proposition 2.9 for the case $\mathrm{SL}_2(\mathbb{Q}_p)$ (or for $\mathrm{SL}_d(\mathbb{Q}_p)$ for $d \geqslant 2$), where $\mathbb{Q}_p$ is the field of p-adic rational numbers. More precisely show that $\mathrm{SL}_2(\mathbb{Q}_p)$ fixes $v_0 \in \mathscr{H}$ if $\mathrm{SL}_2(\mathbb{Q}_p)$ acts unitarily on $\mathscr{H}$ and either
(a) v_0 is fixed by some diagonal element with eigenvalues of absolute value not equal to one, or
(a) v_0 is fixed by a one-parameter[†] unipotent subgroup $\{I + sw \mid s \in \mathbb{Q}_p\}$ defined by some nilpotent $w \in \mathrm{Mat}_2(\mathbb{Q}_p)$.

Exercise 2.3.3. Prove Theorem 2.7 for $\mathrm{SL}_2(\mathbb{Q}_p)$ (or for $\mathrm{SL}_d(\mathbb{Q}_p)$ for $d \geqslant 2$), where $\mathbb{Q}_p$ is the field of p-adic rational numbers, using Exercise 2.3.2 in place of Theorem 2.6. For the analogous KAK-decomposition of $\mathrm{SL}_d(\mathbb{Q}_p)$ set $K = \mathrm{SL}_d(\mathbb{Z}_p)$ and let A consist of all diagonal matrices whose diagonal entries are integer powers of p.

Exercise 2.3.4. Prove (direclty or using Exercise 2.3.3) that an unbounded open subgroup $H < \mathrm{SL}_d(\mathbb{Q}_p)$ necessarily equals $\mathrm{SL}_d(\mathbb{Q}_p)$.

[†] We note that in this p-adic case a single element of this subgroup generates a compact subgroup and so could not satisfy the Mautner phenomenon.

2.3.2 The General Mautner Phenomenon

We will now consider the general Mautner phenomena, which was proven by Moore [?] in 1980.

Elon would use a statement for Lie algebra elements preserving vectors, it simplifies the proof — how much ?

Theorem 2.11 (Mautner phenomenon). *Let G be a connected Lie group with Lie algebra $\mathfrak{g}$. Let $L < G$ be a closed subgroup, and suppose that G acts unitarily on a Hilbert space $\mathscr{H}$ with a non-zero vector v_0 fixed by every element of L. Then there exists a Lie ideal $\mathfrak{f} \lhd \mathfrak{g}$ (the* Mautner ideal*) such that*

- *v_0 is fixed by $\exp(\mathfrak{f}) \leqslant G$ and*
- *the map $A_g : \mathfrak{g}/\mathfrak{f} \to \mathfrak{g}/\mathfrak{f}$ induced by Ad_g for $g \in L$ is diagonalizable with all eigenvalues of absolute value one.*

The proof of Theorem 2.11 will combine the key lemma (Lemma 2.8), the special case of $\mathrm{SL}_2(\mathbb{R})$ from Section 2.3.1, and techniques from the theories of Lie groups and Lie algebras. It subsumes the ergodicity of many natural actions. In particular, it contains Theorem 2.6 (which is needed for the proof of Theorem 2.7). However, we note that the case of G semi-simple will be easier than the general case. We will obtain this case (in a slightly weaker form sufficient for Theorem 2.7) in Section 2.3.4, and the general case only in Section 2.3.7.

2.3.3 Big and Small Eigenvalues

Let G and $\mathfrak{g}$ be as in the statement of Theorem 2.11. In this section we will show a weaker claim, which uses the notion of horospherical algebras. The unstable and stable horospherical Lie subalgebras ($\mathfrak{g}^+$ and $\mathfrak{g}^-$ respectively) for $g \in G$ are defined as follows:

- $\mathfrak{g}^+$ is the sum of all generalized[†] subspaces corresponding to eigenvalues of Ad_g with absolute value bigger than one, so
$$\mathfrak{g}^+ = \{v \in \mathfrak{g} \mid \mathrm{Ad}_g^n(v) \to 0 \text{ as } n \to -\infty\},$$
and
- $\mathfrak{g}^-$ is the sum of all generalized subspaces with eigenvalues of Ad_g with absolute value smaller than one, so
$$\mathfrak{g}^- = \{v \in \mathfrak{g} \mid \mathrm{Ad}_g^n(v) \to 0 \text{ as } n \to \infty\}.$$

[†] Here we allow for Jordan blocks corresponding to eigenvalues of absolute value bigger than one as well as for (generalized) eigenspaces corresponding to pairs of complex eigenvalues of absolute value bigger than one.

To see that $\mathfrak{g}^+$ and $\mathfrak{g}^-$ are subalgebras, the characterization in terms of the adjoint action is most useful. If $v_1, v_2 \in \mathfrak{g}^-$, then

$$\mathrm{Ad}_g^n(v_1 + v_2) = \mathrm{Ad}_g^n(v_1) + \mathrm{Ad}_g^n(v_2) \to 0$$

and

$$\mathrm{Ad}_g^n\left([v_1, v_2]\right) = [\mathrm{Ad}_g^n(v_1), \mathrm{Ad}_g^n(v_2)] \to 0$$

as $n \to \infty$, showing that $v_1 + v_2, [v_1, v_2] \in \mathfrak{g}^-$ also; the same argument (but using $n \to -\infty$) shows that $\mathfrak{g}^+$ is also a subalgebra.

Lemma 2.12 (Auslander ideal). *Let G and $\mathfrak{g}$ be as in Theorem* 2.11*, and let g be an element of G. Then the Lie algebra $\mathfrak{f} = \langle \mathfrak{g}^+, \mathfrak{g}^- \rangle$ generated by the unstable and stable horospherical Lie subalgebras of $\mathfrak{g}$ is a Lie ideal of $\mathfrak{g}$, called the* Auslander ideal *of $\mathfrak{g}$.*

Proof. The proof relies on the Jacobi identity. Let $\mathfrak{g}^0$ be the sum of the generalized eigenspaces for all eigenvalues of absolute value one, so that

$$\mathfrak{g} = \mathfrak{g}^+ + \mathfrak{g}^0 + \mathfrak{g}^-,$$

and we need to show that $[\mathfrak{g}, \mathfrak{f}] \subseteq \mathfrak{f}$. Since $\mathfrak{f}$ is a subalgebra by definition, it is sufficient to show that $[\mathfrak{g}^0, \mathfrak{f}] \subseteq \mathfrak{f}$. Notice first that $[\mathfrak{g}^0, \mathfrak{g}^-] \subseteq \mathfrak{g}^-$ (and similarly $[\mathfrak{g}^0, \mathfrak{g}^+] \subseteq \mathfrak{g}^+$). Indeed, if $u \in \mathfrak{g}^0$ and $v \in \mathfrak{g}^-$, then $\|\mathrm{Ad}_g^n(u)\|$ is either bounded or goes to infinity at most at a polynomial rate as $n \to \infty$, while $\|\mathrm{Ad}_g^n(v)\|$ decays to 0 at exponential speed. It follows that

$$\mathrm{Ad}_g^n\left([u, v]\right) = [\mathrm{Ad}_g^n(u), \mathrm{Ad}_g^n(v)] \to 0$$

as $n \to \infty$, as required.

If now $u \in \mathfrak{g}^+, v \in \mathfrak{g}^-$, so that $[u, v] \in \mathfrak{f}$, then for any $w_0 \in \mathfrak{g}^0$ we have

$$[w_0, [u, v]] + \underbrace{[u, \underbrace{[v, w_0]}_{\in \mathfrak{f}}]}_{\in \mathfrak{f}} + \underbrace{[v, \underbrace{[w_0, u]}_{\in \mathfrak{f}}]}_{\in \mathfrak{f}} = 0$$

by the Jacobi identity, the case above, and the fact that $\mathfrak{f}$ is a subalgebra. It follows that $[\mathfrak{g}^0, [\mathfrak{g}^+, \mathfrak{g}^-]] \subseteq \mathfrak{f}$. Repeating the argument under the assumptions $w \in \mathfrak{g}^0$, $u, v \in \mathfrak{f}$ with $[w_0, u], [w_0, v] \in \mathfrak{f}$ we obtain $[w_0, [u, v]] \in \mathfrak{f}$. Hence $\{u \in \mathfrak{f} : [w_0, u] \in \mathfrak{f}\}$ is a subalgebra and so equals $\mathfrak{f}$. As $w_0 \in \mathfrak{g}^0$ was arbitrary, it follows that $\mathfrak{f}$ is a Lie ideal as claimed. □

Proposition 2.13 (Mautner phenomenon for the Auslander ideal). *Let G and $\mathfrak{g}$ be as in Theorem* 2.11*, and suppose that G acts unitarily on a Hilbert space $\mathscr{H}$ and that $g \in G$ fixes $v_0 \in \mathscr{H}$. Then v_0 is fixed by* $\exp \mathfrak{f}$*, where $\mathfrak{f}$ is the Auslander ideal from Lemma* 2.12*.*

Proof. Lemma 2.8 applied to $h = \exp(v)$ with $v \in \mathfrak{g}^{\pm}$ shows that $v_0 \in \mathscr{H}$ is fixed by $\exp(v)$ for $v \in \mathfrak{g}^{\pm}$. It follows that v_0 is fixed by the closed subgroup F generated by the sets $\exp(\mathfrak{g}^{+})$ and $\exp(\mathfrak{g}^{-})$. In particular, there exists a Lie subalgebra (the Lie algebra of F) containing $\mathfrak{g}^{+}$ and $\mathfrak{g}^{-}$ that fixes v_0. Since $\mathfrak{f}$ is the Lie subalgebra generated by $\mathfrak{g}^{+}$ and $\mathfrak{g}^{-}$, we deduce that every element of $\mathfrak{f}$ fixes v_0. □

Exercises for Section 2.3.3

Exercise 2.3.5. Let $a \in G = \mathrm{SL}_d(\mathbb{R})$ be a diagonal matrix such that

$$G_a^{\pm} = \{u \in G \mid a^n u a^{-n} \to I \text{ as } n \to \mp\infty\}$$

are nontrivial subgroups. Show directly that $\langle G_a^{+}, G_a^{-}\rangle = G$.

Exercise 2.3.6. Show that $\mathfrak{g}^0$ from the proof of Lemma 2.12 is a Lie subalgebra.

Exercise 2.3.7. Let G be a simple Lie group and let $\Gamma < G$ be a lattice. Let $a \in G$ and recall that the Lie algebra of G splits as a direct sum $\mathfrak{g}^{+} + \mathfrak{g}^{0} + \mathfrak{g}^{-}$ as in the proof of Lemma 2.12. Assume that Ad_a is diagonalizable when restricted to $\mathfrak{g}^0$ and that 1 is the only eigenvalue of this restriction (so that $\mathfrak{g}^0$ is the Lie algebra of $C_G(a) = \{g \in G \mid ag = ga\}$). Using the pointwise ergodic theorem (Theorem 2.1) show that for any $x \in X = \Gamma\backslash G$ and $m_{G_a^+}$-a.e. $u \in G_a^{+}$ the forward orbit $\{a^n\raisebox{0.2ex}{\scriptsize$\bullet$}(u\raisebox{0.2ex}{\scriptsize$\bullet$}x) : n \geqslant 0\}$ of $u\raisebox{0.2ex}{\scriptsize$\bullet$}x$ equidistributes[†] in X with respect to the Haar measure m_X.

2.3.4 The case of Semi-simple Lie Algebras

In this subsection we will assume that G is a connected semi-simple Lie group. To study actions of such a group, we will combine the arguments from Section 2.3.3, the Jacobson–Morozov theorem[(8)], and the case of $\mathrm{SL}_2(\mathbb{R})$ from Section 2.3.1. The Jacobson–Morozov theorem (we refer to Knapp [**?**, Sec. X.2] for the proof) is the reason that the special case $G = \mathrm{SL}_2(\mathbb{R})$ is so useful.

Barak says that there isn't a good source for this in textbooks — need to check Knapp and make up my mind about it

Theorem 2.14 (Jacobson–Morozov). *Suppose that $\mathfrak{g}$ is a real semi-simple Lie algebra, and let $x \in \mathfrak{g}$ be a nilpotent element. Then there exist elements $y, h \in \mathfrak{g}$ so that (x, y, h) form an* $\mathfrak{sl}_2$-triple*, meaning that they span a subalgebra of $\mathfrak{g}$ isomorphic to $\mathfrak{sl}_2(\mathbb{R})$:*

[†] We note that the results of this section and Remark 2.2 (3) immediately show that the forward orbit is equidistributed for m_X-a.e. $x \in X$, but the desired statement is stronger as it involves a Haar measure on a subgroup.

$$\begin{aligned}[h, x] &= 2x, \\ [h, y] &= -2y, \ \textit{and} \\ [x, y] &= h.\end{aligned}$$

It may be useful to be more explicit about Theorem 2.14 in two low-dimensional examples. In $\mathfrak{sl}_2(\mathbb{R})$ we have

$$x_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, y_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, h_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In $\mathrm{SL}_3(\mathbb{R})$ there are two (fundamentally different) choices, the first via the most obvious embedding $\mathfrak{sl}_2(\mathbb{R}) \hookrightarrow \mathfrak{sl}_3(\mathbb{R})$ giving

$$x_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, y_2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, h_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The second choice for $\mathrm{SL}_3(\mathbb{R})$ (which is not conjugate to the first) comes from the embedding $\mathfrak{sl}_2(\mathbb{R}) \hookrightarrow \mathfrak{sl}_3(\mathbb{R})$ defined by

$$x_3 = \begin{pmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{pmatrix}, y_3 = \begin{pmatrix} 0 & & \\ 2 & 0 & \\ & 2 & 0 \end{pmatrix}, h_3 = \begin{pmatrix} 2 & & \\ & 0 & \\ & & -2 \end{pmatrix} = [x_3, y_3].$$

One can easily check the fundamental relations from Theorem 2.14:

$$[h_3, x_3] = 2x_3, \ [h_3, y_3] = -2y_3, \ \text{and } [x_3, y_3] = h_3.$$

Proposition 2.15 (Mautner phenomenon for semi-simple groups). *Let G be a connected semi-simple Lie group with Lie algebra $\mathfrak{g}$ which acts unitarily on a Hilbert space $\mathscr{H}$. If $g \in G$ is Ad-diagonalizable with positive eigenvalues or $g = \exp(x)$ for some nilpotent $x \in \mathfrak{g}$, and g fixes some vector $v_0 \in \mathscr{H}$, then there is a normal subgroup of G containing g which also fixes v_0.*

Elon says: the statement shouldn't distinguish here into unipotent and diagonal elements, unbounded is the only thing that matters

Proof. If $g = a \in G$ has the property that Ad_a is diagonalizable with positive eigenvalues, then we can split $\mathfrak{g}$ as before into three spaces

$$\mathfrak{g} = \mathfrak{g}^+ + \mathfrak{g}^0 + \mathfrak{g}^-,$$

where $\mathfrak{g}^0$ is the eigenspace of Ad_a with eigenvalue one. Since the Lie algebra generated by $\mathfrak{g}^+$ and $\mathfrak{g}^-$ is a Lie ideal $\mathfrak{f}$ by Lemma 2.12, $\mathfrak{f}$ is a direct sum of some of the direct simple factors of $\mathfrak{g}$. Hence it has to contain any simple factor of $\mathfrak{g}$ that intersects either of the spaces $\mathfrak{g}^+$ or $\mathfrak{g}^-$ nontrivially. Let $F_1 = \langle \exp(\mathfrak{f}) \rangle$ be the normal subgroup containing these simple factors. Since the eigenvalues of Ad_a are by assumption positive, it follows that (the linear map induced by) Ad_a acts trivially on the Lie algebra of G/F_1 (which may be identified with a sub-algebra of $\mathfrak{g}^0$). Therefore, aF_1 belongs to the center of G/F_1, and so generates a normal subgroup of G/F_1. Let $F = \langle a, F_1\rangle$ be the pre-image in G of this normal subgroup. Then $a \in F$, F is a normal subgroup in G, and F fixes $v_0 \in \mathscr{H}$ as required.

Suppose now that $g = u = \exp(x)$ is unipotent. Then by the Jacobson–Morozov theorem there exists a connected subgroup $H < G$ locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$ containing u such that x corresponds under the isomorphism to an upper nilpotent element of $\mathfrak{sl}_2(\mathbb{R})$. By the case of $\mathfrak{sl}_2(\mathbb{R})$ considered in Section 2.3.1, we see that H fixes v_0. Since H also contains the image of

$$a = \begin{pmatrix} \mathrm{e}^{\alpha} & \\ & \mathrm{e}^{-\alpha} \end{pmatrix}$$

for small $\alpha > 0$ (under the local isomorphism), we have produced the situation of the first case, which was considered above. Let F be again the normal subgroup corresponding to a. Now recall that

$$u = \exp(x) \in \exp(\mathfrak{g}^+)$$

if $\mathfrak{g}^+$ is defined using the element in H corresponding to $a \in \mathrm{SL}_2(\mathbb{R})$. Therefore, it follows from the above that $g \in F$ once again. □

This completes the proof of the Mautner phenomenon for semi-simple Lie groups, and in particular proves the case of Theorem 2.6 ($g = \exp(x)$ for some x in one of the root spaces of the simple Lie algebra $\mathfrak{g}$) that is needed for Theorems 2.4, 2.5, and 2.7.

2.3.5 The Structure of the Inductive Steps

[†]For the solvable and then the general case below, we would like to use an induction process to be outlined in this section. For this, notice first that in proving Theorem 2.11 we may assume that v_0 is a *cyclic vector* in the sense that

$$\mathscr{H} = \overline{\langle \pi(G)v_0\rangle}$$

is the smallest closed subspace containing the orbit of v_0 under the action of G, since if this is not the case we may simply restrict the unitary representation to this subspace.

This remark allow us to use induction on the dimension of G. In the inductive steps we will show that there is a non-trivial Lie ideal $\mathfrak{f} \lhd \mathfrak{g}$ that fixes v_0. Taking exponentials gives a normal subgroup $F \lhd G$ generated by $\exp(\mathfrak{f})$. Let $\bar{F}$ be the closure of F (*a priori* there is no reason for F to be

[†] As the following proof will show, semi-simple groups are easier to work with and are, fortunately, sufficient for many purposes. For this reason the reader may initially skip the remainder of Chapter 2 and return to it when she needs it.

closed), so that $\bar{F} \lhd G$ is a closed normal subgroup that fixes v_0. We claim that $\bar{F}$ acts trivially on $\mathscr{H}$ since $\mathscr{H}$ is the closure of the orbit of v_0. Indeed, if $g \in G$ and $h \in F$ then $hg = gh'$ for some $h' \in F$, and

$$\pi(h)\pi(g)v_0 = \pi(g)\pi(h')v_0 = \pi(g)v_0$$

and since $\mathscr{H} = \overline{\langle \pi(G)v_0 \rangle}$ we see that both F and $\bar{F}$ act trivially. Therefore we may consider the unitary representation of $G/\bar{F}$ on $\mathscr{H}$ induced by the unitary representation of G that we started with. If $\mathfrak{f} \lhd \mathfrak{g}$ was a non-trivial Lie ideal, then the dimension of $\widetilde{G} = G/F$ is smaller.

By induction we may assume that Theorem 2.11 already holds for $\widetilde{G}$ (with the subgroup $\widetilde{L} = LF/F < \widetilde{G} = G/F$) acting on $\mathscr{H}$. This in turn then implies the theorem also for G.

2.3.6 The Inductive Step for Elements in the Radical

Recall from Section 2.1.3 that a real Lie algebra $\mathfrak{g}$ has a Levi decomposition[(9)]

$$\mathfrak{g} = \mathfrak{l} + \mathfrak{r}$$

where $\mathfrak{l}$ is a semi-simple real Lie algebra, and $\mathfrak{r} \lhd \mathfrak{g}$ is the radical (the maximal solvable Lie ideal of $\mathfrak{g}$). Also recall from Knapp [**?**, Prop. 1.40] that

$$\mathfrak{n} = [\mathfrak{r}, \mathfrak{g}] \lhd \mathfrak{g}$$

is a nilpotent Lie ideal. Using this we can prove the following part of the inductive step.

Proposition 2.16 (Mautner phenomenon for nilpotent elements of the radical). *Let $G, \pi, \mathscr{H}, v_0$ be as in Theorem* 2.11*, and suppose these also satisfy the assumptions of Section* 2.3.5*. Suppose there is a nilpotent element $u \in \mathfrak{r} \smallsetminus \{0\}$ (with $\mathrm{Ad}_{\exp(u)}$ unipotent) in the radical of the Lie algebra that fixes v_0. Then there is a non-trivial Lie ideal $\mathfrak{f} \lhd \mathfrak{g}$ that fixes v_0.*

This proposition shows that in the situation above we can always apply the inductive step outlined in Section 2.3.5, so that in particular we can also conclude from the inductive argument that there exists an $\mathfrak{f}$ as in the proposition containing u which fixes v_0.

Proof of Proposition 2.16. Suppose as in the statement of the proposition that the nilpotent element $u \in \mathfrak{r} \smallsetminus \{0\}$ fixes v_0. If u lies in the center of $\mathfrak{g}$ (that is, if $[u, \mathfrak{g}] = 0$), then we can take $\mathfrak{f} = \mathbb{R}u$. Otherwise we claim that we may use Lemma 2.10 finitely many times to find vectors $v_1, \ldots, v_\ell$ that all fix v_0 and such that $v_2, \ldots, v_\ell \in \mathfrak{n} = [\mathfrak{r}, \mathfrak{g}]$ and such that the last vector $v_\ell \neq 0$ lies in the center of $\mathfrak{n}$. Initially set $v_1 = u$. Whenever $[v_j, \mathfrak{n}] \neq 0$ for $j \geqslant 1$ then we may take some $w_j \in \mathfrak{n}$ with

$$v_{j+1} = [v_j, w_j] \neq 0$$

and

$$[v_j, v_{j+1}] = 0.$$

This is possible because v_j (that is, ad_{v_j}) is nilpotent, by assumption for $j = 1$ and also for $j \geqslant 2$ since in that case $v_j \in \mathfrak{n}$. Hence by Lemma 2.10 and induction, v_{j+1} fixes v_0. Clearly by construction

$$v_1 = u \in \mathfrak{r}, v_2 \in [\mathfrak{r}, \mathfrak{n}] \subseteq \mathfrak{n}, v_3 \in [\mathfrak{n}, \mathfrak{n}], \dots .$$

Since $\mathfrak{n}$ is a nilpotent Lie algebra, this sequence stops with $v_\ell \in \mathfrak{n}$ and

$$[v_\ell, \mathfrak{n}] = 0$$

for some ℓ.

Let

$$\mathfrak{c} = \{w \in \mathfrak{n} \mid [w, \mathfrak{n}] = 0\}$$

be the center of $\mathfrak{n}$. This is an abelian Lie ideal of $\mathfrak{g}$. We will define $\mathfrak{f}$ as a subspace of $\mathfrak{c}$ containing v_ℓ; indeed we define $\mathfrak{f}$ to be the Lie ideal of $\mathfrak{g}$ generated by v_ℓ. It remains to show that $\mathfrak{f}$ fixes v_0, and this follows as before: If $w \in \mathfrak{g}$ and we have some $v \in \mathfrak{f}$ that fixes v_0, then $[v, w] \in \mathfrak{f}$ also fixes v_0, because $[v, w] \in \mathfrak{c}$, $[v, [v, w]] = 0$ and we may apply Lemma 2.10 as before. As the Lie ideal $\mathfrak{f}$ generated by v_ℓ is obtained by taking the sum of $\mathbb{R}v_\ell$, $[v_\ell, \mathfrak{g}]$, $[[v_\ell, \mathfrak{g}], \mathfrak{g}], \dots$, the proposition follows. □

2.3.7 The General Case of Theorem 2.11

Let $G, \pi, \mathscr{H}, L, v_0$ be as in Theorem 2.11, and suppose that the allowed assumptions of Section 2.3.5 are satisfied.

Let $g \in L$. If Ad_g has an eigenvalue of absolute value greater than or smaller than 1, then we may apply Section 2.3.3 to find the non-trivial Auslander ideal that fixes v_0, and use induction. Suppose therefore that all the eigenvalues of Ad_g have absolute value equal to 1, but that Ad_g is not diagonalizable over $\mathbb{C}$ (since in that case the theorem already holds trivially for g). Then there exist two vectors $v, w \in \mathfrak{g}$ with

$$\mathrm{Ad}_g(v) = \lambda v,$$
$$\mathrm{Ad}_g(w) = \lambda(w + v),$$

and so for $n \in \mathbb{N}$,

$$\mathrm{Ad}_g^n(w) = \lambda^n (w + nv). \tag{2.9}$$

These expressions have the obvious meaning if $\lambda \in \mathbb{R}$, but if $\lambda \in \mathbb{C}\smallsetminus\mathbb{R}$ then we are using the symbol λ as a convenient shorthand for a rotation of the real linear space corresponding to a complex eigenvalue. There is a sequence (n_k) with $n_k \to \infty$ as $k \to \infty$ along which λ^{n_k} converges to the identity. Using this sequence we can divide (2.9) by n_k and find

$$\lim_{k\to\infty} \mathrm{Ad}_g^{n_k}\left(\frac{1}{n_k}w\right) = v. \tag{2.10}$$

We apply the arguments from Lemma 2.10 again to conclude that $v \in \mathfrak{g}$ fixes v_0 and that ad_v is unipotent.

Now let $\mathfrak{g} = \mathfrak{l} + \mathfrak{r}$ be the Levi decomposition. If $v \in \mathfrak{r}$, then we can apply the argument from Proposition 2.16, thus allowing us to assume that

$$v = x + v_{\mathfrak{r}}$$

with $x \in \mathfrak{l}\smallsetminus\{0\}$ and $v_{\mathfrak{r}} \in \mathfrak{r}$. We note that $x \in \mathfrak{l}\smallsetminus\{0\}$ is a nilpotent element of the semi-simple Lie algebra $\mathfrak{l}$ (because, for example, the adjoint of x on $\mathfrak{l} \cong \mathfrak{g}/\mathfrak{r}$ coincides with the adjoint of v on $\mathfrak{g}/\mathfrak{r}$). Furthermore, we claim that $v_{\mathfrak{r}}$ lies in $\mathfrak{n}$ and so is also nilpotent. This follows from the construction of v. Indeed, since $[\mathfrak{g}, \mathfrak{r}] = \mathfrak{n} < \mathfrak{r}$ it follows that $\mathfrak{g}/\mathfrak{n} \cong \mathfrak{l} + \mathfrak{r}/\mathfrak{n}$ is a direct sum of Lie algebras and that $\mathfrak{r}/\mathfrak{n}$ is in the center of $\mathfrak{g}/\mathfrak{n}$. Therefore, the action of Ad_g is trivial on $\mathfrak{r}/\mathfrak{n}$ for any $g \in G$. Splitting $w + \mathfrak{n}$ into its components in $\mathfrak{l}$ and in $\mathfrak{r}/\mathfrak{n}$, the definition of v in (2.10) shows that $v_{\mathfrak{r}} \in \mathfrak{n}$.

Knowing that $x \in \mathfrak{l}$ is nilpotent and nontrivial, we may apply the Jacobson–Morozov theorem (Theorem 2.14) and choose an $\mathfrak{sl}_2$-triple (x, y, h) in $\mathfrak{l}^3$.

Note that if we would have $v = x$ then we could apply the already established semi-simple case (see below). Our aim is therefore to always reduce the proof via induction to this case.

If $[v, \mathfrak{r}] \neq 0$ then we can apply Lemma 2.10 once again to find a non-trivial element of $\mathfrak{n}$ fixing v_0 after which we may apply Proposition 2.16.

So assume that $[v, \mathfrak{r}] = 0$. Then we have

$$\begin{aligned} [v, h] &= [x, h] + [v_{\mathfrak{r}}, h] && \text{(since } v = x + v_{\mathfrak{r}}) \\ &= -2x + [v_{\mathfrak{r}}, h] && \text{(since } [h, x] = 2x) \end{aligned}$$

and so

$$\begin{aligned} [v, [v, h]] &= [v, -2x + [v_{\mathfrak{r}}, h]] \\ &= [v, -2x] + 0 && \text{(since } [v_{\mathfrak{r}}, h] \in \mathfrak{r}) \\ &= [x + v_{\mathfrak{r}}, -2x] \\ &= -2[v_{\mathfrak{r}}, x] \in \mathfrak{r}. \end{aligned}$$

Furthermore,

$$[v,[v,[v,h]]] = [v, -2\underbrace{[v_{\mathfrak{r}}, x]}_{\in \mathfrak{r}}] = 0.$$

Hence we may apply Lemma 2.10 if $[v,[v,h]] \neq 0$, and then use Proposition 2.16 and induction. So assume that $[v,[v,h]] = 0$. By Lemma 2.10, $[v,h]$ fixes v_0. If $[v,h] \neq -2v$, then (recall that $v = x + v_{\mathfrak{r}}$)

$$[v,h] + 2v = -2x + [v_{\mathfrak{r}}, h] + 2x + 2v_{\mathfrak{r}} \neq 0$$

belongs to $\mathfrak{n}$ and fixes[†] v_0, so we may apply Proposition 2.16 and induction.

So assume now that $[v,h] = -2v$. We claim that this implies (only using structure theory of Lie groups) that $v_{\mathfrak{r}} = 0$, so that $v = x$ is a member of an $\mathfrak{sl}_2$-triple inside $\mathfrak{l}$.

So assume (for the purposes of obtaining a contradiction) that $v_{\mathfrak{r}} \neq 0$. Also recall that $v_{\mathfrak{r}} \in \mathfrak{n}$. There exists a Lie ideal $\mathfrak{f} = \mathfrak{n}_i \lhd \mathfrak{g}$ from the lower central series as in Section 2.1.3 with $v_{\mathfrak{r}} \notin \mathfrak{f}$ but $v_{\mathfrak{r}} \in \mathfrak{n}_{i-1}$ so that $[v_{\mathfrak{r}}, \mathfrak{n}] \subseteq \mathfrak{f} = \mathfrak{n}_i$. Thus $\mathfrak{g}$ acts on $\mathfrak{n}/\mathfrak{f}$, both v (since $[v,\mathfrak{r}] = 0$ by one of our allowed assumptions from above) and $v_{\mathfrak{r}}$ (by construction of $\mathfrak{f}$) act trivially on $\mathfrak{n}/\mathfrak{f}$, and so x also acts trivially on $\mathfrak{n}/\mathfrak{f}$. However, if x acts trivially on the whole space there must be a Lie ideal in $\mathfrak{g}$ (the kernel of the representation) which acts trivially. Therefore, we see that h acts trivially on $\mathfrak{n}/\mathfrak{f}$ and so

$$[h,v] = 2v = 2x + 2v_{\mathfrak{r}} = [h,x] + [h,v_{\mathfrak{r}}] \in 2x + \mathfrak{f}$$

gives the contradiction $v_{\mathfrak{r}} \in \mathfrak{f}$. Therefore, $v_{\mathfrak{r}} = 0$ as claimed.

To finish the proof we wish to apply Proposition 2.15. Since $v = x$ fixes v_0, there exists a Lie ideal $\mathfrak{h} \lhd \mathfrak{l}$ containing v that fixes v_0. As the proof of Proposition 2.15 shows, $\mathfrak{h}$ is the Auslander ideal of $a = \exp(h) \in G$ inside $\mathfrak{l}$ and contains h. By Proposition 2.13 the non-trivial Auslander ideal $\mathfrak{f} \lhd \mathfrak{g}$ defined by a within $\mathfrak{g}$ also fixes v_0, contains $\mathfrak{h}$, and so also v. This concludes the induction and hence also the proof of Theorem 2.11.

Notes to Chapter 2

(5) (Page 46) The main result here is due to Lindenstrauss, who showed that any locally compact amenable group has a Følner sequence along which the pointwise ergodic theorem holds. We refer to a survey of Nevo [**?**] for an overview of both the amenable case and the case of certain non-amenable groups, and to [**?**, Ch. 8] for an accessible discussion of the case of groups with polynomial growth.

[†] Since addition in the Lie algebra and taking products in the Lie group are not quite the same we should add some explanations for this step. Actually in this proof we even know that all elements of $\mathbb{R}v$ and $\mathbb{R}[v,h]$ fix v_0. Hence both $[v,h]$ and v belong to the Lie algebra of the subgroup that fixes v_0, which implies that also $[v,h] + 2v$ fixes v_0.

(6) (Page 54) The Mautner phenomenon was developed for the study of geodesic flows on symmetric spaces by Mautner [?] and has been significantly extended since then, notably by Moore [?].

(7) (Page 57) This argument comes from Margulis [?], and the argument is also presented in [?, Prop. 11.18].

(8) (Page 65) Theorem 2.14 was stated by Morozov [?] and a complete proof was provided by Jacobson [?].

(9) (Page 68) This decomposition, conjectured by Killing and Cartan, was shown by Levi [?], and Malcev [?] later showed that any two Levi factors (the semi-simple Lie algebra viewed as a factor-algebra of $\mathfrak{g}$) are conjugate by a specific form of inner automorphism; we refer to Knapp [?, Th. B.2] for the proof.

Chapter 3
Rationality

In this chapter we generalize some of the phenomena hinted at in Section 1.2. We will define the notion of algebraic groups defined over $\mathbb{Q}$, and show how these often give rise to closed (and sometimes even compact) orbits on $\mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$. We motivate this discussion by studying orthogonal groups, unipotent groups, and orbits arising from number fields. Finally, we will turn this discussion around by proving the Borel density theorem, which implies that finite volume orbits typically arise from algebraic groups defined over $\mathbb{Q}$. For this we also introduce some more basic concepts and results concerning algebraic groups without developing this important theory very far (which cannot be done in a couple of pages).

Manfred wants to restructure this chapter once more to minimize the algebra appearing before the Borel density statement! (without removing any algebra)

3.1 Quadratic Forms, Stabilizer Subgroups, and Orbits

3.1.1 Orthogonal Groups

Let
$$Q(u_1,\dots,u_d)=(u_1,\dots,u_d)A_Q(u_1,\dots,u_d)^{\mathrm{t}}$$
be a rational quadratic form defined by a symmetric matrix $A_Q\in\mathrm{Mat}_d(\mathbb{Q})$. We show now how any such quadratic form gives rise to a closed orbit of its associated special orthogonal subgroup

$$\mathrm{SO}(Q)=\{g\in\mathrm{SL}_d\mid Q\left((u_1,\dots,u_d)g\right)=Q(u_1,\dots,u_d)\text{ for all }u_1,\dots,u_d\}. \tag{3.1}$$

Proposition 3.1 (Closed orbits). *If Q is a rational quadratic form, then the orbit*
$$(\mathrm{SL}_d(\mathbb{Z})I_d)\,\mathrm{SO}(Q)(\mathbb{R})\subseteq d$$
of the identity coset under the real points of $\mathrm{SO}(Q)$ *is closed.*

Notice that the notation $\mathrm{SO}(Q)$ and SL_d used in (3.1) deliberately does not specify any field or ring, and therefore leaves somewhat undetermined the group being discussed; in particular does not specify whether the group is countable or uncountable, for example. For now we should think of this as a convenient shorthand, or a macro, which defines many different groups at once. For example, if we specify the real points, then the notation denotes the closed linear subgroup of $\mathrm{SL}_d(\mathbb{R})$ defined by

$$\begin{aligned}\mathrm{SO}(Q)(\mathbb{R}) = \{g \in \mathrm{SL}_d(\mathbb{R}) \mid Q\left((u_1,\dots,u_d)g\right) = Q(u_1,\dots,u_d)\\ \text{for all } u_1,\dots,u_d\}.\end{aligned}$$

Similarly, we may specify the integer points to obtain a discrete subgroup

$$\begin{aligned}\mathrm{SO}(Q)(\mathbb{Z}) = \{g \in \mathrm{SL}_d(\mathbb{Z}) \mid Q\left((u_1,\dots,u_d)g\right) = Q(u_1,\dots,u_d)\\ \text{for all } u_1,\dots,u_d\}.\end{aligned}$$

of $\mathrm{SO}(Q)(\mathbb{R})$. More generally, for any ring R we obtain the group $\mathrm{SO}(Q)(R)$ of R-points of $\mathrm{SO}(Q)$ (or any similar expression) by taking the R-points of the ambient group, here SL_d, in its definition.

Proof of Proposition 3.1. Notice that $Q\left((u_1,\dots,u_d)g\right)$ is the quadratic form defined by gA_Qg^{t} and that the symmetric matrix A_Q is in a one-to-one correspondence to Q. Therefore, we may also write

$$\mathrm{SO}(Q) = \{g \in \mathrm{SL}_d \mid gA_Qg^{\mathrm{t}} = A_Q\}.$$

Multiplying A_Q by its common denominator if necessary, we may assume that $A_Q \in \mathrm{Mat}_d(\mathbb{Z})$ (without changing $\mathrm{SO}(Q)$). Now suppose that

$$\mathrm{SL}_d(\mathbb{Z})h_n \to \mathrm{SL}_d(\mathbb{Z})g = x \tag{3.2}$$

as $n \to \infty$ with $h_n \in \mathrm{SO}(Q)(\mathbb{R})$ and $g \in \mathrm{SL}_d(\mathbb{R})$. In order to show that the orbit is closed, we need to show that

$$x \in \mathrm{SL}_d(\mathbb{Z})I_d\,\mathrm{SO}(Q)(\mathbb{R}). \tag{3.3}$$

Notice that (3.2) simply means that there exist sequences (γ_n) in $\mathrm{SL}_d(\mathbb{Z})$ and (ε_n) in $\mathrm{SL}_d(\mathbb{R})$ with $\varepsilon_n \to I_d$ as $n \to \infty$, such that $\gamma_n h_n = g\varepsilon_n$ for all $n \geqslant 1$. Applying these matrices to A_Q gives

$$\begin{aligned}\gamma_n A_Q \gamma_n^{\mathrm{t}} &= \gamma_n h_n A_Q (\gamma_n h_n)^{\mathrm{t}}\\ &= g\varepsilon_n A_Q (g\varepsilon_n)^{\mathrm{t}} \to gA_Qg^{\mathrm{t}}\end{aligned}$$

as $n \to \infty$.

However, $\gamma_n A_Q \gamma_n^{\mathrm{t}} \in \mathrm{Mat}_d(\mathbb{Z})$, so the convergent sequence $(\gamma_n A_Q \gamma_n^{\mathrm{t}})$ has to stabilize: there exists some n_0 such that

$$\gamma_{n_0} A_Q \gamma_{n_0}^{\mathrm{t}} = \gamma_n A_Q \gamma_n^{\mathrm{t}} = g A_Q g^{\mathrm{t}}$$

for all $n \geqslant n_0$. This implies that $\gamma_{n_0}^{-1} g \in \mathrm{SO}(Q)(\mathbb{R})$, giving (3.3). $\square$

In some cases it is also relatively straightforward to combine the previous statement with Mahler's compactness criterion (Theorem 1.17) and so deduce compactness of orbits.

Proposition 3.2 (Compact orbits). *If Q is a rational quadratic form such that*[†]

$$0 \notin Q(\mathbb{Q}^d \diagdown \{0\}),$$

then the orbit $(\mathrm{SL}_d(\mathbb{Z}) I_d)\, SO(Q)(\mathbb{R})$ *is compact. Equivalently,*

$$\mathrm{SO}(Q)(\mathbb{Z}) = \{g \in \mathrm{SL}_d(\mathbb{Z}) \mid g A_Q g^{\mathrm{t}} = A_Q\}$$

is a uniform lattice in $\mathrm{SO}(Q)(\mathbb{R})$.

Proof. Just as in the proof of Proposition 3.1, we may assume that A_Q lies in $\mathrm{Mat}_d(\mathbb{Z})$. We need to show that there exists some $\delta > 0$ such that

$$\mathrm{SL}_d(\mathbb{Z}) I_d \,\mathrm{SO}(Q)(\mathbb{R}) \subseteq d(\delta). \tag{3.4}$$

Then Theorem 1.17 and Proposition 3.1 together show that the orbit is compact.

As $Q : \mathbb{R}^d \to \mathbb{R}$ is continuous, there exists some $\delta > 0$ such that $\|x\| < \delta$ implies that $|Q(x)| < 1$. Now suppose that (3.4) does not hold for δ. Then there exists some $h \in \mathrm{SO}(Q)(\mathbb{R})$ such that $\mathbb{Z}^d h$ contains a non-zero δ-short vector mh with $m \in \mathbb{Z}^d$. However, this shows that

$$|Q(m)| = |Q(mh)| < 1 \tag{3.5}$$

which implies that $Q(m) = 0$ since $A_Q \in \mathrm{Mat}_d(\mathbb{Z})$, contradicting our assumption and completing the proof. $\square$

Example 3.3. These examples describe some of the possibilities that may arise in low dimensions.

(1) If $Q_1(u_1, u_2) = u_1 u_2$, then Proposition 3.1 shows that $\mathrm{SL}_2(\mathbb{Z}) A$ is closed since

$$\mathrm{SO}(Q_1)(\mathbb{R}) = A$$

is simply the diagonal subgroup of $\mathrm{SL}_2(\mathbb{R})$ (see Section 1.2.2). However, the orbit is not compact, it is the divergent orbit mentioned on page 24.

(2) If $Q_2(u_1, u_2) = u_1^2 - u_1 u_2 - u_2^2$, then Proposition 3.2 applies (see Exercise 3.1.1), and gives a compact orbit $\mathrm{SL}_2(\mathbb{Z})\,\mathrm{SO}(Q_2)(\mathbb{R})$. As we will see later (in Theorem 3.5), there exists some $g \in \mathrm{SL}_2(\mathbb{R})$ for which

[†] Q is then called anisotropic over $\mathbb{Q}$.

$$Q_2(u_1,u_2) = Q_1\left((u_1,u_2)g\right),$$

which in turn implies that

$$\mathrm{SO}(Q_2)(\mathbb{R}) = g\,\mathrm{SO}(Q_1)(\mathbb{R})g^{-1}.$$

To see this notice that $h \in \mathrm{SO}(Q_1)(\mathbb{R})$ and $u = (u_1,u_2)$ gives

$$Q_2(ughg^{-1}) = Q_1((ughg^{-1})g) = Q_1(ug) = Q_2(u).$$

Hence

$$\mathrm{SL}_2(\mathbb{Z})\,\mathrm{SO}(Q_2)(\mathbb{R})g = \mathrm{SL}_2(\mathbb{Z})gA$$

is also compact. In fact $g = g_{\mathrm{golden}}$ from Section 1.2.2 can be used, recovering the claim made on page 25.

(3) If $Q_3(u_1,u_2,u_3) = 2u_1u_3 - u_2^2$ then Proposition 3.1 applies, and shows that

$$\mathrm{SL}_3(\mathbb{Z})\,\mathrm{SO}(Q_3)(\mathbb{R}) \subseteq 3$$

is closed. However, it is not compact (see Exercise 3.1.2).

(4) If $Q_4(u_1,u_2,u_3) = u_1^2 + u_2^2 - 3u_3^2$ then Proposition 3.2 applies. To see this, assume for the purposes of a contradiction (and without loss of generality by clearing denominators as usual) that $Q_4(m_1,m_2,m_3) = 0$ for some primitive[†] integer vector $(m_1,m_2,m_3) \in \mathbb{Z}^3$. Then using congruences modulo 4 shows that

$$m_1^2 + m_2^2 - 3m_3^2 \equiv m_1^2 + m_2^2 + m_3^2 \pmod 4,$$

is a sum of three squares modulo 4. However, the only squares modulo 4 are 0 and 1, which forces m_1, m_2, m_3 to all be even, contradicting the assumption. Hence the orbit

$$\mathrm{SL}_3(\mathbb{Z})\,\mathrm{SO}(Q_4)(\mathbb{R})$$

is compact.

We now recall some of the basic theory of quadratic forms over the reals[(10)]. Any symmetric matrix $A \in \mathrm{Mat}_d(\mathbb{R})$ can be diagonalized in the sense that there is an orthogonal matrix k for which kAk^{t} is diagonal. In the associated coordinate system $(v_1,\dots,v_d)$ we then have

$$Q'(v_1,\dots,v_d) = Q\left((v_1,\dots,v_d)k\right) = \sum_{i=1}^{d} c_i v_i^2.$$

[†] An integer vector is *primitive* if the entries are co-prime.

The form Q is *non-degenerate* if $c_i \neq 0$ for $i = 1, \ldots, d$ (equivalently, if $\det A_Q \neq 0$), is *indefinite* if there exist i, j with $c_i > 0$ and $c_j < 0$, and is *positive-definite* if $c_i > 0$ for all $i = 1, \ldots, d$.

Taking the square roots of the absolute values of the entries in the diagonal matrix kA_Qk^{t}, we may define a diagonal matrix a for which

$$a^{-1}kA_Qk^{\mathrm{t}}a^{-1}$$

is diagonal with entries in $\{0, \pm 1\}$. Assuming that Q is non-degenerate (so that the entries lie in $\{\pm 1\}$), write p for the number of $+1$s and q for the number of -1s; the *signature*$^{(11)}$ of Q is (p, q). We usually assume that $p \geqslant q$ (this can always be achieved by replacing the form Q with the form $-Q$).

The discussion above shows that if Q and Q' are non-degenerate and of the same signature, then there exists some $g \in \mathrm{GL}_d(\mathbb{R})$ such that

$$Q'(u_1, \ldots, u_d) = Q\left((u_1, \ldots, u_d)g\right).$$

Moreover, we also have

$$Q'(u_1, \ldots, u_d) = \lambda Q\left((u_1, \ldots, u_d)g'\right)$$

for $g' \in \mathrm{SL}_d(\mathbb{R})$ and $\lambda \neq 0$, which implies that $\mathrm{SO}(Q)$ and $\mathrm{SO}(Q')$ are conjugate in $\mathrm{SL}_d(\mathbb{R})$.

Example 3.4. The quadratic forms (from Example 3.3) Q_1 and Q_2 have signature $(1, 1)$; the quadratic forms Q_3 and Q_4 have signature $(2, 1)$. It follows that the orthogonal groups $\mathrm{SO}(Q_1)(\mathbb{R})$ and $\mathrm{SO}(Q_2)(\mathbb{R})$ are conjugate (as claimed earlier), and the orthogonal groups $\mathrm{SO}(Q_3)(\mathbb{R})$ and $\mathrm{SO}(Q_4)(\mathbb{R})$ are conjugate.

We summarize and strengthen our discussion as follows.

Theorem 3.5 (Signature of quadratic forms). *Any non-degenerate quadratic form Q on $\mathbb{R}^d$ can be assigned a signature (p, q) with $p + q = d$. Given a form Q of signature (p, q), the set of quadratic forms of the form Q' with*

$$Q'(u_1, \ldots, u_d) = Q\left((u_1, \ldots, u_d)g\right)$$

obtained from Q by some $g \in \mathrm{GL}_d(\mathbb{R})$, is precisely the set of quadratic forms of signature (p, q). The group of $\mathbb{R}$-points of two orthogonal groups for non-degenerate quadratic forms of the same signature are conjugate in $\mathrm{SL}_d(\mathbb{R})$.

In the following we will always (and sometimes implicitly) assume that the quadratic forms are non-degenerate. Fixing, for a given signature (p, q), some real quadratic form Q of this signature, we define $\mathrm{SO}(p, q) = \mathrm{SO}(Q)$. If $p = d$, then

$$\mathrm{SO}(p, q) = \mathrm{SO}(d)$$

is compact, and if $0 < p < d$ it is not[†]. Our discussion above (and Example 3.3(3),(4)), shows that there are various matrices $g \in \mathrm{SL}_d(\mathbb{R})$ for which

$$\mathrm{SL}_d(\mathbb{Z})g\,\mathrm{SO}(p,q)(\mathbb{R})$$

is closed or even compact – these orbits correspond[‡] to rational quadratic forms with signature (p,q).

3.1.2 Rational Stabilizer Subgroups

It is straightforward to generalize Proposition 3.1. However, setting up the language of linear groups, in which the generalization is naturally phrased, requires more work than does the generalization itself. We start this introduction to linear algebraic groups here, discuss other classes of examples in Sections 3.2 and 3.3, and return to the theory of linear algebraic groups in Section 3.4 and Chapter 7. For a detailed account of algebraic geometry, we refer to the monographs of Hartshorne [**?**] or Shafarevich [**?**], and for linear algebraic groups we refer to those of Humphreys [**?**], Borel [**?**], and Springer [**?**].

An *affine variety* is a subset Z of $\mathbb{C}^n$ or, more generally, of $\overline{\mathbb{K}}^n$ for another field $\mathbb{K}$ with $\overline{\mathbb{K}}$ an algebraic closure, defined by the vanishing of polynomial equations[§]. We will write both Z and $Z(\overline{\mathbb{K}})$ for this variety, so that

$$Z = Z(\overline{\mathbb{K}})$$

will always consist of all solutions to the polynomial equations over the algebraic closure. An important example for us is

$$\mathrm{SL}_d = \{g \in \mathrm{Mat}_d \mid \det g - 1 = 0\},$$

where Mat_d is the d^2-dimensional vector space of $d \times d$ matrices.

A *regular function* is simply the restriction of a polynomial to the variety[¶]. In order to be able to work with this definition, and in particular to have a way to uniquely describe a regular function, we need to know when a polynomial vanishes on the variety. The description of the set of polynomials that vanish

[†] Since, for example, it contains at least one copy of $\mathrm{SO}(1,1) \cong A$ as a closed subgroup.

[‡] At this stage we only know one direction of this correspondence. The second direction will be obtained from the Borel density theorem, see Exercise 3.5.1 and Exercise 4.4.1.

[§] We apologize to the expert for this barbaric and old-fashioned definition, but as our focus will usually be on rather concrete groups comprising $\mathbb{R}$-points, this approach is appropriate here. We will on occasion (indeed, are just about to) avoid mentioning the field we are working over, but we still wish to avoid talking about schemes, spectrum, and using the language of modern algebraic geometry.

[¶] Once again we must apologize for avoiding a more general definition, our excuse being that this is adequate for affine varieties.

on an affine variety is given by the Hilbert Nullstellensatz$^{(12)}$ which we now recall. We refer to Hungerford [**?**, Prop. VIII 7.4] or Eisenbud [**?**, Th. 1.6] for the proof.

Theorem 3.6 (Hilbert Nullstellensatz). *Let $\mathbb{K}$ be an algebraically closed field, and let $J \subseteq \mathbb{K}[x_1,\dots,x_n]$ be an ideal defining the affine variety*

$$Z(J) = \{x \in \mathbb{K}^n \mid f(x) = 0 \textit{ for all } f \in J\}.$$

Then $f \in \mathbb{K}[x_1,\dots,x_n]$ vanishes on $Z(J)$ if and only if there exists a power f^m, $m \geqslant 1$, of f that belongs to J.

The ideal

$$(J) = \{f \in k[x_1,\dots,x_n] \mid f^m \in J \text{ for some } m \geqslant 1\}$$

is called the *radical* of the ideal J. If we now write $\mathbb{K}[Z]$ for the ring of regular functions on the variety $Z = Z(J)$ defined by the ideal J, then we can reformulate the Nullstellensatz by the formula

$$\mathbb{K}[Z(J)] = \mathbb{K}[x_1,\dots,x_n]/(J).$$

Returning to our example

$$\mathrm{SL}_d = Z(\det(\cdot) - 1) \subseteq \mathrm{Mat}_d,$$

we need to establish what the radical of the ideal generated by the polynomial $\det(\cdot) - 1$ in d^2 variables is in order to talk about regular functions. This is explained by the following result.

Lemma 3.7 (SL_d is Zariski connected). *For any $d \geqslant 1$ the polynomial $\det(g) - 1$ is irreducible as a polynomial in the variables g_{ij}, $1 \leqslant i,j \leqslant d$, with coefficients in $\mathbb{C}$ (or in any other field).*

Before proving this, we note that additional background in algebraic groups would make it almost immediate by the following argument. Notice first that the group $\mathrm{SL}_d(\mathbb{C})$ of complex solutions to $\det(g) = 1$ is connected (this may be seen, for example, as a consequence of Lemma 1.24), and that every point of this variety is smooth†. If now $\det(g) - 1 = p_1^{\ell_1}(g)p_2^{\ell_2}(g)\cdots p_\ell^{\ell_k}(g)$ is the decomposition into irreducible polynomials with multiplicities ℓ_i, then the group $\mathrm{SL}_d(\mathbb{C})$ would be the union of the varieties defined by $p_1,\dots,p_k$. By connectedness of $\mathrm{SL}_d(\mathbb{C})$ these varieties would have to intersect, but this contradicts smoothness of the variety SL_d at the intersection points. The only possibility that remains is that $\det(g) - 1 = p^\ell(g)$ for some irreducible

† This is a general property of algebraic groups: Every variety has a smooth point, and as usual for a group any two points in the group have similar properties as the group acts transitively, see Section 3.4.

polynomial p and $\ell > 1$. This is clearly impossible by degree considerations (see the first part of the more elementary proof below).

Proof of Lemma 3.7. Suppose that $\det(g) - 1 = p(g)q(g)$, where p, q are polynomials in the independent variables g_{ij}, $1 \leqslant i, j \leqslant d$. Now notice that the determinant is linear in each of its rows, so for every pair i, j the polynomial $\det(g) - 1$ is of degree one in the variable g_{ij}. It follows that for any i, j either p or q is independent of g_{ij} (that is, of degree zero in the variable g_{ij}). As this holds for every pair i, j, we obtain a partition

$$P \sqcup Q = \{(i, j) \mid 1 \leqslant i, j \leqslant d\}$$

of the indices so that

$$p(g) \in \mathbb{C}[g_{ij} \mid (i, j) \in P]$$

and

$$q(g) \in \mathbb{C}[g_{ij} \mid (i, j) \in Q].$$

If P (or Q) is empty, then $p \in \mathbb{C}$ (respectively $q \in \mathbb{C}$) is a scalar — which is the desired conclusion.

With deg denoting the total degree,

$$d = \deg(\det(g) - 1) = \deg(p(g)q(g)) = \deg(p(g)) + \deg(q(g)). \tag{3.6}$$

Assuming that P and Q are both non-empty, we derive a contradiction by defining

$$\deg_P(g_{ij}) = \begin{cases} 1 & \text{if } (i, j) \in P; \\ -1 & \text{if } (i, j) \in Q, \end{cases}$$

which extends to monomials m in the usual way, and to polynomials by defining

$$\deg_P\left(\sum c_k m_k\right) = \max\{\deg_P(m_k) \mid c_k \neq 0\}.$$

Just as in (3.6), we find that

$$\deg_P(pq) = \deg_P(p) + \deg_P(q).$$

Now q must have a constant term (since $\det(g) - 1$ has a constant term), so $\deg_P(q) = 0$. It follows that $p(g)q(g)$ contains monomials in the variables g_{ij} with $(i, j) \in P$ of total degree $\deg_P(p) = \deg(p)$ only. However, this is a contradiction as $\det(g) - 1$ contains a constant term, and all other monomials have total degree d. □

Let $\mathbb{K}$ be any field. We will often be interested not in the whole variety consisting of all points in $\overline{\mathbb{K}}^n$ defined by an ideal over the algebraic closure of a field, but in fact only in the $\mathbb{K}$-*points* of the variety, meaning those vectors in $\mathbb{K}^n$ on which the polynomials all vanish. In general this set may be empty because $\mathbb{K}$ is not assumed to be algebraically closed, and even if it is non-empty it may not resemble the whole variety. In particular, there is no reason for the set of $\mathbb{K}$-points to remember the ideal at all (in other words, Theorem 3.6 does not hold without the requirement that the field be algebraically closed). Nonetheless, we may define for any affine variety Z its $\mathbb{K}$-points as the set

$$Z(\mathbb{K}) = Z \cap \mathbb{K}^n,$$

where as before $Z = Z(\overline{\mathbb{K}})$ by definition. Moreover, we are often interested in regular functions with 'coefficients' in $\mathbb{K}$, which we formally define as the *ring of* $\mathbb{K}$*-regular functions*

$$\mathbb{K}[Z] = \mathbb{K}[x_1, \dots, x_n]/J \cap \mathbb{K}[x_1, \dots, x_n],$$

under the assumption that Z is *defined over* $\mathbb{K}$, meaning that $J = (J)$ defines Z and $J \cap \mathbb{K}[x_1, \dots, x_n]$ generates $J \subseteq \overline{\mathbb{K}}[x_1, \dots, x_n]$. We will return to these notions in Section 3.4.

Let us return to our main example SL_d which is defined over any field $\mathbb{K}$, since the coefficients of the polynomial $\det(\cdot) - 1$ are integers. Hence it makes sense to consider the ring of $\mathbb{K}$-regular functions

$$\mathbb{K}[\mathrm{SL}_d] = \mathbb{K}[g_{11}, \dots, g_{1d}, g_{21}, \dots, g_{2d}, \dots, g_{d1}, \dots, g_{dd}]/\langle \det(g) - 1 \rangle, \quad (3.7)$$

where $\mathbb{K}$ is the field of coefficients allowed in the polynomials. For us the field $\mathbb{K}$ will often be $\mathbb{R}$, $\mathbb{Q}_p$, or $\mathbb{Q}$.

A D-dimensional *algebraic representation of* SL_d *over* $\mathbb{K}$ is a D^2-tuple of polynomials

$$\phi_{ij}(g) \in \mathbb{K}[\mathrm{SL}_d]$$

for $1 \leqslant i, j \leqslant D$, which we think of as a matrix

$$\phi \in \mathrm{Mat}_D \left(\mathbb{K}[\mathrm{SL}_d] \right)$$

with the properties that $\phi(I_d) = I_D$ and

$$\phi(g)\phi(h) = \phi(gh) \quad (3.8)$$

for all $g, h \in \mathrm{SL}_d$. Equivalently, (3.8) could be required to hold as an abstract identity in the variables $g_{k\ell}, h_{k\ell}$ satisfying the polynomial condition

$$\det(g) = \det(h) = 1.$$

This equivalence follows from Hilbert's Nullstellensatz (Theorem 3.6) and Lemma 3.7.

An example of such a representation has been mentioned: If $A \in \mathrm{Mat}_d$ is symmetric and $g \in \mathrm{SL}_d$, then the map

$$A \longmapsto g A g^{\mathrm{t}} \quad (3.9)$$

is linear in A and polynomial in g. In fact, by identifying the space of symmetric matrices in $\mathrm{Mat}_d(\mathbb{Q})$ with the vector space $\mathbb{Q}^D$, where $D = \frac{d(d+1)}{2}$, we obtain a matrix representation $\phi(g) \in \mathrm{Mat}_D$ of (3.9) for which the matrix entries $\phi_{ij}(g)$ are polynomials of total degree 2 with coefficients in $\mathbb{Q}$. Moreover, $\phi(g)\phi(h)$ is the matrix corresponding to the composition

$$A \longmapsto hAh^{\mathrm{t}} \longmapsto g\left(hAh^{\mathrm{t}}\right)g^{\mathrm{t}} = (gh)A(gh)^{\mathrm{t}},$$

which is also represented by $\phi(gh)$. Therefore, (3.8) holds by uniqueness of matrix representations.

Let us give another representation of SL_d, which will be important in Section 3.3. The conjugation representation is defined by

$$\mathrm{Mat}_d \ni v \longmapsto gvg^{-1}$$

for $g \in \mathrm{SL}_d$. Since $\det(g) = 1$, the matrix g^{-1} has entries which are regular functions (since the inverse is calculated by taking the matrix consisting of the determinants of the minor matrices multiplied by the inverse of the determinant). Therefore, we can again chose a basis and get a $D = d^2$-dimensional representation† (defined over any field $\mathbb{K}$).

Proposition 3.8 (Rational stabilizer groups of points have closed orbits). *Let $\phi : \mathrm{SL}_d \to \mathrm{GL}_D$ be an algebraic representation over $\mathbb{Q}$, and let $v \in \mathbb{Q}^D$. Then the (rational) stabilizer subgroup*

$$\mathrm{Stab}_{\mathrm{SL}_d}(v) = \{g \in \mathrm{SL}_d \mid \phi(g)v = v\}$$

gives rise to a closed *orbit*

$$\mathrm{SL}_d(\mathbb{Z})I_d \,\mathrm{Stab}_{\mathrm{SL}_d}(v)(\mathbb{R}) \subseteq d$$

through the identity coset.

Notice that $\mathrm{Stab}_{\mathrm{SL}_d}(v)$ is itself a subgroup defined by polynomial equations (and hence will be seen to be an algebraic subgroup defined over $\mathbb{Q}$, once we define this notion in Section 3.4). The proof of Proposition 3.8 is much quicker than the discussion above, which was included to familiarize the notion of algebraic representations of SL_d.

Proof of Proposition 3.8. Notice that there are finitely many coefficients in (a representation of) the polynomials in $\phi(g)$. Let N be their common denominator, so that $\phi(\gamma) \in \frac{1}{N}\mathrm{Mat}_D(\mathbb{Z})$ for all $\gamma \in \mathrm{SL}_d(\mathbb{Z})$. Let M be the common denominator of the entries in v. Suppose that

$$\mathrm{SL}_d(\mathbb{Z})h_n \to \mathrm{SL}_d(\mathbb{Z})g = x, \tag{3.10}$$

† As will become more and more clear, part of the art in discussing algebraic groups and their representations will be to not really write down any concrete polynomials or regular functions (as these quickly become quite complicated).

with $h_n \in \mathrm{Stab}_{\mathrm{SL}_d}(v)(\mathbb{R})$ and $g \in \mathrm{SL}_d(\mathbb{R})$. We wish to show that

$$x \in \mathrm{SL}_d(\mathbb{Z})\,\mathrm{Stab}_{\mathrm{SL}_d}(v)(\mathbb{R}). \tag{3.11}$$

As in the proof of Proposition 3.1, we may rewrite (3.10) as $\gamma_n h_n = g\varepsilon_n$ with $\gamma_n \in \mathrm{SL}_d(\mathbb{Z})$, $\varepsilon_n \in \mathrm{SL}_d(\mathbb{R})$, and $\varepsilon_n \to I_d$. Applying these matrices to v via the representation ϕ shows that the sequence $(\phi(\gamma_n)v)$ lies in $\frac{1}{MN}\mathbb{Z}^D$ and converges,

$$\phi(\gamma_n)v = \phi(\gamma_n h_n)v = \phi(g)\phi(\varepsilon_n)v \longrightarrow \phi(g)v.$$

Therefore this sequence must stabilize, and so $\phi(\gamma_n)v = \phi(g)v$ for some n, which shows that $\gamma_n^{-1}g \in \mathrm{Stab}_{\mathrm{SL}_d}(v)(\mathbb{R})$, giving (3.11). □

Although the following is not needed for the proof above, let us try to understand a little more about $\mathrm{SL}_d(\mathbb{K})$ and algebraic representations of SL_d over any field $\mathbb{K}$.

(1) As shown in Lemma 1.24, $\mathrm{SL}_d(\mathbb{K})$ is generated by the elementary unipotent subgroups
$$U_{ij}(\mathbb{K}) = \{u_{ij}(t) = I + tE_{ij} \mid t \in \mathbb{K}\}$$
with $i \neq j$ and E_{ij} being the elementary matrix with (i,j)th entry 1 and all other entries 0.

(2) $\mathrm{SL}_d(\mathbb{K})$ coincides with its commutator subgroup
$$[\mathrm{SL}_d(\mathbb{K}), \mathrm{SL}_d(\mathbb{K})] = \langle [g,h] \mid g,h \in \mathrm{SL}_d(\mathbb{K})\rangle,$$
where $[g,h] = g^{-1}h^{-1}gh$. To see this, notice that if we choose an appropriate diagonal matrix a then
$$[u_{ij}(t), a] = u_{ij}(\alpha t)$$
for some $\alpha \neq 0$. Hence $[\mathrm{SL}_d(\mathbb{K}), \mathrm{SL}_d(\mathbb{K})] \supseteq U_{ij}(\mathbb{K})$ for all $i \neq j$, and the result follows by the remark above.

(3) It follows that $\mathrm{SL}_d(\mathbb{K})$ (resp. $\mathrm{SL}_d(\overline{\mathbb{K}})$) cannot have any abelian factors, and so $\det\phi(g) = 1$ for every algebraic representation over $\mathbb{K}$. By Theorem 3.6 and Lemma 3.7 this must therefore also hold as an identity in
$$\mathbb{K}[\mathrm{SL}_d] = \mathbb{K}[g_{ij} : i,j = 1,\dots,d]/\langle \det g - 1\rangle.$$

Exercises for Section 3.1

Exercise 3.1.1. Prove that $u_1^2 - u_1u_2 - u_2^2 \neq 0$ for $(u_1,u_2) \in \mathbb{Q}^2\smallsetminus\{0\}$ (a fact used in Example 3.3(2)).

Exercise 3.1.2. Prove the claim made in Example 3.3(3), by showing that the closed orbit $\mathrm{SL}_3(\mathbb{Z})\,\mathrm{SO}(Q_3)(\mathbb{R}) \subseteq 3$ has unbounded height.

Exercise 3.1.3. Let $A = \mathrm{SO}(1,1)(\mathbb{R}) \subseteq \mathrm{SL}_2(\mathbb{R})$. Show that every closed A-orbit corresponds (as indicated after Theorem 3.5) to a binary quadratic form with rational coefficients. Notice that this cannot hold for $K = \mathrm{SO}(2)(\mathbb{R})$.

Exercise 3.1.4. For any subspace $V \subseteq \mathbb{R}^d$ we define

$$L_V = \{g \in G \mid V = Vg \text{ and } g|_V \text{ preserves the volume}\}.$$

(1) Show that $\mathrm{SL}_d(\mathbb{Z})L_V \subseteq d$ is closed if V is a rational subspace.
(2) More generally, let $x_0 = \mathrm{SL}_d(\mathbb{Z})g_0$ and let V be a $\mathbb{Z}^d g_0$-rational subspace. Show that $x_0 L_V$ is closed.
(3) Let x_0 and V be as in (2). Let $G < \mathrm{SL}_d(\mathbb{R})$ be a closed subgroup such that $x_0 G$ is closed. Show that $x_0(G \cap L_V)$ is closed.

3.2 Rational Unipotent Subgroups

†In this section we will construct lattices in certain‡ connected, simply connected nilpotent Lie groups. By Ado's theorem (see Ado [**?**] or Knapp [**?**, Th. B.8]) and Engel's theorem (see Knapp [**?**, Th. 1.35],) such a group can be embedded into the upper triangular subgroup§

$$N = \left\{ \begin{pmatrix} 1 & * & * & \dots & * \\ & 1 & * & \dots & * \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix} \right\} \subseteq \mathrm{SL}_d(\mathbb{R}) \tag{3.12}$$

for some $d > 1$. A subgroup $G < \mathrm{SL}_d(\mathbb{R})$ is called *unipotent* if it is conjugated to a subgroup of N.

Theorem 3.9 (Lattices and Mal′cev basis for unipotent $\mathbb{Q}$-groups). *Let $G \leqslant \mathrm{SL}_d(\mathbb{R})$ be a connected unipotent subgroup whose Lie algebra $\mathfrak{g}$ is a rational subspace of $\mathfrak{sl}_d(\mathbb{R})$. Then*

$$\mathbb{G}(\mathbb{Z}) = G \cap \mathrm{SL}_d(\mathbb{Z})$$

is a uniform lattice in G. Moreover, writing $\ell = \dim \mathbb{G}$, there exist elements

$$v_1, \dots, v_\ell \in \mathfrak{g} \cap \mathfrak{sl}_d(\mathbb{Q})$$

† This section gives more examples of compact quotients of nilpotent groups, but otherwise is not essential for most of what follows. It will, however, become part of our proof of the Borel Harish-Chandra theorem in Section 7.4.

‡ Once we have discussed these notions it will be easy to see that the groups we will discuss here are of the form $G = \mathbb{G}(\mathbb{R})$ for a connected unipotent algebraic group $\mathbb{G}$ defined over $\mathbb{Q}$. As the theorem and its proof does not require this language we leave this fact to the reader.

§ Ado's and Engel's theorems are usually stated for a nilpotent Lie algebra instead of for the corresponding simply connected group, but the former implies the latter, see Exercise 3.2.1.

for which

$$\mathbb{G}(\mathbb{Z}) = \{\exp(k_1 v_1)\exp(k_2 v_2)\cdots\exp(k_\ell v_\ell) \mid k_1,\dots,k_\ell \in \mathbb{Z}\},$$

$$G = \mathbb{G}(\mathbb{R}) = \{\exp(s_1 v_1)\exp(s_2 v_2)\cdots\exp(s_\ell v_\ell) \mid s_1,\dots,s_\ell \in \mathbb{R}\},$$

and

$$F = \{\exp(s_1 v_1)\exp(s_2 v_2)\cdots\exp(s_\ell v_\ell) \mid s_1,\dots,s_\ell \in [0,1)\}$$

is a fundamental domain for $\mathbb{G}(\mathbb{Z})$ in G. Moreover, the map

$$s_1,\dots,s_\ell \longmapsto \exp(s_1 v_1)\exp(s_2 v_2)\cdots\exp(s_\ell v_\ell)$$

is a (polynomial) diffeomorphism between $\mathbb{R}^\ell$ and G. The vectors $v_1,\dots,v_\ell \in \mathfrak{g}$ are called a Mal′cev basis.

Proof. As $\mathfrak{g} \subseteq \mathfrak{sl}_d(\mathbb{R})$ is, by assumption, both a nilpotent Lie algebra and a rational subspace, the same holds for all the elements of the lower central series. In particular, $\mathfrak{g}_1 = [\mathfrak{g},\mathfrak{g}]$ is a rational subspace. By assumption $\mathfrak{g}$ can be conjugated into the Lie algebra of N. Therefore, the exponential map

$$\exp(v) = I + v + \tfrac{1}{2}v^2 + \cdots + \tfrac{1}{(d-1)!}v^{d-1}$$

is actually a polynomial map on $\mathfrak{g}$ with the logarithm map

$$\log(g) = g - I - \tfrac{1}{2}(g-I)^2 + \cdots + (-1)^d \tfrac{1}{d-1}(g-I)^{d-1}$$

as a polynomial inverse (which is defined on all of G). From this it follows that the linear group G is isomorphic to its Lie algebra $\mathfrak{g}$ if we equip the latter with the polynomial group operation $v * w = \log(\exp(v)\exp(w))$.

Recall that there is a — possibly immersed — Lie subgroup $G_1 \lhd G$ with Lie algebra $\mathfrak{g}_1$. This shows that for sufficiently small $v, w \in \mathfrak{g}_1$ the product $v * w$ lies in $\mathfrak{g}_1$. However, using the fact that the group product $v * w$ for $v, w \in \mathfrak{g}$ is a polynomial in v and w, we can now conclude[†] that $\mathfrak{g}_1 * \mathfrak{g}_1 \subseteq \mathfrak{g}_1$. Indeed, if ψ is a linear function vanishing on $\mathfrak{g}_1$ and $v \in \mathfrak{g}_1$ is sufficiently small, then $w \mapsto \psi(v * w)$ is a polynomial on $\mathfrak{g}_1$ which vanishes on all sufficiently small w. Hence $\psi(v*w) = 0$ for all $w \in \mathfrak{g}_1$. Reversing the roles of v and w, and using the fact that a linear subspace is defined by the collection of all linear functions that vanish on it, we see that $\mathfrak{g}_1 * \mathfrak{g}_1 \subseteq \mathfrak{g}_1$. However, this shows that $G_1 = \exp(\mathfrak{g}_1)$ is simply the isomorphic image of the Lie ideal $\mathfrak{g}_1$ and so is a normal closed connected subgroup of G. Note furthermore that the Lie algebra of G/G_1 equals $\mathfrak{g}/\mathfrak{g}_1$. Hence G/G_1 is abelian and can be identified with its Lie algebra under the exponential map.

[†] Once we have introduced the notion of Zariski density we will see that this argument uses the fact that the Hausdorff (that is, standard) neighborhood of $(0,0) \in \mathfrak{g}_1 \times \mathfrak{g}_1$ is Zariski dense in $\mathfrak{g}_1 \times \mathfrak{g}_1$

As $m = \dim(G_1) < \ell = \dim(G)$ and the Lie algebra $\mathfrak{g}_1$ of G_1 is rational, we may assume that the theorem already holds for the unipotent subgroup G_1. So let $v_1, \dots, v_m$ be the Mal′cev basis for G_1 and the uniform lattice $G_1(\mathbb{Z}) = G_1 \cap \mathrm{SL}_d(\mathbb{Z})$. Let $F_1 \subseteq G_1$ be a fundamental domain as in the corollary for $G_1(\mathbb{Z})$ in G_1. Let $V \subseteq \mathfrak{g}$ be a rational linear complement to $\mathfrak{g}_1 < \mathfrak{g}$. We claim that the image of $G(\mathbb{Z})$ in the abelian group $G/G_1 \simeq \mathfrak{g}/\mathfrak{g}_1 \simeq V$ is discrete. To see this, suppose that $K \subseteq G/G_1$ is a compact neighborhood of the identity and $G_1\gamma \in K \cap (G(\mathbb{Z})/G_1) \subseteq G/G_1$. Then we may modify the representative γ by elements of $G_1(\mathbb{Z})$ on the left to ensure that $\gamma \in F_1 \exp(V)$ belongs to a fixed compact set. As $G(\mathbb{Z})$ is discrete, this shows that there are only finitely many possibilities for γG_1, and so the image of $G(\mathbb{Z})$ in G/G_1 is discrete.

Next we claim that the image of $G(\mathbb{Z})$ modulo G_1 is a lattice in V. To see this, we have to find $\ell - m = \dim V$ linearly independent elements in the image of $G(\mathbb{Z})$ in $G/G_1 \simeq V$. This follows in turn since for every rational element $v \in V$ we have

$$\exp(Nv) = 1 + Nv + \tfrac{1}{2}N^2v^2 + \cdots + \tfrac{1}{(d-1)!}N^{d-1}v^{d-1} \in G(\mathbb{Z})$$

for a sufficiently divisible N.

We now choose $v_{m+1}, \dots, v_\ell \in \mathfrak{g}$ so that

$$\exp(v_j) \in G(\mathbb{Z})$$

for $j = m+1, \dots, \ell$ and the elements

$$G_1 \exp(v_{m+1}), \dots, G_1 \exp(v_\ell)$$

are a basis of the lattice obtained from $G(\mathbb{Z})$ in G/G_1. The elements

$$v_1, \dots, v_m, v_{m+1}, \dots, v_\ell$$

are now a Mal′cev basis. To see this, let $\gamma \in G(\mathbb{Z})$. Considering γG_1 we find $k_{m+1}, \dots, k_\ell \in \mathbb{Z}$ such that $\gamma G_1 = G_1 \exp(k_{m+1}v_{m+1}) \cdots \exp(k_\ell v_\ell)$, or equivalently $\gamma(\exp(k_{m+1}v_{m+1}) \cdots \exp(k_\ell v_\ell))^{-1} \in G_1$. Applying the inductive assumption it follows that $\gamma = \exp(k_1v_1) \cdots \exp(k_\ell v_\ell)$ for some $k_1, \dots, k_\ell \in \mathbb{Z}$. If $g \in G$ is arbitrary we may argue similarly to obtain unique $s_1, \dots, s_\ell \in \mathbb{R}$ with $g = \exp(s_1v_1) \cdots \exp(s_\ell v_\ell)$. Furthermore, if we consider g as a representative of a coset $G(\mathbb{Z})g$ we may define $k_j = \lfloor s_j \rfloor$ for $j = m+1, \dots, \ell$ and replace g by $(\exp(k_{m+1}v_{m+1}) \cdots \exp(k_\ell v_\ell))^{-1}g$. This ensures after the replacement that we have $s_{m+1}, \dots, s_\ell \in [0, 1)$. Applying the inductive assumption to

$$g(\exp(s_{m+1}v_{m+1}) \cdots \exp(s_\ell v_\ell))^{-1} \in G_1,$$

we deduce that the set F is indeed a fundamental domain. □

Exercises for Section 3.2

Exercise 3.2.1. In Knapp [**?**, Th. B.8, Th. 1.35] it is shown that any nilpotent Lie algebra can be embedded into the Lie algebra $\mathfrak{n}$ of N for some $d > 1$ (where N is defined by (3.12)). Use this (and the discussions regarding the exponential map of this chapter applied to $G = N$) to show that every connected, simply connected nilpotent Lie group can be embedded into N.

Exercise 3.2.2. Let G be a unipotent subgroup of $\mathrm{SL}_d(\mathbb{R})$ (with a rational Lie algebra). Show that G can be defined using polynomial equations (with rational coefficients).

3.3 Dirichlet's Unit Theorem and Compact Torus Orbits

†In this section we study another class of examples of orbits of rational stabilizer groups, which will also lead to a proof of Dirichlet's unit theorem[(13)]. Let

$$K = \mathbb{Q}(\zeta) \cong \mathbb{Q}[T]/\langle m(T)\rangle$$

be an algebraic number field generated by ζ, with minimal polynomial m of degree $d = [K : \mathbb{Q}] = \deg m(T)$. We may assume that $m(T)$ is monic. Let $O \subseteq K$ be an *order* (a subring of K that is isomorphic to $\mathbb{Z}^d$ as a group). Replacing ζ by $n\zeta$ has the effect of multiplying the non-leading coefficients of $m(T)$ by powers of n. Thus we may assume that $m(T) \in \mathbb{Z}[T]$, so that ζ is an algebraic integer‡, and, for example, $O = \mathbb{Z}[\zeta]$ is an order. Even though K can be embedded into $\mathbb{R}$ or $\mathbb{C}$, we prefer not to think of K as a subfield of $\mathbb{C}$ but rather as an abstract field, for instance as $K = \mathbb{Q}[T]/\langle m(T)\rangle$.

Theorem 3.10 (Dirichlet unit theorem). *Let O be an order in an algebraic number field K. The group $O^\times$ of units is isomorphic to $F \times \mathbb{Z}^{r+s-1}$, where $F \subseteq K$ is a finite group of roots of unity, r is the number of real embeddings $K \hookrightarrow \mathbb{R}$, and s is the number of pairs of complex embeddings $K \hookrightarrow \mathbb{C}$.*

The numbers r and s may be described as follows. Splitting $m(T)$ over $\mathbb{C}$ gives

$$m(T) = (T-\zeta_1)\cdots(T-\zeta_r)(T-\zeta_{r+1})(T-\overline{\zeta_{r+1}})\cdots(T-\zeta_{r+s})(T-\overline{\zeta_{r+s}}),$$

with $\zeta_1, \dots, \zeta_r \in \mathbb{R}$ and $\zeta_{r+1}, \dots, \zeta_{r+s} \in \mathbb{C}\smallsetminus\mathbb{R}$. As $\mathbb{Q}[T]/\langle m(T)\rangle \cong K$, the real embeddings $\phi_i : K \to \mathbb{R}$ are then all of the form

† This section provides interesting examples of algebraic groups (more precisely torus subgroups) and compact orbits, and connects these to algebraic number theory. It is not essential for most of the later chapters. It will, however, become part of our proof of the Borel Harish-Chandra theorem in Section 7.4.

‡ An *algebraic integer* is an algebraic number for which the monic minimal polynomial has integer coefficients.

$$\phi_i(f(T)) = f(\zeta_i)$$

for some $i = 1, \dots, r$, and the complex embeddings are all of the form

$$\phi_{r+i}(f(T)) = f(\zeta_{r+i}),$$

respectively

$$\overline{\phi_{r+i}}(f(T)) = f(\overline{\zeta_{r+i}}),$$

for $i = 1, \dots, s$ and $f \in \mathbb{Q}[T]$.

Another point of view is given by studying the multiplication by T map

$$\begin{aligned} \cdot T : \mathbb{Q}[T]/\langle m(T)\rangle &\longrightarrow \mathbb{Q}[T]/\langle m(T)\rangle \\ f(T) + \langle m(T)\rangle &\longmapsto Tf(T) + \langle m(T)\rangle, \end{aligned}$$

or, equivalently, the multiplication map $r \mapsto r\cdot\zeta$ on $K = \mathbb{Q}(\zeta)$. Considered as a linear map over $\mathbb{Q}$, the characteristic polynomial of the linear map $r \mapsto r\cdot\zeta$ is a rational polynomial which annihilates the map. It follows that $m(T)$ is the characteristic and also the minimal polynomial of the map. Therefore, the linear map $r \mapsto r \cdot \zeta$ has eigenvalues

$$\zeta_1, \dots, \zeta_r, \zeta_{r+1}, \overline{\zeta_{r+1}}, \dots, \zeta_{r+s}, \overline{\zeta_{r+s}}.$$

More generally, if $\cdot b$ is the multiplication by $b \in K = \mathbb{Q}(\zeta)$ map, then its eigenvalues (considered as a $\mathbb{Q}$-linear map on the vector space K over $\mathbb{Q}$) are again[†]

$$\phi_1(b), \dots, \phi_r(b), \phi_{r+1}(b), \overline{\phi_{r+1}(b)}, \dots, \phi_{r+s}(b), \overline{\phi_{r+s}(b)}.$$

We now discuss how to obtain a concrete matrix representation of K, which will allow us to use the results of Section 3.1. This is quite similar to how one can consider $\mathbb{C}$ as a field of 2×2 matrices using the correspondence

$$a + \mathrm{i}b \longleftrightarrow \begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

and it is helpful to view the construction below simply as an analogue of this. In order to make the construction a bit more flexible we start with another definition.

A *proper O-ideal* $J \subseteq O$ is an additive subgroup isomorphic to $\mathbb{Z}^d$ and for which

$$O = \{b \in K \mid bJ \subseteq J\}.$$

[†] This follows since $b = f(\xi)$ for some polynomial $f(T)$. If $b \in K\smallsetminus\mathbb{Q}$ then none of the eigenvectors are in $\mathbb{Q}$. In that case the eigenvectors only appear after 'extending the scalars', for example replacing $K \cong \mathbb{Q}^d$ by $K \otimes \mathbb{C} \cong \mathbb{C}^d$.

Clearly $J = O$ is a proper ideal[†]. For Theorem 3.10 $J = O$ would suffice, but it is of independent interest to note that any proper ideal can be used as the basis of the following construction.

So let $a_1, \dots, a_d$ be a $\mathbb{Z}$-basis of a proper O-ideal J. With this basis in mind, we may now identify the linear map $\cdot b$ on K with a matrix

$$\psi(b) \in \mathrm{Mat}_d(\mathbb{Q}).$$

We are again using row vectors so that $\cdot b : K \to K$ corresponds to multiplying row vectors $v \in \mathbb{Q}^n$ on the right by $\psi(b)$. By assumption, for $b \in K$ we have

$$b \in O \iff (\cdot(b))\,(a_i) \in J \text{ for all } i \iff \psi(b) \in \mathrm{Mat}_d(\mathbb{Z}),$$

and so also

$$b \in O^\times \iff \psi(b) \in \mathrm{GL}_d(\mathbb{Z}) = \{g \in \mathrm{Mat}_d(\mathbb{Z}) \mid \det(g) = \pm 1\}. \qquad (3.13)$$

Below we will be studying the subgroup

$$O^1 = \{b \in O^\times \mid \psi(b) \in \mathrm{SL}_d(\mathbb{Z})\};$$

this is either $O^\times$ or an index two subgroup of $O^\times$, and so it suffices to show the desired description for O^1.

Proposition 3.11 (Compact torus orbit). *Let $v_J = \psi(\zeta) \in \mathrm{Mat}_d(\mathbb{Z})$ and consider the stabilizer subgroup*

$$\mathbb{T}_J = \{g \in \mathrm{SL}_d \mid g v_J g^{-1} = v_J\}$$

for the conjugation action (that is, the centralizer *of v_J). Then the orbit*

$$(\mathrm{SL}_d(\mathbb{Z}) I_d)\, \mathbb{T}_J(\mathbb{R})$$

is compact, and

$$\mathbb{T}_J(\mathbb{Z}) = \mathrm{SL}_d(\mathbb{Z}) \cap \mathbb{T}_J(\mathbb{R}) = \psi(O^1).$$

In more technical language, the subgroup $\mathbb{T}_J$ is a special case of an *algebraic torus* (it is in fact a $\mathbb{Q}$-anisotropic $\mathbb{Q}$-torus). Moreover, the algebraic group $\mathbb{T}_J$ is closely related to the group ${}_{K|\mathbb{Q}}\mathbb{G}_m$ obtained by applying *restriction of scalars* to the multiplicative group $\mathbb{G}_m$ — it is the kernel of the $\mathbb{Q}$-split character ${}_{K|\mathbb{Q}}$ on ${}_{K|\mathbb{Q}}\mathbb{G}_m$. Minding our language we will not use these words often, but we will give a short introduction to these terms in Chapter 7.

Proof of Proposition 3.11. By Proposition 3.8, we know that the orbit is closed. We prove compactness along the lines of the proof of Proposition 3.2. For this we need a replacement for the quadratic form, and this is provided by the *norm form*

[†] Moreover, if O is the maximal order, then any ideal is a proper ideal.

$$(b) =_{K|\mathbb{Q}} (b) = \det \psi(b)$$

which is originally defined on K. Since K is a field, ${}_{K|\mathbb{Q}}(b) = 0$ for $b \in K$ if and only if $b = 0$, which is similar to the hypothesis in Proposition 3.2. Let us write

$$\iota(v) = v_1 a_1 + \cdots + v_d a_d$$

for $v \in \mathbb{Q}^d$, so that by assumption ι gives an isomorphism between $\mathbb{Z}^d$ and J as well as between $\mathbb{Q}^d$ and K. We also note that $\psi \circ \iota : \mathbb{Q}^d \to \mathrm{Mat}_d(\mathbb{Q})$ is linear, and so we can extend it to a linear map

$$\Psi : \mathbb{R}^d \to \mathrm{Mat}_d(\mathbb{R})$$

which also extends the norm form to the polynomial map $x \mapsto \det(\Psi(x))$ in d variables $x_1, \ldots, x_d$ of total degree d.

Now suppose that $(\mathrm{SL}_d(\mathbb{Z}) I_d)\, \mathbb{T}_J(\mathbb{R})$ is unbounded. Then for some m in $\mathbb{Z}^d \smallsetminus \{0\}$ and $h \in \mathbb{T}_J(\mathbb{R})$ the vector mh is very small. This implies that

$$|\det \Psi(mh)| < 1.$$

We claim that

$$\Psi(mh) = \Psi(m) h, \tag{3.14}$$

so that (in analogy to (3.5) on page 75) $|\det \Psi(mh)| = |\det \Psi(m)| < 1$, which forces $\det \Psi(m) = 0$ (since $\det \Psi(m) \in \mathbb{Z}$). However, $m \in \mathbb{Z}^d \smallsetminus \{0\}$ corresponding to some $b = \iota(m) \in J \smallsetminus \{0\}$ cannot have

$${}_{K|\mathbb{Q}}(b) = \det \Psi(m) = 0,$$

proving that $(\mathrm{SL}_d(\mathbb{Z}) I_d) \mathbb{T}_J(\mathbb{R})$ is bounded, and hence compact.

To prove the claim (3.14), and the statement $\mathbb{T}_J(\mathbb{Z}) = \psi(O^1)$ in the proposition, we would like to understand $\mathbb{T}_J$ better. Notice that

$$\{g \in \mathrm{Mat}_d \mid g v_J = v_J g\} \tag{3.15}$$

is a linear subspace defined by the requirement to commute with v_J. To analyze the dimension† of this subspace we may conjugate v_J over $\mathbb{C}$ to the diagonal matrix v_{diag} with eigenvalues

$$\zeta_1, \ldots, \zeta_r, \zeta_{r+1}, \overline{\zeta_{r+1}}, \ldots, \zeta_{r+s}, \overline{\zeta_{r+s}}.$$

As these are all different, the only matrices that commute with v_{diag} are diagonal matrices. This shows that the dimension of the subspace in (3.15) is d. Hence

$$\{g \in \mathrm{Mat}_d(\mathbb{Q}) \mid g v_J = v_J g\} = \psi(K)$$

† As the subspace in question is defined by rational equations, the dimension of it as a subspace of $\mathrm{Mat}_d(\mathbb{Q})$ over $\mathbb{Q}$ equals the dimension of it as a subspace of $\mathrm{Mat}_d(\mathbb{R})$ over $\mathbb{R}$ (and similarly for $\mathbb{C}$).

and taking the $\mathbb{R}$-linear hull we get

$$\{g \in \mathrm{Mat}_d(\mathbb{R}) \mid gv_J = v_J g\} = \langle \psi(\mathbb{K}) \rangle_{\mathbb{R}} = \Psi(\mathbb{R}^d). \tag{3.16}$$

The first of these equations implies that

$$\mathbb{T}_J(\mathbb{Z}) = \psi\left(\{b \in K \mid \psi(b) \in \mathrm{SL}_d(\mathbb{Z})\}\right) = \psi(O^1)$$

by (3.13).

Also notice that

$$\psi(ab) = \psi(a)\psi(b) \tag{3.17}$$

for $a, b \in K$ since ψ is giving the matrix representation of multiplication† by elements of K in the given basis. This may also be phrased as

$$\psi(\iota(mh)) = \psi(\iota(m))h \tag{3.18}$$

for $m \in \mathbb{Z}^d$ and $h \in \psi(K)$. Indeed, $a = \iota(m) \in J$ and $h = \psi(b)$ is the matrix which sends m corresponding to a to mh corresponding to ab, so that the left-hand sides of (3.17) and (3.18) agree. The right-hand sides agree tautologically, and so (3.18) follows. Equivalently we have shown $\Psi(mh) = \Psi(m)h$ for $m \in \mathbb{Z}^d$ and $h \in \psi(K)$. However, this is a linear equation in h which therefore also holds for $h \in \Psi(\mathbb{R}^d)$. In summary, we obtain (3.14) and the proposition follows. □

To finish the proof of Theorem 3.10 we need to analyze the structure of $\mathbb{T}_J(\mathbb{R})$.

Proposition 3.12 ($\mathbb{R}$-points of the torus subgroup). *With the notation as above,*

$$\mathbb{T}_J(\mathbb{R}) \cong M \times \mathbb{R}^{r+s-1},$$

where M is a compact linear group with connected component of the identity isomorphic to $(\mathbb{S}^1)^s$.

The pair of numbers (r, s) play a similar role for $\mathbb{T}_J$ as the signature of the associated quadratic form does for an orthogonal group. In this sense the result above is an analog of Theorem 3.5.

Proof of Proposition 3.12. We already did most of the work for this in the proof of Proposition 3.11. In fact, as in that proof, the group

$$\mathbb{T}_J(\mathbb{R}) = \{g \in \mathrm{SL}_d(\mathbb{R}) \mid gv_J = v_J g\}$$

is conjugate to‡

† As K is commutative, we do not have to worry about the order of multiplication of the matrices.

‡ Just as in the theory of Jordan normal forms, this follows quickly from consideration of $\mathbb{R}^d$ as an $\mathbb{R}[T]$-module, where T acts via v_J, which gives

$$\{g \in \mathrm{SL}_d(\mathbb{R}) \mid g v_{\zeta,\mathbb{R}} = v_{\zeta,\mathbb{R}} g\}$$

where $v_{\zeta,\mathbb{R}}$ is the block-diagonal matrix

$$v_{\zeta,\mathbb{R}} = \begin{pmatrix} \zeta_1 & & & & & \\ & \ddots & & & & \\ & & \zeta_r & & & \\ & & & \imath(\zeta_{r+1}) & & \\ & & & & \ddots & \\ & & & & & \imath(\zeta_{r+s}) \end{pmatrix} \in \mathrm{Mat}_d(\mathbb{R})$$

and $\imath$ is the map defined by

$$\imath : x + \mathrm{i}y \to \begin{pmatrix} x & y \\ -y & x \end{pmatrix}.$$

We use $v_{\zeta,\mathbb{R}}$ (instead of v_{diag}) to ensure that the conjugation takes place over $\mathbb{R}$, which is needed to analyze $\mathbb{T}_J(\mathbb{R})$. It is easy to check (for example, by a dimension argument as in the proof of Proposition 3.11) that

$$\{g \in \mathrm{Mat}_d(\mathbb{R}) \mid g v_{\zeta,\mathbb{R}} = v_{\zeta,\mathbb{R}} g\}$$

$$= \left\{ \begin{pmatrix} a_1 & & & & & \\ & \ddots & & & & \\ & & a_r & & & \\ & & & \imath(b_1) & & \\ & & & & \ddots & \\ & & & & & \imath(b_s) \end{pmatrix} \mid a_1, \dots, a_r \in \mathbb{R}, b_1, \dots, b_s \in \mathbb{C} \right\}.$$

Therefore $\mathbb{T}_J(\mathbb{R})$ is isomorphic to the multiplicative group

$$\{(a_1, \dots, a_r, b_1, \dots, b_s) \in \mathbb{R}^r \times \mathbb{C}^s \mid a_1 \cdots a_r |b_1|^2 \cdots |b_s|^2 = 1\}$$

which contains the non-compact part

$$\{(\mathrm{e}^{t_1}, \dots, \mathrm{e}^{t_r}, \mathrm{e}^{t_{r+1}}, \dots, \mathrm{e}^{t_{r+s}}) \mid t_1 + \cdots + t_r + 2t_{r+1} + \cdots + 2t_{r+s} = 0\},$$

and this is isomorphic (as a Lie group) to $\mathbb{R}^{r+s-1}$. The subgroup $M \subseteq \mathbb{T}_J(\mathbb{R})$ is then the subgroup isomorphic to the 'group of signs'

$$\{(\varepsilon_1, \dots, \varepsilon_r, z_1, \dots, z_s) \mid \varepsilon_i \in \{\pm 1\}, |z_i| = 1, \varepsilon_1 \cdots \varepsilon_r = 1\}.$$

$$\mathbb{R}^d \cong \mathbb{R}[T]/\langle T - \zeta_1 \rangle \times \cdots \times \mathbb{R}[T]/\langle T - \zeta_r \rangle \times \mathbb{R}[T]/\langle p_{\zeta_{r+1}}(T) \rangle \times \cdots \times \mathbb{R}[T]/\langle p_{\zeta_{r+s}}(T) \rangle,$$

where $p_{\zeta_{r+1}}(T), \dots, p_{\zeta_{r+s}}(T)$ are the quadratic real minimal polynomials of $\zeta_{r+1}, \dots, \zeta_{r+s} \in \mathbb{C}$. We refer to Hungerford [**?**, Ch. VII] for the details.

□

Proof of Theorem 3.10. By Proposition 3.11, O^1 is isomorphic to a (uniform) lattice in $\mathbb{T}_J(\mathbb{R})$, which by Proposition 3.12 is isomorphic to the abelian group $M \times \mathbb{R}^{r+s-1}$. Taking the quotient by M we obtain a uniform lattice in $\mathbb{R}^{r+s-1}$, which must be generated by $r+s-1$ elements. Suppose that $b_1, \dots, b_{r+s-1} \in O^1$ are elements that give rise to a $\mathbb{Z}$-basis of the lattices in $\mathbb{R}^{r+s-1}$. Then $b_1, \dots, b_{r+s-1}$ generate O^1 up to the kernel of the map from O^1 to $\mathbb{R}^{r+s-1}$. However, this kernel F maps under ψ and the isomorphism to $M \times \mathbb{R}^{r+s-1}$ to the compact group M (with discrete image) and so must be finite. □

3.3.1 More on Compact Orbits for the Diagonal Subgroup

†The proof above, while it fits naturally into our discussion, is certainly not the shortest proof of Dirichlet's unit theorem. However, the set-up used here can be used further to discuss interesting distribution properties of compact orbits arising from number fields. For this we define the *complete Galois embedding*

$$\phi = (\phi_1, \dots, \phi_r, \phi_{r+1}, \dots, \phi_{r+s}) : K \to \mathbb{R}^r \times \mathbb{C}^s \cong \mathbb{R}^{r+2s} \tag{3.19}$$

(which clearly is an embedding, since each ϕ_i is injective). We define (r, s) to be the *type* of the number field (as mentioned this plays the role of the signature of a quadratic form), and define $\mathbb{T}_{r,s}$ to be the centralizer of a regular matrix

$$v_{r,s} = \begin{pmatrix} \alpha_1 &&&&& \\ & \ddots &&&& \\ && \alpha_r &&& \\ &&& \imath(\beta_1) && \\ &&&& \ddots & \\ &&&&& \imath(\beta_s) \end{pmatrix}$$

with pairwise different $\alpha_j \in \mathbb{R}$ and $\beta_j \in \mathbb{C} \smallsetminus \mathbb{R}$.

Proposition 3.13 (Ideal classes and torus orbits). *Let K be a number field of type (r, s), and let O be an order in K. Then for any proper O-ideal $J \subseteq K$ the normalized lattice*

$$x_J = \frac{1}{(\phi(J))^{1/d}} \phi(J) \in d$$

† The remainder of Section 3.3 will not be needed again.

has compact orbit under $\mathbb{T}_{r,s}(\mathbb{R})$. Two ideals J_1, J_2 give rise to the same orbit if and only if they are ideals in the same number field (and order), and are equivalent (that is, there exists some $a \in K \smallsetminus \{0\}$ with $J_1 = aJ_2$).

For the relationship above between ideal classes and compact orbits, we need to agree on the following convention. For a given field K and all its isomorphic copies we agree to pick one representative, say K, and one complete Galois embedding ϕ as in (3.19).

Proof of Proposition 3.13. Let $K = \mathbb{Q}(\zeta)$, O, and J be given. We will use the same notation as used in Proposition 3.11. Recall that $\{a_1, \dots, a_d\}$ is a basis of J, and (replacing a_1 with $-a_1$ if necessary) we may assume that

$$g_J = \frac{1}{(\phi(J))^{1/d}} \begin{pmatrix} \phi(a_1) \\ \vdots \\ \phi(a_d) \end{pmatrix}$$

has determinant one. By construction, $x_J = \mathbb{Z}^d g_J$; also notice that g_J is up to the scalar the matrix representation of the map ϕ from K (with the basis $\{a_1, \dots, a_d\}$) to $\mathbb{R}^r \times \mathbb{C}^s$ (with the standard basis). Furthermore, recall that $v_J = \psi(\zeta)$ is the matrix representation of multiplication by ζ on K (with basis $a_1, \dots, a_d$). In $\mathbb{R}^r \times \mathbb{C}^s$ multiplication by ζ corresponds to multiplying the various coordinates by $\phi_1(\zeta), \dots, \phi_r(\zeta)$ and to applying the matrices corresponding to the complex numbers $\phi_{r+1}(\zeta), \dots, \phi_{r+s}(\zeta)$ respectively; that is, to an application of a block-diagonal matrix $v_{\zeta,\mathbb{R}}$. This shows (as we are using row vectors) that

$$v_J g_J = g_J v_{\zeta,\mathbb{R}}. \tag{3.20}$$

Now $v_{\zeta,\mathbb{R}}$ is of the same type as $v_{r,s}$ and defines the same centralizer $\mathbb{T}_{r,s}$. Therefore,

$$\mathbb{T}_{r,s} = g_J^{-1} \mathbb{T}_J g_J$$

since

$$g_J^{-1} g g_J v_{\zeta,\mathbb{R}} = g_J^{-1} g v_J g_J = g_J^{-1} v_J g g_J = v_{\zeta,\mathbb{R}} g_J^{-1} g g_J$$

for any $g \in \mathbb{T}_J$. Moreover,

$$\mathrm{SL}_d(\mathbb{Z}) g_J \mathbb{T}_{r,s}(\mathbb{R}) = \mathrm{SL}_d(\mathbb{Z}) \mathbb{T}_J(\mathbb{R}) g_J$$

is compact by Proposition 3.11.

Notice that if we choose a different basis of J, then this does not change the point $x_J \in d$. Also notice that if $J' = bJ$ for some $b \in K^\times$ then $ba_1, \dots, ba_d$ is a basis of J', and using this basis we see by (3.20) (which by the same argument also holds for $\cdot b$ instead of $\cdot \zeta$) that

$$g_{J'} = \psi(b) g_J = g_J v_{b,\mathbb{R}}.$$

Since $v_{b,\mathbb{R}} \in \mathbb{T}_{r,s}(\mathbb{R})$ this shows that

$$x_{J'} \in x_J \mathbb{T}_{r,s}(\mathbb{R}),$$

which is the first direction of the second claim in the proposition.

Let now J (and J') be a proper O (respectively O')-ideal in a number field K (respectively K'), let x_J, $x_{J'}$ be the corresponding elements of d, and assume that

$$x_{J'} = x_J t$$

for some $t \in \mathbb{T}_{r,s}(\mathbb{R})$. By the definition of properness for an O-ideal J we have

$$\begin{aligned} O &= \{a \in K \mid aJ \subseteq J\} \\ &\cong \{v \in \langle \psi(\mathbb{K}) \rangle_{\mathbb{R}} \mid \mathbb{Z}^d v \subseteq \mathbb{Z}^d\} && \text{(via } v = \psi(a)) \\ &= \{v \in \mathrm{Mat}_d(\mathbb{R}) \mid v v_J = v_J v \text{ and } \mathbb{Z}^d v \subseteq \mathbb{Z}^d\} \\ &\cong \{v \in \mathrm{Mat}_d(\mathbb{R}) \mid v v_{r,s} = v_{r,s} v \text{ and } x_J v \subseteq x_J\}, \end{aligned}$$

via conjugation by g_J. The latter set comprises all block diagonal matrices with entries $\phi(a)$ for all $a \in O$. For the lattices $x_{J'}$ and x_J, this implies that $O \cong O'$ and hence $K \cong K'$. By the convention discussed just after the proposition, this means that $K = K'$, and that the same complete Galois embedding ϕ is used. By the argument above, this also implies that $O = O'$. Suppose that $a_1, \dots, a_d$ is a basis of J, so that $x_J = \mathbb{Z}^d g_J$ as before. Choosing the basis $a'_1, \dots, a'_d$ of J' correctly gives $x_{J'} = \mathbb{Z}^d g_{J'}$ and $g_{J'} = g_J t$. This shows that $\phi_i(a'_j) = \phi_i(a_j) t_i$ where t_i (in $\mathbb{R}$ or $\mathbb{C}$) is the ith entry of the block-diagonal matrix $t \in \mathbb{T}_{r,s}(\mathbb{R})$. This implies that

$$t_i = \phi_i \left(\frac{a'_j}{a_j} \right)$$

is independent of j. Hence there exists some $b \in K$ with

$$t_i = \phi_i(b)$$

for $i = 1, \dots, r+s$, and it follows that $J' = bJ$. □

We remark that for a given order O there are only finitely many inequivalent ideal classes of proper O-ideals. This observation makes the following folklore conjecture (generalizing results and conjectures of Linnik) well-formulated.

Conjecture 3.14. For a given order O in an algebraic number field K of type (r, s), let μ_O be the probability measure on d obtained from normalizing the sum of the $\mathbb{T}_{r,s}(\mathbb{R})$-invariant probability measures on $x_J \mathbb{T}_{r,s}(\mathbb{R})$ for the various ideal classes of proper O-ideals. Then, as the *discriminant* $D = ((\phi(J)))^2$ goes to infinity, all of the weak*-limit of the measures μ_O are Haar measures of finite volume orbits xL for some closed linear subgroup $L \subseteq \mathrm{SL}_d(\mathbb{R})$.

This conjecture has been shown for $d = 2$ by Duke [?] (using subconvexity of L-functions, building on a breakthrough of Iwaniec [?]), and for $d = 3$ and type $r = 3$, $s = 0$ by Einsiedler, Lindenstrauss, Michel and Venkatesh [?] (by combining subconvexity bounds for L-functions with ergodic methods). More accessible but weaker results are contained in [?] and [?].

Exercises for Section 3.3

Exercise 3.3.1. (a) Let $d \geqslant 2$. Show that the compact orbits of $\mathbb{T}_{(d,0)}(\mathbb{R})$ (of type $(d, 0)$) in d are all of the form $x_J\mathbb{T}$ for some proper O-ideal and some order $O \subseteq K$ in a totally real number field.
(b) Show that this is not the case for the type $(0, d/2)$ (with d even).
(c) Decide the same question for the remaining cases.

3.4 Linear Algebraic Groups

In this section and in Chapter 7 we will introduce linear algebraic groups, and will link this concept to the theory of linear Lie groups, pointing out the obvious similarities as well as some of the more subtle differences between the theories. We start with the basic definitions, but in order to avoid being too diverted by this important (and large) theory, we will be brief at times.

3.4.1 Basic Notions of Algebraic Varieties

Let $\mathbb{K}$ be a field† and let $\overline{\mathbb{K}}$ denote an algebraic closure of $\mathbb{K}$. A subset $S \subseteq \overline{\mathbb{K}}^d$ is called *Zariski closed* if $S = Z(\mathscr{J})$ is the variety $Z(\mathscr{J})$ defined by a subset or, without loss of generality, an ideal $\mathscr{J} \subseteq \overline{\mathbb{K}}[x_1, \dots, x_d]$. A subset $S \subseteq \overline{\mathbb{K}}^d$ is also called Zariski $\mathbb{K}$-closed if $\mathscr{J}$ can be chosen in $\mathbb{K}[x_1, \dots, x_d]$. The Zariski closed subsets are the closed sets of a topology, which is called the *Zariski topology*. This is easily checked:

- If $S_1 = Z(\mathscr{J}_1)$ and $S_2 = Z(\mathscr{J}_2)$ then $S_1 \cup S_2 = Z(\mathscr{J}_1\mathscr{J}_2)$.
- If $S_\alpha = Z(\mathscr{J}_\alpha)$ for $\alpha \in A$, then

$$\bigcap_{\alpha \in A} S_\alpha = Z\left(\bigcup_{\alpha \in A} \mathscr{J}_\alpha\right).$$

† We will generally be interested in the cases $\mathbb{R}$, $\mathbb{Q}_p$ and $\mathbb{Q}$, but will only assume that the field has characteristic zero a little later.

If $\mathbb{K} = \mathbb{R}$, $\mathbb{K} = \mathbb{C}$, or $\mathbb{K} = \mathbb{Q}_p$, then clearly every Zariski closed (or Zariski open) subset is also closed (or open) in the usual sense. For most of the derived properties (density, connectedness) this is not clear and indeed is often false. We will always say Zariski open, Zariski closed, Zariski dense, and so on, if we refer to properties of the Zariski topology. When we use the words open, closed, dense, and so on, then this will refer to the metric (often also referred to as the *Hausdorff*) topology of $\mathbb{R}^d$, $\mathbb{C}^d$, or $\mathbb{Q}_p^d$ derived from the norms on these spaces.

A variety (equivalently, a Zariski closed set) is called *Zariski connected*† or *irreducible* if it is not a union of two proper Zariski closed subsets. Equivalently, a variety Z is irreducible if its ring of regular functions

$$\overline{\mathbb{K}}[Z] = \overline{\mathbb{K}}[x_1, \dots, x_d]/J(Z)$$

is a principal ideal domain (that is, without zero divisors).

Assume now that $Z = Z(\mathscr{J})$ is a connected variety. Then we can form the field of rational functions $\overline{\mathbb{K}}(Z)$ comprising all quotients $\frac{f}{g}$ with $f, g \in \overline{\mathbb{K}}[Z]$ and $g \neq 0$. The *transcendence degree*‡ (see Hungerford [**?**, Sec. VI.1]) of $\overline{\mathbb{K}}(Z)$ is the *dimension* $\dim(Z)$ of the variety Z. Notice that if $Z = \overline{\mathbb{K}}^d$ then the dimension of Z is d, and if Z is defined by a single irreducible polynomial

$$f \in \overline{\mathbb{K}}[x_1, \dots, x_d]$$

(in which case Z is called a *hypersurface*), then the dimension of Z is $(d-1)$. The following lemma further reinforces our intuition concerning this notion of dimension.

Lemma 3.15 (Strict monotonicity of dimension). *Suppose that $Z_2 \subseteq Z_1$ is a proper connected subvariety of a connected variety $Z_1 \subseteq \overline{\mathbb{K}}^d$. Then*

$$\dim Z_2 < \dim Z_1.$$

Proof. By definition

$$\overline{\mathbb{K}}[Z_1] = \overline{\mathbb{K}}[x_1, \dots, x_d]/\mathscr{J}_1,$$

with $\mathscr{J}_1 = J(Z_1)$, has transcendence degree $k = \dim Z_1$. By reordering the variables if necessary, we may assume that

† This definition does not match the topological definition of connectedness, but it will come closer to doing so so in the context of algebraic subgroups.

‡ A field extension $\mathbb{F}|\mathbb{K}$ has transcendence degree n if $\mathbb{F}$ contains n mutually transcendental elements $f_1, \dots, f_n \in \mathbb{F}$ (that is, elements with the property that the evaluation map

$$\mathbb{K}[T_1, \dots, T_n] \ni g \mapsto g(f_1, \dots, f_n)$$

is injective) but does not contain $n+1$ mutually transcendental elements.

$$x_1 + \mathscr{J}_1, \dots, x_k + \mathscr{J}_1 \in \overline{\mathbb{K}}[Z_1] \tag{3.21}$$

are algebraically independent, and

$$x_{k+1} + \mathscr{J}_1, \dots, x_d + \mathscr{J}_1$$

are algebraically dependent on the elements in (3.21). All other regular or rational functions in $\overline{\mathbb{K}}(Z_1)$ are then algebraically dependent on the elements in (3.21). It follows that

$$\overline{\mathbb{K}}(Z_1) \cong \overline{\mathbb{K}}(x_1, \dots, x_k)\,[x_{k+1} + \mathscr{J}_1, \dots, x_d + \mathscr{J}_1]$$

is a finite field extension of the field of rational functions in the first k variables.

Since $Z_2 \subseteq Z_1$ is a proper subvariety, there exists some $f \in J(Z_2) \smallsetminus J(Z_1)$. As $f + \mathscr{J}_1$ is non-zero in $\overline{\mathbb{K}}(Z_1)$, there exists some

$$g + \mathscr{J}_1 \in \overline{\mathbb{K}}(x_1, \dots, x_k)\,[x_{k+1} + \mathscr{J}_1, \dots, x_d + \mathscr{J}_1]$$

such that $fg + \mathscr{J}_1 = 1 + \mathscr{J}_1$. Clearing the denominators (which belong to $\overline{\mathbb{K}}[x_1, \dots, x_k]$) in this relation, we find that there exists some $g_1 \in \overline{\mathbb{K}}[Z_1]$ such that

$$fg_1 + \mathscr{J}_1 = h + \mathscr{J}_1$$

for some non-zero $h \in \overline{\mathbb{K}}[x_1, \dots, x_k] \cap \mathscr{J}(Z_2)$. This shows that the transcendence degree of $\overline{\mathbb{K}}(Z_2)$ is less than or equal to $k - 1$. $\square$

Assume again that $Z \subseteq \overline{\mathbb{K}}^d$ is a connected k-dimensional variety. A point $x^{(0)} \in Z$ is called *smooth* if the 'tangent space' in the variables $u_1, \dots, u_d$ defined by

$$\sum_{j=1}^{d} u_j \partial_{x_j} f(x^{(0)}) = (u_1, \dots, u_d) \cdot \nabla f(x^{(0)}) = 0$$

for all $f \in J(Z)$, is k-dimensional. The partial derivatives are defined as abstract linear maps on the space of polynomials (so that the definition matches the usual maps if $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$). It satisfies the usual properties (the product and chain rules, for example) over any field $\mathbb{K}$. The reader may quickly decide which points of the variety defined by the equation $y^2 = x^3$ are smooth in this sense (and thus see why the definition makes sense and accords with geometrical intuition; see also Lemma 3.18). A variety is called *smooth* if every point of the variety is a smooth point.

Lemma 3.16 (Most points are smooth). *Let $Z \subseteq \overline{\mathbb{K}}^d$ be a connected variety and suppose the characteristic $\mathbb{K}$ of the field $\mathbb{K}$ is zero. Then the set of smooth points of Z is a non-empty Zariski open subset of Z. Moreover, the tangent space has at no point of Z a dimension smaller than* $\dim Z$.

The lemma should indeed be interpreted as saying that most points of a connected variety are smooth. This is because a non-empty Zariski open subset of a connected variety is automatically Zariski dense. Moreover, Zariski dense and Zariski open subsets of any variety have a nice intersection property†: every finite intersection of Zariski dense and open subsets is again Zariski dense and open.

Proof of Lemma 3.16. Let $k = \dim Z$, and assume again that

$$x_1 + J(Z), \ldots, x_k + J(Z) \in \overline{\mathbb{K}}(Z)$$

are algebraically independent while

$$x_{k+1} + J(Z), \ldots, x_d + J(Z)$$

are algebraically dependent on

$$x_1 + J(Z), \ldots, x_k + J(Z).$$

Thus there exists, for every $\ell \in \{k+1, \ldots, d\}$ a non-zero polynomial

$$f_\ell \in \overline{\mathbb{K}}[x_1, \ldots, x_\ell] \cap J(Z)$$

of minimal degree in x_ℓ for which (viewed as a polynomial in x_ℓ) the non-zero coefficients do not belong to $\overline{\mathbb{K}}[x_1, \ldots, x_{\ell-1}] \cap J(Z)$. Since $\mathbb{K} = 0$, we get‡

$$g_\ell = \partial_{x_\ell} f_\ell \notin J(Z).$$

Using the derivative $\nabla(f_\ell)$ (for $\ell = k+1, \ldots, d$) of these polynomials (as equations that define the tangent space) we see that every point outside the proper subvariety defined by the ideal

$$\langle g_{k+1} \cdots g_d, J(Z) \rangle$$

(that is, every point in a non-empty Zariski open subset O) has a tangent space of dimension less than or equal to k. To see that these points are smooth points of the variety we have to show that the tangent space is indeed k-dimensional. We show this first§ on an even smaller Zariski open subset O'.

We claim that there exists some non-zero $h \in \overline{\mathbb{K}}[x_1, \ldots, x_d] \smallsetminus J$ with

$$hJ \subseteq (f_{k+1}, \ldots, f_d).$$

† For a connected variety this is easy to see. For a general variety this follows for example from the decomposition discussed in Lemma 3.17.

‡ If $\mathbb{K} = p$ and it so happens that f_ℓ is a polynomial in $x_1, \ldots, x_{\ell-1}, x_\ell^p$ then $\partial_{x_\ell} f_\ell = 0$. With more care this problem can be dealt with — we refer to Hartshorne [**?**] for the details.

§ We use this step below to show that we can never have a tangent space of dimension strictly less than k, hence we cannot rely on this fact here.

Assuming the claim we see that $hf = g'_1 f_{k+1} + \cdots + g'_{d-k} f_d$, so that

$$\nabla(hf) = \nabla(h)f + h\nabla(f) = \\ \nabla(g'_1)f_{k+1} + g'_1\nabla(f_{k+1}) + \cdots + \nabla(g'_{d-k})f_d + g'_{d-k}\nabla(f_d).$$

After evaluation at any point $x \in Z$ we then get

$$h(x)\nabla(f)(x) = g'_1(x)\nabla(f_{k+1})(x) + \cdots + g'_{d-k}(x)\nabla(f_d)(x)$$

which expresses $\nabla(f)(x)$ as a linear combination of $\nabla(f_j)(x)$ for

$$j = k+1, \ldots, d$$

if only $h(x) \neq 0$. This shows that on the Zariski open set

$$O' = Z \smallsetminus Z(hg_{k+1} \cdots g_d)$$

every tangent space is exactly k-dimensional.

We now prove the claim. As J is finitely generated and prime, we only have to show that for every $f \in J$ there is some $h \notin J$ with $hf \in (f_{k+1}, \ldots, f_d)$. If $f \in \overline{\mathbb{K}}[x_1, \ldots, x_{k+1}] \cap J$, then we can take h to be a power of the leading coefficient of f_{k+1} (considered as a polynomial in x_{k+1} with coefficients in $\overline{\mathbb{K}}[x_1, \ldots, x_k]$). In fact, with this choice of h we ensure that we can apply division with remainder† to obtain $hf = af_{k+1} + b$ where $b = 0$ as it has smaller degree in x_{k+1} than f_{k+1} does and belongs to J. By induction on ℓ the same argument applies for any $f \in \overline{\mathbb{K}}[x_1, \ldots, x_{\ell+1}] \cap J$ (where we will have $b \in \overline{\mathbb{K}}[x_1, \ldots, x_\ell] \cap J$ by the same argument).

It remains to show that the set of smooth points is Zariski open and that at no point of Z the tangent space has dimension $< k$. If now

$$x^{(0)} = \left(x_1^{(0)}, \ldots, x_d^{(0)}\right) \in Z$$

is an arbitrary smooth point, or more generally a point whose tangent space has dimension $K \leqslant k$, then we may reorder the variables so that the tangent space projects onto the subspace spanned by the first K basis vectors, and so that for each $\ell \in \{K+1, \ldots, d\}$ there exists some $f_\ell \in J(Z)$ such that

$$(\nabla f_\ell)_\ell \neq 0$$

but

$$(\nabla f_\ell)_j = 0$$

for $j \in \{K+1, \ldots, d\} \smallsetminus \{\ell\}$. It follows that the determinant

---

† Formally we apply division with remainder in the Euclidean domain $\overline{\mathbb{K}}(x_1, \ldots, x_k)[x_{k+1}]$, and later in the argument in the Euclidean domain $\overline{\mathbb{K}}(x_1 + J, \ldots, x_\ell + J)[x_{\ell+1}]$.

$$g = \det\left(\nabla f_\ell\right)_j,$$

where $\ell, j \in \{K+1, \dots, d\}$, does not vanish at the point $x^{(0)}$. Unfolding the definition shows that any other point

$$x \in O_g = Z \smallsetminus Z(g)$$

is also a point at which the tangent space has dimension less than or equal to K, which is less than or equal to k.

If $K < k$ at some point $x^{(0)}$, then we have found a non-empty Zariski open subset O_g on which all points have tangent spaces of dimension less than or equal to K. However, as Z is irreducible this set would have to intersect the non-empty Zariski open subset O' (on which the tangent spaces are known to be k-dimensional) nontrivially, which would give a contradiction.

Therefore, there is no point where the tangent space has dimension strictly less than k, and so applying the argument for $K = k$ we see that the set of smooth points is Zariski open (and Zariski dense). □

To generalize the notion of smoothness to general varieties we need another lemma.

Lemma 3.17 (Decomposition into Zariski connected components). *Let Z be a variety. Then Z is a finite union*

$$Z = \bigcup_{i=1}^{n} Z_i$$

of connected varieties $Z_1, \dots, Z_n$, where we may and will assume that $Z_i \not\subseteq Z_j$ for $i \neq j$. We will refer to $Z_1, \dots, Z_n$ as the Zariski connected components. *We claim furthermore that the decomposition into Zariski connected components is (up to their order) unique.*

We note that if Z is a hypersurface, then the claimed existence and uniqueness follow quickly from the statement that $\overline{\mathbb{K}}[x_1, \dots, x_d]$ is a unique factorization domain.

Sketch of Proof of Lemma 3.17. The existence of the decomposition follows from the fact that $\overline{\mathbb{K}}[x_1, \dots, x_d]$ is Noetherian. We sketch the argument. If $J = J(Z) \subseteq \overline{\mathbb{K}}[x_1, \dots, x_d]$ is not a prime ideal, then there exist

$$f_1, f_2 \in \overline{\mathbb{K}}[x_1, \dots, x_d] \smallsetminus J$$

with $f_1 f_2 \in J$. We may define $J_1 = \langle J, f_1 \rangle$ and $J_2 = \langle J, f_2 \rangle$. Notice that $J_1 J_2 \subseteq J \subseteq J_1 \cap J_2$. If both of these are prime ideals, then we are done (see below). If not, then we may assume that J_1 is not a prime ideal, and repeating the argument gives ideals $J_{1,1}, J_{1,2}$. We do the same for J_2 if J_2 is not a prime ideal, and repeat as necessary. By the Noetherian property this construction has to terminate after finitely many steps. In other words, we can always find a finite tree with $\mathscr{J}$ at the top and prime ideals at the bottom, as illustrated in Figure 3.1.

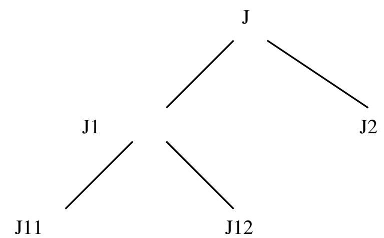

Fig. 3.1 Ideals inside $\mathscr{J}$.

If the prime ideals found are denoted $P_1, \dots, P_n$, then we have (by construction of the prime ideals) that

$$P_1 \cdots P_n \subseteq J \subseteq \bigcap_{i=1}^{n} P_i. \tag{3.22}$$

This translates to the statement

$$Z = \bigcup_{i=1}^{n} Z(P_i).$$

If the list of prime ideals has repetitions, we simply remove the repetitions. Also, if $P_i \subseteq P_j$ for $i \neq j$ then $Z(P_i) \supseteq Z(P_j)$ and we remove P_j from the list. Using $J = (J)$, we can now show that (3.22) still holds for the shortened list. Finally, uniqueness follows directly from the definitions: If $P_1, \dots, P_n$ and $P'_1, \dots, P'_m$ both satisfy (3.22) (and are minimal lists), then for every

$$j \in \{1, \dots, m\}$$

we have

$$P_1 \cdots P_n \subseteq P'_j$$

and since P'_j is a prime ideal there exists some $i(j)$ with $P_{i(j)} \subseteq P_j$. Similarly, there exists for every $i \in \{1, \dots, n\}$ some $j(i)$ with $P'_{j(i)} \subseteq P_i$. Since now $P_{i(j(i))} \subseteq P_i$ for every i and $P'_{j(i(j))} \subseteq P'_j$ for every j, it follows that $i(\cdot)$ and $j(\cdot)$ are inverses of each other, $m = n$, and $P'_{j(i)} = P_i$. $\square$

A point $x^{(0)} \in Z$ of a (not necessarily connected) variety is *smooth* if $x^{(0)}$ belongs to precisely one of the connected varieties $Z_i \subseteq Z$ as above, and $x^{(0)}$ is a smooth point of Z_i. Lemma 3.18 now says that inside every variety Z the subset of points that are smooth points of Z is a Zariski open and dense subset of Z.

3.4.2 Properties concerning the field

One smooth $\mathbb{K}$-point of a variety already gives rise to many other $\mathbb{K}$-points, if $\mathbb{K}$ is a local field.

Lemma 3.18 (Neighborhoods of smooth points). *Let $Z \subseteq \mathbb{C}^d$ be a k-dimensional connected variety defined over $\mathbb{R}$. Let $x^{(0)} \in Z(\mathbb{R})$ be a smooth point. Then there exists an analytic function defined on an open subset in $\mathbb{R}^k$ which is a homeomorphism to a neighborhood of $x^{(0)} \in Z(\mathbb{R})$. The same holds over $\mathbb{C}$ or over $\mathbb{Q}_p$ for a prime $p < \infty$.*

Proof. Choose some $f_1, \dots, f_{d-k} \in J(Z)$ such that $\nabla(f_j)(x^{(0)})$ are linearly independent for $j = 1, \dots, d-k$. By choosing a new coordinate system $x_1, \dots, x_k, y_1, \dots, y_{d-k}$ (which we will abbreviate to x, y) we can assume without loss of generality that

$$\partial_{y_i}(f_j)(x^{(0)}, y^{(0)}) = \delta_{ij}$$

for $i, j = 1, \dots, d-k$, and furthermore

$$\partial_{x_i}(f_j)(x^{(0)}, y^{(0)}) = 0$$

for $i = 1, \dots, k$ and $j = 1, \dots, d-k$.

Applying the implicit function theorem (over $\mathbb{R}$, $\mathbb{C}$, or[14] $\mathbb{Q}_p$) on a neighborhood of $(x^{(0)}, y^{(0)})$ to the equations $f_1(x, y) = \dots = f_{d-k}(x, y) = 0$, we obtain $(d-k)$ analytic functions $\phi_1(x), \dots, \phi_{d-k}(x)$ which are all defined on a neighborhood U of $x^{(0)}$ such that

$$f_j\left(x, \phi_1(x), \dots, \phi_{d-k}(x)\right) = 0$$

for $j = 1, \dots, d-k$. It remains to see why the points

$$(x, \phi_1(x), \dots, \phi_{d-k}(x))$$

belong to Z (this is in question because we do not know whether $f_1, \dots, f_{d-k}$ generate $\mathscr{J}(Z)$) in some possibly smaller neighborhood $U' \subseteq U$.

Let

$$\mathscr{J}' = \langle f_1, \dots, f_{d-k} \rangle \subseteq \mathscr{J} = \mathscr{J}(Z)$$

and $Z(\mathscr{J}') = Z \cup Z'$, where Z' is the union of all connected components of $Z(\mathscr{J}')$ other than Z. Here Z cannot be contained properly in a connected component of $Z' \subseteq Z(\mathscr{J}')$ since the tangent space of $Z(\mathscr{J}')$ at $\left(x^{(0)}, y^{(0)}\right)$ has dimension k, which would contradict Lemma 3.15 and Lemma 3.16.

We claim that $(x^{(0)}, y^{(0)}) \notin Z'$. Assuming this claim, there exists some polynomial $g \in \mathscr{J}(Z')$ with $g(x^{(0)}, y^{(0)}) \neq 0$. Suppose now that $f \in \mathscr{J}$. Then the product $f \cdot g$ vanishes on

$$Z \cup Z' = Z(\mathscr{J}'),$$

and so there exists some ℓ with $(fg)^\ell \in \mathscr{J}'$ by Hilbert's Nullstellensatz (Theorem 3.6). Let

$$U' = \{x \in U \mid g(x, \phi_1(x), \dots, \phi_{d-k}(x)) \neq 0\},$$

and suppose that $x \in U'$. Then

$$(fg)^\ell (x, \phi_1(x), \dots, \phi_{d-k}(x)) = 0$$

since all elements in $\mathscr{J}'$ vanish on such vectors (by definition of $\phi_1, \dots, \phi_{d-k}$). However, since $x \in U'$, this shows that

$$f(x, \phi_1(x), \dots, \phi_{d-k}(x)) = 0$$

for all $f \in \mathscr{J}$ and $x \in U'$, as required.

To prove the claim† we will show that after removing all connected components of Z' that do not contain $(x^{(0)}, y^{(0)})$ from $Z(J') = Z \cup Z'$ we obtain a connected variety Z'' (see below for a more formal definition). Since Z is not removed, this then implies that $Z'' = Z$ and that Z' does not contain $(x^{(0)}, y^{(0)})$ as claimed.

Let $R = \mathbb{R}[x, y]$ be the ring of polynomials, let

$$M = \{f \in R \mid f(x^{(0)}, y^{(0)}) = 0\}$$

be the maximal ideal in R corresponding to $(x^{(0)}, y^{(0)})$, and let

$$R_M = \left\{ \frac{f}{g} \mid f, g \in \mathbb{R}[x, y] \text{ with } g \notin M \right\}$$

be the local ring‡ corresponding to M (consisting of rational functions that are well-defined at $(x^{(0)}, y^{(0)})$).

Let $Z_1, \dots, Z_a$ be the connected components of $Z \cup Z' = Z(J')$ that contain $(x^{(0)}, y^{(0)})$, and let $Z_1', \dots, Z_b'$ be those that do not contain $(x^{(0)}, y^{(0)})$. Making our definition above more precise, we set $Z'' = Z_1 \cup \dots \cup Z_a$. We now show that

$$J'' = (J' R_M) \cap R = \left\{ f = \frac{p}{q} \in R \mid p \in J', q \notin M \right\}$$

defines the variety Z''. Pick a polynomial $q_j \in J(Z_j')$ with $q_j(x^{(0)}, y^{(0)}) \neq 0$ for $j = 1, \dots, b$. Choose any polynomials $F_i \in J(Z_i)$ for $i = 1, \dots, a$. Then by Hilbert's Nullstellensatz (Theorem 3.6) there exists some $\ell \geqslant 1$ with

† A slight warning is in order. The remainder of this argument is surprisingly long, and quite algebraic. The reader who wishes to only get a glimpse of the algebraic background may decide to skip it, we will not need this type of argument again.

‡ A *local ring* is a ring with a unique maximal ideal. The ring R_M is the *localization* of R at M, it is a local ring with maximal ideal MR_M.

$$(F_1 \cdots F_a q_1 \cdots q_b)^\ell \in J'.$$

Using only the definition of J'', this implies that $(F_1 \cdots F_a)^\ell \in J''$. Using the Noetherian property of R we find some $\ell \geqslant 1$ with $J(Z_1)^\ell \cdots J(Z_a)^\ell \subseteq J''$ and so $Z(J'') \subseteq Z'' = Z_1 \cup \cdots \cup Z_a$.

For the opposite inclusion, fix some $i \in \{1, \dots, a\}$ and notice that by definition any $f \in J''$ is of the form $f = \frac{p}{q}$ with

$$p \in J' \subseteq J(Z_i)$$

and

$$q \notin M \supseteq J(Z_i),$$

which gives $f \in J(Z_i)$. This shows that $J'' \subseteq J(Z_i)$ and so $Z_i \subseteq Z(J'')$ for $i = 1, \dots, a$.

We will show the claim by showing that J'' is a prime ideal (which then gives the claim that $a = 1$ and $Z'' = Z$). For this we prove that $f \in J''$ if and only if there exists a neighborhood O of $(x^{(0)}, y^{(0)})$ in

$$M = \{(x, \phi_1(x), \dots, \phi_{d-k}(x)) \mid x \in U\}$$

such that the restriction of f to O is zero. By the properties of $\phi_1, \dots, \phi_{d-k}$ any such restriction can be identified with an analytic function on a neighborhood of $x^{(0)}$ inside U, and so the restriction is uniquely determined by its Taylor expansion at $x^{(0)}$. Since the Cauchy product of Taylor series has no zero-divisors†, this equivalence then shows that J'' is a prime ideal.

Suppose first that $f \in J''$. Then $f = \frac{p}{q}$, where $p \in J'$ vanishes on M (by definition of M), and q does not vanish at $(x^{(0)}, y^{(0)})$. This shows that there is a neighborhood O on which f is well-defined and identical to zero.

Now suppose that f is a polynomial for which there exists a neighborhood O of $(x^{(0)}, y^{(0)})$ in M on which f vanishes. By our assumptions from the beginning of the proof we have $f_j(x, y) \in y_j - y_j^{(0)} + M^2$ (where M^2 consists of all polynomials that vanish with order 2 or more at $(x^{(0)}, y^{(0)})$). Let $n \geqslant 1$ be arbitrary. We can now use the polynomials f_j to express the polynomial f as above in the form

$$f \in \sum_{\ell=0}^{n} F_\ell(x - x^{(0)}) + R_{n+1}(x, y) + J',$$

where F_ℓ is a homogeneous polynomial of degree ℓ for $\ell = 0, \dots, n$ and the polynomial $R_{n+1}(x, y) \in M^{n+1}$ only has terms that vanish of order $n + 1$ or higher. This shows that $\sum_{\ell=0}^{n} F_\ell(x - x^{(0)})$ is the Taylor approximation

† The Cauchy product of $\sum_{n=0}^{\infty} a_n x^n$ and $\sum_{n=0}^{\infty} b_n x^n$ is the series $\sum_{n=0}^{\infty} c_n x^n$ with $c_n = \sum_{j=0}^{n} a_j b_{n-j}$ for $n \geqslant 0$; viewed either as formal power series or as functions where they converge, the product will only vanish if one of the series vanishes.

of $f(x, \phi_1(x), \ldots, \phi_{d-k}(x))$ at $x^{(0)}$ of degree n. Since f vanishes in a neighborhood of $(x^{(0)}, y^{(0)})$ in M we have $\sum_{\ell=0}^{n} F_\ell = 0$. This shows that $f \in J' + M^n$ for all $n \geqslant 1$.

A corollary of Nakayama's lemma states that

$$\bigcap_{n=1}^{\infty} (I + M^n) = I$$

in any local Noetherian ring with an ideal I and a maximal ideal M. We refer to Hungerford [**?**, Cor. VIII.4.7] and Matsumura [**?**] for convenient sources for this result. Switching from the ring R to the local ring R_M we see that

$$f \in \bigcap_{n=1}^{\infty} \left(J' R_M + M^n R_M\right),$$

which gives $f \in J' R_M$ and so $f \in J''$. This establishes the above equivalence, and hence shows that J'' is a prime ideal, the claim, and so also the lemma. $\square$

In Section 3.1 we considered two notions of '$\mathbb{K}$ varieties': A variety Z is *defined over* $\mathbb{F}$, for some subfield† $\mathbb{F} \subseteq \overline{\mathbb{K}}$, if its complete ideal of relations (as in the Hilbert Nullstellensatz Theorem 3.6) is generated by polynomials with coefficients in $\mathbb{F}$. On the other, a variety is $\mathbb{F}$-*closed* if it can be defined by polynomials with coefficients in $\mathbb{F}$.

As in any topological space, we can define a notion of *closure*: the Zariski closure of a subset $S \subseteq \overline{\mathbb{K}}^d$ is the smallest Zariski closed subset $Z \subseteq \overline{\mathbb{K}}^d$ containing S. This notion has many convenient properties, including good behavior with regards to subfields. Note however, that the Zariski closure of a subset in $\mathbb{R}^d$ is frequently much bigger than the closure in the Hausdorff topology.

Lemma 3.19 (Closures of subsets of $\mathbb{F}^d$). *Let $\mathbb{F} \subseteq \overline{\mathbb{K}}$ be any subfield and $S \subseteq \mathbb{F}^d$. Then the Zariski closure of S is defined over $\mathbb{F}$.*

Proof. Suppose that f is a polynomial in $x_1, \ldots, x_d$ that vanishes on S. Let V be the vector space generated by the coefficients of f over $\mathbb{F}$. Let $a_1, \ldots, a_n$ be a basis of V over $\mathbb{F}$, and write

$$f = \sum_{i=1}^{n} f_i a_i$$

with $f_i \in \mathbb{F}[x_1, \ldots, x_d]$. For any $x \in S$ we now have

† We introduce this extra field for example in order to set $\mathbb{K} = \mathbb{R}$, $\overline{\mathbb{K}} = \mathbb{C}$, and $\mathbb{F} = \mathbb{Q}$.

$$f(x) = \sum_{i=1}^{n} \underbrace{f_i(x)}_{\in \mathbb{K}} a_i = 0,$$

and so $f_i(x) = 0$ for $i = 1, \ldots, n$. This shows that the ideal of polynomials that vanish on S is generated by those that have coefficients in $\mathbb{F}$. $\square$

Clearly a variety that is defined over $\mathbb{K}$ is also $\mathbb{K}$-closed. In general the converse is not true, but fortunately this problem only manifests itself over fields of positive characteristic.

Lemma 3.20 ($\mathbb{K}$-closed vs. defined over $\mathbb{K}$). *Suppose that $\mathbb{K}$ has characteristic zero. Then a $\mathbb{K}$-closed variety (or a variety that is stable under all Galois automorphisms of $\overline{\mathbb{K}}|\mathbb{K}$) is also defined over $\mathbb{K}$.*

Proof. Let $Z = Z(f_1, \ldots, f_n)$ be the variety defined by the polynomials

$$f_1, \ldots, f_n \in \mathbb{K}[x_1, \ldots, x_d],$$

and suppose that $f \in \overline{\mathbb{K}}[x_1, \ldots, x_d]$ vanishes on Z (that is, suppose that f lies in $J(Z)$). Then there exists a finite Galois field extension $\mathbb{L}|\mathbb{K}$ such that f has coefficients in $\mathbb{L}$.

Let σ be any Galois automorphism of the extension $\mathbb{L}|\mathbb{K}$. We now claim that the polynomial $\sigma(f)$ obtained by applying σ to all coefficients of f also belongs to $J(Z)$. This is straightforward to check as follows. Since Z is $\mathbb{K}$-closed, any Galois automorphism of $\overline{\mathbb{K}}|\mathbb{K}$ maps $Z = Z(\overline{\mathbb{K}})$ onto Z. Extending the automorphism σ of $\mathbb{L}|\mathbb{K}$ in some way to an automorphism of $\overline{\mathbb{K}}|\mathbb{K}$ we get

$$(\sigma(f))\,(x) = (\sigma(f))\left(\sigma(\sigma^{-1}(x))\right) = \sigma\Bigg(\underbrace{f(\underbrace{\sigma^{-1}(x)}_{\in Z})}_{=0}\Bigg) = 0$$

for all $x \in Z$.

The claim now implies that $\mathrm{tr}(f) = \sum_\sigma \sigma(f)$, where the sum is taken over the finite list of Galois automorphisms of $\mathbb{L}|\mathbb{K}$, belongs to $J(Z)$. Clearly $\mathrm{tr}(f)$ has coefficients in $\mathbb{L}$ and is fixed by all Galois automorphisms of $\mathbb{L}|\mathbb{K}$. Therefore, $\mathrm{tr}(f) \in \mathbb{K}[x_1, \ldots, x_d]$ (this requires the assumption that $(\mathbb{K}) = 0$).

We next claim that there exist elements

$$a_1, \ldots, a_{[\mathbb{L}|\mathbb{K}]} \in \mathbb{L}$$

and

$$a_1^*, \ldots, a_{[\mathbb{L}|\mathbb{K}]}^* \in \mathbb{L}$$

that are dual bases in the sense that

$$\mathrm{tr}(a_i^* a_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

for all i, j. We then have

$$a = \sum_i \operatorname{tr}(a_i^* a) a_i,$$

which also holds for the polynomial f instead of $a \in \mathbb{L}$. Since

$$\operatorname{tr}(a_i^* f) \in J(Z) \cap \mathbb{K}[x_1, \ldots, x_d]$$

by the argument above, the lemma follows from the claim.

It remains to construct the dual basis. Let $a_1, \ldots, a_{[\mathbb{L}:\mathbb{K}]} \in \mathbb{L}$ be any basis of $\mathbb{L}$ over $\mathbb{K}$. By linear algebra there exists a dual basis for the dual vector space $\mathbb{L}^*$ over $\mathbb{K}$. We claim that the map sending $a \in \mathbb{L}$ to $\phi(a) \in \mathbb{L}^*$ defined by

$$\phi(a)(b) = \operatorname{tr}(ab)$$

is an isomorphism of vector spaces. This may be seen as follows:

- $\phi(1)(1) = \operatorname{tr}(1) = [\mathbb{L} : \mathbb{K}]$, so ϕ is non-trivial (again since $(\mathbb{K}) = 0$);
- if $\phi(a) = 0$ then also $\phi(aa')(b) = \operatorname{tr}(a(a'b)) = 0$ for all $a', b \in \mathbb{L}$, so the kernel of ϕ is an ideal, and the field $\mathbb{L}$ has no non-trivial ideals.

Thus the pre-image under ϕ of the dual basis in $\mathbb{L}^*$ gives a dual basis in the above sense in $\mathbb{L}$.

If the variety Z is only assumed to be invariant under all Galois automorphisms, then once more $J(Z)$ is invariant under all Galois automorphisms and so the above argument shows again that Z is defined over $\mathbb{K}$. □

In the arguments above there is always an implied coordinate system in $\overline{\mathbb{K}}^d$ (corresponding to the variables $x_1, \ldots, x_d$). We note that it is customary to write $\mathbb{A}^d$ for the d-dimensional affine space without a preferred origin, coordinate system, or base field (so that $\mathbb{A}^d(\mathbb{L}) \simeq \mathbb{L}^d$ for any field $\mathbb{L}$). For us the ambient affine space will be $\operatorname{Mat}_d \simeq \mathbb{A}^{d^2}$, and on this space very few coordinate changes make sense with regards to the existing (and to us important) multiplicative structure. For that reason and also because we are often interested in subgroups of SL_d (and the orbits of the group of their $\mathbb{R}$-points), we are happy with choosing one coordinate system and discussing subvarieties and algebraic subgroups of SL_d instead of general varieties and general algebraic groups. We will however, switch frequently from one field to another, and as before will write $Z(\mathbb{K}) = Z(\overline{\mathbb{K}}) \cap \operatorname{Mat}_d(\mathbb{K})$ for the $\mathbb{K}$-points of a subvariety $Z < \operatorname{Mat}_d$ defined over $\mathbb{K}$.

3.4.3 Linear Algebraic Groups

A variety $\mathbb{G} \subseteq \mathrm{SL}_d$ is a *(linear) algebraic subgroup* (of SL_d) if $\mathbb{G}(\overline{\mathbb{K}}) \subseteq \mathrm{SL}_d(\overline{\mathbb{K}})$ is a subgroup. Notice that for any subvariety $Z \subseteq \mathrm{SL}_d$ and $g \in \mathrm{SL}_d(\overline{\mathbb{K}})$ we can define the translated variety gZ by the ideal

$$\lambda(g)J(Z) = \{f(g^{-1}x) \mid f \in J(Z)\}.$$

Here λ denotes the left representation

$$\lambda(g)f(x) = f(g^{-1}x)$$

on the space of all polynomials.

Lemma 3.21 (Smoothness). *Every point of a linear algebraic subgroup is smooth.*

The tangent space at the identity is called the *Lie algebra* of the algebraic subgroup.

PROOF OF LEMMA 3.21. Suppose that $g \in \mathbb{G}(\overline{\mathbb{K}})$ is a smooth point of the variety $\mathbb{G}$. Then one can quickly check that $I = g^{-1}g$ is a smooth point of the left-translate variety $g^{-1}\mathbb{G}$. However, since $g^{-1}\mathbb{G} = \mathbb{G}$ we see that I is a smooth point of $\mathbb{G}$. By the same argument, any other point is also smooth. $\square$

Lemma 3.22 (Connected components). *Let $\mathbb{G} \subseteq \mathrm{SL}_d$ be an algebraic subgroup. The connected component $\mathbb{G}^o < \mathbb{G}$ is by definition the unique Zariski connected component of $\mathbb{G}$ that contains the identity, it is an algebraic normal subgroup. There are points $g_1, \dots, g_n \in \mathbb{G}$ where $n = [\mathbb{G}(\overline{\mathbb{K}}) : \mathbb{G}^o(\overline{\mathbb{K}})]$ with*

$$\mathbb{G} = \bigsqcup_{i=1}^{n} g_i \mathbb{G}^o.$$

If $\mathbb{G}$ is defined over $\mathbb{K}$, and $\mathbb{K}$ has zero characteristic, then $\mathbb{G}^o$ is also defined over $\mathbb{K}$.

As a corollary of the lemma we mention that it makes sense to talk about the dimension of a (not necessarily Zariski connected) algebraic subgroup. Since all Zariski connected components are translates of the connected component $\mathbb{G}^o$, they all have the same dimension.

PROOF OF LEMMA 3.22. The first statement is essentially an extension of the argument in the previous lemma. If

$$\mathbb{G} = \bigcup_{i=1}^{n} Z_i$$

is the decomposition into connected components, then there exists a point which is contained in only one component. Translating $\mathbb{G}$ by one of its elements $g \in \mathbb{G}(\overline{\mathbb{K}})$ we may permute the connected components

$$\mathbb{G} = g^{-1}\mathbb{G} = \bigcup_{i=1}^{n} g^{-1} Z_i.$$

but would leave the subvariety $\bigcup_{i\neq j} Z_i \cap Z_j$ consisting of all points that are contained in more than one of the connected components invariant. Therefore, we have

$$\mathbb{G} = \bigsqcup_{i=1}^{n} Z_i.$$

Suppose that $Z_1 = \mathbb{G}^o$. If now $g \in \mathbb{G}^o$ then $I \in g^{-1}\mathbb{G}^o$, which by uniqueness of the decomposition gives $\mathbb{G}^o = g^{-1}\mathbb{G}^o$ for all $g \in \mathbb{G}^o$.

We have shown that $\mathbb{G}^o$ is a linear algebraic subgroup. If now $g \in Z_i$ for $i > 1$, then the same argument gives $g^{-1}Z_i = \mathbb{G}^o = Z_i g^{-1}$. In other words,

$$Z_i = g\mathbb{G}^o = \mathbb{G}^o g$$

is a coset of $\mathbb{G}^o$ in $\mathbb{G}$.

Now suppose that $\mathbb{G}$ is defined over $\mathbb{K}$, and let σ be a Galois automorphism of $\overline{\mathbb{K}}|\mathbb{K}$. Then σ induces a permutation of the cosets $g_i\mathbb{G}^o(\overline{\mathbb{K}})$ with

$$\sigma\left(\mathbb{G}^o(\overline{\mathbb{K}})\right) = \mathbb{G}^o(\overline{\mathbb{K}})$$

since $\sigma(I) = I$. As this holds for all Galois automorphisms we see that $\mathbb{G}^o$ is defined over $\mathbb{K}$ if $\mathbb{K}$ has characteristic zero by Lemma 3.20. $\square$

For completeness we mention another (more general but, up to isomorphisms, equivalent) definition: A *linear algebraic group* is an affine variety equipped with multiplication and inverse maps such that

- the multiplication and inverse maps are regular functions (from the group to the group);
- the variety is isomorphic to a linear algebraic subgroup of SL_d for some d such that the multiplication and inverse maps correspond to multiplication and inversion for matrices.

We note that the standard definition does not make the second requirement above, and instead derives this property from the first via a construction similar to the proof of Chevalley's theorem in Section 3.4.5.

Example 3.23. We list some standard examples of linear algebraic groups.

(a) $\mathbb{G}_a$ denotes the additive group structure of the field. This is a linear algebraic group because (for example) it is isomorphic to the algebraic subgroup $U < \mathrm{SL}_2$ with

$$U(\overline{\mathbb{K}}) = \left\{ \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \mid x \in \overline{\mathbb{K}} \right\},$$

which we saw earlier is associated to the horocycle flow if $\mathbb{K} = \mathbb{R}$.

(b) $\mathbb{G}_m$ stands for the multiplicative group structure of the field. This is a linear algebraic group because (for example) it is isomorphic to the algebraic subgroup $A < \mathrm{SL}_2$ with

$$A(\overline{\mathbb{K}}) = \left\{ \begin{pmatrix} a & \\ & a^{-1} \end{pmatrix} \mid a \in \overline{\mathbb{K}} \right\},$$

which we saw earlier is associated to the geodesic flow if $\mathbb{K} = \mathbb{R}$.

3.4.4 $\mathbb{K}$*-points of Linear Algebraic Groups*

As noted before, a variety Z defined over a field $\mathbb{K}$ does not have to contain any $\mathbb{K}$-points (that is, $Z(\mathbb{K})$ may be empty[†]), and even if it is non-empty it may not be Zariski dense in the variety. Since a subgroup always contains the identity the former problem cannot arise for linear algebraic subgroups. Even more is true, as a result of the following lemma, which relies on the fact that $\mathbb{G}$ is smooth at the identity.

Lemma 3.24 (Density of $\mathbb{R}$-points and $\mathbb{Q}_p$-points). *If $\mathbb{G} \subseteq \mathrm{SL}_d$ is a Zariski connected linear algebraic subgroup defined over $\mathbb{R}$, then $\mathbb{G}(\mathbb{R})$ is Zariski dense in $\mathbb{G}$. The same holds for $\mathbb{K} = \mathbb{Q}_p$ for a prime number $p < \infty$.*

We note that the above holds much more generally, see [**?**, Th. 18.3]. We will come back to this problem for the special case $\mathbb{K} = \mathbb{Q}$ later.

PROOF OF LEMMA 3.24. By Lemma 3.21, $x^{(0)} = I \in \mathbb{G}$ is a smooth point. By Lemma 3.18 $\mathbb{G}(\mathbb{R})$ contains the image of an analytic function of the form

$$\Phi : U \ni (x_1, \dots, x_k) \longmapsto (x_1, \dots, x_k, \phi_{k+1}(x_1, \dots, x_k), \dots, \phi_{d^2}(x_1, \dots, x_k)). \tag{3.23}$$

Let Z be the Zariski closure of these real points. By Lemma 3.17 we may write

$$Z = \bigcup_{i=1}^{n} Z_i$$

as a union of irreducible varieties. By Lemma 3.15, either $Z = \mathbb{G}$ or

$$\dim Z_i < k = \dim \mathbb{G}$$

for $i = 1, \dots n$. However, the latter case cannot happen since a finite union of varieties of dimension strictly less than k cannot contain all points

[†] A trivial example to have in mind here is the variety defined by the equation

$$x^2 + y^2 = -1,$$

defined over $\mathbb{R}$, and a less trivial example is the variety defined by the equation

$$x^3 + y^3 = 1,$$

defined over $\mathbb{Q}$.

$$\Phi(x_1, \dots, x_k)$$

in (3.23). Specifically, in this case each $\mathscr{I}(Z_i)$ must contain some

$$f_i \in \mathbb{C}[x_1, \dots, x_k],$$

so that every point $\Phi(x_1, \dots, x_k)$ in (3.23) would have to satisfy the equation $f_1 \cdots f_n = 0$. This is a contradiction, since every non-empty open subset of $\mathbb{R}^k$ in the Hausdorff topology is Zariski dense (since all the partial derivatives, including the 0th, of a polynomial at a point determine the polynomial). The p-adic case is similar. □

Clearly, the group $\mathbb{G}(\mathbb{R})$ of $\mathbb{R}$-points of an algebraic subgroup $\mathbb{G} \subseteq \mathrm{SL}_d$ is a linear Lie group with a real Lie algebra $\mathfrak{g}_{\mathbb{G}(\mathbb{R})}$. For the algebraic subgroup $\mathbb{G}$ in SL_d we have already defined a Lie algebra $\mathfrak{g}$, which by definition is a complex vector space. Assuming that $\mathbb{G}$ is defined over $\mathbb{R}$, this complex vector space

$$\mathfrak{g} \subseteq \mathfrak{sl}_d(\mathbb{C})$$

can be defined by linear equations with real coefficients so that

$$\mathfrak{g}(\mathbb{R}) = \mathfrak{g} \cap \mathfrak{sl}_d(\mathbb{R})$$

has the same dimension over $\mathbb{R}$ as $\mathfrak{g} = \mathfrak{g}(\mathbb{C})$ has over $\mathbb{C}$.

Lemma 3.25 (Lie algebras of Lie groups and algebraic groups). *Let $\mathbb{G} \subseteq \mathrm{SL}_d$ be an algebraic subgroup defined over $\mathbb{R}$. Then the $\mathbb{R}$-points*

$$\mathfrak{g}(\mathbb{R}) = \mathfrak{g} \cap \mathfrak{sl}_d(\mathbb{R})$$

of the Lie algebra $\mathfrak{g}$ of the algebraic subgroup comprise precisely the Lie algebra of the Lie subgroup $\mathbb{G}(\mathbb{R}) \subseteq \mathrm{SL}_d(\mathbb{R})$. The same holds over $\mathbb{C}$ or $\mathbb{Q}_p$ for a prime $p < \infty$.

Proof. Using the same notation and setup as in the proofs of Lemmas 3.18 (with d replaced by d^2) and Lemma 3.24, we see that the tangent space of $\mathbb{G}(\mathbb{R})$ (in the sense of manifolds or of Lie groups) is the image of the total derivative of Φ at $(x_1, \dots, x_k)$. However, by the implicit function theorem, this image is precisely the real subspace defined by the equations

$$(u_1, \dots, u_{d^2}) \cdot \nabla f_j(I) = 0$$

for $j = 1, \dots, d^2 - k$. This proves the lemma in the real case, and the complex and p-adic cases are similar. □

The following discussion is not essential for later developments, but it may be useful to bear it in mind. By [**?**, Ch. VII, Sect. 2.2, Th. 1] the set of $\mathbb{C}$-points $Z(\mathbb{C})$ of a Zariski connected variety Z is connected in the Hausdorff topology. For the $\mathbb{R}$-points $Z(\mathbb{R})$ of a Zariski connected variety Z over $\mathbb{R}$

this is not true. However, for algebraic groups $\mathbb{G}$ defined over $\mathbb{R}$, the connected component $\mathbb{G}(\mathbb{R})^o$ (in the Hausdorff topology) only has finite index. We will discuss this again for particular algebraic subgroups later (where it will usually be easy to see). For now, notice that $A(\mathbb{R})^o < A(\mathbb{R})$ for A as in Example 3.23(b) has index two. Over $\mathbb{Q}_p$ the analogous question does not make sense, so Zariski connected is *a priori* the only sensible notion of connectedness.

3.4.5 Chevalley's theorem: subgroups and representations

Clearly, every algebraic representation gives rise to many algebraic subgroups by defining stabilizer subgroups (as in Section 3.1.2). Chevalley's theorem[(15)] is turning this construction almost around: Given an algebraic subgroup there exists an algebraic representation so that the subgroup can be defined via the representation as a stabilizer of a line (instead of a point as in Section 3.1.2).

Theorem 3.26 (Chevalley). *Let $\mathbb{H} < \mathrm{SL}_d$ be an algebraic subgroup. Then there exists an algebraic representation $\rho : \mathrm{SL}_d \to \mathrm{SL}_D$ and a D-dimensional vector v such that*

$$\mathbb{H} = \{g \in \mathrm{SL}_d \mid \rho(g)v \sim v\},$$

where $\sim$ denotes proportionality[†]*. If H is defined*[‡] *over $\mathbb{K}$, then the algebraic representation ρ is also defined over $\mathbb{K}$, and we may choose $v \in \mathbb{K}^D$.*

As we will see the theorem is proved by transforming the defining ideal of $\mathbb{H}$ (which is finitely-generated) into a single vector in a high-dimensional vector space.

Proof of Theorem 3.26. For any $g \in \mathbb{H}$ we have $g\mathbb{H} = \mathbb{H}$ and equivalently

$$\lambda(g)J(\mathbb{H}) = J(\mathbb{H}).$$

Moreover, we also have that $\lambda(g)J(\mathbb{H}) = J(\mathbb{H})$ for some $g \in \mathrm{SL}_d$ implies that $g \in \mathbb{H}$. As the ideal is infinite dimensional we cannot use it directly. However, by the Noetherian property we know that $J(\mathbb{H}) \subseteq \overline{\mathbb{K}}[\mathrm{Mat}_d]$ is finitely generated (as an ideal). Thus we can assume it is generated by polynomials of degree less than or equal to m for some m. Write $P_{\leqslant m}$ for the space of all polynomials in $\overline{\mathbb{K}}[\mathrm{Mat}_d]$ of degree $\leqslant m$, and define

$$J_{\leqslant m} = J(\mathbb{H}) \cap P_{\leqslant m}.$$

† Notice that proportionality is itself a polynomial condition, defined by requiring the vanishing of all 2×2 determinants corresponding to pairs of components of $\rho(g)v$ and of v.

‡ That is, if the ideal of relations that are satisfied on $\mathbb{H}$ is generated by polynomials with coefficients in $\mathbb{K}$.

Now notice that $\lambda(g)P_{\leqslant m} = P_{\leqslant m}$ for all $g \in \mathrm{SL}_d$ and that

$$\lambda(g)J_{\leqslant m} = J_{\leqslant m}$$

is equivalent to $g \in \mathbb{H}$ (since $J_{\leqslant m}$ generates $J(\mathbb{H})$). In other words, we have found a finite-dimensional representation of SL_d and a subspace so that $\mathbb{H}$ is precisely the subgroup of SL_d that sends the subspace into itself. The representation is also an algebraic representation (which the reader can quickly check).

What is not quite as in the theorem is that the subspace might not be a single line. However, even that can quickly be rectified. Let $\ell = \dim J_{\leqslant m}$ and define $V = \bigwedge^{\ell} P_{\leqslant m}$ and let $v \in \bigwedge^{\ell} J_{\leqslant m} \smallsetminus \{0\}$. The algebraic representation of SL_d on $P_{\leqslant m}$ induces an algebraic representation ρ on V (check this) and for any $g \in \mathrm{SL}_d$ the condition $\rho(g)v \sim v$ is equivalent to $\lambda(g)J_{\leqslant m} = J_{\leqslant m}$ and hence to $g \in \mathbb{H}$.

If $\mathbb{H}$ is now additionally defined over $\mathbb{K}$, then $J_{\leqslant m} \cap \mathbb{K}[\mathrm{Mat}_d]$ generates $J(\mathbb{H})$ and we can choose v as the wedge of ℓ elements in $J_{\leqslant m} \cap \mathbb{K}[\mathrm{Mat}_d]$. Since the regular representation (and its ℓth wedge power) are defined over any field, this proves the last claim of the theorem. □

Lemma 3.27 (Zariski closures of groups). *If $S \subseteq \mathrm{SL}_d(\mathbb{K})$ is a subgroup, then the Zariski closure $\mathbb{G} = \overline{S}^Z$ is a linear algebraic subgroup defined over $\mathbb{K}$.*

Proof. By Lemma 3.19 we know that $\mathbb{G}$ is defined over $\mathbb{K}$, so it is enough to show that $\mathbb{G}(\overline{\mathbb{K}})$ is a subgroup.

For any $g \in S$ we have $gS = S$ by the assumption on S, so $g\mathbb{G} = \mathbb{G}$ for all $g \in S$. However, as in the proof of Theorem 3.26 this property of preserving the variety is equivalent to the property of preserving the ideal $J(\mathbb{G})$ of relations defining $\mathbb{G}$ or equivalently a particular line inside an algebraic representation of SL_d.

As this is a polynomial condition (see one of the footnotes to Theorem 3.26) which holds for all $g \in S$ it must also hold for all $g \in \mathbb{G}$. In other words, we have shown that $g\mathbb{G} = \mathbb{G}$ also holds for g in the Zariski closure of S, that is for all $g \in \mathbb{G} = \mathbb{G}(\overline{\mathbb{K}})$. □

3.4.6 Jordan Decomposition, Algebraic Subgroups and Representations

Algebraic representations and algebraic groups have some striking differences to the theory of Lie groups, which we will now start to discuss.

Let ρ be an algebraic representation of SL_d (or more generally of an algebraic subgroup $\mathbb{H}$). Then we have the following facts:

- if $u \in \mathrm{SL}_d$ $(u \in \mathbb{H})$ is nilpotent, then so is $\rho(u)$;
- if $a \in \mathrm{SL}_d(\mathbb{R})$ $(a \in \mathbb{H})$ is diagonalizable (when considered as an element $a \in \mathrm{SL}_d$) and has only real and positive eigenvalues, then the same holds for $\rho(a)$.

The first property is readily proved for the case SL_d and $\mathbb{K} = \mathbb{Q}$ or $\mathbb{K}$ a local field. Indeed, if $u \in \mathrm{SL}_d(\overline{\mathbb{K}})$ is unipotent, then there exists some a with $a^n u a^{-n} \to I$ as $n \to \infty$, which implies that

$$\rho(a)^n \rho(u) \rho(a)^{-n} = \rho(a^n u a^{-n}) \longrightarrow I$$

as $n \to \infty$, so the eigenvalues of $\rho(u)$ (which are not changed by conjugation) must all be 1, so $\rho(u)$ must be unipotent.

The second property requires a bit more work. We also note that if the algebraic representation is only defined on the subgroup $\mathbb{H}$ then neither claim would be correct in the context of Lie theory. For this notice that the Lie groups $U(\mathbb{R})$ and $A(\mathbb{R})$ are not that much different. On the one hand, the former is connected and the latter is not, so they are not isomorphic. However, there is a surjective group homomorphism from $A(\mathbb{R})$ onto $U(\mathbb{R})$, and an injective homomorphism from $U(\mathbb{R})$ into $A(\mathbb{R})^o < A(\mathbb{R})$. This does not contradict the above claims, since the two maps are basically the logarithm and the exponential map, which are not algebraic homomorphisms.

Recall from linear algebra that every matrix $g \in \mathrm{SL}_d(\overline{\mathbb{K}})$ has a *Jordan decomposition*

$$g = g_{\mathrm{ss}} g_{\mathrm{u}}$$

into a $\overline{\mathbb{K}}$-*diagonalizable* or *semi-simple* matrix $g_{\mathrm{ss}} \in \mathrm{SL}_d(\overline{\mathbb{K}})$ and a *unipotent* $g_{\mathrm{u}} \in \mathrm{SL}_d(\overline{\mathbb{K}})$. The two components g_{ss} and g_{u} commute with each other, and under this requirement the decomposition is unique. If $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, then $g_{\mathrm{ss}} = g_{\mathrm{pos}} g_{\mathrm{comp}}$ can be further decomposed into a product of two commuting semi-simple elements $g_{\mathrm{pos}}, g_{\mathrm{comp}} \in \mathrm{SL}_d(\mathbb{C})$, where the positive semi-simple part g_{pos} has only real and positive eigenvalues, and all the eigenvalues of the compact semi-simple part g_{comp} have absolute value one. This decomposition is also unique, and if $g \in \mathrm{SL}_d(\mathbb{R})$ then $g_{\mathrm{u}}, g_{\mathrm{pos}}, g_{\mathrm{comp}}$ lie in $\mathrm{SL}_d(\mathbb{R})$. If $\mathbb{K} = \mathbb{Q}_p$, then a similar decomposition can be shown, and the following results hold in that case also (see Exercise 3.4.1).

The following two results contain the claims made in the beginning of this section in greater generality.

Proposition 3.28 (Jordan decomposition and subgroups). *Let $\mathbb{H}$ be an algebraic subgroup of SL_d, and let g be an element of $\mathbb{H}$. If $g = g_{\mathrm{ss}} g_{\mathrm{u}}$ is the Jordan decomposition of g in $\mathrm{SL}_d(\overline{\mathbb{K}})$, then $g_{\mathrm{ss}}, g_{\mathrm{u}} \in \mathbb{H}$ also. If $\mathbb{H}$ is defined over $\mathbb{K} = \mathbb{R}$ (or $\mathbb{K} = \mathbb{C}$) and $g_{\mathrm{ss}} = g_{\mathrm{pos}} g_{\mathrm{comp}}$ is the decomposition into positive semi-simple and compact semi-simple parts, then once again $g_{\mathrm{pos}}, g_{\mathrm{comp}} \in \mathbb{H}$.*

Proposition 3.29 (Jordan decomposition and representations). *Let $\mathbb{H}$ be an algebraic subgroup of SL_d, and let $\rho : \mathbb{H} \to \mathrm{GL}_D$ be an algebraic rep-*

resentation. Then $\rho(g)_{\mathrm{u}} = \rho(g_{\mathrm{u}})$ *and* $\rho(g)_{\mathrm{ss}} = \rho(g_{\mathrm{ss}})$ *for all* $g \in \mathbb{H}$*. If* $\mathbb{K}$ *is* $\mathbb{R}$ *or* $\mathbb{C}$*, then we also have* $\rho(g)_{\mathrm{pos}} = \rho(g_{\mathrm{pos}})$ *and* $\rho(g)_{\mathrm{comp}} = \rho(g_{\mathrm{comp}})$*.*

The proof of these results is intertwined. We will first prove Proposition 3.29 in a special case, then prove Proposition 3.28, and finally obtain Proposition 3.29 as a corollary.

Proof of Proposition 3.29 for a Chevalley representation. Suppose that ρ is the representation of SL_d obtained in the proof of Theorem 3.26 for a subgroup $\mathbb{H} \leqslant \mathrm{SL}_d$. Let $g = g_{\mathrm{ss}}$ be semi-simple, and assume (without loss of generality, by applying any necessary conjugation to $\mathbb{H}$ and g) that g is diagonal. Then it is easy to see[†] that $\lambda(g)$ restricted to $P_{\leqslant m}$ is diagonal, with eigenvalues given by monomials in the standard variables. Therefore all eigenvalues of $\lambda(g)$ are simply products of powers of eigenvalues of g. Taking the ℓth wedge representation, the same holds for $\rho(g) = \wedge^{\ell}\lambda(g)$. Let $g = g_u$ be unipotent. If $\mathbb{K}$ is $\mathbb{Q}$ or a local field (which is where our main interest lies), we have already shown that $\rho(g)$ is unipotent. In general we may argue again step by step as above. First, show that $\lambda(g)$ restricted to $P_{\leqslant m}$ is unipotent by considering monomials corresponding to the eigendirections (resp. generalized eigendirections). Then we can show that $\rho(g) = \wedge^{\ell}\lambda(g)$ is also unipotent.

If now $g = g_{\mathrm{ss}}g_{\mathrm{u}}$ is any element of SL_d, then $\rho(g_{\mathrm{ss}})$ is semi-simple, $\rho(g_{\mathrm{u}})$ is unipotent, $\rho(g) = \rho(g_{\mathrm{ss}})\rho(g_{\mathrm{u}})$, and $\rho(g_{\mathrm{ss}})$, $\rho(g_{\mathrm{u}})$ commute with each other. This proves the claim.

If $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, then the argument above also shows that the eigenvalues of $\rho(g_{\mathrm{pos}})$ are positive and the eigenvalues of $\rho(g_{\mathrm{comp}})$ have absolute value one, giving the theorem. □

Proof of Proposition 3.28. Let $\mathbb{H} \leqslant \mathrm{SL}_d$ be an algebraic subgroup and let $\rho, v \in \overline{\mathbb{K}}^D$ be as in Theorem 3.26. Let $g \in \mathbb{H}$ so that $v \in \overline{\mathbb{K}}^D$ is an eigenvector of $\rho(g)$ for the Chevalley representation. By the properties of the Jordan decomposition, v is therefore also an eigenvector of $\rho(g)_{\mathrm{ss}} = \rho(g_{\mathrm{ss}})$ and of $\rho(g)_{\mathrm{u}} = \rho(g_{\mathrm{u}})$. It follows that $g_{\mathrm{ss}}, g_{\mathrm{u}} \in \mathbb{H}$. If $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, and $g_{\mathrm{ss}} = g_{\mathrm{pos}}g_{\mathrm{comp}}$ then $\rho(g)_{\mathrm{pos}} = \rho(g_{\mathrm{pos}})$ has v as an eigenvalue. Thus $g_{\mathrm{pos}}, g_{\mathrm{comp}} \in \mathbb{H}$ as well. □

Proof of Proposition 3.29. Let $\mathbb{H} \leqslant \mathrm{SL}_d$ and let $\rho : \mathbb{H} \to \mathrm{GL}_D$ be an arbitrary algebraic representation. Then

$$\mathbb{L} = (\rho) \subseteq \mathbb{H} \times \mathrm{GL}_D \subseteq \mathrm{SL}_{d+D+1}$$

is an algebraic subgroup in the following way. We require the elements of $\mathbb{L}$ to be of block form

$$\begin{pmatrix} h & & \\ & g & \\ & & \det(g)^{-1} \end{pmatrix}$$

[†] We use the notation from the proof of Theorem 3.26.

with $h \in \mathrm{SL}_d$ and $g \in \mathrm{GL}_D$ (by using linear equations, the condition $\det h = 1$, and the polynomial equation that the last entry should be the inverse of the determinant of the middle block), require $h \in \mathbb{H}$ (by the known relations of $\mathbb{H}$), and finally $g = \rho(h)$ (which is a polynomial condition by assumption on ρ).

Now let $h = h_{\mathrm{ss}} \in \mathbb{H}$ be semi-simple, so that

$$g = \begin{pmatrix} h & & \\ & \rho(h) & \\ & & \det(\rho(h))^{-1} \end{pmatrix} \in \mathbb{L}$$

and hence by Proposition 3.28 we also have

$$g_{\mathrm{u}} = \begin{pmatrix} h_{\mathrm{u}} & & \\ & \rho(h)_{\mathrm{u}} & \\ & & 1 \end{pmatrix} \in \mathbb{L}.$$

However, since $h_{\mathrm{u}} = I_d$ and $\mathbb{L}$ is a graph of a homomorphism, we also have $\rho(h)_{\mathrm{u}} = I_D$. This shows that $\rho(h)$ is semi-simple if h is semi-simple. The same argument also applies to unipotent elements (respectively, to positive or compact semi-simple elements if $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$). The proposition follows from the uniqueness of the Jordan decomposition. □

Exercises for Section 3.4

Exercise 3.4.1. Let $\mathbb{K} = \mathbb{Q}_p$. Show that every matrix $g \in \mathrm{SL}_d(\overline{\mathbb{K}})$ is the product of commuting elements $g_{\mathrm{pos}}, g_{\mathrm{comp}} \in \mathrm{SL}_d(\overline{\mathbb{K}})$ where the eigenvalues of g_{pos} are rational powers of p, and the eigenvalues of g_{comp} have absolute value one. Generalize the results of Section 3.4.6 to include this p-adic case.

3.5 Borel Density Theorem

We will show in this section a version of the Borel density theorem,[(16)] which will show another relationship between finite volume orbits and rationally defined subgroups. It is the generalization of the basic observation that a lattice $\Lambda < \mathbb{R}^d$ cannot be contained in a proper subspace to the setting of lattices in linear algebraic groups.

For the proof we will need two basic theorems, each of them fundamental to its own subject. However, the two subjects concerned are often — in the context of this book wrongly — considered far from each other. Concretely, we will need Poincaré recurrence from ergodic theory (in some sense the pigeonhole principle for ergodic theory, see Theorem 1.8 and Exercise 1.1.7), and

Chevalley's theorem from the theory of algebraic groups (see Theorem 3.26), and will combine these with the facts derived in Section 3.4.6. This approach goes back to work of Furstenberg [**?**] and Dani [**?**].

Theorem 3.30 (Borel density theorem). *Suppose that* $\mathbb{H} < \mathrm{SL}_d$ *is an algebraic subgroup defined over* $\mathbb{R}$ *and suppose that* $\Gamma < \mathbb{H}(\mathbb{R})$ *is a lattice. Then*

(1) *If* $\mathbb{H}$ *is semi-simple*[†] *such that* $\mathbb{H}(\mathbb{R})^o$ *has no compact factors then* Γ *is Zariski dense in* $\mathbb{H}$*. If* $\mathbb{H}$ *is only assumed to be semi-simple then the Zariski closure of* Γ *contains all non-compact factors of* $\mathbb{H}(\mathbb{R})^o$ *(and possibly some or all of the compact factors).*
(2) *In the general case, the Zariski closure* $\mathbb{L} < \mathbb{H}$ *of* Γ *contains all unipotent elements of* $\mathbb{H}(\mathbb{R})$ *and more generally all elements of* $\mathbb{H}(\mathbb{R})$ *that only have positive real eigenvalues.*

For the proof we will also need the following simple observation from linear algebra.

Lemma 3.31 (Convergence to some eigenvector). *Let* $g \in \mathrm{SL}_d(\mathbb{R})$ *have the property that all its eigenvalues are real and positive, and let*

$$\rho : \mathrm{SL}_d(\mathbb{R}) \to \mathrm{SL}_D(\mathbb{R})$$

be a finite-dimensional algebraic representation (obtained, for example, from Chevalley's theorem). Then for any $w \in \mathbb{R}^D \smallsetminus \{0\}$ *there is some* $v \in \mathbb{R}^D$ *with*

$$\frac{1}{\|\rho(g^n)w\|}\rho(g^n)w \longrightarrow v \in \mathbb{R}^D$$

as $n \to \infty$*, and* v *is an eigenvector of* $\rho(g)$*.*

Proof. By Proposition 3.28 if g is unipotent then $\rho(g)$ is also, and if g has only positive eigenvalues then the same holds for $\rho(g)$. Given $w \in \mathbb{R}^D \smallsetminus \{0\}$, we may write

$$w = \sum_{\lambda > 0} w_\lambda \neq 0,$$

where each w_λ is a generalized eigenvector for the eigenvalue λ and the map $\rho(g)$. Then there is some largest eigenvalue λ_L with $w_{\lambda_L} \neq 0$ (and hence $w_\lambda = 0$ for any $\lambda > \lambda_L$). Also notice that $\|\rho(g^n)w_\lambda\|$ is asymptotic to $\lambda^n n^{k(\lambda)}$ for some $k(\lambda) \geqslant 0$ (this may be seen by looking at the Jordan normal form of $\rho(g)$, see also the argument below). Thus

$$\frac{1}{\|\rho(g^n)w\|}\rho(g^n w) - \frac{1}{\|\rho(g^n)w_{\lambda_L}\|}\rho(g^n w_{\lambda_L}) \longrightarrow 0$$

[†] A linear algebraic group $\mathbb{H}$ is semi-simple if it is Zariski connected and its Lie algebra is semi-simple. Notice that this does not imply that $\mathbb{H}(\mathbb{R})$ is connected as a manifold.

as $n \to \infty$. This reduces the problem to the case of a single eigenvalue, and hence (by canceling the eigenvalue) to the case of a unipotent matrix

$$A = \frac{1}{\lambda_L}\rho(g)|_{V_{\lambda_L}}$$

acting on the generalized eigenspace V_{λ_L} of $\rho(g)$ for the eigenvalue λ_L. Choosing a Jordan basis of A, we may assume that A is a block matrix

$$A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_\ell \end{pmatrix}$$

where each

$$A_i = \begin{pmatrix} 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}.$$

We split $w = w_{\lambda_L}$ into components $\sum_i w^{(i)}$ corresponding to this block decomposition, and apply A_i to the vector

$$w^{(i)} = \begin{pmatrix} w_1^{(i)} \\ \vdots \\ w_k^{(i)} \end{pmatrix}$$

to obtain

$$A_i^n \begin{pmatrix} w_1^{(i)} \\ \vdots \\ w_k^{(i)} \end{pmatrix} = \begin{pmatrix} w_1^{(i)} + w_2^{(i)} n + w_3^{(i)} \binom{n}{2} + \cdots + w_k^{(i)} \binom{n}{k} \\ \vdots \\ w_{k-1}^{(i)} + w_k^{(i)} n \\ w_k^{(i)} \end{pmatrix}.$$

If now $w^{(i)} \neq 0$, then the above is a vector-valued polynomial whose entry of highest degree is in any case the first row corresponding to the eigenspace of A_i. Since this holds for each i, the lemma follows. □

Proof of Theorem 3.30, Part (2). Let $g \in \mathbb{H}(\mathbb{R})$ have positive real eigenvalues, let $\mathbb{L}$ be the Zariski closure of $\Gamma \leqslant \mathbb{H}(\mathbb{R}) \leqslant \mathrm{SL}_d(\mathbb{R})$ and let

$$\rho : \mathrm{SL}_d \to \mathrm{SL}_D$$

and $w \in \mathbb{R}^D$ be the Chevalley representation for $\mathbb{L} = \mathrm{Stab}_{\mathrm{SL}_d}(\mathbb{R}w)$ as in Theorem 3.26. By Poincaré recurrence we have for almost every $x \in \Gamma\backslash\mathbb{H}(\mathbb{R})$ a sequence $n_k \to \infty$ with $xg^{n_k} \to x$ as $k \to \infty$. We now switch this convergence to the group level as follows: for almost every $h \in \mathbb{H}(\mathbb{R})$ there exist sequences $n_k \to \infty$, $\varepsilon_k \to e$, and $\gamma_k \in \Gamma$ with $\gamma_k h g^{n_k} = h\varepsilon_k$ for all $k \geqslant 1$, or equivalently with

$$\gamma_k = \underbrace{h\varepsilon_k h^{-1}}_{\to e} h g^{-n_k} h^{-1}.$$

Applying this group element to w gives

$$\frac{1}{\|w\|} w = \frac{1}{\|\rho(\gamma_n) w\|} \rho(\gamma_n) w = \lim_{k\to\infty} \frac{1}{\|\rho(h g^{-n_k} h^{-1}) w\|} \rho(h g^{-n_k} h^{-1}) w = v_h,$$

where we have used the fact that $\Gamma \leqslant \mathbb{L}(\mathbb{R})$ fixes $\mathbb{R}w$ by definition, and Lemma 3.31. It follows by the same lemma that w is an eigenvector of $\rho(hgh^{-1})$ for almost every h. Taking $h \to e$ shows that w is an eigenvector of $\rho(g)$ also and so $g \in \mathbb{L}(\mathbb{R})$. □

PROOF OF THEOREM 3.30, PART (1). Let $H^o = \mathbb{H}(\mathbb{R})^o$ be the connected component of the set of real points of $\mathbb{H}$. Let F be a non-compact almost direct simple factor of H^o. Then F contains a one-parameter unipotent subgroup U, and we can apply Part (2) of the theorem to U and to all its conjugates, which together generate a normal connected subgroup of F (and hence all of F). Thus $\mathbb{L}(\mathbb{R})$ contains F. We may apply this for all non-compact almost direct factors of $\mathbb{H}$, which then proves the second claim in Part (1).

This also proves the first claim in Part (1) since by the above $\mathbb{L}$ and $\mathbb{H}$ have the same Lie algebra and hence have the same dimension. However, $\mathbb{H}$ is by assumption connected and so $\mathbb{L} = \mathbb{H}$ follows. □

Exercises for Section 3.5

Exercise 3.5.1. Let Q be a real non-degenerate quadratic form of signature (p, q) in d variables with $p \geqslant q \geqslant 1$. Suppose the orbit $\mathrm{SL}_d(\mathbb{Z})\,\mathrm{SO}(Q)(\mathbb{R})$ has finite volume. Show that a multiple of Q has integer coefficients.

3.6 Irreducible Quotients

In this section we classify lattices in semi-simple groups into reducible and irreducible lattices, and derive interesting density results (in the standard topology) from the Borel density theorem (which gives weak Zariski density).

Definition 3.32. Let G be a connected semi-simple Lie group. A lattice $\Gamma < G$ is called *reducible* if $G = H_1 \cdot H_2$ can be written as an almost direct product of nontrivial connected Lie subgroups $H_1, H_2 \leqslant G$ such that $\Gamma_1 = \Gamma \cap H_1$

is a lattice in H_1 and $\Gamma_2 = \Gamma \cap H_2$ is a lattice in H_2. The lattice is called *irreducible* if it is not reducible.

Examples of reducible lattices are of course very easy to find, for example $\mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ is a reducible lattice in $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$. Irreducible lattices are a bit more difficult to find†, but for now we only note that, for example, $\mathrm{SL}_2(\mathbb{Z}[\sqrt{2}])$ can be made into an irreducible lattice in $\mathrm{SL}_2(\mathbb{R})\times\mathrm{SL}_2(\mathbb{R})$, see Exercise 3.6.1.

Corollary 3.33 (Dense projections of irreducible lattices). *Let*

$$G = H_1 H_2 \subseteq \mathrm{SL}_d(\mathbb{R})$$

be an almost direct product of the connected components of the groups of $\mathbb{R}$-points $H_1, H_2 \subseteq \mathrm{SL}_d(\mathbb{R})$ of two semi-simple algebraic groups defined over $\mathbb{R}$. We assume furthermore that G has no compact factors. Let $\Gamma < G$ be an irreducible lattice in G, and suppose that H_2 is non-trivial. Then the projection of Γ to

$$G/C(G)H_2 \cong H_1/C(H_1)$$

is dense in $H_1/C(H_1)$.

Proof. We note that G is also the connected component of the group of $\mathbb{R}$-points of its Zariski closure. In fact if $\mathbb{H}_1, \mathbb{H}_2$ are the algebraic groups giving rise to H_1, H_2 then $\mathbb{G} = \mathbb{H}_1\mathbb{H}_2$ is an algebraic group defined over $\mathbb{R}$ with $G = \mathbb{G}(\mathbb{R})^o$. Also if $F \lhd G$ is any connected normal subgroup, then $F = \mathbb{F}(\mathbb{R})^o$ for a normal algebraic subgroup $\mathbb{F} \lhd \mathbb{G}$. In fact, if $\mathfrak{g} = \mathfrak{f} + \mathfrak{f}'$ is a decomposition of the Lie algebra $\mathfrak{g}$ of G into the Lie algebra $\mathfrak{f}$ of F and a transversal Lie ideal $\mathfrak{f}'$ of $\mathfrak{g}$, then $\mathbb{F} = C_{\mathbb{G}}(\mathfrak{f}')^o$. Therefore we may apply the Borel density theorem (Theorem 3.30) for $\mathbb{G}$ or any of its normal subgroups.

Write

$$\pi_1 : G \longrightarrow G/C(G)H_2 \cong H_1/C(H_1)$$

for the projection map. There are two cases to consider: either $\pi_1(\Gamma)$ is discrete or it is not.

Discrete image implies reducibility. If $\pi_1(\Gamma)$ is discrete and $B_1 \subseteq H_1$ is a fundamental domain for the discrete pre-image of $\pi_1(\Gamma)$ in H_1 and $B_2 \subseteq H_2$ is a fundamental domain for $\Gamma \cap H_2$ in H_2, then we claim that $B_1 B_2 \subseteq G$ is an injective domain for Γ. Indeed, if $\gamma \in \Gamma$, $b_1, b_1' \in B_1$, and $b_2, b_2' \in B_2$ satisfy $\gamma b_1 b_2 = b_1' b_2'$, then this identity modulo $H_2 \lhd G$ gives

$$(\gamma H_2)\,(b_1 H_2) = b_1' H_2.$$

Interpreting this in H_1 gives

$$\eta b_1 = b_1'$$

† By definition any lattice in a simple group is irreducible, but let us discuss a more interesting example.

for some η with $\pi_1(\eta) = \pi_1(\gamma)$. By our assumption that B_1 is a fundamental domain, it follows that $b_1 = b_1'$. Multiplying $\gamma b_1 b_2 = b_1' b_2'$ with b_1^{-1} we get $\gamma b_2 = b_2'$ and $\gamma \in H_2$. Now $b_2 = b_2'$ and $\gamma = I$ by the injectivity assumption on B_2. Hence $B_1 B_2 \subseteq G$ is an injective domain for Γ, and has finite Haar measure since Γ is a lattice by assumption. This also implies that[†] each of B_1 and B_2 has finite Haar measure. This implies that $\Gamma \cap H_2$ is a lattice in H_2.

By the Borel density theorem (Theorem 3.30) applied to $\Gamma \cap H_2 \subseteq H_2$ there is a finite collection $\{\gamma_1, \dots, \gamma_n\} \subseteq \Gamma \cap H_2$ such that

$$C(\gamma_1, \dots, \gamma_n) = \{h \in H_2 \mid \gamma_i h = h \gamma_i \text{ for } i = 1, \dots, n\}$$

is the center $C(H_2)$ of H_2. In fact, we may choose $\gamma_1 \in \Gamma \cap H_2 \smallsetminus C(H_2)$ and then successively choose $\gamma_2, \dots$ so that at each stage

$$C(\gamma_1, \dots, \gamma_m) \subsetneq C(\gamma_1, \dots, \gamma_{m-1}).$$

By the Noetherian property, we must find some n with

$$C(\gamma_1, \dots, \gamma_n) = C(\Gamma \cap H_2).$$

Since $\Gamma \cap H_2$ is Zariski dense in H_2 we deduce that

$$C(\gamma_1, \dots, \gamma_n) = C(H_2)$$

as required.

We claim that this implies that

$$\pi_2(\Gamma) \subseteq H_2 / C(H_2)$$

must be discrete as well. In fact, if $\pi_2(\gamma)$ is sufficiently small but non-trivial, then by construction

$$[\pi_2(\gamma), \pi_2(\gamma_m)] \neq I$$

for some $m \in \{1, \dots, n\}$, and then

$$[\gamma, \gamma_m] \in H_2 \cap \Gamma$$

is very close to an element of $C(H_2)$ but does not belong to $C(H_2)$. However, $C(H_2)$ is finite (it is zero-dimensional because its Lie algebra is trivial). This contradicts the assumed discreteness of Γ, so $\pi_2(\Gamma)$ must be discrete as claimed.

The claim establishes a symmetry between H_1 and H_2 in the above discussion. Applying the argument above to H_1 we also see that $\Gamma \cap H_1$ is a lattice in H_1. In other words, we have shown that Γ is a reducible lattice.

[†] As G is the almost direct product of H_1 and H_2 the Haar measure m_G is, in the same sense, also almost the product of the Haar measures $m_{H_1} \times m_{H_2}$.

Showing density. This shows that we may assume that $\pi_1(\Gamma)$ is not discrete. Let

$$F = \pi_1^{-1}\left(\overline{\pi_1(\Gamma)}\right) \cap H_1$$

be the pre-image in H_1 of the closure of $\pi_1(\Gamma)$. Clearly Γ stabilizes the Lie algebra $\mathfrak{f}$ of F. By the Borel density theorem (Theorem 3.30) applied to the lattice Γ in G, the same holds for $G \geqslant H_1$. It follows that $\mathfrak{f} \lhd \mathfrak{h}_1$ is a Lie ideal in the Lie algebra $\mathfrak{h}_1$ of H_1.

If $\mathfrak{f} = \mathfrak{h}_1$, then we get the desired density of $\pi_1(\Gamma)$ in $H_1/C(H_1)$. So suppose that $\mathfrak{f} \neq \mathfrak{h}_1$, and define H_1' to be the almost direct product of all factors of H_1 whose Lie algebra is not contained in $\mathfrak{f}$. Also define H_2' to be the almost direct product of H_2 and all factors of H_1 whose Lie algebra is contained in $\mathfrak{f}$. Since $\mathfrak{f} \neq \mathfrak{h}_1$, the group H_1' is non-trivial. If π_1' denotes the analogous projection for the almost direct product $G = H_1'H_2'$ then we see that

$$\pi_1'(\Gamma) \subseteq H_1'/C(H_1')$$

is discrete. By the first argument in the proof, this implies that Γ is a reducible lattice. Therefore irreducibility of the lattice implies that $\mathfrak{f} = \mathfrak{h}_1$ and the result follows. □

While Corollary 3.33 gives interesting results for irreducible lattices, it can also apply in a weaker (potentially trivial) form to reducible lattices. This is because every reducible lattice can be 'reduced', or 'almost decomposed' into irreducible lattices as follows. If $\Gamma < H_1H_2$ is a reducible lattice such that $\Gamma \cap H_i < H_i$ is a lattice for $i = 1, 2$, then

$$(\Gamma \cap H_1)(\Gamma \cap H_2) \subseteq \Gamma$$

is also a lattice in H_1H_2 and so has finite index in Γ. Studying now

$$\Gamma \cap H_i < H_i$$

we may obtain an irreducible lattice, and if not we may repeat the decomposition step as before. Ultimately we find finitely many irreducible lattices (that are potentially lattices in simple groups).

In this context the following notion is useful. Let $\Gamma, \Lambda < G$ be two subgroups. Then we say that Γ and Λ are *commensurable* if $\Gamma \cap \Lambda$ has finite index in both Γ and Λ.

Our interest in the notion of irreducibility is clearly explained in the following corollary.

Corollary 3.34 (Mixing of semi-simple groups). *Let G be the connected component of the group of $\mathbb{R}$-points of a semisimple algebraic group defined over $\mathbb{R}$. Suppose that G has no compact factors. Let $X = \Gamma\backslash G$ be the quotient by an irreducible quotient of G. Then every almost direct factor of G acts ergodically and the action of G is mixing with respect to the Haar measure m_X on X.*

Proof. By the Howe–Moore theorem for semi-simple groups (Theorem 2.7), it is sufficient to show that every simple factor acts ergodically.

So let $F \lhd G$ be a (non-trivial) simple factor of G, and suppose that

$$F\centerdot B = B \subseteq X$$

is F-invariant†.

Let

$$\pi_X : G \longrightarrow X = \Gamma\backslash G$$

be the natural factor map, and let $B_G = \pi_X^{-1}(B) \subseteq G$ be the set in G corresponding to B. By the properties of B we have $B_G F = B_G$, or equivalently $B_G = \pi^{-1}(\pi(B_G))$ if $\pi : G \to G/F$ denotes the projection map. By construction, $\Gamma B_G = B_G$ and so $\pi(\Gamma)\pi(B_G) = \pi(B_G)$.

Recall from [**?**, Prop. 8.6] that, for any two Borel sets $B_1, B_2 \subseteq G/F$ with $m_{G/F}(B_1)m_{G/F}(B_2) > 0$, the set

$$\bigl\{gF \in G/F \mid m_{G/F}\left(gFB_1 \cap B_2\right) > 0\bigr\}$$

is non-empty and open.

We may apply this to the sets $B_1 = \pi(B_G)$ and $B_2 = G/F\smallsetminus\pi(\Gamma)$. Since $\pi(\Gamma)$ is dense in G/F by Corollary 3.33, we deduce that either $\pi(B_G)$ has zero measure or its complement does. Since G is the almost direct product of F and G/F, we see that either B_G or its complement has zero measure in G. It follows that either $m_X(B) = 0$ or $m_X(X\smallsetminus B) = 0$ as required. □

Exercises for Section 3.6

Exercise 3.6.1. Let $D > 1$ be a non-square integer, and for

$$\alpha = a + b\sqrt{D} \in \mathbb{Q}(\sqrt{D})$$

let $\overline{\alpha} = a - b\sqrt{D}$ denote its Galois conjugate. Now let

$$\mathrm{SL}_2(\mathbb{Z}[\sqrt{D}]) = \left\{ g = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \mid \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22} \in \mathbb{Z}[\sqrt{D}], \det(g) = 1 \right\},$$

and consider $\mathrm{SL}_2(\mathbb{Z}[\sqrt{D}])$ as a subgroup of $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ using the *diagonal embedding*

$$\begin{aligned} \imath : \mathrm{SL}_2(\mathbb{Z}[\sqrt{D}]) &\longrightarrow \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}) \\ g = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} &\longmapsto (g, \overline{g}) \end{aligned}$$

where

† By [**?**, Prop. 8.3], we may assume the strict invariance $F\centerdot B = B$ rather than the *a priori* weaker invariance in the measure algebra $m_X(g\centerdot B \triangle B) = 0$ for all $g \in F$.

$$\overline{g} = \begin{pmatrix} \overline{\alpha}_{11} & \overline{\alpha}_{12} \\ \overline{\alpha}_{21} & \overline{\alpha}_{22} \end{pmatrix}.$$

(a) Show that $\Gamma = \imath\left(\mathrm{SL}_2(\mathbb{Z}[\sqrt{D}])\right) \leqslant \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ is a discrete subgroup.
(b) Show that Γ is a lattice in $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$.

Notes to Chapter 3

(10)(Page 76) Almost any algebra text will cover this material, for example Gerstein [**?**] or, for the more sophisticated aspects of the algebraic theory, see Lam [**?**].
(11)(Page 77) The word signature is used in various ways, all meaning that the number of $+1$s, -1s (and in the degenerate case 0s) can be reconstructed from the signature (and the dimension). The fact that the signature is a property of the form itself is Sylvester's law of inertia [**?**].
(12)(Page 79) Hilbert [**?**] proved this in his development of invariant theory.
(13)(Page 87) This was shown by Dirichlet [**?**] in 1846 for the ring $\mathbb{Z}[\zeta]$ (the understanding that this may not be the ring of integers in $\mathbb{Q}(\zeta)$ for an algebraic integer ζ came later, and of course the rank is not affected as $\mathbb{Z}[\zeta]$ has finite index in the ring of integers). We refer to the paper of Elstrodt [**?**] for an account of the history.
(14)(Page 103) The history, and various generalizations, of the implicit function theorem may be found in the account by Krantz and Parks [**?**]. The p-adic implicit function theorem may be found in the notes of Serre [**?**, p. 83].
(15)(Page 113) A modern proof from a sophisticated point of view is given by Conrad [**?**], and the original proof in Chevalley [**?**]. Any book book on algebraic groups will contain a version of the theorem (possibly not under this name).
(16)(Page 117) Borel [**?**] proved this for semi-simple Lie groups without compact factors; generalizations and simplifications have been provided by Furstenberg [**?**], Moskowitz [**?**] and Dani [**?**] among others.

Chapter 4
Quantitative Non-Divergence

In this chapter we will show that a unipotent trajectory cannot diverge to infinity in $\mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$. In fact we will show that unipotent orbits have 'no escape of mass', or in other words have 'quantitative non-divergence'. The former result was shown by Margulis [**?**] in his work on the arithmeticity of lattices, and the latter is Dani's refinement [**?**], [**?**], [**?**]. About 20 years later the argument was further refined by Kleinbock and Margulis [**?**] and Kleinbock [**?**], and applied to various Diophantine problems. As a corollary we will also obtain a special case of the Borel Harish-Chandra theorem [**?**]: $\mathbb{G}(\mathbb{Z})$ is a lattice in $\mathbb{G}(\mathbb{R})$ if $\mathbb{G}$ is a semi-simple algebraic group defined over $\mathbb{Q}$.

4.1 The Case of $\mathbf{SL_2}(\mathbb{Z})\backslash \mathbf{SL_2}(\mathbb{R})$.

We first describe a case that is both easy and familiar: horocycle orbits on

$$2 = \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R}).$$

We refer to Section 1.2 or [**?**, Ch. 9] for the background and to [**?**, Ch. 11] for a more detailed proof.

4.1.1 A Topological Claim

In the hyperbolic description of 2, the topological non-divergence claim is particularly easy to see.

Lemma 4.1 (Non-divergence for 2**).** *For any $x \in 2$ the horocycle orbit $u_t \centerdot x$ does not go to infinity as $t \to \infty$, nor as $t \to -\infty$.*

Proof. Every $x \in 2$ corresponds to a point $(z, v) \in \mathrm{T}^1(\mathbb{H})$ with z chosen in the usual fundamental domain, which we denote by F, for $\mathrm{SL}_2(\mathbb{Z})$ in $\mathbb{H}$ (see Section 1.2). To prove the lemma we find for a given x a compact set K and a sequence $t_n \to \infty$ with $u_{t_n} \cdot x \in K$ for all $n \geqslant 1$. If x is periodic under the action of $\{u_t \mid t \in \mathbb{R}\}$ then the orbit is compact and we may take

$$K = \{u_t \cdot x \mid t \in \mathbb{R}\}$$

and obtain the claim trivially. Otherwise, we may take

$$K = \{(z, v) \mid z \in F, \Im(z) \leqslant 1\}.$$

Then it is easy to see (from the geometric picture of the horocycle flow) that there exists some $t_1 \geqslant 0$ with $u_{t_1} \cdot x \in K$, as illustrated in Figure 4.1. In fact,

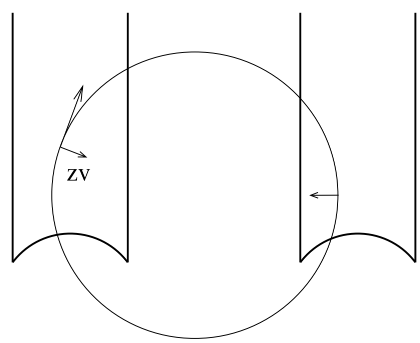

Fig. 4.1 A horocycle orbit returns to K.

the horocycle orbit is a circle touching $\mathbb{R}$. Hence it moves up and then down again, returning to K. Now consider the point $u_{t_1+1} \cdot x$, and apply the same argument to find some $t_2 \geqslant t_1 + 1$ with $u_{t_2} \cdot x \in K$. Repeating the argument proves the lemma by induction. □

4.1.2 Non-escape of Mass

While the topological statement in Lemma 4.1 above was easy to derive from the hyperbolic geometry of horocycle orbits, the quantitative claim is more difficult to see from this geometric picture. Hence we will switch the description and think of 2 as the space of unimodular lattices in $\mathbb{R}^2$.

Proposition 4.2 (Quantitative non-divergence for 2). *A point $x \in 2$ is either periodic for the horocycle flow or*[†] *has the property that there exists some $T_x \geqslant 0$ such that for all $\varepsilon > 0$ and all $T \geqslant T_x$ we have*

$$\frac{1}{T}\left|\{t \in [0,T] \mid u_t \bullet x \notin 2(\varepsilon)\}\right| \ll \varepsilon. \tag{4.1}$$

Here we are using $|A|$ as a shorthand for the Lebesgue measure of a subset $A \subseteq \mathbb{R}$, and the notation

$$2(\varepsilon) = \{x \in 2 \mid \lambda_1(x) \geqslant \varepsilon\}$$

introduced in Section 1.3.3.

Proof of Proposition 4.2. Suppose that x is not periodic, and assume first that the lattice Λ_x associated to x has no vectors of length less than 1. Fix $T \geqslant 0$ and define, for every vector $v \in \Lambda_x \smallsetminus \{0\}$ a 'protecting' intervals

$$P_v = \{t \in [0,T] \mid \|vu_t^{-1}\| < 1\}.$$

Notice that if $v = (v_1, v_2)$, then

$$\begin{aligned}\|vu_{-t}\| &= \|(v_1, v_2 - tv_1)\| \\ &= \sqrt{v_1^2 + (v_2 - tv_1)^2},\end{aligned} \tag{4.2}$$

and so P_v is a subinterval of $[0,T]$. If $v \in \Lambda_x \smallsetminus \{0\}$ is large enough (how large depends on T), then P_v is trivial. Hence there are only finitely many non-trivial intervals. As the unimodular lattice $\Lambda_x u_{-t}$ cannot contain two linearly independent vectors of length strictly less than 1, these intervals can only intersect if they are associated to linearly dependent vectors. To rule even this out, we choose within every Λ-rational line (that is, every line $\mathbb{R}v$ with $v \in \Lambda_x \smallsetminus \{0\}$) one and only one primitive vector in the lattice (that is, a vector $v \in \Lambda_x \smallsetminus \{0\}$ with $\mathbb{R}v \cap \Lambda_x = \mathbb{Z}v$). Let $v^{(1)}, \dots, v^{(n)}$ be the resulting list of pairwise linearly independent primitive vectors, so that $P_i = P_{v^{(i)}}$ and

$$P_1 \sqcup \cdots \sqcup P_n = \{t \in [0,T] \mid \lambda_1(u_t \bullet x) < 1\}. \tag{4.3}$$

Now let $\varepsilon \geqslant 0$ and define the 'bad' set

$$B_i^\varepsilon = \{t \in [0,T] \mid \|v^{(i)}u_{-t}\| \leqslant \varepsilon\}$$

for $i = 1, \dots, n$. We see that

$$B_1^\varepsilon \sqcup \cdots \sqcup B_n^\varepsilon = \{t \in [0,T] \mid \lambda_1(u_t \bullet x) \leqslant \varepsilon\}$$

[†] Note that the distinction of the two cases is absolutely necessary here: If $U \bullet x$ is a periodic orbit that is stuck high up in the cusp (equivalently a periodic orbit of short period), then the estimate (4.1) cannot hold uniformly for all $\varepsilon \leqslant 1$.

is precisely the set whose measure we wish to estimate. For this, we claim that

$$|B_i^\varepsilon| \ll \varepsilon \, |P_i| \tag{4.4}$$

for $i = 1, \dots, n$.

Summing this up, and using the disjointness in (4.3), the estimate (4.1) follows at once (for the case at hand, $\lambda_1(x) \geqslant 1$, and with $T_x = 0$).

To see the claim (4.4) we estimate both $|B_i^\varepsilon|$ and $|P_i|$ in terms of $|v^{(i)}|$. To simplify the notation we fix some $i \in \{1, \dots, n\}$ and drop the sub- and super-scripts. Notice first that we may assume $\varepsilon \leqslant \frac{1}{2}$ (for otherwise (4.4) is trivial) and hence $|v_1| \leqslant \frac{1}{2}$ (for otherwise B_i is empty and (4.4) is trivial). Thus, since $\lambda_1(x) \geqslant 1$ by definition and (4.2) we have $|v_2| \geqslant \sqrt{\frac{3}{4}}$ and

$$\begin{aligned} P_i &= \left\{t \in [0, T] \mid \|(v_1, v_2) u_t^{-1}\| < 1\right\} \\ &= \left\{t \in [0, T] \mid |v_2 - t v_1| < \sqrt{1 - (v_1)^2}\right\} \\ &\supseteq \left\{t \in [0, T] \mid |v_2 - t v_1| < \sqrt{\tfrac{3}{4}}\right\}. \end{aligned}$$

On the other hand we clearly have

$$B_i \subseteq \{t \in [0, T] \mid |v_2 - t v_1| \leqslant \varepsilon\}.$$

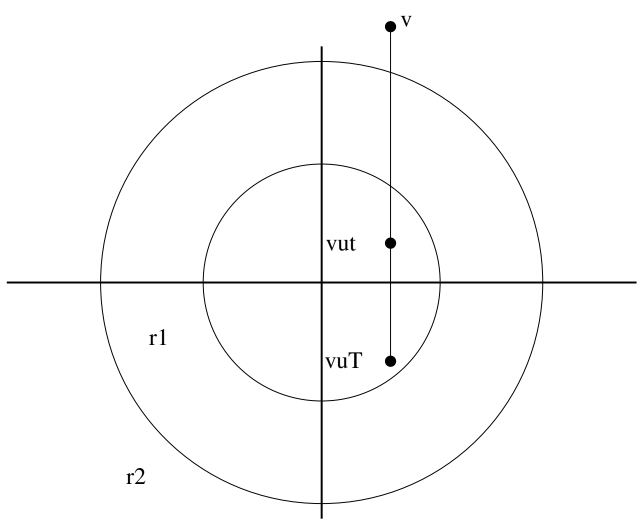

Fig. 4.2 The u_t-orbits of points $v \in \mathbb{R}^2$ travel at linear speed (determined by v_1). Thus the set B of bad times where $\|v u_t\| \leqslant \varepsilon$ is always a $\ll \varepsilon$-fraction of the protecting set P where $\|v u_t\| \leqslant 1$.

Since $v_2 - tv_1$ is linear in t with slope $-v_1$ as a function of t, it follows that either $B_i = \varnothing$ (this happens, for example, if $|v_1| > \varepsilon$) or

$$|P_i| \geqslant \left(\sqrt{\tfrac{3}{4}} - \tfrac{1}{2}\right)|v_1|^{-1}.$$

Here we subtract $\frac{1}{2}$ to also handle the case where $\|vu_{-T}\| < 1$, i.e. the right end point of the interval $\{t \mid |v_2 - tv_1| < \sqrt{\frac{3}{4}}\}$ is to the right of $[0, T]$. Similarly we get

$$|B_i| \leqslant 2\varepsilon|v_1|^{-1}.$$

Therefore (4.4) (and so also (4.1)) follows for any $T > 0$ and any $x \in 2$ with $\lambda_1(x) \geqslant 1$.

If now $x_0 \in 2$ is non-periodic but otherwise arbitrary, then there exists some $T_0 > 0$ for which $x = u_{T_0}{\scriptstyle\bullet}x_0$ has $\lambda_1(x) \geqslant 1$ by choosing T_0 such that the (unique up to sign) primitive vector $v_0 \in \Lambda_{x_0}$ with $\|v_0\| < 1$ has $\|v_0 u_{-T_0}\| = 1$. Let

$$\varepsilon_0 = \min_{t\in[0,T_0]} \lambda_1(u_t{\scriptstyle\bullet}x),$$

and let T_x be chosen with

$$\frac{T_0}{T_x} \leqslant \varepsilon_0.$$

Now let $\varepsilon \in (0, 1]$ be arbitrary and $T \geqslant T_x$. If $\varepsilon < \varepsilon_0$, then

$$\{t \in [0, T] \mid u_t{\scriptstyle\bullet}x \notin 2(\varepsilon)\} = \{t \in [T_0, T] \mid u_t{\scriptstyle\bullet}x \notin 2(\varepsilon)\},$$

and applying the first case to $u_{T_0}{\scriptstyle\bullet}x$ gives (4.1) in that case. If on the other hand $\varepsilon \geqslant \varepsilon_0$ then the first case applied to $u_{T_0}{\scriptstyle\bullet}x$,

$$\{t \in [0, T] \mid u_t{\scriptstyle\bullet}x \notin 2(\varepsilon)\} \subseteq [0, T_0] \cup \{t \in [T_0, T] \mid u_t{\scriptstyle\bullet}x \notin 2(\varepsilon)\},$$

and $\frac{T_0}{T} \leqslant \varepsilon_0 \leqslant \varepsilon$ again gives (4.1), completing the proof. □

Corollary 4.3 (Non-escape of mass for 2). *If $x \in 2$, then every weak*-limit of the collection of measures*

$$\left\{\frac{1}{T}\int_0^T (u_t)_*\delta_x \,\mathrm{d}t\right\}$$

is a probability measure on 2.

Exercises for Section 4.1

Exercise 4.1.1. Prove Corollary 4.3.

4.2 The Case of 3 $= \mathbf{SL_3}(\mathbb{Z})\backslash \mathbf{SL_3}(\mathbb{R})$

The proof for the generalizations of Proposition 4.2 and Corollary 4.3 becomes significantly more involved for d with $d \geqslant 3$. We start with the case $d = 3$ because it is easier to envision and because it already contains all the main ingredients of the general case.

4.2.1 Non-Escape of Mass for Polynomial Trajectories

Even though we are primarily interested in unipotent trajectories, we will prove a more general claim allowing for general *polynomial orbits* of the shape

$$\mathrm{SL}_3(\mathbb{Z})p(t)$$

for $t \geqslant 0$ or for $t \in [0, T]$ for some $T \geqslant 0$, where

$$p : \mathbb{R} \to \mathrm{Mat}_3(\mathbb{R})$$

is a polynomial map taking values in $\mathrm{SL}_3(\mathbb{R})$. We say that p has degree no more than D if each matrix entry is a polynomial of degree no more than D. Notice that if $\{u_t \mid t \in \mathbb{R}\}$ is a one-parameter unipotent subgroup (of which there are precisely two up to conjugation in $\mathrm{SL}_3(\mathbb{R})$) with Lie algebra $\mathbb{R}v$ then $p(t) = gu_{-t} = g\exp(-tv)$ is a polynomial in t for any $g \in \mathrm{SL}_3(\mathbb{R})$. Hence a unipotent trajectory is also a polynomial trajectory. This generalization comes more or less for free in the sense that it does not complicate the proof significantly, while the generalization does have interesting consequences.

Much like a short periodic orbit for the horocycle flow on 2, there is always the possibility that there are 'rational reasons' for a polynomial trajectory to remain stuck in the cusp in the following sense. There could be a vector $v \in \mathbb{Z}^3$ with $vp(t) = vp(0)$ for all t and with $\|vp(0)\|$ being small, or there could be a rational plane $V \subseteq \mathbb{R}^3$ with $Vp(t) = Vp(0)$ for all t such that the co-volume of $Vp(0) \cap \mathbb{Z}^3$ is small†, in which case there always exists for every $t \in \mathbb{R}$ a short vector in $(V \cap \mathbb{Z}^3)p(t)$, which may depend on t.

---

† If $Vp(t) = Vp(0)$ for all $t \in \mathbb{R}$, then

$$\mathrm{volume}\bigl(Vp(t)/(\mathbb{Z}^3 \cap V)p(t)\bigr) = \mathrm{volume}\bigl(Vp(0)/(\mathbb{Z}^3 \cap V)p(0)\bigr)$$

for all $t \in \mathbb{R}$. If $p(t) = gu_{-t}$ is the parametrization of an orbit under a one-parameter unipotent subgroup, this is clear as the restriction of the unipotent subgroup to the invariant subspace Vg is again unipotent. In general, $\bigwedge^2 p(t)$ sends by assumption the line in $\bigwedge^2 \mathbb{R}^3$ corresponding to V to one and the same line for every t. If the co-volume of $(\mathbb{Z}^3 \cap V)p(t)$ inside $Vp(t)$ is not constant, or equivalently if $\bigwedge^2 p(t)$ applied to an element w of $\bigwedge^2 V \subseteq \bigwedge^2 \mathbb{R}^3$ is not constant, then $w \bigwedge^2 p(t) = (w \bigwedge^2 p(0))h(t)$ for a non-constant $\mathbb{R}$-valued polynomial $h(t)$. As $h(t)$ has a complex root, we get a contradiction to $\bigwedge^2 p(\mathbb{C}) \subseteq \mathrm{SL}(\bigwedge^2 \mathbb{C}^3)$.

Similarly, a finite piece

$$\Gamma p(t)$$

for $t \in [0,T]$ of the trajectory would surely be entirely far out if there were a vector $v \in \mathbb{Z}^3$ with

$$\|vp(t)\| \leqslant \eta$$

for all $t \in [0,T]$, or if there is a rational plane $V \subseteq \mathbb{R}^3$ for which

$$\text{volume}\big(Vp(t)/(V \cap \mathbb{Z}^3)p(t)\big) \leqslant \eta^2$$

for all $t \in [0,T]$.

As the last volume expression looks quite complicated but expresses the simple concept that we are studying the volume of the deformed plane with respect to the deformed lattice inside it, we now define some abbreviations for such expressions. For any $d \geqslant 2$ and any given discrete subgroup $\Lambda \leqslant \mathbb{R}^d$ (possibly of smaller rank) we write (Λ) as shorthand for the volume of $\mathbb{R}\Lambda/\Lambda$. Also if a polynomial $p(t) \in \mathrm{SL}_d(\mathbb{R})$ is given, we define for the study of the polynomial orbit $\mathbb{Z}^d p(t)$ the expression

$$(V,t) = \big((V \cap \mathbb{Z}^d)p(t)\big)$$

for any rational subspace $V \subseteq \mathbb{R}^d$.

To avoid the above mentioned 'rational constraints' for $d = 3$ we assume that there is some $\eta \leqslant 1$ such that

$$\sup_{t\in[0,T]} \|vp(t)\| \geqslant \eta \tag{4.5}$$

for all $v \in \mathbb{Z}^3 \smallsetminus \{0\}$, and

$$\sup_{t\in[0,T]} \big((V \cap \mathbb{Z}^3)p(t)\big) \geqslant \eta^2 \tag{4.6}$$

for all rational planes $V \subseteq \mathbb{R}^3$. Using our abbreviation we could combine these two estimates into the assumption that

$$\sup_{t\in[0,T]} (V,t) \geqslant \eta^{\dim V}$$

for any rational subspace $V \subseteq \mathbb{R}^3$. This unified treatment of all intermediate subspaces will be our view point in the general case, see Section 4.3, but will also play a role in the proof of the following theorem[(17)].

Theorem 4.4 (Quantitative non-divergence for 3). *Suppose that the piece $\Gamma p(t)$, $t \in [0,T]$ of a polynomial trajectory satisfies* (4.5) *and* (4.6) *for some $\eta \leqslant 1$. Then, for $\varepsilon \in (0,\eta]$,*

$$\frac{1}{T}\left|\{t\in[0,T]\mid \Gamma p(t)\notin 3(\varepsilon)\}\right| \ll_D \left(\frac{\varepsilon}{\eta}\right)^{1/2D}$$

where p is a polynomial of degree no more than D.

Remark 4.5. (1) The alternating tensor product $\bigwedge^2(\mathbb{R}^3)$ may be identified with $\mathbb{R}^3$ by choosing (for example) the basis $e_2\wedge e_3$, $e_3\wedge e_1$ and $e_1\wedge e_2$ where as usual e_1,e_2,e_3 is the standard basis of $\mathbb{R}^3$. This way the map

$$(v,w)\in\mathbb{R}^3\times\mathbb{R}^3\to v\wedge w\in\bigwedge^2\mathbb{R}^3$$

is identified with the exterior product

$$(v,w)\in\mathbb{R}^3\times\mathbb{R}^3\to v\times w\in\mathbb{R}^3.$$

The linear map

$$\bigwedge^2 p(t):\bigwedge^2(\mathbb{R}^3)\longrightarrow\bigwedge^2(\mathbb{R}^3)$$

is then the linear map with

$$e_i\wedge e_j\longmapsto(e_ip(t))\wedge(e_jp(t))$$

for $1\leqslant i,j\leqslant 3$. It is again a polynomial (of at most doubled degree) with values in $\mathrm{SL}\left(\bigwedge^2(\mathbb{R}^3)\right)$. Moreover, note that the co-volume of $\mathbb{Z}v_1+\mathbb{Z}v_2$ equals the area of the parallelogram spanned by v_1 and v_2, or equivalently the length of $v_1\wedge v_2$ (identified with the exterior product $v_1\times v_2$).

(2) As we will see in the course of the proof, the exponent $2D$ can be replaced by any $\ell\geqslant 1$ with the property that $\|vp(t)\|^2$ for $v\in\mathbb{R}^3$ and $\|w\bigwedge^2p(t)\|^2$ for $w\in\bigwedge^2(\mathbb{R}^3)$ are polynomials of degree no more than 2ℓ. Notice that the choice $\ell=2D$ has this property. In the case of the orbit of the one-parameter unipotent subgroup given by

$$p(t)=g\begin{pmatrix}1&-t&\\&1&\\&&1\end{pmatrix}$$

we may take $\ell=1$, while for that defined by

$$p(t)=g\begin{pmatrix}1&-t&\frac12t^2\\&1&-t\\&&1\end{pmatrix}$$

we may take $\ell=2$.

(3) There are two ways in which one can establish the assumptions (4.5) and (4.6), and both are important in applying Theorem 4.4.

(a) Given p and T, one can find $\eta>0$ with the desired property, for example by taking

$$\eta = \min\left\{\lambda_1(\mathbb{Z}^3 p(0)), \sqrt{\alpha_2(\mathbb{Z}^3 p(0))}\right\}.$$

(b) Given p such that $vp(t)$ is non-constant for any $v \in \mathbb{Z}^3$ and also $Vp(t)$ is a non-constant subspace for any rational plane $V \subseteq \mathbb{R}^2$, one can find some $T_0 > 0$ such that for $T \geqslant T_0$ we can use $\eta = 1$. In fact, there are only finitely many vectors $v \in \mathbb{Z}^3$ with $\|vp(0)\| \leqslant 1$, and for each of them $vp(t)$ is non-constant and hence there must be some T_0 such that (4.5) holds for $T \geqslant T_0$ and $\eta = 1$. The argument to establish (4.6) is similar.

4.2.2 A Lemma About Polynomials

We now prove a lemma which replaces the argument involving the linear function $v_2 - tv_1$ in the proof of Proposition 4.2 (see in particular Figure 4.1).

Lemma 4.6 (Small values of polynomials). *Let $p \in \mathbb{R}[t]$ be a polynomial of degree L, and fix $T > 0$. Then for every $\varepsilon > 0$,*

$$\frac{1}{T}\left|\left\{t \in [0,T] \mid |p(t)| < \varepsilon \|p\|_{T,\infty}\right\}\right| \ll_L \varepsilon^{1/L}, \tag{4.7}$$

where

$$\|p\|_{T,\infty} = \sup_{t\in[0,T]} |p(t)|.$$

The situation is illustrated in Figure 4.3 for the polynomial $p(t) = t^4$.

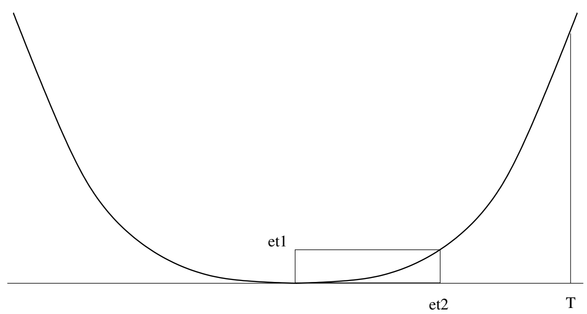

Fig. 4.3 The graph of $p(t) = t^L$ for $L = 4$ shows that the left-hand side of (4.7) can indeed be of the size $\varepsilon^{1/L}$.

The main property of polynomials that will be used in the proof of Theorem 4.4 is Lemma 4.6. A function or family of functions $p : [0,T] \to \mathbb{R}$ is† *polynomial-like of degree no more than L*, or simply is *of degree no more*

† The more common, but less informative, terminology is (C,α)-good, where $\alpha = \frac{1}{L}$ and C is the implied constant.

than L if p satisfies the conclusion of Lemma 4.6, and the implied constant does not depend on the particular function p if a whole family of such functions is being considered. We will not pursue this generality here, and instead refer to the papers of Kleinbock and Margulis [**?**] and of Kleinbock [**?**].

PROOF OF LEMMA 4.6. By induction we may assume that the lemma already holds for all polynomials of degree less than L. The claim of the lemma is invariant under the following transformations:

- Replacing p by $\frac{1}{\|p\|_{T,\infty}}p$.
- Replacing T by 1 and at the same time $p(t)$ by $p(tT)$ for $t \in [0,1]$.

Thus we may assume without loss of generality that $T = 1$ and $\|p\|_{1,\infty} = 1$. Let $a_1, \dots, a_r \in \mathbb{R}$ and $z_1, \overline{z_1}, \dots, z_s, \overline{z_s} \in \mathbb{C} \smallsetminus \mathbb{R}$ be the list of real zeros and pairs of complex conjugate zeros of p, listed with multiplicity so that

$$r + 2s = L.$$

Let $b \in \mathbb{R}$ be the leading coefficient of p, so that

$$p(t) = b(t - a_1) \cdots (t - a_r)(t - z_1)(t - \overline{z_1}) \cdots (t - z_s)(t - \overline{z_s}).$$

Suppose first that $|a_1| \geqslant 2$. Then we have

$$1 \ll \left| \frac{t - a_1}{a_1} \right| \ll 1$$

for all $t \in [0,1]$. Hence the claim for p is equivalent to the claim for the polynomial

$$\tilde{p}(t) = b a_1 (t - a_2) \cdots (t - a_r)(t - z_1)(t - \overline{z_1}) \cdots (t - z_s)(t - \overline{z_s})$$

of degree $L-1$ (with different multiplicative constants). Similarly, if $|z_1| \geqslant 2$, then

$$1 \ll \left| \frac{(t - z_1)(t - \overline{z_1})}{z_1 \overline{z_1}} \right| \ll 1$$

for all $t \in [0,1]$, and we may reduce the claim to a polynomial of degree $L-2$.

Thus we may assume that

$$|a_1|, \dots, |a_r|, |z_1|, \dots, |z_s| \leqslant 2.$$

Now for $t \in [0,1]$ we have

$$|t - a_i| \leqslant 3 \text{ and } |t - z_j| \leqslant 3$$

for $i = 1, \dots, r$ and $j = 1, \dots, s$. It follows that

$$1 = \|p\|_{1,\infty} \leqslant |b| 3^L. \tag{4.8}$$

Suppose now

$$|\{t \in [0,1] \mid |q(t)| < \varepsilon\}| \ll_L \varepsilon^{1/L} \tag{4.9}$$

holds for all $\varepsilon > 0$ and the polynomial

$$q(t) = (t-a_1)\cdots(t-a_r)(t-z_1)(t-\overline{z_1})\cdots(t-z_s)(t-\overline{z_s}) = \frac{1}{b}p(t).$$

Then, since $|p(t)| < \varepsilon$ implies $|q(t)| < \frac{\varepsilon}{b}$, we get from (4.8)–(4.9) that

$$|\{t \in [0,T] \mid |p(t)| < \varepsilon\}| \ll_L \left(\frac{\varepsilon}{b}\right)^{1/L} \leqslant 3\varepsilon^{1/L}$$

and so the lemma.

It remains to prove (4.9). Suppose that $t \in [0,1]$ has distance at least $\varepsilon^{1/L}$ from any of the zeros

$$a_1, \dots, a_r, z_1, \overline{z_1}, \dots, z_s, \overline{z_s}.$$

Then clearly $|q(t)| \geqslant \varepsilon$. On the other hand the elements $t \in [0,1]$ with distance less than $\varepsilon^{1/L}$ from a_i (respectively z_i) lie in an interval of length at most $2\varepsilon^{1/L}$. This gives (4.9), with $2(r+s) \leqslant 2L$ as the implied constant. As discussed above, the lemma follows. □

4.2.3 Protection Arising From a Flag

The most important feature that makes the proof of Proposition 4.2 easier than the case of $\mathrm{SL}_3(\mathbb{R})$ considered here is the fact that a unimodular lattice $\Lambda \leqslant \mathbb{R}^2$ cannot have two linearly independent vectors of length less than one. This gave automatic 'protection' from short vectors: if there is a Λ-primitive vector of length less than one, and this vector is not tiny, then no tiny non-zero vector can exist in Λ. Using this we defined protecting intervals which were automatically disjoint.

This property of only one short vector is manifestly false for unimodular lattices in $\mathbb{R}^3$. For example, the lattice

$$\Lambda_n = \tfrac{1}{n}\mathbb{Z}e_1 + \tfrac{1}{n}\mathbb{Z}e_2 + n^2\mathbb{Z}e_3$$

is unimodular for any $n \geqslant 1$, and contains two linearly independent vectors of length $\frac{1}{n}$. What we need to discuss in order to get a similar protection phenomenon in $\mathbb{R}^3$ are flags.

A *flag* in $\mathbb{R}^d$ is a collection comprising a line

$$V_1 = \mathbb{R}v_1,$$

a plane

$$V_2 = \mathbb{R}v_1 + \mathbb{R}v_2 \supseteq V_1,$$

and so on up to a hyperplane

$$V_{d-1} = V_{d-2} + \mathbb{R}v_{d-1} \supseteq V_{d-2}.$$

We also write $V_0 = \{0\}$ and $V_d = \mathbb{R}^d$.

Lemma 4.7 (Protection coming from flags). *Let $\Lambda \leqslant \mathbb{R}^d$ be a unimodular lattice, and let*

$$V_0 = \{0\} \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_d = \mathbb{R}^d$$

be a flag of Λ-rational subspaces. Then

$$\lambda_1(\Lambda) \geqslant \min_{i=1,\dots,d} \left\{ \frac{(\Lambda \cap V_i)}{(\Lambda \cap V_{i-1})} \right\},$$

where $(\{0\}) = (\Lambda) = 1$.

This gives the desired protection in the following sense, illustrated for the case $d = 3$: If $(\Lambda \cap V_1)$ is of size roughly ε, and $(\Lambda \cap V_2)$ is of size roughly ε^2, then Λ does not contain vectors that are much shorter than ε.

Proof of Lemma 4.7. Let $v \in \Lambda$ be chosen with norm $\|v\| = \lambda_1(\Lambda)$. If v does not lie in V_{d-1}, then the co-volume of

$$\Lambda \cap V_{d-1} + \mathbb{Z}v \subseteq \mathbb{R}^d$$

is equal to $(\Lambda \cap V_{d-1}) \cdot \|\pi(v)\|$, where $\pi : \mathbb{R}^d \to V_{d-1}^{\perp}$ is the orthogonal projection. In particular, since the co-volume of $\Lambda \supseteq \Lambda \cap V_{d-1} + \mathbb{Z}v$ is 1, we have

$$\begin{aligned} 1 = (\Lambda) &\leqslant (\Lambda \cap V_{d-1} + \mathbb{Z}v) \\ &= (\Lambda \cap V_{d-1})\|\pi(v)\| \\ &\leqslant (\Lambda \cap V_{d-1})\|v\|, \end{aligned}$$

which implies the lemma in the case $v \notin V_{d-1}$.

Suppose now that $v \in \Lambda \cap V_{i+1}$ but $v \notin V_i$ for some $i \in \{1, \dots, d\}$. As before,

$$\begin{aligned} (\Lambda \cap V_{i+1}) &\leqslant (\Lambda \cap V_i + \mathbb{Z}v) \\ &= (\Lambda \cap V_i)\|\pi(v)\| \\ &\leqslant (\Lambda \cap V_i)\|v\|, \end{aligned}$$

where π is the appropriate projection, and the lemma follows immediately. □

To handle the lack of disjointness of the protecting intervals for individual vectors or subspaces we are also going to use a simple[(18)] *covering lemma*.

Lemma 4.8 (A covering lemma on intervals). *Let $I \subseteq \mathbb{R}$ be a compact interval, and let $P_1, \dots, P_K \subseteq I$ be a finite collection of compact sub-intervals. Then there exists a subcollection of these intervals $P_{j(1)}, \dots, P_{j(k)}$ which are* nearly disjoint *in the sense that*

$$\sum_{\ell=1}^{k} \mathbb{1}_{P_{j(\ell)}} \leqslant 2 \tag{4.10}$$

while still having the same union,

$$\bigcup_{\ell=1}^{k} P_{j(\ell)} = \bigcup_{n=1}^{K} P_n. \tag{4.11}$$

None of the selected intervals $P_{j(\ell)}$ is strictly contained in any of the other intervals $P_1, \dots, P_K$.

Proof. Let

$$U = \bigcup_{n=1}^{K} P_n,$$

and note that we may remove from the finite collection of intervals any interval P_n which is properly contained in any other interval of the list without affecting the set U. Below we construct our subcollection of intervals from the remaining ones with a simple greedy algorithm, which will immediately imply the last claim in the lemma.

Let $P_n = [a_n, b_n]$ for $n = 1, \dots, K$, and let $d_0 = \min U$. Among all intervals P_n with $a_n = d_0$, there is one with maximal b_n. We select this interval first, writing d_1 for its endpoint so

$$P_{j(1)} = [d_0, d_1].$$

If d_1 is an interior point of U then we list all intervals containing d_1 and choose once again an interval whose right-hand end point is largest among all those containing d_1. Formally,

$$\begin{aligned} d_1 &= \max\{b_n \mid a_n = d_0\}, \\ d_2 &= \max\{b_n \mid a_n \leqslant d_1 < b_n\} > d_1, \end{aligned}$$

and

$$\begin{aligned} P_{j(1)} &= [d_0, d_1], \\ P_{j(2)} &= P_n = [a_n, d_2], \end{aligned}$$

where we pick some n as in the definition of d_2.

If d_2 is again an interior point of U we continue in the same way until we come to a boundary point of U. At this stage we may restart the process with the next biggest element of U if there is one, or finish the process of selecting intervals if U has no remaining points to the right of the right end point of the last selected interval.

By construction, at each stage of the selection of the intervals

$$P_{j(1)}, \dots, P_{j(\ell)}$$

we have

$$\bigcup_{r=1}^{\ell} P_{j(r)} = U \cap (-\infty, b_{j(\ell)}],$$

giving (4.11). Moreover, the only selected interval other than $P_{j(1)}$ that can intersect non-trivially with $P_{j(1)}$ is $P_{j(2)}$, since if

$$P_{j(1)} \cap P_{j(\ell)} \neq \varnothing$$

for some $\ell \geqslant 2$ then $a_{j(\ell)} \leqslant d_1$, d_1 is an interior point of U, and

$$b_{j(\ell)} > b_{j(1)} = d_1$$

(since $\ell \geqslant 2$). Hence $b_{j(\ell)} \leqslant d_2$ by our choice of d_2, which forces ℓ to be 2. Repeating this argument gives (4.10) as required. □

4.2.4 Non-Divergence for 3 — Obtaining Protecting Flags

In the course of the proof we will treat 1- and 2-dimensional subspaces on the same footing, so we will use the notation V uniformly for both from now on.

PROOF OF THEOREM 4.4. Assume that $p : [0, T] \to \mathrm{SL}_3(\mathbb{R})$ has the property that

$$(V, t)^2$$

is polynomial of degree† no more than $2D$ for every rational subspace $V \subseteq \mathbb{R}^3$. Furthermore, let $\eta \leqslant 1$ satisfy (4.5) and (4.6), and fix $\varepsilon \in (0, \eta]$.

FIRST STAGE PROTECTION INTERVALS: Notice that there are only finitely many rational subspaces $V \subseteq \mathbb{R}^3$ for which

$$(V, t) \leqslant \eta^{\dim V}$$

† We will here use this meaning of the parameter D, rather than the one used in the theorem (see Remark 4.5).

for some $t \in [0, T]$. For each of those subspaces V we define the intervals $P_{V,i}$ for $i = 1, \ldots, \ell_V$ to be the set of maximal subintervals† of

$$\{t \in [0, T] \mid (V, t) \leqslant \eta^{\dim V}\}.$$

Notice that by maximality of the subintervals and the assumptions (4.5) and (4.6) we have $(V, t) = \eta^{\dim V}$ for at least one of the endpoints of each of the intervals $P_{V,i}$. In particular

$$\sup_{t \in P_{V,i}} (V, t) = \eta^{\dim V}. \tag{4.12}$$

This defines a collection of closed intervals $P_{V,i}$ where we vary both V and i. Applying Lemma 4.8 to this collection and the interval $[0, T]$, we obtain a nearly disjoint subcollection

$$P_1, \ldots, P_m$$

of these intervals with

$$\bigcup_V \bigcup_i P_{V,i} = \bigcup_{r=1}^m P_r$$

and with

$$\sum_{i=1}^m \mathbb{1}_{P_i} \leqslant 2.$$

We write V_r for the subspace that gave rise to the interval $P_r = P_{V_r, i_r}$ for some $i_r \in \{1, \ldots, \ell_{V_r}\}$. As this subspace alone does not give protection (since Lemma 4.7 needs a complete flag and we only have one subspace), we need to do another search for a compatible subspace as follows.

Second stage protection intervals: Suppose first that V_r for

$$r \in \{1, \ldots, m\}$$

is a line. Consider now the intervals

$$P_{V,i} \cap P_r$$

for all rational planes $V \subseteq \mathbb{R}^3$ that are compatible‡ with V_r. Now apply the covering lemma on P_r to this collection to obtain nearly disjoint subintervals

$$P_{r,1}, \ldots, P_{r,n(r)} \subseteq P_r$$

with

† As each such subinterval accounts for two roots of the polynomial equation $(V, t)^2 = \eta^{2 \dim V}$, there can be at most D such intervals.

‡ That is, with $V_r \subseteq V$.

$$\bigcup_{\substack{V_r \subseteq V,\\ \text{plane}}} \bigcup_i P_{V,i} \cap P_r = \bigcup_{s=1}^{n(r)} P_{r,s} \cap P_r.$$

Similarly, if V_r for $r \in \{1, \dots, m\}$ is a plane, then we obtain nearly disjoint subintervals

$$P_{r,1}, \dots, P_{r,n(r)} \subseteq P_r$$

defined by compatible rational lines $V \subseteq V_r$ with

$$\bigcup_{\substack{V \subseteq V_r,\\ \text{line}}} \bigcup_i P_{V,i} \cap P_r \subseteq \bigcup_{s=1}^{n(r)} P_{r,s} \cap P_r.$$

In both cases $n(r) = 0$ is possible.

Just as we denote by V_r the subspace that gave rise to the interval P_r, we also write $V_{r,s}$ for the subspace giving rise to $P_{r,s}$ and $i_{r,s}$ for the corresponding index so that $P_{r,s} = P_{V_{r,s}, i_{r,s}} \cap P_r$.

By construction V_r and $V_{r,s}$ are compatible (that is, they define a complete flag in $\mathbb{R}^3$) for all r and s. We will show that the intervals

$$P_1, \dots, P_m, P_{1,1}, \dots, P_{1,n(1)}, \dots, P_{m,1}, \dots, P_{m,n(m)}$$

together give the desired protection.

Bad subsets: We now define for $\varepsilon > 0$ the associated bad subsets of the intervals above:

$$\begin{aligned}(r, \varepsilon) &= \left\{ t \in P_r \mid (V_r, t) \leqslant \varepsilon \eta^{\dim V_r - 1} \right\}, \\ (r, s, \varepsilon) &= \left\{ t \in P_{r,s} \mid (V_{r,s}, t) \leqslant \varepsilon \eta^{\dim V_{r,s} - 1} \right\}\end{aligned}$$

and the union

$$(\varepsilon) = \bigcup_{r=1}^{m} \left((r, \varepsilon) \cup \bigcup_{s=1}^{n(r)} (r, s, \varepsilon) \right).$$

Estimate of bad subset: By Lemma 4.6 applied to[†]

$$(V_r, t)^2$$

on the interval P_r we get

$$|(r, \varepsilon)| \ll \left(\frac{\varepsilon}{\eta} \right)^{1/D} |P_r|, \tag{4.13}$$

by (4.12).

[†] Note that (V_r, t) is in general not a polynomial, but $(V_r, t)^2$ is a polynomial.

To prove the same for (r, s, ε) we need to show an analogue of (4.12) for $P_{r,s} = P_{V_{r,s},i_{r,s}} \cap P_r$. For this, notice that by Lemma 4.8 (from the first application that gave rise to $P_1, \dots, P_r, \dots, P_m$) none of the intervals $P_{V_{r,s},i_{r,s}}$ can contain P_r properly — let us refer to this as the *non-containment.* If both end points t of $P_{V_{r,s},i_{r,s}}$ satisfy $(V_{r,s}, t) = \eta^{\dim V_{r,s}}$ (because they are in $(0, T)$, for example) then (due to the non-containment) one of them must be in P_r, and so

$$\sup_{t \in P_{r,s}} (V_{r,s}, t) = \eta^{\dim V_{r,s}}. \tag{4.14}$$

If, on the other hand, we have $(V_{r,s}, t) < \eta^{\dim V_{r,s}}$ for one of the endpoints of $P_{V_{r,s},i_{r,s}}$ (this endpoint would have to be 0 or T), then the other will have to be in P_r (due to the non-containment) and we again get (4.14). Therefore, using Lemma 4.6 together with (4.14) as a replacement for (4.12) gives as before

$$|(r, s, \varepsilon)| \ll \left(\frac{\varepsilon}{\eta}\right)^{1/D} |P_{r,s}|\,. \tag{4.15}$$

Since the intervals $P_{r,s} \subseteq P_r$ are all nearly disjoint we get

$$\sum_{s=1}^{n(r)} |P_{r,s}| \leqslant 2\,|P_r| \ll |P_r|\,. \tag{4.16}$$

Thus we may take the union and use (4.13), (4.15) and (4.16) to obtain the estimate

$$\begin{aligned} |(\varepsilon)| &\leqslant \sum_{r=1}^{m} \left(|(r, \varepsilon)| + \sum_{s=1}^{n(r)} |(r, s, \varepsilon)| \right) \\ &\ll \left(\frac{\varepsilon}{\eta}\right)^{1/D} \sum_{r=1}^{m} |P_r| \\ &\leqslant 2 \left(\frac{\varepsilon}{\eta}\right)^{1/D} T, \end{aligned}$$

since the intervals

$$P_1, \dots, P_m \subseteq [0, T]$$

are nearly disjoint.

Protection: We now show that

$$\{t \in [0, T] \mid \mathrm{SL}_3(\mathbb{Z})p(t) \notin 3(\varepsilon)\} \subseteq (\varepsilon), \tag{4.17}$$

for all $\varepsilon \leqslant \eta$, so that the estimate above then implies the theorem.

Suppose therefore that $t \in [0, T]$ has the property that $\mathbb{Z}^3 p(t)$ contains an ε-short vector $vp(t)$. Since $\varepsilon \leqslant \eta$, this shows that t belongs to one of the protecting intervals defined by $V=\mathbb{R}v$. Hence we must have $t \in P_r$ for some $r \in \{1,\dots,m\}$ by choice of these intervals.

If $V = V_r$ then we have $t \in (r,\varepsilon) \subseteq (\varepsilon)$. If V_r is a line but $V \neq V_r$, then $V+V_r$ is a subspace compatible with V_r and

$$(V+V_r,t) \leqslant (V,t)\,(V_r,t) \leqslant \varepsilon\eta \leqslant \eta^2.$$

Therefore $t \in P_r \cap P_{V+V_r,i}$ (for some i) and so $t \in P_{r,s}$ for some $s \in \{1,\dots,n(r)\}$ by construction. We have obtained a complete flag: $V_r \subseteq V_{r,s}$ with

$$t \in P_r \cap P_{r,s}.$$

Suppose now that V_r is a plane. Recall that

$$(V,t) \leqslant \varepsilon,$$
$$(V_r,t) \leqslant \eta.$$

We may assume that $V \subseteq V_r$. For if $V+V_r=\mathbb{R}^3$, $\eta \leqslant 1$ and $\varepsilon < 1$ (which we may assume) then we get a contradiction to the unimodularity of the three-dimensional lattice. Therefore, $t \in P_r \cap P_{V,i}$ for some i and so there must exist some $s \in \{1,\dots,n(r)\}$ with $t \in P_{r,s}$. Once more we have obtained a complete flag: $V_{r,s} \subseteq V_r$ with $t \in P_r \cap P_{r,s}$.

Hence it remains to consider the case $t \in P_r$ and $t \in P_{r,s}$. Let us also assume, for the purposes of a contradiction, that

$$t \notin (r,\varepsilon) \cup (r,s,\varepsilon)\,.$$

Hence

$$\varepsilon\eta^{\dim V_r - 1} \leqslant (V_r,t) \leqslant \eta^{\dim V_r}$$

and

$$\varepsilon\eta^{\dim V_{r,s} - 1} \leqslant (V_{r,s},t) \leqslant \eta^{\dim V_{r,s}},$$

and together V_r and $V_{r,s}$ define a flag in $\mathbb{R}^3$. Lemma 4.7 may now be applied to show that

$$\lambda_1\left(\mathbb{Z}^3 p(t)\right) \geqslant \min\left(\varepsilon, \tfrac{1}{\eta^2}\right) = \varepsilon,$$

in contradiction to the assumption on t. This proves the claim (4.17), and hence the theorem. □

4.3 The General Case of $d = \mathrm{SL}_d(\mathbb{Z})\backslash\,\mathrm{SL}_d(\mathbb{R})$

Let us now state and prove the general version of the non-divergence theorem (using the abbreviations and tools introduced in the last section).

Theorem 4.9 (Quantitative non-divergence for d by Margulis, Dani and Kleinbock[(19)]). *Suppose that*

$$p : \mathbb{R} \to \mathrm{SL}_d(\mathbb{R})$$

is a polynomial and $T > 0$ is such that

$$\sup_{t\in[0,T]} (V,t) \geqslant \eta^{\dim V} \tag{4.18}$$

for some $\eta \in (0,1]$ and all rational subspaces $V \subseteq \mathbb{R}^d$. Assume furthermore that $2D$ is an upper bound for the degrees of $(V,t)^2$ for all rational subspaces $V \subseteq \mathbb{R}^d$. Then, for $\varepsilon \in (0,\eta]$,

$$\frac{1}{T}\left|\{t \in [0,T] \mid \varGamma p(t) \notin d(\varepsilon)\}\right| \ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{1/D}. \tag{4.19}$$

Proof. The proof comprises the following steps:

(1) iterated construction of protecting intervals and partial flags;
(2) definition and estimate of the bad subsets;
(3) reaching the conclusion by combining the established properties.

Inductive step to construct protecting intervals: Suppose we are given an interval $I \subseteq \mathbb{R}$ and a 'partial flag'

$$\mathscr{F} = \bigl\{\{0\} \subsetneq V_1 \subsetneq V_2 \subsetneq \cdots \subsetneq V_k \subsetneq \mathbb{R}^d\bigr\}$$

of rational subspaces of $\mathbb{R}^d$ with $0 \leqslant k < d-1$ such that

$$\sup_{t\in I}(V_j,t) \leqslant \eta^{\dim V_j}$$

for $j = 1,\dots,k$ and

$$\sup_{t\in I}(V,t) \geqslant \eta^{\dim V} \tag{4.20}$$

for any rational subspace $V \leqslant \mathbb{R}^d$ that is compatible† with $\mathscr{F}$. Initially we have $k = 0$ and (4.20) is precisely the assumption of Theorem 4.9.

Now consider all rational subspaces $V \leqslant \mathbb{R}^d$ that are compatible with the partial flag $\mathscr{F}$. For each such subspace split

$$\{t \in I \mid (V,t) \leqslant \eta^{\dim V}\}$$

into its connected components, giving rise to subintervals

$$P_{V,1},\dots,P_{V,\ell_V}.$$

† This means that $V \notin \mathscr{F}$ and $\mathscr{F} \cup \{V\}$ is again a partial flag or a flag.

Varying both V and the second index, we may apply Lemma 4.8 to obtain a finite nearly disjoint subcollection

$$P_1, \dots, P_m$$

of these intervals with the same union, so

$$\bigcup_{\substack{V \text{ compatible} \\ \text{with } \mathscr{F}}} \{t \in I \mid (V,t) \leqslant \eta^{\dim V}\} = \bigcup_{r=1}^{m} P_r$$

and and union is nearly disjoint so that

$$\sum_{r=1}^{m} |P_r| \leqslant 2|I|.$$

Let us write V_r for the subspace that gave rise to the interval $P_r = P_{V_r, i_r}$ for some i_r.

On each of those sub-intervals P_r we have the new (maybe partial or complete) flag

$$\mathscr{F} \cup \{V_r\}$$

with

$$\sup_{t \in P_r} (V,t) \leqslant \eta^{\dim V}$$

for all $V \in \mathscr{F} \cup \{V_r\}$. Now let V be either V_r or a rational subspace that is compatible with $\mathscr{F} \cup \{V_r\}$. In particular, V is compatible with $\mathscr{F}$ and so was considered in the construction of the subintervals

$$P_1, \dots, P_m \subseteq I.$$

By Lemma 4.8 this shows that P_r is not strictly contained in any of the subintervals

$$P_{V,1}, \dots, P_{V,\ell_V}$$

defined by V, so that (by the same argument that lead to (4.14))

$$\sup_{t \in P_r} (V,t) \geqslant \eta^{\dim V}$$

respectively

$$\sup_{t \in P_r} (V_r,t) \geqslant \eta^{\dim V_r}. \tag{4.21}$$

Iterating the construction: As hinted at before, we start the iterative construction with $I = [0,T]$, $\mathscr{F} = \{\}$ and $k = 0$. By (4.18) the inductive hypothesis is satisfied, and the inductive step above defines intervals

$$P_1, \dots, P_r$$

and subspaces
$$V_1, \ldots, V_r.$$

On each of the intervals P_{i_1} for $i_1 = 1, \ldots, r$, the partial flag $\mathscr{F}_{i_1} = \{V_{i_1}\}$ satisfies the inductive hypothesis so that the inductive step can be repeated, giving rise to intervals
$$P_{i_1,i_2} \subseteq P_{i_1}$$
and partial flags
$$\mathscr{F}_{i_1,i_2} = \mathscr{F}_{i_1} \cup \{V_{i_1,i_2}\}$$
for a compatible subspace V_{i_2}. In general, let us write
$$\bar{\imath} = (i_1, \ldots, i_k)$$
for the multi-index arising,
$$P_{\bar{\imath}} = P_{i_1,\ldots,i_k}$$
for the intervals arising, and
$$\mathscr{F}_{\bar{\imath}} = \mathscr{F}_{i_1,\ldots,i_k}$$
for the flags or partial flags arising. The construction stops when, for a given interval $P_{\bar{\imath}}$ and partial or complete flag $\mathscr{F}_{\bar{\imath}}$ there is no compatible rational subspace V for which
$$\{t \in P_{\bar{\imath}} \mid (V, t) < \eta^{\dim V}\}$$
is non-empty, and certainly stops if $\mathscr{F}_{\bar{\imath}}$ is a (complete) flag. This may be thought of as a finite graded tree labeled by the intervals and the flags or partial flags, as illustrated in Figure 4.4.

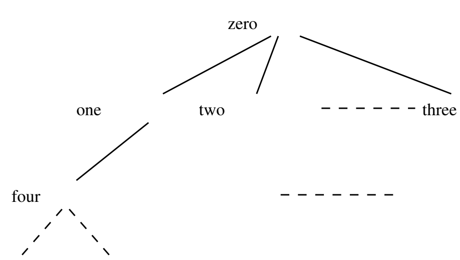

Fig. 4.4 Inductive construction of the intervals and flags.

Definition of Bad subsets: For any $(P_{\bar{\imath}}, \mathscr{F}_{\bar{\imath}})$ as constructed above, we define the following bad subset
$$(\bar{\imath}, \varepsilon) = \{t \in P_{\bar{\imath}} \mid (V_{\bar{\imath}}, t) \leqslant \varepsilon\eta^{\dim V}\}.$$

Taking the union we define

$$(\varepsilon) = \bigcup_{\substack{\bar\imath=(i_1,\dots,i_k),\\ k\geqslant 1}} (\bar\imath,\varepsilon)$$

Estimate for Bad subset: Applying Lemma 4.6 to the interval $P_{\bar\imath}$ and the polynomial $(V_{\bar\imath},t)^2$ (using (4.21), and the definition of $(\bar\imath,\varepsilon)$), we get

$$|(\bar\imath,\varepsilon)| \ll_D \left(\frac{\varepsilon}{\eta}\right)^{1/D} |P_{\bar\imath}|\,. \tag{4.22}$$

We now have to induct backwards to obtain the desired estimate for (ε). In fact we claim that

$$\left|\bigcup_{(j_1,\dots,j_s)} ((\bar\imath,j_1,\dots,j_s),\varepsilon)\right| \ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{1/D} |P_{\bar\imath}|\,. \tag{4.23}$$

If $\{P_{\bar\imath},\mathscr{F}_{\bar\imath}\}$ is a bottom leaf of the tree in Figure 4.4, then this is the same bound as (4.22). If, on the other hand, it is not then we may assume that (4.23) already holds for $(\bar\imath,j_1)$ for all $j_1=1,2,\dots$. Therefore

$$\begin{aligned}\left|\bigcup_{(j_1,\dots,j_s)} ((\bar\imath,j_1,\dots,j_s)\varepsilon)\right| &\leqslant |(\bar\imath,\varepsilon)| + \sum_{j_1}\left|\bigcup_{(j_2,\dots,j_s)} ((\bar\imath,j_1,\dots,j_s)\varepsilon)\right|\\ &\ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{1/D} |P_{\bar\imath}| + \left(\frac{\varepsilon}{\eta}\right)^{1/D} \sum_{j_1} |P_{\bar\imath,j_1}|\end{aligned}$$

by (4.22) for $(\bar\imath,\varepsilon)$ and the inductive hypothesis. Since the intervals

$$P_{\bar\imath,1},\dots,P_{\bar\imath,m} \subseteq P_{\bar\imath}$$

are nearly disjoint we also have

$$\sum_{j_1} |P_{\bar\imath,j_1}| \ll |P_{\bar\imath}|,$$

which concludes the inductive step. For $\bar\imath=\varnothing$ (the root at the top of the graded tree) this shows

$$|(\varepsilon)| \ll_{d,D} \left(\frac{\varepsilon}{\eta}\right)^{1/D} T. \tag{4.24}$$

Conclusion of the argument: It remains to show that

$$\{t\in[0,T] \mid \Gamma p(t) \notin d(\varepsilon)\} \subseteq (\varepsilon), \tag{4.25}$$

since (4.24) then proves the theorem. Suppose therefore that

$$\Gamma p(t) \notin d(\varepsilon),$$

or equivalently that there exists some vector $w \in \mathbb{Z}^d \smallsetminus \{0\}$ with $\|wp(t)\| < \varepsilon$. Since $\varepsilon \leqslant \eta$ we have $t \in P_{W,j}$ for $W = \mathbb{R}w$ and some j. Hence t lies in P_{i_1} for some i_1. If $t \in (i_1, \varepsilon) \subseteq (\varepsilon)$ then we have shown (4.25) for this value of t. So we may assume that $t \notin (i_1, \varepsilon)$. For the sake of the induction to come we continue the argument in greater generality.

Suppose we have shown (or rather, reduced the problem to the case) $t \in P_{\bar{\imath}}$ but $\varepsilon\eta^{\dim V - 1} < (V, t) \leqslant \eta^{\dim V}$ for all $V \in \mathscr{F}_{\bar{\imath}}$. Write

$$\mathscr{F}_{\bar{\imath}} = \{V_1 \subsetneq V_2 \subsetneq \cdots \subsetneq V_k\}$$

and assume that $a \in \{1, \dots, k\}$ is maximal with respect to the property

$$W = \mathbb{R}w \not\subseteq V_a.$$

This implies that

$$(V_a + W, t) \leqslant \eta^{\dim V_a}\varepsilon \leqslant \eta^{\dim V_a}$$

and so $V_a + W \notin \mathscr{F}_{\bar{\imath}}$ and $V_a + W$ is compatible with $\mathscr{F}_{\bar{\imath}}$ (since it contains V_a and is contained in V_{a+1}). In other words, $\mathscr{F}_{\bar{\imath}}$ is not a complete flag and t belongs to one of the intervals defined by $V_a + W$, so that $t \in P_{(\bar{\imath}_1, i_{k+1})}$ for some i_{k+1}. If $t \in (\bar{\imath}, i_{k+1}, \varepsilon)$ then we are again done. That is, we have the same situation as before and can repeat the argument.

The iterative argument above will only stop when t lies in (ε). Since every time the argument repeats we know that we only can have had a partial flag at the last stage, it can take at most d iterations to reach the conclusion. $\square$

Corollary 4.10 (Non-escape of mass for d). *If $x \in d$ and*

$$\{u_t \mid t \in \mathbb{R}\} < \mathrm{SL}_d(\mathbb{R})$$

is a one-parameter unipotent subgroup, then every weak-limit of the collection of measures*

$$\left\{ \tfrac{1}{T} \int_0^T (u_t)_* \delta_x \, \mathrm{d}t \mid T > 0 \right\}$$

is a probability measure on d.

Proof. Let $\Lambda_x < \mathbb{R}^d$ be the lattice corresponding to $x \in d$, and define

$$\eta = \min\left\{ \sqrt[k]{\alpha_k(\Lambda_x)} \mid 1 \leqslant k \leqslant d \right\}.$$

Fix an arbitrary $\varepsilon \in (0, \eta]$ and choose some $f \in C_c(d)$ with

$$\mathbb{1}_{d(\varepsilon)} \leqslant f \leqslant \mathbb{1} = \mathbb{1}_d.$$

By Theorem 4.9 we have

$$1 - c\left(\frac{\varepsilon}{\eta}\right)^{1/D} \leqslant \frac{1}{T}\int_0^T f\left(u_t \centerdot x\right)\,\mathrm{d}t \leqslant 1$$

for some constant $c = c_{d,D}$. Now choose a weak*-convergent subsequence of the measures

$$\frac{1}{T}\int_0^T \left(u_t\right)_* \delta_x \,\mathrm{d}t$$

to obtain the bound

$$1 - c\left(\frac{\varepsilon}{\eta}\right)^{1/D} \leqslant \int_d f\,\mathrm{d}\mu$$

for the limit measure μ. Since $f \leqslant 1$ this shows that

$$\mu\left(d\right) \geqslant 1 - c\left(\frac{\varepsilon}{\eta}\right)^{1/D}.$$

As $\varepsilon \in (0, \eta]$ was arbitrary, the corollary follows. □

4.4 Closed Orbits (often) Have Finite Volume

In this section we return to the discussion of orbits $H \centerdot x$ for a connected subgroup $H \leqslant \mathrm{SL}_d(\mathbb{R})$ and point $x \in d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$. Recall that H is called *semi-simple* if its Lie algebra is semi-simple, and that this implies that H is an almost direct product of normal simple subgroups (which may be compact or non-compact; see Section 2.1). We say that the subgroup H is *unipotent* if H can be conjugated into the strict upper-triangular subgroup

$$N = \left\{\begin{pmatrix} 1 & * & \cdots & * \\ & 1 & * & * \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}\right\},$$

which implies that its Lie algebra is nilpotent. For these subgroups we can give another connection between the property of having a closed orbit and the property of having an orbit of finite volume. In fact we will prove a partial converse to Proposition 1.12.

ELon suggests: the second proof probably works for all groups generated by unipotents

Theorem 4.11 (Borel Harish-Chandra theorem, Part I). *Let $x \in d$, and let $H < \mathrm{SL}_d(\mathbb{R})$ be a connected subgroup which is semi-simple or unipo-*

tent. If the orbit $H\bullet x$ is closed, then it has finite volume†*. In the case H is unipotent, the orbit is compact.*

We refer to Exercise 4.4.2 for an immediate corollary (which is the standard way of phrasing the Borel Harish-Chandra theorem) and to Section 7.4 for the general case of the theorem (which requires a few more definitions from the theory of algebraic groups).

Proof of Theorem 4.11 for semi-simple subgroups. Let

$$H = H_1 \cdots H_\ell H_{\mathrm{compact}}$$

be the almost direct product of simple non-compact normal factors $H_1, \dots, H_\ell$ and a compact normal semi-simple subgroup H_{compact} as in Section 2.1. Now choose, for each H_i, a non-trivial unipotent one-parameter subgroup

$$U_i = \{u_i(t) \mid t \in \mathbb{R}\}$$

and define the diagonally embedded unipotent subgroup

$$U = \{u_1(t)u_2(t)\cdots u_n(t) \mid t \in \mathbb{R}\}.$$

By Theorem 2.11 this subgroup $U \leqslant H$ satisfies the following form of the Mautner phenomenon: If H acts unitarily on a Hilbert space‡ $\mathscr{H}$ and a vector is fixed by U, then the same vector is fixed by $H_1 \cdots H_\ell$.

Now choose a compact set $K \subseteq H\bullet x$ of positive volume with respect to the H-invariant Haar measure $m_{H\bullet x}$ on the orbit $H\bullet x \subseteq d$ (as in Proposition 1.9 applied to $\mathrm{Stab}_H(x)\backslash H$). Since $K \subseteq d$ is compact, we can find some $\eta \in (0,1]$ such that

$$\alpha_k(\varLambda_x) \geqslant \eta^k$$

for $k = 1, \dots, d$ and any $x \in K$. Now apply Theorem 4.9 to find some $\varepsilon \in (0,\eta]$ with

$$\frac{1}{T}\left|\{t \in [0,T] \mid u_t\bullet x \notin d(\varepsilon)\}\right| < \frac{1}{2} \tag{4.26}$$

for all $T > 0$. Since $H_{\mathrm{compact}} \subseteq H$ is compact and $H\bullet x$ is closed, we have that (see Proposition 1.13)

$$d(\varepsilon)H_{\mathrm{compact}} \cap H\bullet x$$

is a compact subset of the orbit $H\bullet x$, and so

$$f = \mathbb{1}_{d(\varepsilon)H_{\mathrm{compact}}} \in L^2(H\bullet x, m_{H\bullet x})$$

is square-integrable with respect to $m_{H\bullet x}$. We define

† That is, the orbit supports a finite H-invariant measure.

‡ In this instance, the Hilbert space will be $L^2(H\bullet x, m_{H\bullet x})$.

$$\underline{f}(y) = \liminf_{n\to\infty} \frac{1}{n} \int_0^n f(u_t\boldsymbol{\cdot} y)\,\mathrm{d}t.$$

Notice that

$$\begin{aligned}\left\|\frac{1}{n}\int_0^n f(u_t\boldsymbol{\cdot} x)\,\mathrm{d}t\right\|^2_{L^2(m_{H\boldsymbol{\cdot} x})} &= \frac{1}{n^2}\int_0^n\int_0^n \underbrace{\int_{H\boldsymbol{\cdot} x} f(u_{t_1}\boldsymbol{\cdot} y) f(u_{t_2}\boldsymbol{\cdot} y)\,\mathrm{d}m_{H\boldsymbol{\cdot} x}(y)}_{\leqslant \|f\|^2_{L^2(m_{H\boldsymbol{\cdot} x})}}\\ &\leqslant \|f\|^2_{L^2(m_{H\boldsymbol{\cdot} x})}.\end{aligned}$$

Hence, by Fatou's lemma, we get

$$\begin{aligned}\|\underline{f}\|^2_{L^2(m_{H\boldsymbol{\cdot} x})} &= \int_{H\boldsymbol{\cdot} x} \liminf_{n\to\infty}\left(\frac{1}{n}\int_0^n f(u_t\boldsymbol{\cdot} y)\,\mathrm{d}t\right)^2 \mathrm{d}m_{H\boldsymbol{\cdot} x}(y)\\ &\leqslant \liminf_{n\to\infty}\int_{H\boldsymbol{\cdot} x}\left(\frac{1}{n}\int_0^n f(u_t\boldsymbol{\cdot} y)\,\mathrm{d}t\right)^2 \mathrm{d}m_{H\boldsymbol{\cdot} x}(y)\\ &\leqslant \|f\|^2_{L^2(m_{H\boldsymbol{\cdot} x})} < \infty,\end{aligned}$$

or equivalently $\underline{f} \in L^2(m_{H\boldsymbol{\cdot} x})$.

We can now finish the proof quite quickly. Since $f \in L^2(m_{H\boldsymbol{\cdot} x})$ is u_t-invariant for all $t \in \mathbb{R}$ by construction, it is also $H_1 \cdots H_\ell$-invariant by the Mautner phenomenon (Theorem 2.11). Furthermore, $f = \mathbb{1}_{d(\varepsilon)H_{\mathrm{compact}}}$ is invariant under H_{compact} by definition. Since u_t commutes with H_{compact}, it follows that $\underline{f}$ is also invariant under H_{compact}. Since $H = H_1 \cdots H_\ell H_{\mathrm{compact}}$ and $\underline{f} \in L^2(m_{H\boldsymbol{\cdot} x})$ we see that $\underline{f} \equiv c$ is equal $m_{H\boldsymbol{\cdot} x}$-almost everywhere to some constant c. By definition and (4.26) we have $c \geqslant \frac{1}{2}$ and so

$$c^2 m_{H\boldsymbol{\cdot} x}(H\boldsymbol{\cdot} x) = \|\underline{f}\|^2_{L^2(m_{H\boldsymbol{\cdot} x})} < \infty$$

implies that $H\boldsymbol{\cdot} x$ has finite volume. □

Proof of Corollary 4.11 for unipotent subgroups. In the proof of Corollary 4.11 for the semi-simple case it was convenient that we could find one one-parameter unipotent subgroup that satisfied the hypothesis of the Mautner phenomenon for 'most' of H. In the unipotent case we have instead to use finitely many one-parameter unipotent subgroups $U_j = \{u_j(t) \mid t \in \mathbb{R}\}$ for $j = 1, \dots n$ that together generate H.

Let $K \subseteq H\boldsymbol{\cdot} x$ be a compact set. Then, finding first $\eta > 0$ and then $\varepsilon \in (0, \eta]$ as above, there exists a compact subset $L \subseteq H\boldsymbol{\cdot} x$ (where $L = d(\varepsilon) \cap H\boldsymbol{\cdot} x$, relying on the assumption that $H\boldsymbol{\cdot} x$ is closed) such that

$$\frac{1}{T}\left|\{t \in [0, T] \mid u_1(t)\boldsymbol{\cdot} y \notin L\}\right| < \frac{1}{2} \tag{4.27}$$

for all $y \in K$. Now let $f = \mathbb{1}_L \in L^2(m_{H\boldsymbol{\cdot} x})$ and

$$f_1(y) = \underline{f}(y) = \liminf_{n\to\infty} \frac{1}{n}\int_0^n f(u_1(t)\boldsymbol{\cdot} y)\,\mathrm{d}t$$

so that $f_1 \in L^2(m_{H\boldsymbol{\cdot} x})$, f_1 is U_1-invariant, and $f_1(y) \geqslant \frac{1}{2}$ for all $y \in K$.

Suppose now that for $j \leqslant n$ we have already shown that for any compact set $K \subseteq H\boldsymbol{\cdot} x$ there exists some $f_j \in L^2(m_{H\boldsymbol{\cdot} x})$ which is U_1-invariant, U_2-invariant, and so on up to U_j-invariant, and satisfies $f_j(y) \geqslant (\frac{1}{2})^j$ for all $y \in K$. If $j = n$ then the function is H-invariant and the theorem follows as before.

So suppose that $j < n$ and let $K \subseteq H\boldsymbol{\cdot} x$ be a compact subset. Now choose $L \subseteq H\boldsymbol{\cdot} x$ as in (4.27) but for $u_j(t)$ instead of $u_1(t)$. Next apply the inductive hypothesis to L to find a function $f_j \in L^2(m_{H\boldsymbol{\cdot} x})$ which is invariant under $U_1, U_2, \dots, U_j$ and satisfies $f_j(y) \geqslant (\frac{1}{2})^j$ for all $y \in L$. We define

$$f_{j+1}(y) = \underline{f_j}(y) = \liminf_{n\to\infty} \frac{1}{n}\int_0^n f_j(u_t\boldsymbol{\cdot} y)\,\mathrm{d}t.$$

By construction of f_j, L, and f_{j+1} we know that $f_{j+1} \in L^2(m_{H\boldsymbol{\cdot} x})$, that f_{j+1} is U_{j+1}-invariant, and that $f_{j+1}(y) \geqslant (\frac{1}{2})^{j+1}$ for all $y \in K$. However, at first sight it may not be clear why f_{j+1} is still invariant under U_i for $i = 1, \dots, j$ (since U_i may not commute with U_{j+1}). Here the Mautner phenomenon (Theorem 2.11) comes to the rescue. In fact, by Theorem 2.11, f_j is actually invariant under a normal subgroup $N \lhd H$ containing $U_1, \dots, U_j$. Therefore

$$u_{j+1}(t)u_i(s) = n_{s,t}u_{j+1}(t)$$

for all $i = 1, \dots, j$, and $s, t \in \mathbb{R}$, and some $n_{s,t} \in N$. This shows that

$$f_j(u_{j+1}(t)u_i(s)\boldsymbol{\cdot} y) = f_j(n_{s,t}u_{j+1}(t)\boldsymbol{\cdot} y) = f_j(u_{j+1}(t)\boldsymbol{\cdot} y)$$

for $m_{H\boldsymbol{\cdot} x}$-almost every y. Integrating over $t \in [0, n]$ and taking the limit infimum as in the definition of f_{j+1}, we get

$$f_{j+1}(u_i(s)\boldsymbol{\cdot} y) = f_{j+1}(y).$$

This concludes the induction and so also the proof of the first statement Corollary 4.11 for unipotent subgroups.

It remains to show that $H\boldsymbol{\cdot} x$ is compact. Note that any unipotent subgroup of $\mathrm{SL}_d(\mathbb{R})$ has a nilpotent Lie algebra and so can be conjugated into the upper triangular unipotent subgroup by Engel's theorem ([**?**, Thm. 1.35]). On the upper triangular unipotent subgroup the logarithm map is a polynomial with a polynomial inverse, and H consists of all points that image in the Lie algebra of H. Hence we see that a unipotent connected subgroup H consists of the $\mathbb{R}$-points $H = \mathbb{H}(\mathbb{R})$ of an algebraic subgroup $\mathbb{H}$ over $\mathbb{R}$. Now assume by conjugating H by $g \in G$, where $x = \mathrm{SL}_d(\mathbb{Z})g$, that $x = \mathrm{SL}_d(\mathbb{Z})$. Then the Borel density theorem (Theorem 3.30) implies that $H \cap \mathrm{SL}_d(\mathbb{Z})$ is Zariski dense in $\mathbb{H}$, which in turn implies that $\mathbb{H}$ is an algebraic group over $\mathbb{Q}$. Hence the Lie algebra of H is a rational subspace of $\mathfrak{sl}_d(\mathbb{R})$ and by Theorem 3.9 we see that $H \cdot \mathrm{SL}_d(\mathbb{Z})$ is compact. $\square$

Exercises for Section 4.4

Exercise 4.4.1. Let Q be a real non-degenerate quadratic form of signature (p,q) in $d \geqslant 3$ variables with $p \geqslant q \geqslant 1$. Suppose that the orbit $\mathrm{SL}_d(\mathbb{Z})\,\mathrm{SO}(Q)(\mathbb{R})$ is closed. Show that a multiple of Q has integer coefficients.

Exercise 4.4.2. (20) Let $\mathbb{G} < \mathrm{SL}_d$ be a semi-simple or unipotent algebraic group defined over $\mathbb{Q}$. Show that $\mathbb{G}(\mathbb{Z}) = \mathbb{G}(\mathbb{R}) \cap \mathrm{SL}_d(\mathbb{Z})$ is a lattice in $\mathbb{G}(\mathbb{R})$.

Notes to Chapter 4

(17)(Page 133) This result, or rather its higher-dimensional counterpart in Section 4.3, has a long history; see Margulis [?], [?]; Dani [?], [?]; Kleinbock and Margulis [?]; Kleinbock [?].
(18)(Page 139) This is a simple special case of the Besicovitch covering lemma (see [?]).
(19)(Page 145) As mentioned before, this result has a long history; see Margulis [?], [?]; Dani [?], [?]; Kleinbock and Margulis [?]; Kleinbock [?].
(20)(Page 154) This is a special case of the Borel Harish-Chandra theorem [?].

Chapter 5
Action of Horospherical Subgroups

The inheritance property of ergodicity in the Mautner phenomenon (Theorem 2.6) established in Chapter 2 already gives the equidistribution of many orbits.

Indeed, if a simple Lie group G acts ergodically on (X, μ) and

$$\{g_t \mid t \in \mathbb{R}\} \subseteq G$$

is an unbounded one-parameter subgroup, then

$$\frac{1}{T}\int_0^T f(g_t \cdot x)\,\mathrm{d}t \to \int_X f\,\mathrm{d}\mu$$

for μ-almost every $x \in X$, for any $f \in C_c(X)$. This is a straightforward application of the pointwise ergodic theorem (see [**?**, Cor. 8.15 and Sec. 4.4.2] and Lemma 6.10). A point $x \in X$ with this property is called *generic* for μ and the one-parameter subgroup $\{g_t : t \in \mathbb{R}\}$.

In this short chapter we start the discussion of unipotent dynamics by considering the case of horospherical actions. For those actions we will show unique ergodicity, and sometimes 'almost unique ergodicity', and we will understand the distribution of orbits of any given point.

5.1 Dynamics on Hyperbolic Surfaces

Let us start by discussing briefly the case of the geodesic flow and the horocycle flow on quotients of $\mathrm{SL}_2(\mathbb{R})$ as introduced in Section 1.2

We note first that for the geodesic flow it is not possible to make a more general statement about the equidistribution of orbits by relaxing the requirement that the point be μ-typical. Indeed, if $g_t = a_t$ is diagonalizable, then the flow is partially hyperbolic and as a result X contains many irregular orbits. As this result can be considered of negative type we will not prove it here, but see Example 5.1 resp. [**?**, Sec. 9.7.2] for a more detailed discussion of the case of the geodesic flow on the modular surface.

Example 5.1. [(21)]For a compact quotient $X = \Gamma\backslash \mathrm{SL}_2(\mathbb{R})$, the action of the one-parameter subgroup

$$\left\{a_t = \begin{pmatrix} \mathrm{e}^{-t/2} & \\ & \mathrm{e}^{t/2} \end{pmatrix} \mid t \in \mathbb{R}\right\}$$

has many orbits that stay on one side of the dotted line in Figure 5.1.

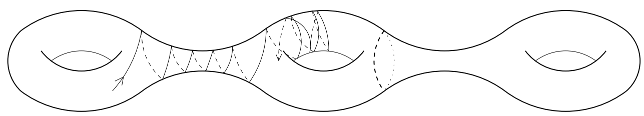

Fig. 5.1 There are many orbits under the action of A that stay on one side of the dotted line furthest to the right.

We also refer to Exercises 5.1.1–5.1.4 for the behavior of the geodesic flow and higher dimensional analogues.

This is in stark contrast to the behavior of orbits of the horocycle subgroup

$$\left\{u_s = \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R}\right\}$$

in the compact surface X: the orbit of every point under this group action visits the right-hand side at some point (much more is true, as we will show below).

Hedlund [**?**] showed in 1936 that the horocycle flow on any compact quotient $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ is minimal (that is, has no non-trivial closed invariant subsets) and that Haar measure is ergodic. This was strengthened by Furstenberg [**?**] in 1972 and by Dani [**?**] in 1978, who showed the following theorems.

Theorem 5.2 (Unique ergodicity of horocycle flow). *If $X = \Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ is compact, then the horocycle flow (i.e. the action of the subgroup)*

$$\left\{u_s = \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R}\right\}$$

is uniquely ergodic.

Theorem 5.3 (Almost unique ergodicity of horocycle flow). *If*

$$X = \mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$$

(or another non-compact quotient of finite volume) then a probability measure m *on* X *that is invariant and ergodic for the action of*

$$\left\{u_s = \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R}\right\}$$

is either

- *the Haar measure* m_X *on* X *inherited from the Haar measure on* $\mathrm{SL}_2(\mathbb{R})$*; or*
- *a one-dimensional Lebesgue measure supported on a periodic orbit of the action.*

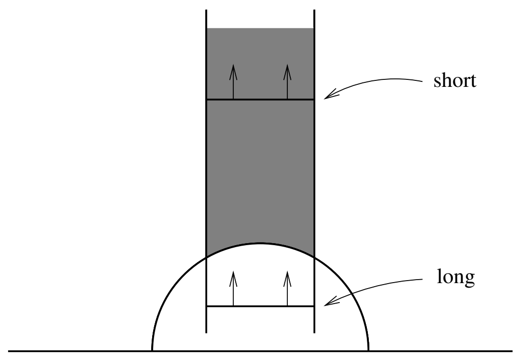

Fig. 5.2 In the upper half-plane model of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$, the speed of a periodic horocycle orbit increases with the height, so the two different periodic orbits shown are of different lengths. The longer periodic orbit could also be drawn in the fundamental domain, but it would look very complicated.

We will prove Theorem 5.2 in Section 5.2 and give an outline of the proof of Theorem 5.3 in Section 5.3.

Exercises for Section 5.1

Exercise 5.1.1. (Anosov shadowing for $\mathrm{SL}_2(\mathbb{R})$) Let $X = \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ be the quotient of $\mathrm{SL}_2(\mathbb{R})$ by a discrete subgroup $\Gamma < \mathrm{SL}_2(\mathbb{R})$.
(a) Let $x \in X$, $T > 0$, $\varepsilon > 0$ and $y \in X$ be chosen with $\mathsf{d}(a_T{\bullet}x, y) < \varepsilon$. Then there exists a point $z \in X$ with $\mathsf{d}(x, z) \ll \mathrm{e}^{-T}\varepsilon$ (and so $\mathsf{d}(a_t{\bullet}x, a_t{\bullet}z) \ll \varepsilon$ for $t \in [0, T]$) and $\mathsf{d}(a_t{\bullet}y, a_{T+t}{\bullet}z) \ll \varepsilon$ for all $t \geqslant 0$. Also show that there exists some δ with $|\delta| \ll \varepsilon$ such that $\mathsf{d}(a_{t+\delta}{\bullet}y, a_{T+t}{\bullet}z) \ll \mathrm{e}^{-t}$ for all $t \geqslant 0$.
(b) Assume now that X is compact (for example, as in Figure 5.1) and use (a) to construct non-periodic orbits as in Example 5.1.

Exercise 5.1.2. (Anosov closing for $\mathrm{SL}_2(\mathbb{R})$) Let $X = \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ be as in Exercise 5.1.1. Let $x \in X$ and $T \geqslant 1$ be chosen so that $\mathsf{d}(a_T\bullet x, x) \leqslant \varepsilon < 1$. Show that there exists a point $z \in X$ which is periodic with period T_z satisfying

$$|T_z - T| \ll \varepsilon$$

and

$$\mathsf{d}(a_t\bullet x, a_t\bullet z) \ll \varepsilon$$

for all $t \in [0, T]$.

Exercise 5.1.3. (Anosov shadowing for G) Let G be a connected Lie group, let $\Gamma < G$ be a discrete subgroup, let $X = \Gamma\backslash G$, and let $a \in G$ be such that Ad_a is diagonalizable with positive eigenvalues.
(a) Let $x \in X$, $N > 1$, $\varepsilon > 0$ and $y \in X$ be such that $\mathsf{d}(a^N\bullet x, y) < \varepsilon$. Then there exists a point $z \in X$, some $\lambda < 1$ (independent of x, y and Γ) with

$$\mathsf{d}(a^n\bullet x, a^n\bullet z) \ll \lambda^{N-n}\varepsilon$$

for $n = 0, \ldots, N$ and

$$\mathsf{d}(a^{N+n}\bullet z, a^n\bullet y) \ll \varepsilon$$

for all $n \geqslant 0$.
(b) Assume that X has finite volume and a acts mixingly on X with respect to m_X. Construct non-periodic irregular orbits by iterating (a).

Exercise 5.1.4. (Anosov closing for $X = \Gamma\backslash\mathrm{SL}_d(\mathbb{R})$) We let $X = \Gamma\backslash\mathrm{SL}_d(\mathbb{R})$ be any quotient by a discrete subgroup $\Gamma < G = \mathrm{SL}_d(\mathbb{R})$, and let A be the subgroup of G of positive diagonal matrices. Let $a \in A$ be a non-trivial element.
(a) Suppose that $x \in X$ and $N \geqslant 1$ are such that $\mathsf{d}(a^N, I) \geqslant 1$ but $\mathsf{d}(a^N\bullet x, x) \leqslant \varepsilon < 1$. Assume that ε is sufficiently small and that N is sufficiently large. Show that there exists some $z \in X$ and some $a' \in \mathrm{SL}_d(\mathbb{R})$ with $aa' = a'a$, $\mathsf{d}(a^N, a') \ll \varepsilon$, $a'\bullet z = z$ and $\mathsf{d}(a^n\bullet x, a^n\bullet z) \ll \varepsilon$ for $n = 0, \ldots, N$.
(b) Suppose that a is generic (that is, no two eigenvalues are the same) and X is compact. Show that z as in (a) is a periodic point for A.
(c) Suppose $d = 3$ and a is generic and a does not have 1 as an eigenvalue, and $X = 3 = \mathrm{SL}_3(\mathbb{Z})\backslash\mathrm{SL}_3(\mathbb{R})$. Show again that the point z as in (a) is periodic for A.
(d) In the setting of (b) and of (c), show that periodic A-orbits are dense in X.
(e) Generalize the statement in (b) to any semi-simple group[†].

realized how to do the Anosov closing for $SL_d(Z)$ $SL_d(R)$ so that one does get compact A-orbits instead of just closed ones for every d. After taking with Uri, he told me that the solution is containing ideas that are also in his paper with Barak $http://arxiv.org/abs/$1207.6343 I will at some point make the appropriate changes to the exercise and the hints in the back. Added reference to the hint for 5.1.4.(c).

5.2 Horospherical Actions on Compact Quotients

As we will show now the unique ergodicity of the horocycle flow on compact quotients of $\mathrm{SL}_2(\mathbb{R})$ as in Theorem 5.2 holds for other Lie groups and their horospherical subgroups as well.

Suppose that G is a connected Lie group and let $a \in G$ be an $\mathbb{R}$-diagonalizable element that acts as a mixing transformation[‡] on all the quo-

[†] In that sense Poincaré recurrence can be used to construct anisotropic tori (see Section 7.3).

[‡] Unless a specific other probability measure is identified, a property of a transformation on a homogeneous space like ergodicity, mixing, and so on, is meant with respect to the measure induced by the Haar measure on G.

tients of G appearing below. Let

$$G_a^- = \{g \in G \mid a^n g a^{-n} \to I \text{ as } n \to \infty\}$$

be the stable horospherical subgroup of a. The 'general method' discussed below gives a way to classify the G_a^--invariant ergodic probability measures on X.

Theorem 5.4 (Unique ergodicity of horospherical actions[(22)]). *Let G be a linear Lie group, $\Gamma < G$ be a uniform lattice, and let $a \in G$ be $\mathbb{R}$-diagonalizable. Suppose a acts mixingly on $X = \Gamma\backslash G$. Then the action of G_a^- is uniquely ergodic.*

Proof of Theorem 5.4 (and hence of 5.2). Let us assume compatibility of the Haar measures in the sense that $m_X(\pi(B)) = m_G(B)$ for any injective Borel subset $B \subseteq G$, and that $m_X(X) = 1$.

Since a is diagonalizable and G is linear, the subgroups G_a^- and

$$P_a = \{g \in G \mid a^n g a^{-n} \text{ stays bounded as } n \to -\infty\}$$

can easily be defined in terms of the vanishing of certain matrix entries, and so are closed subgroups. Together they define a local coordinate system, in the sense that $P_a G_a^-$ contains an open neighborhood of the identity†, and the implied representation of elements of G in that neighborhood is unique. In fact, if $u_1 p_1 = u_2 p_2$ with $u_1, u_2 \in G_a^-$ and $p_1, p_2 \in P_a$ then

$$g = u_2^{-1} u_1 = p_2 p_1^{-1}$$

has $a^n g a^{-n} \to I$ as $n \to \infty$ and stays bounded as $n \to -\infty$, which together show‡ that $g = I$. Moreover, the Haar measure of G restricts to the product of a Haar measure on G_a^- and a left Haar measure on P_a (see Lemma 1.22).

We let $B_0 \subseteq G_a^-$ be a neighborhood of the identity with compact closure such that $m_{G_a^-}(\partial B_0) = 0$ and define $B_n = a^{-n} B_0 a^n$. We claim that

$$\frac{1}{m_{G_a^-}(B_n)} \int_{B_n} f(u \cdot x) \, \mathrm{d}m_{G_a^-}(u) \longrightarrow \int_X f \, \mathrm{d}m_X \tag{5.1}$$

for any $f \in C(X)$ and any $x \in X$.

Assuming this for now, it follows that $\mu = m_X$ is the only G_a^--invariant probability measure. Indeed if μ is another such measure then

† This can be quickly checked using the Lie algebras of G_a^- (and of P_a), which are simply the sum of the eigenspaces of Ad_a for all eigenvalues of absolute value less than one (respectively greater or equal than one).

‡ This is a consequence of considering the eigenvalue decomposition of the matrix $g - I \in \mathrm{Mat}_d(\mathbb{R})$ for the linear map

$$\mathrm{Mat}_d(\mathbb{R}) \ni v \longmapsto a v a^{-1}.$$

$$\int_X f\,\mathrm{d}\mu = \int_X \frac{1}{m_{G_a^-}(B_n)} \int_{B_n} f(u\boldsymbol{\cdot}x)\,\mathrm{d}m_{G_a^-}(u)\,\mathrm{d}\mu(x) \longrightarrow \int_X f\,\mathrm{d}m_X$$

by dominated convergence. As this would hold for any $f \in C(X)$ we deduce that $\mu = m_X$, as claimed.

Now fix a point $x \in X = \Gamma\backslash G$ and a function $f \in C(X)$. By compactness f is uniformly continuous, so given $\varepsilon > 0$ there is a $\delta > 0$ for which

$$\mathsf{d}(h, I) < \delta \implies |f(h\boldsymbol{\cdot}y) - f(y)| < \varepsilon \tag{5.2}$$

where d is a left-invariant metric on G (giving rise to the metric on X). Now we can choose a compact neighborhood $V \subseteq P_a$ of the identity whose boundary has measure zero with

$$\mathsf{d}(a^{-n}ha^n, I) < \delta$$

for $h \in V$ and $n \geqslant 0$. Then

$$\frac{1}{m_{G_a^-}(B_n)} \int_{B_n} f(u\boldsymbol{\cdot}x)\,\mathrm{d}m_{G_a^-}(u)$$

is within ε of

$$\frac{1}{m_{G_a^-}(B_n)m_{P_a}(a^{-n}Va^n)} \int_{B_n} \int_{a^{-n}Va^n} f(hu\boldsymbol{\cdot}x)\,\mathrm{d}m_{P_a}(h)\,\mathrm{d}m_{G_a^-}(u)$$

because of (5.2). Using $B_n = a^{-n}B_0a^n$, the latter may in turn be written as

$$\frac{1}{m_G(VB_0)} \int_{VB_0} f(a^{-n}ga^n\boldsymbol{\cdot}x)\,\mathrm{d}m_G(g), \tag{5.3}$$

since m_G is locally the product of m_{P_a} and $m_{G_a^-}$. Now notice that

$$G_a^- \ni u \mapsto u\boldsymbol{\cdot}x$$

is injective for all $x \in X$, for otherwise the injectivity radius at $a^n\boldsymbol{\cdot}x$ would shrink to zero, contradicting the compactness of X [MLE] (see Proposition 1.11 and Lemma 10.8). By a simple compactness argument, we may assume that the above δ is small enough to ensure that the map

[MLE] Rene suggested the forward ref — not sure, but it also doesn't make sense to move the lemma forward as it has a completely different setup than what we do now — maybe this is a compromise.

$$VB_0 \ni g \mapsto g\boldsymbol{\cdot}x$$

is injective for all $x \in X$. Thus (5.3) can also be written as

$$\frac{1}{m_G(VB_0)} \int_X f(a^{-n}y)\mathbb{1}_{VB_0a^n\boldsymbol{\cdot}x}(y)\,\mathrm{d}m_X(y). \tag{5.4}$$

In the sequence (or in any of its subsequences) $(a^n \bullet x)_{n\geqslant 1}$ we can find (by compactness) a subsequence $(a^{n_k} \bullet x)_{k\geqslant 1}$ converging to some $z \in X$. Since†

$$\|\mathbb{1}_{VB_0 a^{n_k}\bullet x} - \mathbb{1}_{VB_0\bullet z}\|_2 \longrightarrow 0$$

by dominated convergence as $k \to \infty$, we see that the expression (5.4) converges to

$$\frac{1}{m_G(VB_0)} \int_X f \,\mathrm{d}m_X \int \mathbb{1}_{VB_0\bullet z} \,\mathrm{d}m_X$$

as $n \to \infty$ because a defines a mixing transformation. This proves (5.1) for the given function f up to an error of ε. Since f and $\varepsilon > 0$ were both arbitrary, the theorem follows. □

Notice that once unique ergodicity is proved (by using the Følner sequence $(a^{-n}B_0a^n)$) then the pointwise everywhere convergence of the ergodic averages also follows for other Følner sets (see Exercises 5.2.1–5.2.2).

Exercises for Section 5.2

Exercise 5.2.1. We let $B_n = a^{-n}B_0a^n$ be as in the proof of Theorem 5.4 (with $m_{G_a^-}(\partial B_0) = 0$). Show that this sequence is a Følner sequence in G_a^-, that is a sequence satisfying

$$\frac{m_{G_a^-}(B_n \triangle (KB_n))}{m_{G_a^-}(B_n)} \longrightarrow 0$$

as $n \to \infty$ for every compact subset $K \subseteq G_a^-$.

Exercise 5.2.2. Let a and X be as in Theorem 5.4. Let $F_n \subseteq G_a^-$ be any Følner sequence and show that

$$\frac{1}{m_{G_a^-}(F_n)} \int_{F_n} f(u\bullet x) \,\mathrm{d}m_{G_a^-}(u) \to \int_X f \,\mathrm{d}m_X$$

as $n \to \infty$, for any $f \in C(X)$ and any $x \in X$.

5.3 Almost Unique Ergodicity on Non-Compact Quotients with Finite Volume

We now explain, guided by examples, how the presence of a cusp (that is, the lack of compactness of the quotient) and the presence of horospherical invariant measures other than the Haar measure are related to each other.

† Here we use $m_G(\partial(VB_0)) = 0$ which follows since m_G is the product measure in the local coordinate system $G_a^- P_a$ of G that we use and since $m_{G_a^-}(\partial B_0) = 0$ and $m_{P_a}(\partial V) = 0$.

5.3.1 Horocycle Action on Non-Compact Quotients

Indeed for the horocycle flow a quotient $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ is non-compact with finite volume if and only if Γ is a lattice and contains a unipotent element $\gamma\in\Gamma$. In that case the unipotent γ is conjugated to an element of the horocycle subgroup U, that is there exists some $g\in\mathrm{SL}_2(\mathbb{R})$ with $g\gamma g^{-1}\in U$. This shows that Γg^{-1} is periodic under U and hence there exists an invariant measure other than the Haar measure.

Outline Proof of Theorem 5.3. The argument used for the proof of Theorem 5.4 may also be used for non-compact quotients. Indeed, if μ is an invariant and ergodic probability measure on $X=\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ for the horocycle flow $U=G_a^-$ and $x\in X$ is generic for μ. Then either x is periodic under U and the geodesic orbit $a^n\boldsymbol{\cdot}x$ necessarily diverges into one of the cusps of X as $n\to\infty$ (because a shrinks the length of the periodic orbit and so the injectivity radius at $a^n\boldsymbol{\cdot}x$ goes to zero), or x is not periodic under U and the geodesic orbit $a^n\boldsymbol{\cdot}x$ visits a fixed compact set of X infinitely often. Using this subsequence of times n_j and the corresponding pieces of the horocycle orbit $a^{-n_j}B_0a^{n_j}\boldsymbol{\cdot}x$ for the argument we see by the argument in the proof of Theorem 5.4 that the ergodic average for the horocycle orbit and x converges to the integral of the test function with respect to the Haar measure. As x was chosen to be generic for μ we also have that the averages converge to the integral with respect to μ.

Therefore μ is either the Lebesgue measure on a periodic orbit or the Haar measure on X. (See [**?**, Ch. 11] for more details.) □

We would also like to point out that the same argument can be used to prove the following theorem† of Sarnak [**?**] (see Exercise 5.3.1).

Theorem 5.5 (Equidistribution of long periodic orbits of the horocycle flow). *Let $X=\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ be a non-compact quotient of finite volume. Let $a=a_1$ be the element corresponding to the time-one map for the geodesic flow corresponding the diagonal subgroup $A=\{a_t:t\in\mathbb{R}\}$. Let $x\in X$ be a periodic orbit for the horocycle flow $U=G_a^-$ and let μ be the normalized Lebesgue measure on the one-dimensional orbit xU. Then the periodic orbit measures $(a_{t_n})_*\mu$ diverge to infinity if $t_n\to\infty$ (in which case the periodic orbit $a_{t_n}\boldsymbol{\cdot}(xU)$ becomes shorter and shorter) and equidistribute with respect to the Haar measure if $t_n\to-\infty$ (in which case the periodic orbit $a_{t_n}\boldsymbol{\cdot}(xU)$ become longer and longer).*

† Sarnak also gives an error rate in this equidistribution result — obtaining this (or even any) error estimate requires more sophisticated methods than we will discuss here.

5.3.2 The General Case

By analyzing the proof of Theorem 5.4 more carefully we identify the places where we used that X is compact:

- We used test functions $f \in C(X)$ and that these are uniformly continuous.
- We used that any subsequence of the sequence of points $a^n \bullet x$ converges along some subsequence to some $z \in X$ (equivalently that the injectivity radius at these points is bounded away from zero).

The first point is trivial to fix: we just work with functions $f \in C_c(X)$. These are still uniformly continuous and together are still dense in $L^2_\mu(X)$ for any probability measure on X.

The second point is of course the reason why the argument fails to prove unique ergodicity for non-compact spaces: It is entirely possible for the sequence $a^n \bullet x$ to go to infinity. In fact, for $X = \Gamma \backslash \mathrm{SL}_2(\mathbb{R})$ this happens precisely for points with a periodic orbit for the horocycle flow. Even so, for the horocycle flow we found that for a non-periodic point at least one has some subsequence that stays within a compact subset and that this was sufficient for our purposes. However, for other spaces the divergence properties of a sequence of the form $a^n \bullet x$ are potentially much more complicated, and in particular a clear equivalence to rational properties of the starting point x may not hold. For that reason we are going to invoke the *Margulis–Dani–Kleinbock non-divergence* (Theorem 4.9) to obtain rational constraints on μ (rather than on[†] x).

Before we state the (somewhat inductive) description of invariant measures for horospherical subgroups we state the general version of the argument that was used in the proof of Theorem 5.4.

Proposition 5.6 (Mixing argument for G_a^-). *Let $X = G \bullet x_0 \subseteq d$ be a finite volume orbit for a closed connected subgroup $G \leqslant \mathrm{SL}_d(\mathbb{R})$. Let $a \in G$ be diagonalizable over $\mathbb{R}$, and suppose that a acts as a mixing transformation on X with respect to m_X. Let G_a^- be the stable horospherical subgroup for a, and let B_0 be a neighborhood of $I \in G_a^-$ with compact closure and a boundary of zero Haar measure. For any $f \in C_c(X)$, any compact set $K \subseteq X$, and every $\varepsilon > 0$ there exists an integer k_0 such that*

$$\left| \frac{1}{m_{G_a^-}(a^{-k} B_0 a^k)} \int_{a^{-k} B_0 a^k} f(u \bullet x) \, \mathrm{d}m_{G_a^-}(u) - \int_X f \, \mathrm{d}m_X \right| < \varepsilon$$

for all $k \geqslant k_0$ whenever $a^k \bullet x \in K$.

Proof. We show how, after minor modifications, the argument for the proof of Theorem 5.4 also proves the proposition.

[†] See the next section where we go further and describe properties of a given x.

As $f \in C_c(X)$ is uniformly continuous, we can choose $V \subseteq P_a$ as in the proof of Theorem 5.4, again with a boundary of zero Haar measure. This shows that $VB_0 \subseteq P_a G_a^- \subseteq G$ has a boundary of zero measure. Therefore, there exists a compact set $C \subseteq (VB_0)^o$ and an open set $O \supseteq \overline{VB}$ such that $m_G(O \smallsetminus C) < \varepsilon^2$. If now $\delta > 0$ is sufficiently small, then $CB_\delta^G \subseteq VB_0$ and $\overline{VB_0} B_\delta^G \subseteq O$. By compactness of K it follows that

$$K \subseteq \bigcup_{i=1}^{n} B_\delta^G \bullet x_i$$

for some finite collection $x_1, \dots, x_n \in K$. This implies that for every $x \in K$ there is some x_i with $x \in B_\delta^G \bullet x_i$,

$$C \bullet x \subseteq CB_\delta^G \bullet x_i \subseteq VB_0 \bullet x_i$$

and

$$O \bullet x \supseteq VB_0 B_\delta^G \bullet x \supseteq VB_0 \bullet x_i.$$

We also have trivially $C \bullet x \subseteq VB_0 \bullet x \subseteq O \bullet x$, and so we get

$$\left\| \mathbb{1}_{VB_0 \bullet x} - \mathbb{1}_{VB_0 \bullet x_i} \right\|_2 < \varepsilon.$$

To summarise, we have shown that the set of characteristic functions $\mathbb{1}_{VB_0 \bullet x} \in L^2(X, m_X)$ with $x \in K$ is totally bounded. By applying the mixing property to f and to $\mathbb{1}_{VB_0 \bullet x_i}$ for $i = 1, \dots, n$, we may assume that mixing holds uniformly for f and for all $\mathbb{1}_{VB_0 \bullet x}$ for $x \in K$. Now the argument used in the proof of Theorem 5.4 gives the proposition. □

[MLE] need to decide whether we really call this a theorem — it is in a sense the inductive step of the real theorem (if we can make the induction work)

Theorem 5.7. *Let $X = G \bullet x_0 \subseteq d$ be a finite volume orbit of some closed connected subgroup $G < \mathrm{SL}_d(\mathbb{R})$ and some point $x_0 \in d$. Let $a \in G$ and assume that the action of a on X is mixing with respect to the Haar measure m_X. Let $U = G_a^- < G$ be a horospherical subgroup defined by $a \in G$. Then any U-invariant and ergodic probability measure on X other than m_X is supported on a closed orbit $L \bullet x$ for some closed connected proper subgroup $L < G$ and some $x \in X$.*

For the proof the following will be helpful(23).

Lemma 5.8. *Let U be a nilpotent connected Lie group acting ergodically on a locally compact metric space X with respect to some invariant probability measure μ. Then there exists a one-parameter subgroup of U that also acts ergodically.*

Proof. We consider first the case where $U \cong \mathbb{R}^d$ is abelian, with $d \geqslant 1$, where we will apply the spectral theory of locally compact abelian groups (we refer to Folland [?] for the details). Let U act on X, and let μ be a U-invariant and ergodic probability measure. For each $u \in U$ let

$$\pi_u : \mathscr{H} = L^2(X,\mu) \longrightarrow \mathscr{H}$$

be the associated unitary representation defined by

$$(\pi_u(f))(x) = f(u^{-1}\bullet x).$$

By [**?**, Sec. 1.4] there exists a sequence† of spectral measures that completely describe the unitary representations. Specifically, there exists a sequence (ν_n) of finite measures on $\mathbb{R}^d$ such that π_u is unitarily isomorphic to the operator

$$\begin{aligned} M_u : \bigoplus_{n=0}^{\infty} L^2(\mathbb{R}^d,\nu_n) &\longrightarrow \bigoplus_{n=0}^{\infty} L^2(\mathbb{R}^n,\nu_n) \\ (f_n) &\longmapsto (M_{u,n}(f))_n \end{aligned}$$

where

$$M_{u,n}(f_n)(t) = \mathrm{e}^{2\pi\mathrm{i}\langle u,t\rangle} f_n(t).$$

We may suppose that $\nu_0 = \delta_0$, with $M_{u,0}$ the trivial representation of U on $\mathbb{C}$ corresponding to the invariant subspace of constant functions in the Hilbert space $\mathscr{H} = L^2(X,\mu)$. By ergodicity of the U-action, we must have $\nu_n(\{0\}) = 0$ for $n \geqslant 1$, for otherwise we could find a non-trivial U-invariant L^2-function in the orthogonal complement of the constant functions. Therefore, we may push forward the measures ν_n for $n \geqslant 1$ from $\mathbb{R}^d \smallsetminus \{0\}$ to the projective space $\mathbb{P}^{d-1}(\mathbb{R}) \cong \mathbb{R}^d \smallsetminus \{0\}/\sim$, where $x \sim y$ if and only if there is some $\lambda \in \mathbb{R}\smallsetminus\{0\}$ with $x = \lambda y$. As these countably many finite measures can only have countably many atoms in total, and $\mathbb{P}^{d-1}(\mathbb{R})$ is uncountable as $d \geqslant 2$, there must be a point in $\mathbb{P}^{d-1}(\mathbb{R})$ of zero measure for all of the measures. Hence there is a line $\mathbb{R}v \leqslant \mathbb{R}^d$ with $\nu_n(\mathbb{R}v) = 0$ for $n \geqslant 1$. We claim that the restriction of the ergodic $U \cong \mathbb{R}^d$-action on (X,μ) to the hyperplane $v^\perp \leqslant \mathbb{R}^d$ is still ergodic. Indeed, if $f \in L^2(X,\mu)$ is orthogonal to the constant functions, then it corresponds to an element

$$(f_n) \in \bigoplus_{n \geqslant 1} L^2(\mathbb{R}^d,\nu_n),$$

and if f is $v^\perp$-invariant then the functions f_n for $n \geqslant 1$ would have to be supported on $\mathbb{R}v$ (since this is the subset of $\mathbb{R}^d$ where all the operators M_u with $u \in v^\perp$ act trivially). However, by the choice of v this forces $f_n = 0$ for $n \geqslant 1$ and hence forces f to be 0 (since it is assumed to be orthogonal to the constant functions).

The argument above shows that every ergodic action of $\mathbb{R}^d$ with $d \geqslant 2$ can be restricted to a hyperplane with the property that the restriction remains ergodic. By induction, this shows that the lemma holds for $U \cong \mathbb{R}^d$.

† Here we are using the fact that $\mathscr{H}$ is separable, which in turn follows from the fact that X is a locally compact σ-compact metric space.

Now suppose that U is nilpotent but not abelian. We define

$$U' = [U, U]$$

so that $G = U/U' \cong \mathbb{R}^d$ for some $d \geqslant 2$. By assumption U acts ergodically by measure-preserving transformations of (X, μ). Recall from [**?**, Th. 8.20] (also see [**?**, Sec. 5.3]) that μ has an ergodic decomposition

$$\mu = \int_X \mu_x^{\mathscr{E}'} \, \mathrm{d}\mu(x) \tag{5.5}$$

into U'-ergodic components $\mu_x^{\mathscr{E}'}$ given by the conditional measures for the σ-algebras

$$\mathscr{E}' = \{B \subseteq X \mid \mu(u' \raisebox{0.3ex}{\scriptsize$\bullet$} B \triangle B) = 0 \text{ for all } u' \in U'\}.$$

We now show that $\mathscr{E}'$ is U-invariant in the sense that $B \in \mathscr{E}'$ implies that $u \raisebox{0.3ex}{\scriptsize$\bullet$} B \in \mathscr{E}'$ for all $u \in U$. Indeed, for $B \in \mathscr{E}'$, $u \in U$, and $u' \in U'$, we have

$$\mu\left(u' \raisebox{0.3ex}{\scriptsize$\bullet$} (u \raisebox{0.3ex}{\scriptsize$\bullet$} B) \triangle (u \raisebox{0.3ex}{\scriptsize$\bullet$} B)\right) = \mu\left((u \raisebox{0.3ex}{\scriptsize$\bullet$} ((u'' \raisebox{0.3ex}{\scriptsize$\bullet$} B) \triangle B))\right) = 0,$$

where $u'' = u^{-1} u' u \in U''$. By [**?**, Cor. 5.24] this shows that

$$u_* \mu_x^{\mathscr{E}'} = \mu_{u \raisebox{0.3ex}{\scriptsize$\bullet$} x}^{\mathscr{E}'}$$

for almost every $x \in X$ and all $u \in U$. Hence the map

$$\begin{aligned} \phi : (X, \mu) &\longrightarrow \left(\mathscr{M}(X)^{U'}, \phi_* \mu\right) \\ x &\longmapsto \mu_x^{\mathscr{E}'} \end{aligned}$$

is a μ-almost everywhere well-defined factor map intertwining the U-action on (X, μ) and the induced U-action on $(\mathscr{M}(X)^{U'}, \phi_* \mu)$, where $\mathscr{M}(X)^{U'}$ denotes the space of U'-invariant measures with total mass no more than 1, and the induced action is defined by $u \raisebox{0.3ex}{\scriptsize$\bullet$} \nu = u_* \nu$ for $u \in U$ and $\nu \in \mathscr{M}(X)^{U'}$. Since by construction U' acts trivially on $\mathscr{M}(X)^{U'}$ we have obtained an ergodic[†] action of $G = U/U'$ on $(\mathscr{M}(X)^{U'}, \phi_* \mu)$. By the abelian case, there is a subgroup $H \leqslant G$ isomorphic to $\mathbb{R}$ which still acts ergodically on $(\mathscr{M}(X)^{U'}, \phi_* \mu)$. Let HU' denote the pre-image of H in U. We claim that the subgroup HU' still acts ergodically on (X, μ). Assuming this, the lemma follows by induction as HU' is a nilpotent group of smaller dimension than U.

Suppose now that $B \subseteq X$ is an HU'-invariant measurable set. In particular, it is U'-invariant and so we must have

$$\mu_x^{\mathscr{E}'}(B) \in \{0, 1\} \tag{5.6}$$

for μ-almost every $x \in X$. We define

[†] Ergodicity is automatic, as the system is exhibited as a factor of an ergodic system.

$$B' = \{\nu \in \mathscr{M}(X)^{U'} \mid \nu(B) = 1\},$$

and notice that $\nu \in \mathscr{M}(X)^{U'}$ and $h \in HU'$ implies that

$$h_*\nu(B) = \nu(h^{-1}\boldsymbol{\cdot} B) = \nu(B),$$

so that B' is an H-invariant set. By ergodicity of H on $(\mathscr{M}(X)^{U'}, \phi_*\mu)$, this shows that

$$\phi_*\mu(B') = \mu(\{x \mid \mu_x^{\mathscr{E}'}(B) = 1\}) \in \{0,1\}.$$

By (5.5) and (5.6), this shows that $\mu(B) \in \{0,1\}$ and so HU acts ergodically, which concludes the induction. □

Proof of Theorem 5.7. Let $X = G\boldsymbol{\cdot} x_0$ and $a \in G$ be as in the theorem. Let μ be a G_a^--invariant ergodic probability measure on X. By Lemma 5.8 there is a one-parameter unipotent subgroup $U < G_a^-$ that acts ergodically with respect to μ.

Without loss of generality we may assume that $x_0 \in \operatorname{Supp}\mu$ and that x_0 is U-generic (see Section 6.3.1 and [**?**, Sec. 4.4.2]). In particular, we have

$$\overline{x_0 U} = \operatorname{Supp}\mu.$$

Let $g_0 \in \mathrm{SL}_d(\mathbb{R})$ be chosen with $x_0 = \Gamma g_0$, and let $\Lambda_0 = \mathbb{Z}^d g_0$ be the corresponding unimodular lattice in $\mathbb{R}^d$. For every Λ_0-rational subspace† $V \subseteq \mathbb{R}^d$ we define

$$L_V = \{g \in G \mid V = Vg \text{ and } g|_V \text{ preserves the volume}\}.$$

Applying Exercise 3.1.4 we see that $x_0 L_V$ is closed. If there exists one such proper subgroup $L_V \subsetneq G$ such that $\operatorname{Supp}\mu \subseteq x_0 L_V$, then the theorem already holds for μ. Therefore, we assume that $\operatorname{Supp}\mu \not\subseteq x_0 L_V$ for every $L_V \subsetneq G$.

We define

$$\eta = \min\Big\{(\Lambda_0 \cap V, V)^{1/\dim V} \mid V \text{ is } \Lambda_0\text{-rational}\Big\}. \tag{5.7}$$

Applying the quantitative non-divergence theorem (Theorem 4.9) with D chosen for U and η as above, we find some $\varepsilon > 0$ with the following property: For any $x = \Gamma g \in d$, any one-parameter subgroup U' with the same D as worked for U, we have that either U' fixes a $\mathbb{Z}^d g$-rational subspace V with

$$V < \eta^{\dim V},$$

or there exists some T_x such that

$$\frac{1}{T}\left|\{t \in [0,T] \mid u'_t\boldsymbol{\cdot} x \in d(\varepsilon)\}\right| > \frac{9}{10}. \tag{5.8}$$

† A subspace V is called Λ_0-rational if $V \cap \Lambda_0$ is also a lattice.

for $T > T_x$. We set $K = X \cap d(\varepsilon)$. Applying (5.8) to $x = x_0$, $U' = U$, and $T \to \infty$, for example, gives $\mu(K) > \frac{9}{10}$.

We now let $n \geqslant 1$ and set $\mu_n = a_*^n \mu$. Notice that the subgroup $a^n U a^{-n}$ acts ergodically with respect to μ_n, and that $x_n = a^n {\scriptstyle\bullet} x_0$ is generic for $a^n U a^{-n}$ and μ_n. Suppose that the lattice $\Lambda_n = \Lambda_0 a^{-n}$ corresponding to x_n has a Λ_n-rational subspace V with

$$(V) < \eta^{\dim V}.$$

If the one-parameter subgroup $a^n U a^{-n}$ stabilizes V then $a^n U a^{-n} \leqslant L_V$, the orbit $x_n L_V$ is closed by Exercise 3.1.4, and so

$$\overline{a^n U a^{-n} {\scriptstyle\bullet} x_n} = \operatorname{Supp} \mu_n \subseteq x_n L_V$$

gives

$$\overline{U {\scriptstyle\bullet} x_0} = \operatorname{Supp} \mu \subseteq x_0 L_{V a^n}.$$

Since $a \in G \setminus L_{V a^n}$, this would contradict our assumption on μ. As this holds for all subspaces as above, we see that $x = x_n$ and $U' = a^n U a^{-n}$ satisfy the assumptions that lead to (5.8). Letting $T \to \infty$ gives $\mu_n(K) \geqslant \frac{9}{10}$.

In other words, we have shown that $\frac{9}{10}$ of all $x \in X$ with respect to μ satisfy $a^n {\scriptstyle\bullet} x \in K$. For any such x, any $f \in C_c(X)$, and any $\varepsilon > 0$, we may apply Proposition 5.6 to see that

$$\frac{1}{m_{G_a^-}(a^{-n} B_0 a^n)} \int_{a^{-n} B_0 a^n} f(h {\scriptstyle\bullet} x) \, \mathrm{d} m_{G_a^-}(h) = \int_X f \, \mathrm{d} m_X + \mathrm{O}(\varepsilon)$$

if n is large enough. By Exercise 5.2.1, $(a^{-n} B_0 a^n)$ is a Følner sequence in G_a^-. Therefore, we may apply the mean ergodic theorem (see [**?**, Th. 8.13]) to see that for large enough n and $\frac{9}{10}$ of all $x \in X$ with respect to μ we have

$$\frac{1}{m_{G_a^-}(a^{-n} B_0 a^n)} \int_{a^{-n} B_0 a^n} f(h {\scriptstyle\bullet} x) \, \mathrm{d} m_{G_a^-}(h) = \int_X f \, \mathrm{d}\mu + \mathrm{O}(\varepsilon).$$

As $\varepsilon > 0$ and $f \in C_c(X)$ were arbitrary, we deduce that

$$\int_X f \, \mathrm{d}\mu = \int_X f \, \mathrm{d} m_X,$$

and conclude that $\mu = m_X$. □

Let us give a complete description in the case of 3.

Corollary 5.9. *Let* $3 = \mathrm{SL}_3(\mathbb{Z}) \backslash \mathrm{SL}_3(\mathbb{R})$. *Let* $U = G_a^-$ *be a horospherical subgroup of* $\mathrm{SL}_3(\mathbb{R})$ *defined by some* $\mathbb{R}$*-diagonalizable element* $a \in \mathrm{SL}_3(\mathbb{R})$. *Then any* U*-invariant and ergodic probability measure is* algebraic*, meaning that it is the Haar measure on a closed orbit of a closed connected subgroup* L *in* $\mathrm{SL}_3(\mathbb{R})$. *In fact, we could have either* $L = \mathrm{SL}_3(\mathbb{R})$, $L \simeq \mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$ *(which can be embedded into* $\mathrm{SL}_3(\mathbb{R})$ *in two non-conjugate ways), or* L *could be unipotent.*

Sketch of Proof. Let μ be a U-invariant and ergodic probability measure on 3 as in the corollary. If $\mu \neq m_3$ then there exists a subgroup $L < \mathrm{SL}_3(\mathbb{R})$ such that μ is supported by the closed orbit $L\boldsymbol{\cdot}x$ for some $x \in 3$. As the proof of Theorem 5.7 shows we may assume that L is the stabilizer of a vector v or the stabilizer of a volume element $v_1 \wedge v_2$ in a plane in $\mathbb{R}^3$. Conjugating a and U we may suppose that $x = \mathrm{SL}_d(\mathbb{Z})$ is the identity coset, v resp. v_1, v_2 are integer vectors, and so $L\boldsymbol{\cdot}x$ is isomorphic to $_2(\mathbb{Z})\backslash(\mathbb{R})$. (This quotient can be embedded into 3 in two ways depending on whether the radical is chosen to be represented by row vectors or by column vectors.)

If U is a two-dimensional horospherical subgroup, then either U can still be defined as a horospherical subgroup $U = L_{a'}^{-}$ for some $a' \in L$ or U equals the radical (which has a closed orbit for all points in xL). Hence we may repeat the argument or are done. If U is the Heisenberg group, then $U < L$ is not a horospherical subgroup within L, but U contains the full radical. Taking the quotient by the radical we obtain a measure on the modular surface that is invariant under the horocycle flow. In each of these cases we obtain that μ is algebraic and find that L is one of the mentioned subgroups. □

Exercises for Section 5.3

Exercise 5.3.1. Prove Theorem 5.5 using the method of proof from Theorem 5.4.

5.4 Equidistribution for Non-Compact Quotients

Dani and Smillie showed in [**?**] that even for non-compact quotients

$$X = \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$$

a rather strong equidistribution theorem holds: a horocycle orbit is either periodic or it equidistributes with respect to the uniform measure m_X.

For higher dimensional non-compact quotients $X = \Gamma\backslash G$ and their horospherical actions other possibilities can occur. For the following characterization of whether or not a horospherical orbit equidistributes we specialize to the case where the horospherical subgroup is abelian.

Before we can prove Theorem 5.11 we need to extend the non-divergence result to actions of more general unipotent groups. For simplicity we give only the version needed for the case at hand.

Corollary 5.10. *Let U be an abelian unipotent subgroup of $\mathrm{SL}_d(\mathbb{R})$, and fix some coordinate system identifying U with $\mathbb{R}^k$ and with respect to which we can describe 'blocks' whose edges are parallel to the coordinate axes. Then for* *every $\eta > 0$ and every $\delta > 0$ there exists a compact subset $K \subseteq d$ with the property that for any $x \in d$ either*

- *there is a Λ_x-rational U-invariant subspace V with $(V) < \eta^{\dim V}$, or*
- *for any symmetric block $F \subseteq U$ with sufficiently large width we have*

$$\frac{1}{m_U(F)} m_U\left(\{u \in F \mid u{\bullet}x \in K\}\right) > 1 - \delta.$$

Proof. The corollary follows quite directly from Theorem 4.9. Let D be chosen so that we may apply Theorem 4.9 for any one-parameter subgroup of U, and let $x \in d$ be arbitrary.

If U fixes a Λ_x-rational subspace $V \subseteq \mathbb{R}^d$ with $(V) < \eta^{\dim V}$, then there is nothing to prove. So suppose that this is not the case. As there are only finitely many Λ_x-rational subspaces with co-volume less than $\eta^{\dim V}$, and for each such subspace the subgroup of U that fixes V is of codimension at least 1, there exists a one-parameter subgroup

$$U' = \{u'(t) \mid t \in \mathbb{R}\} \subseteq U$$

that does not fix any of these subspaces. Applying Theorem 4.9 to $p(t) = gu'(t)$ with $x = \mathrm{SL}_d(\mathbb{Z})g$, η as above, some $\varepsilon_0 > 0$ (depending on η and the implicit constant in (4.19) only), and some possibly very large T (depending on x), it follows that there exists at least one $t \in \mathbb{R}$ with

$$x' = u'(t){\bullet}x \in d(\varepsilon_0).$$

Now let $\{u_1(t_1) \mid t_1 \in \mathbb{R}\}, \{u_2(t_2) \mid t_2 \in \mathbb{R}\}, \dots, \{u_k(t_k) \mid t_k \in \mathbb{R}\}$ be the one-parameter subgroups of U corresponding to the chosen coordinate system in $U \cong \mathbb{R}^k$.

We can now split $F_n{\bullet}x$ as indicated in Figure 5.3 above into 2^k sets of the form $F{\bullet}x'$ where each F is a block with the origin as one corner. For simplicity we consider only the block in the positive quadrant where

$$F = \{u_1(t_1)u_2(t_2)\cdots u_k(t_k) \mid 0 \leqslant t_i \leqslant T_i, i = 1, \dots, k\} \subseteq U,$$

the blocks in the other quadrants can be dealt with in the same way. We now successively choose $\varepsilon_1, \dots, \varepsilon_k$ (depending only on ε) such that

$$\frac{1}{T_1}\left|\{t_1 \in [0, T_1] \mid u_1(t_1){\bullet}x' \notin d(\varepsilon_1)\}\right| < \frac{\delta}{k},$$

and, if $u_1(t_1){\bullet}x' \in d(\varepsilon_1)$,

$$\frac{1}{T_2}\left|\{t_2 \in [0, T_2] \mid u_1(t_1)u_2(t_2){\bullet}x' \notin d(\varepsilon_2)\}\right| < \frac{\delta}{k},$$

and so on, ending with

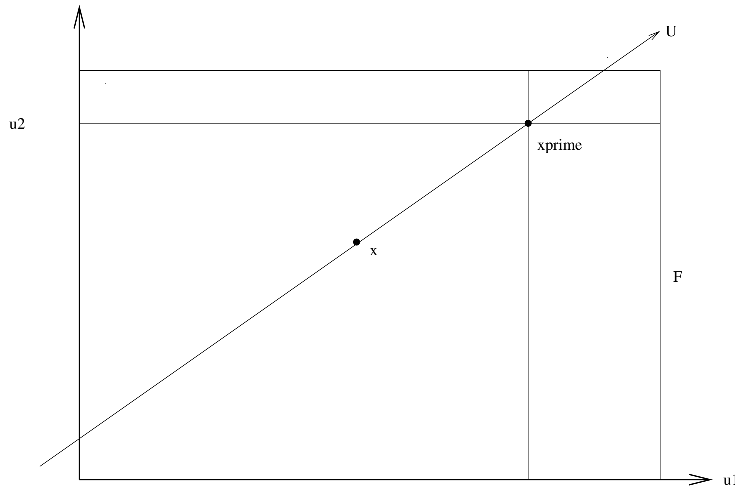

Fig. 5.3 The symmetric box $F_n \cdot x$ inside the U-orbit has to contain $x' \in d(\varepsilon)$ if the width of F_n is sufficiently large.

$$\frac{1}{T_k} \left|\{t_k \in [0, T_k] \mid u_1(t_1) \cdots u_k(t_k) \text{•} x' \notin d(\varepsilon_k)\}\right| < \frac{\delta}{k}$$

if $u_1(t_1) \cdots u_{k-1}(t_{k-1}) \in d(\varepsilon_{k-1})$. We set $K = d(\varepsilon_k)$, and the corollary follows. $\square$

Theorem 5.11. *Let $G \text{•} x_0 \subseteq d = \mathrm{SL}_d(\mathbb{Z}) \backslash \mathrm{SL}_d(\mathbb{R})$ be a finite volume orbit for some closed connected subgroup $G \leqslant \mathrm{SL}_d(\mathbb{R})$ and some point $x_0 \in d$. Let $a \in G$ be diagonalizable over $\mathbb{R}$ so that the action of a is mixing with respect to $m_{G\text{•}x_0}$. Let $U = G_a^-$ be the stable horospherical subgroup of a and suppose that it is abelian. Let (F_n) be a Følner sequence in U consisting of blocks whose sides are parallel to some fixed coordinate system spanned by some eigenvectors for the conjugation map by a. Then for every $x \in G\text{•}x_0$ the following are equivalent*

(1) *The U-orbit through x is equidistributed, meaning that*

$$\frac{1}{m_U(F_n)} \int_{F_n} f(u_t) \, \mathrm{d}m_U(t) \to \int_d f \, \mathrm{d}m_d$$

for any $f \in C_c(d)$.

(2) *The orbit $U\text{•}x$ is not contained in a closed orbit $L\text{•}x$ for some proper connected subgroup $L < \mathrm{SL}_d(\mathbb{R})$.*

If, in addition, $G = \mathrm{SL}_d(\mathbb{R})$ then we also have the equivalence to the next property.

(3) *Let $x = \mathrm{SL}_d(\mathbb{Z})g$ for some $g \in \mathrm{SL}_d(\mathbb{R})$. Then there is no rational subspace $V \subseteq \mathbb{R}^d$ for which Vg is fixed by U and contracted by a.*

Proof. We let $x \in G{\boldsymbol\cdot}x_0$ be as in the theorem. If the G_a^--orbit of x is contained in a closed orbit of a proper connected subgroup $L < G$ as in (2), then clearly we cannot have equidistribution of the G_a^--orbit as in (1). This shows that (1) implies (2), so we now assume that (2) holds.

Fix some $f \in C_c(d)$ and $\varepsilon > 0$. We let $x_0 = \Gamma g_0$, $\Lambda_0 = \mathbb{Z}^d g_0$ be the lattice corresponding to x_0, and define η as in (5.7). By quantitative non-divergence for the action of $U = G_a^-$ there exists some compact set $K \subseteq d$ with the property as in Corollary 5.10 with $\delta = \varepsilon$. We let B_0 be the symmetric unit cube (that is, centered at the origin) in $G_a^- \cong \mathbb{R}^k$. Applying Proposition 5.6 to f, KB_0, and ε we find some $k \geqslant 1$ such that

$$\left| \frac{1}{m_{G_a^-}(a^{-k}B_0a^k)} \int_{a^{-k}B_0a^k} f(u{\boldsymbol\cdot}y)\,\mathrm{d}m_{G_a^-}(u) - \int_X f\,\mathrm{d}m_X \right| < \varepsilon \tag{5.9}$$

whenever $a^k{\boldsymbol\cdot}y \in (KB_0) \cap X$ (or equivalently whenever $B_0a^k{\boldsymbol\cdot}y$ intersects K).

Now let $x' = a^k{\boldsymbol\cdot}x$ and notice that it may not belong to K. Since (F_n) is chosen to be a Følner sequence consisting of symmetric blocks, the same is true for $a^kF_na^{-k}$. If $U = G_a^-$ fixes a $\Lambda_{x'}$-rational subspace V of co-volume $< \eta^{\dim V}$, then we can define the subgroup

$$L' = \mathrm{Stab}_G^1(V) = \{g \in G \mid Vg = V \text{ and } g|_V \text{ has determinant } 1\} \lneq G.$$

Exercise 3.1.4 shows that $x'L'$ is closed, which shows that

$$x'L'a^k = x'a^ka^{-k}L'a^k = xL,$$

where $L = a^{-k}L'a^k$, is a closed orbit which contradicts the assumption that (2) holds if L is a proper subgroup of G.

It follows that U does not fix any Λ_x-rational subspaces that are not already fixed by G. Applying Corollary 5.10 we see now that for large enough n we have

$$\frac{1}{m_{G_a^-}(a^kF_na^{-k})} m_{G_a^-}\left(\{u \in a^kF_na^{-k} \mid u{\boldsymbol\cdot}x' \notin K\}\right) < \varepsilon. \tag{5.10}$$

We now split $a^kF_na^{-k}$ into translates of the form B_0u_ℓ for $\ell = 1, \ldots, L$ of the unit cube B_0. Ignoring the effects of the boundary which contribute no more than $\mathrm{o}_f(1)$ to the ergodic average as $n \to \infty$, we now have

$$\frac{1}{m_{G_a^-}(F_n)} \int_{F_n} f(u{\boldsymbol\cdot}x)\,\mathrm{d}m_{G_a^-}$$

$$= \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{m_{G_a^-}(a^{-k}B_0a^k)}\int_{a^{-k}B_0a^k} f\left(ua^{-k}u_\ell a^k\boldsymbol{\cdot} x\right)\,\mathrm{d}m_{G_a^-} + \mathrm{o}_f(1).$$

For all those ℓ for which $B_0u_\ell a^k\boldsymbol{\cdot} x$ intersects K the corresponding average is ε-close to $\int_X f\,\mathrm{d}m_X$ by (5.9). However, the number of boxes $B_0u_\ell\boldsymbol{\cdot} x'$ that do not intersect K is controlled by (5.10), and gives

$$\frac{1}{m_{G_a^-}(F_n)}\int_{F_n} f(u\boldsymbol{\cdot} x)\,\mathrm{d}m_{G_a^-}(u) = \int_X f\,\mathrm{d}m_X + \mathrm{o}_f(1) + \mathrm{O}_f(\varepsilon).$$

As $\varepsilon > 0$ and $f \in C_c(X)$ were arbitrary, this shows (1).

Now suppose that $G = \mathrm{SL}_d(\mathbb{R})$ and

$$a = \begin{pmatrix}\lambda^n I_m & \\ & \lambda^{-m} I_n\end{pmatrix} \in \mathrm{SL}_d(\mathbb{R})$$

for some $\lambda > 1$ so that

$$G_a^- = \left\{\begin{pmatrix} I_m & \\ * & I_n\end{pmatrix}\right\}$$

is indeed abelian. (Up to signs in the entries of a and the choice of m and n this is the only choice of a for which G_a^- is abelian.) If

$$L = \mathrm{Stab}^1_{\mathrm{SL}_d(\mathbb{R})}(V)$$

for some proper subspace V is G_a^--invariant, then either $V \subseteq \mathbb{R}^m \times \{0\}^n$ or V contains some $v = (v_m, v_n)$ with $v_m \in \mathbb{R}^m$ and $v_n \in \mathbb{R}^n \smallsetminus \{0\}$, which implies that $\mathbb{R}^m \times \{0\}^n \subseteq V$. In both cases $Va^{-1} = V$ and the restriction of a^{-1} to V has determinant smaller than 1. □

In the exercises we outline how one can remove the assumptions on commutativity of U.

Exercises for Section 5.4

Exercise 5.4.1. Let $G\boldsymbol{\cdot} x_0 \subseteq d = \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})$ be a finite volume orbit for some closed connected subgroup $G \leqslant \mathrm{SL}_d(\mathbb{R})$ and some point $x_0 \in d$.

Let U be a unipotent subgroup and let $U_1, \dots, U_\ell < U$ be one-parameter subgroups such that $U = U_1 \cdots U_\ell$. Suppose $F_n = F_{1,n} \cdots F_{\ell,n}$ is a Følner sequence with respect to left and right translation where $F_{i,n} \subseteq U_i$ corresponds to an interval in U_i.

Prove that

$$\frac{1}{m_U(F_n)} m_U(\{u \in F_n \mid u\boldsymbol{\cdot} x \notin d(\delta)\}) \ll \delta^\kappa + o(1)$$

for $n \to \infty$ and some $\kappa > 0$ (depending on $U_1, \dots, U_\ell$) in the following two cases:

(a) x belongs to a fixed compact subset and the implicit constant is allowed to depend on the compact subset, and

(b) x is arbitrary but U does not fix any Λ_x-rational subspace V of co-volume $\eta^{\dim V}$ and the implicit constant is allowed to depend on η.

Exercise 5.4.2. Let $G\bullet x_0 \subseteq d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$ be a finite volume orbit for some closed connected subgroup $G \leqslant \mathrm{SL}_d(\mathbb{R})$ and some point $x_0 \in d$. Let $a \in G$ be diagonalizable over $\mathbb{R}$ so that the action of a is mixing with respect to $m_{G\bullet x_0}$. Let $U = G_a^-$ be the stable horospherical subgroup of a and let F_n be as in Exercise 5.4.1. Let $x \in G\bullet x_0$. Suppose that U does not fix any Λ_x-rational subspace which is not also fixed by G. Show that $F_n\bullet x$ equidistributes in $X = G\bullet x_0$, i.e.

$$\frac{1}{m_U(F_n)}\int_{F_n} f(u\bullet x)dm_U(u) \longrightarrow \int_X f dm_X$$

as $n \to \infty$ for any $f \in C_c(X)$.

Notes to Chapter 5

[21](Page 156) We refer to [**?**, Ex. 3.3.1, 9.6.3] for one example of such a construction. McMullen [**?**] gives explicit constructions of bounded geodesics of arbitrary length associated to elements of any given quadratic field, and relates the construction to continued fractions.

[22](Page 159) This is an example of a circle of results developed among others by Dani [**?**], [**?**] and Veech [**?**].

[23](Page 164) A much stronger form of this result is obtained by Pugh and Shub [**?**], where it is shown that if T is an ergodic measure-preserving action of $\mathbb{R}^d$ on a Borel probability space, then there is a countable collection $\{H_n \mid n \in \mathbb{N}\}$ of hyperplanes with the property that for any $g \in \mathbb{R}^d \smallsetminus \bigcup_{n\in\mathbb{N}} H_n$ the measure-preserving transformation T_g is ergodic. In our setting, Lemma 5.8 can be avoided by using finitely many one-parameter subgroups as in the proof of Theorem 4.11 for unipotent subgroups on page 152.

Chapter 6
Ratner's Theorems in Unipotent Dynamics

In this chapter we discuss Ratner's theorems concerning unipotent dynamics and prove some special cases. We will not discuss the history in detail, and refer to the survey papers of Kleinbock, Shah and Starkov [**?**], Ratner [**?**], Margulis [**?**], and Dani [**?**] for that. In particular, the order in which the material is developed is not historical but instead emphasizes a logical development with the benefit of hindsight. We will also postpone any discussion of applications of unipotent dynamics, including the solution of the Oppenheim conjecture by Margulis that motivated some of the original theorems, to later chapters.

In this volume we will prove many special cases of the results of Section 6.1 and Section 6.2, potentially[†] reaching the general case for spaces of the form $X = \mathbb{G}(\mathbb{Z})\backslash\mathbb{G}(\mathbb{R})$.

6.1 The Main Theorems

We let $X = \Gamma\backslash G$, where G is a connected Lie group and $\Gamma < G$ a lattice. Let

$$U = \{u_s \mid s \in \mathbb{R}\} < G$$

be a one-parameter unipotent subgroup of G. Then the U-invariant probability measures on X can be completely classified. This was conjectured by Dani (in [**?**, Conjecture I], as an analog of Raghunathan's conjecture, which will be described below) and proved by Ratner [**?**], [**?**], [**?**].

Theorem 6.1 (Dani's conjecture; Ratner's measure classification). *If $X = \Gamma\backslash G$ and $U = \{u_s \mid s \in \mathbb{R}\} < G$ is a one-parameter unipotent subgroup, then every U-invariant ergodic probability measure μ on X is*[‡]

[†] Currently none of the first five theorems of this chapter are proven in this volume.

[‡] An alternative term that is used is *homogeneous*.

algebraic. *That is, there exists a closed connected unimodular subgroup L with $U \leqslant L \leqslant G$ such that μ is the L-invariant normalized probability measure (that is, the normalized Haar measure) on a closed orbit $L \centerdot x_0$ (for any $x_0 \in \operatorname{Supp} \mu$).*

In this result (unlike the following ones), it is sufficient to assume that $\varGamma$ is discrete or even just closed. Theorem 6.1, the theorem of Dani and Smillie [**?**], (resp. its generalization from Section 5.4), and the general non-divergence property of unipotent orbits, suggest other results which we now start to describe. Ratner [**?**] generalized all of these results in the following theorem.

Theorem 6.2 (Ratner's equidistribution theorem). *Let $X = \varGamma\backslash G$ where $\varGamma$ is a lattice, and let $U = \{u_s \mid s \in \mathbb{R}\} < G$ be a one-parameter unipotent subgroup. Then for any $x_0 \in X$ there exists some closed connected unimodular subgroup $L \leqslant G$ such that $U \leqslant L$,*

- *$L \centerdot x_0$ is closed with finite L-invariant volume, and*
- $\displaystyle \frac{1}{T}\int_0^T f(u_s \centerdot x_0)\,\mathrm{d}s \longrightarrow \frac{1}{\operatorname{volume}(L \centerdot x_0)}\int_{L\centerdot x_0} f\,\mathrm{d}m_{L\cdot x_0}$ *as $T \to \infty$.*

It is interesting to note that Theorem 6.2 in particular implies that any point $x \in X$ returns close to itself under a unipotent flow. That is, for any one-parameter unipotent subgroup $\{u_s \mid s \in \mathbb{R}\}$ and any $x \in X$ there is a sequence $(t_k)_{k\geqslant 1}$ for which $t_k \to \infty$ and $\mathsf{d}(x, u_{t_k} \centerdot x) \to 0$ as $k \to \infty$. This close return statement is of course incomparably weaker than Ratner's equidistribution theorem, but even this weak statement does not seem to have an independent proof to our knowledge.

Theorem 6.2 also suggests that the closures of orbits under the action of a unipotent one-parameter subgroup should be algebraic. A more general version of that statement is the famous conjecture of Raghunathan(24) that motivated all of the theorems above and was proved by Ratner [**?**].

Theorem 6.3 (Raghunathan's conjecture; Ratner's orbit closure theorem). *Suppose that $X = \varGamma\backslash G$, with G a connected Lie group and $\varGamma$ a lattice. Let $H < G$ be a closed subgroup generated by one-parameter unipotent subgroups. Then for any $x_0 \in X$ the orbit closure is*[†] algebraic*, meaning that there exists some closed connected unimodular subgroup L with $H \leqslant L \leqslant G$ such that*

$$\overline{H \centerdot x_0} = L \centerdot x_0$$

and $L \centerdot x_0$ supports a finite L-invariant measure.

It is also interesting to ask what the structure of the set of all probability measures that are invariant and ergodic under some unipotent flow really is. This generalizes the theorem of Sarnak (Theorem 5.5) concerning periodic horocycle orbits. At first sight, one might only ask this out of curiosity

[†] Again this is also called *homogeneous*.

or to satisfy the urge to complete our understanding of this aspect of these dynamical systems. However, this line of enquiry turns out to be useful for applications to number-theoretic problems. A satisfying answer to this question is given by Mozes and Shah [**?**].

Theorem 6.4 (Mozes–Shah equidistribution theorem). † *Let $X = \Gamma\backslash G$ with G a connected Lie group and Γ a lattice, and let $H_n < G$ be a sequence of subgroups generated by unipotent one-parameter subgroups. Let μ_n be an invariant ergodic probability measure for the action of H_n for all $n \geqslant 1$. Assume that‡ $\mu_n \to \mu$ in the weak*-topology as $n \to \infty$. Then one of the following two possibilities holds.*

(1) *$\mu = 0$, and $\operatorname{Supp}\mu_n \to \infty$ as $n \to \infty$ in the sense that for every compact set $K \subseteq X$ there is an N with $\operatorname{Supp}\mu_n \cap K = \varnothing$ for $n \geqslant N$.*
(2) *$\mu = m_{L\bullet y}$ is the L-invariant probability measure on a closed finite volume orbit $L\bullet y$ for the closed connected group $L = \operatorname{Stab}_G(\mu)^\circ \leqslant G$. Moreover, μ is invariant and ergodic for the action of a one-parameter unipotent subgroup. Furthermore, suppose that $x_n = \varepsilon_n\bullet x \in \operatorname{Supp}\mu_n$ for $n \geqslant 1$ and some $x \in X$ with $\varepsilon_n \to 1$ as $n \to \infty$, and suppose the connected subgroups (L_n) satisfy $\mu_n = m_{L_n\bullet x_n}$ for $n \geqslant 1$. Then $xL = yL = \operatorname{Supp}\mu$ and there exists some N with $\varepsilon_n^{-1}L_n\varepsilon_n \subseteq L$ for $n \geqslant N$.*

The additional information in each case is useful in applying this theorem. According to (1), once we know that for every measure μ_n there exists some point $x_n \in \operatorname{Supp}\mu_n$ within a fixed compact set, the limit measure is a probability measure.

In (2), if we know that $H_n = H$ for all $n \geqslant 1$, then L has to contain H and the conjugates $\varepsilon_n^{-1}H\varepsilon_n$ as in (2). Together this often puts severe limitations on the possibilities that $L \leqslant G$ can take, and sometimes forces L to be G. This situtation arises, for example, if we study long periodic horocycle orbits, or orbits of a maximal subgroup $H < G$. In any case, the final claim of (2) says that the convergence to the limit measure $m_{L\bullet x}$ is almost from within the orbit $L\bullet x$. In fact, after modifying the measures in the sequence only slightly by the elements ε_n we get

$$\operatorname{Supp}\left((\varepsilon_n)_*^{-1}\mu_n\right) = \varepsilon_n^{-1}L_n\bullet x_n = \varepsilon_n^{-1}L_n\varepsilon_n\bullet x \subseteq L\bullet x = L\bullet y = \operatorname{Supp}\mu$$

for $n \geqslant N$.

† This version differs from the theorem in the paper, but should follow from it. Awaiting a decision: will it be proven here from scratch or using their theorem?

‡ By Tychonoff-Alaoglu there always exists a subsequence that converges.

6.2 Rationality Questions

A natural question is to ask which subgroups $L < G$ appear for a certain choice of one-parameter unipotent subgroup $U < G$ and $x \in X = \Gamma\backslash G$. In this section we explain how this kind of question is intimately related to questions of rationality.

This relationship is elementary in the abelian case $G = \mathbb{R}^d$, $\Gamma = \mathbb{Z}^d$, and $U = \mathbb{R}v$ for some $v \in \mathbb{R}^d$. In this case L is independent of

$$x \in X = \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$$

(and one should only expect this independence for abelian Lie groups). Moreover, L is the smallest subspace of $\mathbb{R}^d$ that can be defined by rational linear equations and contains $U = \mathbb{R}v$. This claim follows quickly from the special case where no such $L \neq \mathbb{R}^d$ exists. Under this assumption, $\{tv \mid t \in \mathbb{R}\}$ is equidistributed, as may be shown for example by integrating the characters of $\mathbb{T}^d$.

To start to see the possibilities in the general case, consider the special case

$$U = \left\{ \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\} < \mathrm{SL}_2(\mathbb{R})$$

and $2 = \mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$, which we already understand in some detail (see Section 1.2, Chapter 5, and [**?**, Sec. 11.7]). If $x = \Gamma g$ for some

$$g = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix},$$

then $L = U$, and otherwise $L = \mathrm{SL}_2(\mathbb{R})$. In order to be able to phrase this in terms of a rationality question, notice that $x \in X$ determines a geodesic in the upper half-plane (where we choose for example the base point in our fundamental domain, as illustrated in Figure 6.1). Then $L = U$ if the forward end point of the geodesic $\alpha \in \mathbb{R} \cup \{\infty\}$ is rational, meaning $\alpha \in \mathbb{Q} \cup \{\infty\}$, and $L = \mathrm{SL}_2(\mathbb{R})$ otherwise. This dichotomy is independent of the chosen representative within the orbit $\mathrm{SL}_2(\mathbb{Z})\boldsymbol{\cdot}(z, v)$.

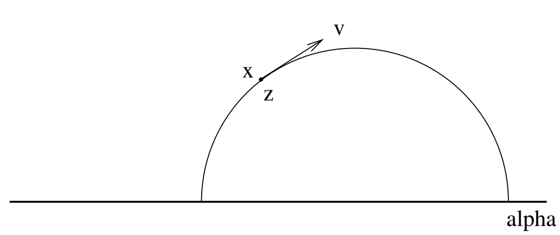

Fig. 6.1 The geodesic determined by x.

In general the answer is given by the following result found by Borel and Prasad [**?**]. A more general version of this result was obtained more recently by Tomanov [**?**].

Theorem 6.5. *Let $d = \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})$, $x = \Gamma g \in X$, and $U < G$ a one-parameter unipotent subgroup (or $H < G$ a closed subgroup generated by one-parameter unipotent subgroups). Then the group L appearing in Theorems* 6.1 *and* 6.2 *(respectively Theorem* 6.3*) is the connected component of $g^{-1}\mathbb{F}(\mathbb{R})g$, where $\mathbb{F}(\mathbb{R})$ is the group of $\mathbb{R}$-points of the smallest algebraic group $\mathbb{F} \leqslant \mathrm{SL}_d$ defined over $\mathbb{Q}$ for which $g^{-1}\mathbb{F}(\mathbb{R})g$ contains U (respectively H). Similarly, the group L in Theorem* 6.4 *is the connected component of $g^{-1}\mathbb{F}(\mathbb{R})g$ where $x = \Gamma g$ and $\mathbb{F}$ is the smallest algebraic group $\mathbb{F} \leqslant \mathrm{SL}_d$ defined over $\mathbb{Q}$ for which $g^{-1}\mathbb{F}(\mathbb{R})g$ contains $\varepsilon_n^{-1}L_n\varepsilon_n$ for $n \geqslant N$, where N is as in Theorem* 6.4.

For this result, one needs some understanding of the mechanisms that make orbits $\mathrm{SL}_d(\mathbb{Z})\mathbb{F}(\mathbb{R})$ of $\mathbb{Q}$-groups closed or not closed, and the Borel density theorem. In this setting of $\Gamma = \mathrm{SL}_d(\mathbb{Z}) < G = \mathrm{SL}_d(\mathbb{R})$, which contains *all* other arithmetic quotients even over number fields if we allow d to vary, the connection to algebraic group theory described above puts additional constraints on the possible structure of the subgroup L.

For instance, the algebraic group $\mathbb{F}$ over $\mathbb{Q}$ must have the property that the radical of $\mathbb{F}$ is equal to the unipotent radical of $\mathbb{F}$. In the language of Lie groups this implies that the radical of L, which by definition is only solvable, is nilpotent. Another restriction is, for example, that L cannot be isomorphic to $\mathrm{PSL}_2(\mathbb{R}) \times \mathrm{SO}(5)(\mathbb{R})$. This is because the unipotent group has to be contained in $\mathrm{PSL}_2(\mathbb{R})$ and the induced lattice $L \cap g^{-1}\mathrm{SL}_d(\mathbb{Z})g$ cannot give an irreducible lattice in $\mathrm{PSL}_2(\mathbb{R}) \times \mathrm{SO}(5)(\mathbb{R})$ as the direct factors are simple groups of different types in the classification of complex Lie algebras and they cannot be exchanged by a Galois action. On the other hand $L = \mathrm{PSL}_2(\mathbb{R}) \times \mathrm{SO}(3)(\mathbb{R})$ *is* a possibility since $\mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{SO}(2,1)(\mathbb{R})^o$, and a simple switch in the sign of the quadratic forms (via a Galois automorphism) can interchange these groups.

6.3 First Ideas in Unipotent Dynamics

The structure of proof of Theorem 6.1 is to study

$$\mathrm{Stab}(\mu) = \{g \in G \mid g_*\mu = \mu\}$$

and to show that the measure μ on $X = \Gamma\backslash G$ is supported on a single orbit of this subgroup. This is achieved indirectly; if μ is not supported on a single orbit of a particular subgroup $H < G$ that leaves the measure invariant then one shows that the subgroup can be enlarged to some $H' > H$ so that the new subgroup H' also preserves μ.

We also note that, in the setting of Theorem 6.1, once we have shown that μ is supported on a single orbit of $\mathrm{Stab}(\mu)$, we actually obtain that μ is supported on a single closed orbit of $\mathrm{Stab}(\mu)^\circ$.

Lemma 6.6. *Let $X = \Gamma\backslash G$ be a quotient of a Lie group by a discrete subgroup Γ. Let H be a connected subgroup of G and let μ be an H-invariant and ergodic probability measure. If μ gives full measure to a single orbit of its stabilizer subgroup $\mathrm{Stab}(\mu)$, then μ is the Haar measure on a closed orbit of the subgroup $\mathrm{Stab}(\mu)^\circ$.*

PROOF. Suppose that μ is the Haar measure on $\mathrm{Stab}(\mu)\boldsymbol{\cdot}x_0$ so that this orbit has finite volume. Since the Haar measure on $\mathrm{Stab}(\mu)^\circ$ is simply the restriction of the Haar measure on $\mathrm{Stab}(\mu)$ to $\mathrm{Stab}(\mu)^\circ$ and a fundamental domain for the orbit map for $\mathrm{Stab}(\mu)^\circ$ is an injective domain for the orbit map for $\mathrm{Stab}(\mu)$, we obtain that $\mathrm{Stab}(\mu)^\circ\boldsymbol{\cdot}x_0$ also has finite volume. Since H is connected, $H \subseteq \mathrm{Stab}(\mu)^\circ$ and so $\mathrm{Stab}(\mu)^\circ\boldsymbol{\cdot}x_0$ is a H-invariant subset of positive measure. Hence μ is the $\mathrm{Stab}(\mu)^\circ$-invariant Haar measure on $\mathrm{Stab}(\mu)^\circ\boldsymbol{\cdot}x_0 = \mathrm{Stab}(\mu)\boldsymbol{\cdot}x_0$. Finally by Proposition 1.12 this orbit is also closed. □

6.3.1 Generic Points

We present in this section the basic idea for using generic points to show an 'additional invariance', which in a more specialized context goes back to work of Furstenberg on the unique ergodicity of skew product extensions, leading to the equidistribution of the fractional parts of the sequence $(n^2\alpha)_{n\geqslant 1}$ for α irrational[(25)].

Recall that $x \in X$ is said to be *generic* with respect to μ and a one-parameter flow $\{u_s : s \in \mathbb{R}\}$ if

$$\frac{1}{T}\int_0^T f(u_s\boldsymbol{\cdot}x)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $T \to \infty$ for all $f \in C_c(X)$. Using the pointwise ergodic theorem [**?**, Cor. 8.15] and separability of $C_0(X)$ one can easily show that μ-almost every point is generic if only μ is invariant and ergodic under the one-parameter flow

$$U = \{u_s \mid s \in \mathbb{R}\}$$

(see Lemma 6.10).

Lemma 6.7 (Centralizer Lemma). *If $x, y = h\boldsymbol{\cdot}x \in X$ are generic for μ and $h \in C_G(U) = \{g \in G \mid gu = ug \text{ for all } u \in U\}$, then h preserves μ.*

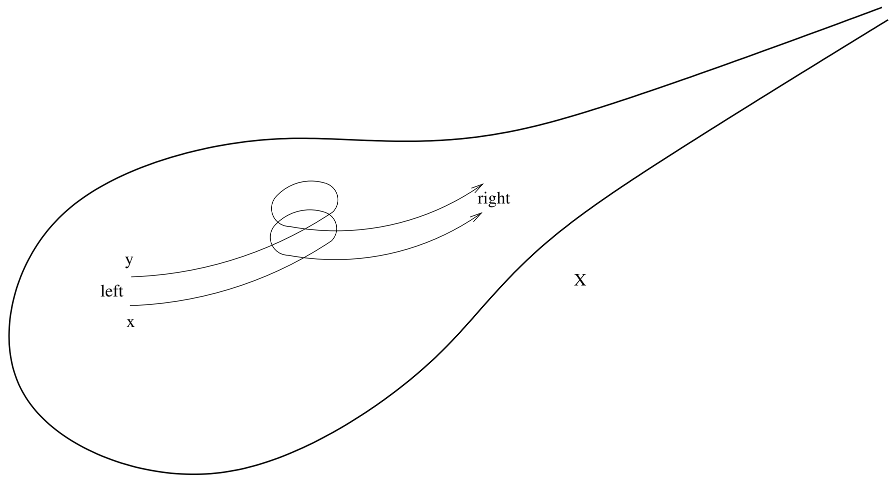

Fig. 6.2 If $y = xh^{-1}$ with $h \in C_G(V)$, then the two orbits are parallel. If in addition both x and y are generic, then the orbits equidistribute that is, approximate μ, which gives Lemma 6.7.

We refer to Figure 6.2 for the proof of Lemma 6.7.

Proof of Lemma 6.7. We know that

$$\frac{1}{T}\int_0^T f(u_s{\scriptstyle\bullet}y)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

for any $f \in C_c(X)$. On the other hand

$$\begin{aligned}\frac{1}{T}\int_0^T f(u_s{\scriptstyle\bullet}y)\,\mathrm{d}s &= \frac{1}{T}\int_0^T f(u_s{\scriptstyle\bullet}(h{\scriptstyle\bullet}x))\,\mathrm{d}s\\ &= \frac{1}{T}\int_0^T \underbrace{f(h{\scriptstyle\bullet}(u_s{\scriptstyle\bullet}x))}_{f^h(u_s\bullet x)}\,\mathrm{d}s \qquad \text{(since } h \in C_G(\{u_s\})\text{)}\\ &\longrightarrow \int_X f^h\,\mathrm{d}\mu = \int_X f(h{\scriptstyle\bullet}z)\,\mathrm{d}\mu\end{aligned}$$

so μ is h-invariant. □

Lemma 6.7 seems (and is) useful, but it can only be applied in very special circumstances as the centralizer is usually very small and we would need to be fortunate to find two generic points bearing such a special relation to each other.

6.3.2 Polynomial divergence leading to invariance

A much more useful observation, due to Ratner, that leads to additional invariance in more circumstances, is the following observation[†] which is based on the polynomial divergence property of unipotent flows. In fact, as we have seen before, the action of an element $u \in G$ on $\Gamma\backslash G$ is locally described by conjugation and hence can also be described by the adjoint representation of u on the Lie algebra $\mathfrak{g}$ of G. More precisely, if $y = \varepsilon \centerdot x$ is close to x, and $\varepsilon \in G$ is the local displacement between x and y, then $u \centerdot y = u \centerdot \varepsilon \centerdot x = u\varepsilon u^{-1} \centerdot (u \centerdot x)$ and so a displacement between $u \centerdot x$ and $u \centerdot y$ is given by the conjugated element $u\varepsilon u^{-1}$. If the displacement ε was not small enough, then $u\varepsilon u^{-1}$ may not be the smallest displacement between $u \centerdot x$ and $u \centerdot y$. However, if ε is very small, then the calculation leading to the conjugated element as the displacement may be iterated several times. Thus, in order to compare the orbit of points close to x to the orbit of x we will need to study conjugation by u (or equivalently its adjoint representation on the Lie algebra).

If $\{u(t) \mid t \in \mathbb{R}\}$ is a unipotent one-parameter subgroup of G, then $\mathrm{Ad}_{u(t)}$ is unipotent for all $t \in \mathbb{R}$ also, and is a (matrix-valued) polynomial in t. This polynomial structure (as opposed to exponential) of unipotent subgroups has the following consequence. Given a nearby pair of points x and $y = \varepsilon \centerdot x$, let $v = \log \varepsilon$ and consider the $\mathfrak{g}$-valued polynomial $\mathrm{Ad}_{u(t)}(v)$. For very small values of ε, this polynomial is close to zero in the space of all polynomials. However, if we choose a large 'speeding up' parameter T then we may consider the polynomial

$$p(r) = \mathrm{Ad}_{u(rT)}(v)$$

in the rescaled variable $r \in \mathbb{R}$. Assuming the original polynomial is nonconstant (equivalently, ε does not lie in $C_G(\{u(s)\})$), we can choose T precisely so that the polynomial p above in the variable r belongs to a compact set of polynomials not containing the zero polynomial. In fact, if $T > 0$ is the smallest number with[‡] $\| \mathrm{Ad}_{u(T)}(v)\| = 1$, then

$$\sup_{r \in [0,1]} \|p(r)\| = 1.$$

Moreover, p is a polynomial of bounded degree. Notice that this feature — that this acceleration or renormalization of a polynomial is again a polynomial from the same finite-dimensional space — is specific[§] to polynomials and hence to unipotent flows.

[†] This is often called the H-principle. Our presentation of the idea will be closer to the work of Margulis and Tomanov [?].

[‡] It does not matter which norm on $\mathfrak{g}$ is used; for concreteness we use the norm derived from the Riemannian metric.

[§] In contrast, diagonalizable flows leading in the same way to exponential maps do not have this property, as the acceleration would change the base of the exponential functions involved.

In order to state the principle that gives additional invariance, we will need the following refinement of the notion of genericity.

Definition 6.8. A set $K \subseteq X$ is called a set of *uniformly generic points* if for any $f \in C_c(X)$ and $\varepsilon > 0$ there is some $T_0 = T_0(f, \varepsilon)$ with

$$\left| \frac{1}{T} \int_0^T f(u_s \bullet x) \, \mathrm{d}s - \int_X f \, \mathrm{d}\mu \right| < \varepsilon$$

for all $T \geqslant T_0$ and all $x \in K$.

Proposition 6.9 (Polynomial divergence leads to invariance). *Suppose that $(x_n), (y_n)$ are two sequences of uniformly generic points with*

$$x_n \to z, y_n \to z$$

and $y_n = \varepsilon_n \bullet x_n$ with $\varepsilon_n \to I$ as $n \to \infty$ and $\varepsilon_n \notin C_G(U)$ for $n \geqslant 1$. Define $v_n = \log \varepsilon_n$ and polynomials

$$p_n(r) = \mathrm{Ad}_{u(T_n r)}(v_n),$$

where the speeding up parameter $T_n \to \infty$ is chosen so that

$$\sup_{r \in [0,1]} \|p_n(r)\| = 1$$

for each $n \geqslant 1$. Suppose that $p_n(r) \to p(r)$ as $n \to \infty$ for all $r \in [0, 1]$, where

$$p : \mathbb{R} \to \mathfrak{g}$$

is a polynomial with entries in the Lie algebra $\mathfrak{g}$. Then μ is preserved by $\exp(p(r))$ for all $r \in \mathbb{R}_{\geqslant 0}$.

Notice that the assumption that the sequence of polynomials converges is a mild one. The polynomials all lie in a compact subset of a finite-dimensional space, so there is a subsequence that converges with respect to any norm on that space. Also the assumption $\varepsilon_n \notin C_G(U)$ is somewhat unproblematic as in the case $\varepsilon_n \in C_G(U)$ one may be able to apply Lemma 6.7. Part of the argument for Proposition 6.9 is illustrates in Figure 6.3.

Proof of Proposition 6.9. Fix $r_0 \in \mathbb{R}_{>0}$, $f \in C_c(X)$, and $\varepsilon > 0$. By uniform continuity of f there exists some $\delta = \delta(f, \varepsilon) > 0$ with

$$\mathsf{d}(h_1, h_2) < \delta \implies |f(h_1 \bullet x) - f(h_2 \bullet x)| < \varepsilon$$

for all $x \in X$. Furthermore, choose $\kappa > 0$ so that

$$\mathsf{d}(\exp p(r), \exp p(r_0)) < \delta/2$$

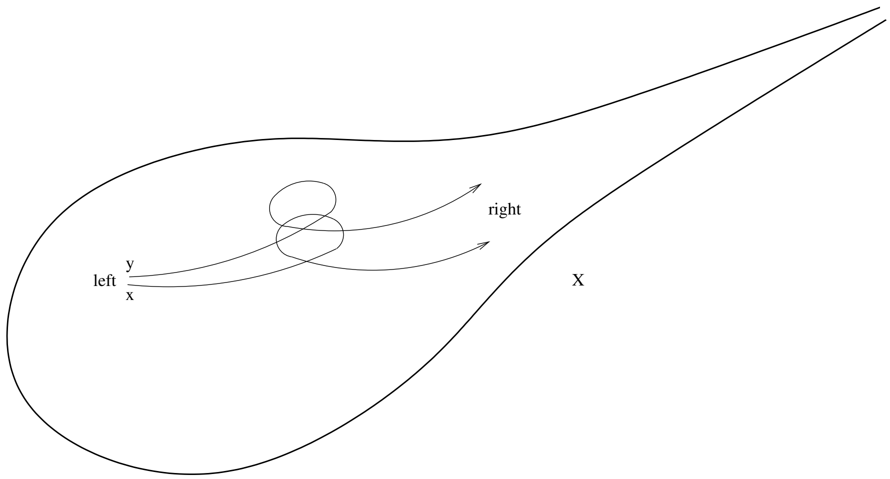

Fig. 6.3 If $y = x\varepsilon^{-1}$ with $\varepsilon \notin C_G(V)$ close to the identity, then the orbits of x and y move away from each other at polynomial speed. If x and y are generic then the last 1% of these pieces of orbits are almost parallel and equidistribute.

for $r \in [r_0 - \kappa, r_0]$. Then there is an N such that we also have[†]

$$\mathsf{d}(\exp p_n(r), \exp p(r_0)) < \delta \tag{6.1}$$

for $n \geqslant N$ and $r \in [r_0 - \kappa, r_0]$. We know by the uniform genericity of x_n that

$$\frac{1}{r_0 T_n} \int_0^{r_0 T_n} f(u_s \centerdot x_n)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $n \to \infty$, and

$$\frac{1}{(r_0-\kappa) T_n} \int_0^{(r_0-\kappa) T_n} f(u_s \centerdot x_n)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $n \to \infty$. Taking the correct linear combination ($\kappa > 0$ is fixed) and replacing f by $f^{\exp(p(r_0))}$, we get[‡]

$$\frac{1}{\kappa T_n} \int_{(r_0-\kappa) T_n}^{r_0 T_n} f^{\exp p(r_0)}(u_s \centerdot x_n)\,\mathrm{d}s \longrightarrow \int_X f^{\exp p(r_0)}\,\mathrm{d}\mu$$

and by the same argument we also have

[†] This is the formal version of the statement in Figure 6.3 that the last 1% are parallel.

[‡] In Figure 6.3 we referred to this as the equidistribution of the last 1% of the orbit.

$$\frac{1}{\kappa T_n}\int_{(r_0-\kappa)T_n}^{r_0T_n} f(u_s\boldsymbol{\cdot} y_n)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu.$$

However, using the definition of v_n and p_n we have

$$\begin{aligned} u_s\boldsymbol{\cdot} y_n &= u_s\exp(v_n)\boldsymbol{\cdot} x_n\\ &= \exp\left(\mathrm{Ad}_{u_s}(v_n)\right)u_s\boldsymbol{\cdot} x_n\\ &= \exp\left(p_n(s/T_n)\right)u_s\boldsymbol{\cdot} x_n \end{aligned}$$

for all $x \in \mathbb{R}$.

We now restrict ourself to the range of $s \in \mathbb{R}$ with $\frac{s}{T_n} \in [r_0-\kappa, r_0]$. Together with (6.1), we deduce that

$$\mathsf{d}(u_s\boldsymbol{\cdot} y_n, \exp p(r_0)u_s\boldsymbol{\cdot} x_n) < \delta,$$

and so

$$|f(u_s\boldsymbol{\cdot} y_n) - f(\exp p(r_0)u_s\boldsymbol{\cdot} x_n)| < \varepsilon$$

for every $s \in [(r_0-\kappa)T_n, r_0T_n]$. Using this estimate in the integrals above gives

$$\left|\frac{1}{\kappa T_n}\int_{(r_0-\kappa)T_n}^{r_0T_n} f^{\exp p(r_0)}(u_s\boldsymbol{\cdot} x_n)\,\mathrm{d}s - \frac{1}{\kappa T_n}\int_{(r_0-\kappa)T_n}^{r_0T_n} f(u_s\boldsymbol{\cdot} y_n)\,\mathrm{d}s\right| < \varepsilon,$$

and so

$$\left|\int_X f^{\exp p(r_0)}\,\mathrm{d}\mu - \int_X f\,\mathrm{d}\mu\right| \leqslant \varepsilon.$$

Since this holds for any $\varepsilon > 0$ and $f \in C_c(X)$ we deduce that μ is invariant under $\exp p(r_0)$. As $r_0 > 0$ was arbitrary, the proposition follows. □

Because of the results above, we are interested in finding large sets of uniformly generic points. It is too much to expect that almost every point with respect to an invariant measure will have this property (due to the requested uniformity), but we can get close to this statement as follows.

Lemma 6.10 (Almost full sets consisting of uniformly generic points). *Let μ be an invariant and ergodic probability measure on X for the action of a one-parameter flow $\{u_s : s \in \mathbb{R}\}$. For any $\rho > 0$ there is a compact set $K \subseteq X$ with $\mu(K) > 1-\rho$ consisting of uniformly generic points.*

PROOF. Let $D = \{f_1, f_2, \dots\} \subseteq C_c(X)$ be countable and dense. Then by the pointwise ergodic theorem [**?**, Cor. 8.15] for every $f_\ell \in D$ we have

$$\frac{1}{T}\int_0^T f_\ell(u_s\boldsymbol{\cdot} x)\,\mathrm{d}s \longrightarrow \int_X f_\ell\,\mathrm{d}\mu$$

μ-almost everywhere, or equivalently for every $\varepsilon > 0$

$$\mu\left(\left\{x\in X\,\middle|\,\sup_{T>T_0}\left|\frac{1}{T}\int_0^T f_\ell(u_s\bullet x)\,\mathrm{d}s-\int_X f_\ell\,\mathrm{d}\mu\right|>\varepsilon\right\}\right)\longrightarrow 0$$

as $T_0\to\infty$. Now choose, for every $f_\ell\in D$ and for every $\varepsilon=\frac{1}{n}$ a time $T_{\ell,n}$ so that

$$\mu\left(\left\{x\in X\,\middle|\,\sup_{T>T_{\ell,n}}\left|\frac{1}{T}\int_0^T f_\ell(u_s\bullet x)\,\mathrm{d}s-\int_X f_\ell\,\mathrm{d}\mu\right|>\frac{1}{n}\right\}\right)<\frac{\rho}{2^{\ell+n}}.$$

Let $K'\subseteq X$ be the complement of the union of these sets, so that $\mu(K')>1-\rho$ by construction. It is clear that the points in K' are uniformly generic for all $f\in D$. Moreover, since $D\subseteq C_c(X)$ is dense in the uniform norm, this extends to all functions by a simple approximation argument. Finally we may choose a compact $K\subseteq K'$ with $\mu(K)>1-\rho$ by regularity of μ. □

The principle outlined above is sufficient to prove the measure classification theorem for 2-step nilpotent groups (see Exercise 6.3.2; as we will see in the next section with more effort the same holds for more general nilpotent groups). However, in general this use is limited — for example, in the above form it does not even allow us to give a new proof of measure classification for the horocycle flow. This will be discussed again in Section 6.6, where we discuss the second more powerful refinement of the use of generic points to show additional invariance leading to a strengthening of Dani's theorem (Theorem 5.3) due to Ratner.

Exercises for Section 6.3

Exercise 6.3.1. Show that the limit polynomial in Proposition 6.9 takes only values in the centralizer $C_{\mathfrak{g}}(U)=\{v\in\mathfrak{g}\mid \mathrm{Ad}_u(v)=v \text{ for all } u\in U\}$ of U in the Lie algebra $\mathfrak{g}$ of G.

Exercise 6.3.2. Use the results from Section 6.3.2 to prove the measure classification theorem (Theorem 6.1) under the assumption that G is a 2-step nilpotent group.

6.4 Unipotent Dynamics on Nilmanifolds

In this section we will assume that G is a nilpotent Lie group and $\Gamma<G$ a discrete subgroup. In this case $X=\Gamma\backslash G$ is called a *nilmanifold*.

6.4.1 Measure Classification for Nilmanifolds

Theorem 6.11. *Let $\Gamma < G$ be a discrete subgroup of a connected nilpotent Lie group G and let $X = \Gamma\backslash G$. Let $U \leqslant G$ be a one-parameter subgroup. Then any U-invariant and ergodic probability measure μ on G is algebraic.*

PROOF. As we will see, the result follows from a (double) induction argument and Proposition 6.9. First, notice that the theorem is trivial if $\dim G = 1$.

A second special case is obtained by assuming in addition that U belongs to the center C_G of G. In this case, if

$$X' = \{x \in X \mid x \text{ is generic for } \mu\},$$

$x_0 \in X'$, and $y = g\boldsymbol{\cdot}x_0 \in X'$, then $g \in C_G(U) = G$ so $g \in \operatorname{Stab}_G(\mu)$ by Lemma 6.7 and $y \in \operatorname{Stab}_G(\mu)\boldsymbol{\cdot}x_0$ also. It follows that $X' \subseteq \operatorname{Stab}_G(\mu)\boldsymbol{\cdot}x_0$ has full measure, and we deduce that μ must be the Haar measure on $\operatorname{Stab}_G(\mu)\boldsymbol{\cdot}x_0$ as required.

We assume now that G is a nilpotent connected Lie group of nilpotency degree k, meaning that

$$G_0 = G \geqslant G_1 = [G, G_0] \geqslant \cdots \geqslant G_{k-1} = [G, G_{k-2}] \geqslant G_k = [G, G_{k-1}] = \{I\}.$$

We also assume that $U \leqslant G_j$ for some $j \in \{0, \dots, k-1\}$. We may also assume that $U \nsubseteq C_G$. The inductive hypothesis is then the following statement: the theorem holds for any $X' = \Gamma'\backslash G'$, $U' \leqslant G'$ and any U'-invariant and ergodic probability measure μ' if either

- $\dim G' < \dim G$, or
- $G' = G$, $\Gamma' = \Gamma$, and $U' \leqslant G_{j+1}$.

Now let $K \subseteq X$ be a set of uniformly generic points of measure $\mu(K) > 0.9$ as in Lemma 6.10. Choose some

$$x_0 \in K \cap \operatorname{Supp}(\mu|_K). \tag{6.2}$$

We distinguish between two possible scenarios.

It could be that there is some $\delta > 0$ such that $y = \varepsilon\boldsymbol{\cdot}x_0 \in K$ with

$$\mathsf{d}(\varepsilon, I) < \delta$$

implies that $\varepsilon \in C_G(U)$. In this case (6.2) implies that the U-invariant set $C_G(U)\boldsymbol{\cdot}x_0$ has positive measure, and so by ergodicity we have

$$\mu(C_G(U)\boldsymbol{\cdot}x_0) = 1.$$

Now set $G' = C_G(U) \lneqq G$ and $\Gamma' = \operatorname{Stab}_{G'}(x_0)$ so that we may also consider μ as a U-invariant probability measure on

$$\Gamma'\backslash G' \cong G'\centerdot x_0$$

(with $\Gamma' g'$ corresponding to $x_0 g'$). By the above the theorem follows in this case.

In the second case we find a sequence $(y_n = \varepsilon_n\centerdot x_0)$ in K with $\varepsilon_n \to e$ as $n \to \infty$ but $\varepsilon_n \notin C_G(U)$ for all $n \geqslant 1$. Choosing a subsequence, we may assume that the sequence of polynomials $(p_n(r))$ from Proposition 6.9 converges to a non-constant polynomial $p : \mathbb{R} \to \mathfrak{g}$. By Proposition 6.9 we deduce that μ is preserved by $\exp(p(r))$ for all $r \geqslant 0$.

We claim that $\exp(p(r))$ takes values in G_{j+1}. Indeed, since $U \subseteq G_j$ we have (in the notation of Proposition 6.9)

$$p_n(r) = \mathrm{Ad}_{u(T_n r)}(\log \varepsilon_n) \in \log \varepsilon_n + \mathfrak{g}_{j+1}$$

for all r, where

$$\mathfrak{g}_{j+1} = \mathrm{Lie}\, G_{j+1} = [\mathfrak{g}, \mathfrak{g}_j].$$

Since $\varepsilon_n \to e$ as $n \to \infty$ this gives $p(r) \in \mathfrak{g}_{j+1}$ for all $r \geqslant 0$ as claimed.

The argument above shows that

$$(\mathrm{Stab}_G(\mu) \cap G_{j+1})^o$$

is a non-trivial subgroup. Clearly U normalizes this subgroup, and since $\mathrm{Ad}_{u(t)}$ is unipotent for all $t \in \mathbb{R}$, it follows that there exists a one-parameter unipotent subgroup

$$U' = \{u'_t \mid t \in \mathbb{R}\} \leqslant \mathrm{Stab}_G(\mu) \cap G_{j+1} \cap C_G(U).$$

We are going to apply the inductive hypothesis to $G' = G$, $\Gamma' = \Gamma$, and U'. However, as μ may not be[†] ergodic with respect to U' we first have to decompose μ into U'-ergodic components. Recall from [**?**, Th. 6.2, 8.20] that the ergodic decomposition allows us to write

$$\mu = \int_X \mu_x^{\mathscr{E}'}\, \mathrm{d}\mu, \tag{6.3}$$

where $\mu_x^{\mathscr{E}'}$ is the conditional measure for the σ-algebra

$$\mathscr{E}' = \{B \in \mathscr{B}_X \mid \mu(u'_t\centerdot B \triangle B) = 0 \text{ for all } t\}$$

and that for μ-almost every x the conditional measure $\mu_x^{\mathscr{E}'}$ is a U'-invariant and ergodic probability measure on X with $x \in \mathrm{Supp}\, \mu_x^{\mathscr{E}'}$.

By applying the inductive hypothesis to μ-almost every $\mu_x^{\mathscr{E}'}$ we obtain a function $x \mapsto L_x$ that assigns to x the connected subgroup L_x for which $\mu_x^{\mathscr{E}'}$ is the L_x-invariant probability measure on the closed orbit $L_x\centerdot x$. We claim that there is a connected subgroup L such that $L_x = L$ for μ-almost every x.

[†] In fact U' never acts ergodically with respect to μ.

Indeed, since $U = \{u(t) \mid t \in \mathbb{R}\}$ preserves μ and leaves the σ-algebra $\mathscr{E}'$ invariant (since U' and U commute) we get

$$(u_t)_*\mu_x^{\mathscr{E}'} = \mu_{u_t\boldsymbol{\cdot}x}^{\mathscr{E}'} \tag{6.4}$$

for every $t \in \mathbb{R}$ and μ-almost every x by [**?**, Cor. 5.24]. Since $\mu_x^{\mathscr{E}'}$ is L_x-invariant, it follows from (6.4) that $(u_1)_*\mu_x^{\mathscr{E}'}$ is $u_1L_xu_1^{-1}$-invariant, which implies that

$$u_1L_xu_1^{-1} \subseteq L_{u(1)\boldsymbol{\cdot}x}$$

and, by a similar argument for the reverse inclusion,

$$u_1L_xu_1^{-1} = L_{u_1\boldsymbol{\cdot}x}.$$

Iterating this relationship shows that

$$u_1^nL_xu_1^{-n} = L_{u(n)\boldsymbol{\cdot}x} \tag{6.5}$$

for μ-a.e. x Now either L is normalized by u_1, or the sequence of subgroups in (6.5) converges to a subgroup that is normalized by u_1 (to see this, apply the argument from the proof of Lemma 3.31 to any element of $\bigwedge^{\dim L_x}(\operatorname{Lie} L_x)$. Hence Poincaré recurrence shows that we must have

$$u_1L_xu_1^{-1} = L_x$$

for μ-almost every x. Notice that for any such x we also get

$$u(t)L_xu(t)^{-1} = L_x$$

for all $t \in \mathbb{R}$. By ergodicity it follows that $L_x = L$ is constant μ-almost everywhere. The cautious reader will have noticed that the argument above has assumed implicitly that the function $x \mapsto L_x$ is measurable, which we will show in Lemma 6.12 below. Equation (6.3) now shows that μ is a convex combination of L-invariant measures and hence is itself L-invariant.

To summarize, we have shown that there exists a non-trivial connected subgroup $L \leqslant \operatorname{Stab}_G(\mu)$ containing U' such that the orbit $L\boldsymbol{\cdot}x$ is for μ-almost every x closed, with finite L-invariant measure and with the property that $U' \leqslant L$ acts ergodically on $L\boldsymbol{\cdot}x$. Since $L \leqslant G$ is nilpotent, simply connected and connected, $M = C_L(L)$ is a non-trivial connected subgroup. We claim that the orbit $M\boldsymbol{\cdot}x$ is compact for μ-almost every x and postpone the proof to Lemma 6.14.

Next we claim that $N_G^1(M)\boldsymbol{\cdot}x$ is a closed orbit for μ-a.e. x, see Lemma 6.15. This implies that μ is supported on a single orbit $x_0N_G^1(M)$ of the unimodular normalizer. In fact we note first that

$$U \leqslant N_G(L) \leqslant N_G(M),$$

and since U is unipotent we also have $U \leqslant N_G^1(M)$. If now x_0 is generic for μ and U, then

$$\operatorname{Supp}\mu = \overline{x_0 U} \subseteq x_0 N_G^1(M).$$

Therefore, without loss of generality we may assume $x_0 = \Gamma$, $G = N_G^1(M)^o$ and hence $M \lhd G$ and that the orbit ΓM is compact.

Let $\pi_M : G \to G/M$ denote the canonical projection $\pi_M(g) = gM$. We claim that $\pi_M(\Gamma) \leqslant G/M$ is again discrete. Suppose that

$$\pi_M(\gamma_n) \longrightarrow I$$

in G/M as $n \to \infty$ with $\gamma_n \in \Gamma$, or equivalently $\gamma_n m_n \to I$ as $n \to \infty$ in G for $\gamma_n \in \Gamma$ and $m_n \in M$ for all $n \geqslant 1$. Since $M \cap \Gamma$ is co-compact in M, we may simultaneously modify γ_n and m_n by elements of $M \cap \Gamma$ and assume that m_n lies in a pre-compact fundamental domain for Γ for all $n \geqslant 1$. Choosing a subsequence, we may also now assume that $m_n \to m \in M$ as $n \to \infty$. This implies that $\gamma_n \to \gamma \in \Gamma$ as $n \to \infty$ for some γ, and so $\gamma_n = \gamma$ for all large $n \geqslant 1$. This shows that $\pi_M(\gamma_n) = \pi_M(\gamma) = I$ for large enough n, and hence that $\pi_M(\Gamma)$ is discrete.

There is also an associated factor map

$$\pi_X : \Gamma\backslash G \longrightarrow \pi_M(\Gamma)\backslash\pi_M(G)$$

defined by

$$\pi_X : \Gamma g \longmapsto \pi_M(\Gamma)\pi_M(g).$$

The fibers of this map are precisely the M-orbits in the sense that

$$\pi_X^{-1}\left(\pi_X(\Gamma g)\right) = \{\Gamma h \mid \pi_M(\Gamma)hM = \pi_M(\Gamma)gM\} = \Gamma g M$$

for all $g \in G$.

We set $G' = \pi_M(G)$, $\Gamma' = \pi_M(\Gamma)$, $U' = \pi_M(U)$, $\mu' = (\pi_X)_*\mu$ and deduce from the inductive hypothesis that μ' is an algebraic measure. Let $H' \leqslant G'$ be a connected subgroup, so that μ' is the H'-invariant probability measure on a finite volume orbit

$$\pi_M(\Gamma)\pi_M(g)H'$$

for some $\pi_M(g) \in G'$. Finally, we claim that μ is the H-invariant probability measure on the closed orbit $\Gamma g H$ where $H = \pi_M^{-1}(H')$.

Since $\pi_M(\Gamma)\pi_M(g)H'$ is closed we also obtain that

$$\Gamma g H = \pi_X^{-1}\left(\pi_M(\Gamma)\pi_M(g)H'\right)$$

is closed. Now let $f \in C(X)$. Then

$$\int_X f(x)\,\mathrm{d}\mu(x) = \int_X f(m\raisebox{0.2ex}{\scriptsize$\bullet$}x)\,\mathrm{d}\mu(x) \tag{6.6}$$

for all $m \in M$. Now take a Følner sequence (F_n) in M and notice that

$$\frac{1}{m_M(F_n)} \int_{F_n} f(m\bullet x)\,\mathrm{d}m_M(m) \longrightarrow \int_{M\bullet x} f(z)\,\mathrm{d}m_{M\bullet x}(z) = \overline{f}(\pi_X(x))$$

for all $x \in X$, where the expression on the right defines a function $\overline{f}$ in $C(\pi_X(X))$. Applying this convergence to the average of (6.6) over the Følner sequence gives

$$\int_X f(x)\,\mathrm{d}\mu(x) = \int_{\pi_X(X)} \underbrace{\int_{M\bullet x} f(z)\,\mathrm{d}m_{M\bullet x}(z)}_{\overline{f}(\pi_X(x))}\,\mathrm{d}\mu'$$

Now fix $h \in H$ and define f^h by $f^h(x) = f(h\bullet x)$ so that

$$\overline{f^h}(\pi_X(x)) = \int_{M\bullet x} f(h\bullet z)\,\mathrm{d}m_{M\bullet x}(z) = \int_{M\bullet(h\bullet x)} f(z)\,\mathrm{d}m_H(z) = \overline{f}(h\bullet\pi_X(x)),$$

and

$$\int_X f^h\,\mathrm{d}\mu = \int_{\pi_X(X)} \overline{f^h}\,\mathrm{d}\mu' = \int_{\pi_X(X)} \left(\overline{f}\right)^h\,\mathrm{d}\mu' = \int_{\pi_X(X)} \overline{f}\,\mathrm{d}\mu' = \int_X f\,\mathrm{d}\mu.$$

Therefore μ is supported on $H\bullet x$ and is H-invariant. This concludes the induction, and the theorem follows. □

In the course of the proof we made use of several lemmas which we now prove.

Lemma 6.12 (Measurability of stabilizer). *Let G be a Lie group, $\Gamma \leqslant G$ a discrete subgroup, and let $X = \Gamma\backslash G$. Then the map*

$$\mathscr{M}(X) \ni \mu \longmapsto \mathrm{Stab}_G(\mu)^o$$

from the space $\mathscr{M}(X)$ of Borel probability measures on X is measurable.

Implicit in the statement of the lemma is a measurable structure on the space of connected subgroups, and this is achieved as follows. We identify a connected subgroup $L \leqslant G$ with its Lie algebra $\mathrm{Lie}\,L$, and if $L \neq \{I\}$ with the corresponding point of the Grassmannian of G. In other words, we consider the map in the lemma as a map from $\mathscr{M}(X)$ to

$$\{e\} \sqcup \bigsqcup_{\ell=1}^{\dim G} {}_{\ell}(\mathrm{Lie}\,G),$$

which is a compact metric space and hence has a measurable structure via the Borel σ-algebra.

Proof of Lemma 6.12. Let $d = \dim G$, so that

$$\mathscr{M}_d = \{\mu \in \mathscr{M}(X) \mid \dim \operatorname{Stab}_G(\mu) = d\} = \{m_X\}$$

and $\mathscr{M}_d \ni \mu \longmapsto \operatorname{Stab}_G(\mu)^o$ is trivially measurable.

Fix k with $0 \leqslant k \leqslant d$ and suppose that we have already shown that the sets

$$\mathscr{M}_\ell = \{\mu \mid \dim \operatorname{Stab}_G(\mu) = \ell\}$$

for $\ell \geqslant k$ and the map $\mathscr{M}_{k+1} \ni \mu \longmapsto \operatorname{Stab}_G(\mu)^o$ are measurable.

Let $\mu_n \in \mathscr{M}_{\geqslant k} = \mathscr{M}_k \cup \cdots \cup \mathscr{M}_d$ for $n \geqslant 1$ and suppose that $\mu_n \to \nu$ in the weak*-topology as $n \to \infty$. Let $\mathfrak{h}_n$ be the Lie algebra of $\operatorname{Stab}_G(\mu)^o$. As

$$\{I\} \cup \bigcup_{1 \leqslant \ell \leqslant d} {}_{\ell}(\operatorname{Lie} G)$$

is compact, we may choose a subsequence and assume that $\mathfrak{h}_n \to \mathfrak{h} \leqslant \mathfrak{g}$ as $n \to \infty$ with $\dim \mathfrak{h} \geqslant k$. We will prove below that μ is invariant under $\exp(\mathfrak{h})$ and so $\mu \in \mathscr{M}_{\geqslant k}$. It follows that $\mathscr{M}_{\geqslant k}$ is closed and hence measurable, which implies that $\mathscr{M}_k = \mathscr{M}_{\geqslant k} \smallsetminus \mathscr{M}_{\geqslant k+1}$ is also measurable.

The argument above also shows that the assumption $\mu_n \in \mathscr{M}_k$ for all $n \geqslant 1$ and $\mu_n \to \mu \in \mathscr{M}_k$ as $n \to \infty$ implies that $\mathfrak{h}_n \to \mathfrak{h}$ as $n \to \infty$, with $\dim \mathfrak{h} = k$. Therefore

$$\mathscr{M}_k \ni \mu \longmapsto \operatorname{Stab}_G(\mu)^o$$

is actually continuous on the measurable set $\mathscr{M}_k$.

Iterating the argument until we reach $k = 0$ proves the lemma.

It remains to prove the invariance of $\mu = \lim_{n\to\infty} \mu_n$ under $\mathfrak{h} = \lim_{n\to\infty} \mathfrak{h}_n$. For $v \in \mathfrak{h}$ there exists a sequence (v_n) with $v_n \in \mathfrak{h}_n$ for $n \geqslant 1$ with $v_n \to v$ as $n \to \infty$. Then, by uniform continuity,

$$\left\| f^{\exp(v_n)} - f^{\exp(v)} \right\|_\infty \longrightarrow 0$$

as $n \to \infty$ for $f \in C_c(X)$. As μ_n is a probability measure for $n \geqslant 1$ this also shows that

$$\left| \underbrace{\int f^{\exp(v_n)} \,\mathrm{d}\mu_n}_{=\int f \,\mathrm{d}\mu_n} - \int f^{\exp(v)} \,\mathrm{d}\mu_n \right| \leqslant \left\| f^{\exp(v_n)} - f^{\exp(v)} \right\|_\infty \longrightarrow 0$$

as $n \to \infty$. Taking limits gives

$$\int f \,\mathrm{d}\mu = \int f^{\exp(v)} \,\mathrm{d}\mu,$$

so $\exp(v)$ preserves μ. As $v \in \mathfrak{h}$ was arbitrary, the lemma follows. □

Lemma 6.13. *Let G be a σ-compact, locally compact group equipped with a left-invariant metric. Let $\Gamma < G$ be a discrete subgroup and $\eta_1, \dots, \eta_k \in \Gamma$ arbitrary elements. Then $\Gamma C_G(\eta_1, \dots, \eta_k)$ is closed in $X = \Gamma \backslash G$.*

Proof. The proof is similar to the proof of Proposition 3.1 or Proposition 3.8. So suppose that $\Gamma g_n \to \Gamma g$ as $n \to \infty$ with $g_n \in C_G(\eta_1, \dots, \eta_k)$ for $n \geqslant 1$ and some $g \in G$. Choose $\gamma_n \in \Gamma$ for $n \geqslant 1$ with $\gamma_n g_n \to g$ as $n \to \infty$. Fix some $i \in \{1, \dots, k\}$ and notice that

$$\Gamma \ni \gamma_n \eta_i \gamma_n^{-1} = \gamma_n g_n \eta_i (\gamma_n g_n)^{-1} \longrightarrow g \eta_i g^{-1}$$

as $n \to \infty$ has to become eventually stable. So assume that

$$\gamma_N \eta_i \gamma_N^{-1} = \gamma_n \eta_i \gamma_n^{-1} = g \eta_i g^{-1}$$

for all $n \geqslant N$ and all i. However, this shows that $\gamma_N^{-1} g \in C_G(\eta_1, \dots, \eta_k)$ and

$$\Gamma g = \Gamma \gamma_N^{-1} g \in \Gamma C_G(\eta_1, \dots, \eta_k)$$

as required. □

Lemma 6.14. *Let $G \leqslant \mathrm{SL}_d(\mathbb{R})$ be a closed linear group and let $\Gamma < G$ be a discrete subgroup. Suppose that $L < G$ is a unipotent subgroup such that xL has finite volume. Then $xC_L(L)$ is compact.*

Proof. Clearly $xL \cong \Lambda \backslash L$ for a lattice $\Lambda < L$, so it suffices to consider the case $G = L$ and $x = \Lambda \in \Lambda \backslash L$. By Borel density (Theorem 3.30; also see the argument on p. 122) there exist elements $\lambda_1, \dots, \lambda_k \in \Lambda$ with

$$C_L(L) = C_L(\lambda_1, \dots, \lambda_k).$$

Thus Lemma 6.13 shows that $\Lambda C_L(L)$ is closed.

Finally, notice that if $\Lambda g_n \to \infty$ for some $g_n \in C_L(L)$ as $n \to \infty$, then the injectivity radius at Λg_n has to approach zero. In fact, by Proposition 1.11 there exist $\lambda_n \in \Lambda \smallsetminus \{I\}$ for which $g_n^{-1} \lambda_n g_n \to I$ as $n \to \infty$. However, for $g_n \in C_L(L)$ we have $g_n^{-1} \lambda_n g_n = \lambda_n \in \Lambda \smallsetminus \{I\}$ which contradicts the stated convergence. Therefore $\Lambda C_L(L)$ is a bounded closed set in ΛL, and so is compact. □

Lemma 6.15. *Suppose that $G \leqslant \mathrm{SL}_d(\mathbb{R})$ is a closed linear group, $\Gamma < G$ is a discrete subgroup, and $M < G$ is a unipotent abelian subgroup. If xM is compact for some $x \in X = \Gamma \backslash G$, then $xN_G^1(M)$ is closed, where*

$$N_G^1(M) = \{g \in G \mid gMg^{-1} = M \text{ and } gm_M g^{-1} = m_M\}$$

is the unimodular normalizer of M in G.

Proof. Let $x = \Gamma g$. By conjugating M with g we may assume without loss of generality that $x = I$. As in the proof of Lemma 6.13, we assume that $\gamma_n g_n \to g$ as $n \to \infty$ for $g_n \in N_G^1(M)$, $\gamma_n \in \Gamma$ and $g \in G$. We wish to show that $\gamma g \in N_G^1(M)$ for some $\gamma \in \Gamma$.

Notice that

$$\Gamma g_n M \cong \bigl((g_n^{-1}\Gamma g_n) \cap M\bigr)\backslash M,$$

which is isomorphic to $(\Gamma \cap M)\backslash M$ via conjugation by $g_n \in N_G^1(M)$. This implies that $\Gamma g_n M$ has the same volume as ΓM since conjugation by g_n in $N_G^1(M)$ preserves the Haar measure on M by definition. Moreover, since

$$\Gamma g_n \longrightarrow \Gamma g$$

as $n \to \infty$, we see that the injectivity radius of Γg_n stays bounded away from zero. By Minkowski's theorem on successive minima (Theorem 1.15, equivalently via the argument in the proof of Mahler's compactness criteria in Theorem 1.17) there exist elements

$$\eta_{n,1}, \dots, \eta_{n,\dim M} \in \Gamma$$

such that

$$(\gamma_n g_n)^{-1}\, \eta_{n,i}\, (\gamma_n g_n) \in M \tag{6.7}$$

is of bounded size (independent of n) and gives a basis of $(g_n^{-1}\Gamma g_n) \cap M$ for $i = 1, \dots, \dim(M)$. Therefore, we may choose a subsequence such that for every $i = 1, \dots, \dim(M)$ we have (after renaming the indexing variable in the sequence) that

$$(\gamma_n g_n)^{-1}\, \eta_{n,i}\, (\gamma_n g_n) \longrightarrow m_i \in M. \tag{6.8}$$

Since we also have $\gamma_n g_n \to g$ we may conjugate by $\gamma_n g_n$ in (6.8) to obtain

$$\eta_{n,i} \longrightarrow g m_i g^{-1}$$

as $n \to \infty$. However, since $\eta_{n,i} \in \Gamma$ this shows that we must have

$$\eta_{N,i} = \eta_{n,i} = g m_i g^{-1}$$

for $i = 1, \dots, \dim(M)$ and all $n \geqslant N$ for some large enough N. Conjugating by γ_n we obtain from (6.7) that

$$\underbrace{\gamma_n^{-1}\eta_{n,i}\gamma_n}_{\in M} = \gamma_n^{-1} g \underbrace{m_i}_{\in M} g^{-1}\gamma_n,$$

by the definition of $\eta_{n,i}$ for $i = 1, \dots, \dim(M)$ and all $n \geqslant N$. Since

$$(\gamma_n g_n)^{-1}\, \eta_{n,i}\gamma_n g_n$$

gives a basis of the lattice
$$\left(g_n^{-1}\Gamma g_n\right)\cap M$$
by definition of $\eta_{n,i}$, and a lattice in M is Zariski dense, it follows that
$$\langle m_1,\dots,m_{\dim M}\rangle$$
is also Zariski dense in M and
$$\gamma_n^{-1}g\in N_G(M)$$
for all $n\geqslant N$.

In particular,
$$\gamma_N^{-1}g\left(\gamma_n^{-1}g\right)^{-1}=\gamma_N^{-1}\gamma_n\in N_G(M)$$
for all $n\geqslant N$. We claim that $\gamma_N^{-1}\gamma_n\in N_G^1(M)$. For if $\eta=\gamma_N^{-1}\gamma_n$ (or its inverse) were to contract the Haar measure on M then $\eta^\ell(\Gamma\cap M)\eta^{-\ell}$ would have to contain shorter and shorter vectors as $\ell\to\infty$ by Minkowski's first theorem (Theorem 1.14). As $\eta^\ell(\Gamma\cap M)\eta^{-\ell}\subseteq\Gamma$ this is impossible, proving the claim.

It follows that
$$\gamma_N^{-1}g=\lim_{n\to\infty}\gamma_N^{-1}\gamma_n g_n\in N_G^1(M)$$
as required. □

6.4.2 Equidistribution and Orbit Closures on Nilmanifolds

Using Theorem 6.11 we can establish the equidistribution theorem (Theorem 6.2) and the orbit closure theorem (Theorem 6.3) on nilmanifolds. In the case of unipotent flows on nilmanifolds this step of the proof is significantly easier due to the following special feature of unipotent flows on nilmanifolds (which we know is false for the horocycle flow on a non-compact quotient, for example).

Corollary 6.16. *Let $\Gamma<G$ be a discrete subgroup of a connected nilpotent Lie group G and let $X=\Gamma\backslash G$. Let $U\leqslant G$ be a one-parameter subgroup and $x_0\in X$. Then the orbit closure $\overline{U\bullet x}=L\bullet x$ is algebraic and the U-action on $L\bullet x_0$ is uniquely ergodic.*

6.5 Invariant Measures for Semi-simple Groups

Using Section 6.3.2 we are also ready to prove the special case of Ratner's measure classification theorem where the acting group is semi-simple†. We are going to use the Mautner phenomenon to find an ergodic one-parameter unipotent flow. This is possible due to the results of Chapter 2, but requires that the group H has no compact factors. While almost all of the ideas of the proof certainly go back to the work of Ratner, and in particular to the paper [**?**], the observation that this particular case has a short and relatively easy proof was made in [**?**].

Theorem 6.17 (Ratner measure classification; the semi-simple case). *Let G be a connected Lie group, $\Gamma < G$ a discrete subgroup, and assume that $H < G$ is a semi-simple subgroup without compact factors. Suppose that μ is an H-invariant and ergodic probability measure on X. Then μ is algebraic.*

Proof. Define the closed subgroup

$$\mathrm{Stab}(\mu) = \{g \in G \mid g_*\mu = \mu\},$$

the connected component

$$L = \mathrm{Stab}(\mu)^o,$$

and its Lie algebra $\mathfrak{l}$. We need to prove that μ is supported on a single L-orbit. So let us assume (for the purposes of a contradiction) that this is not the case. Then by ergodicity of μ, each L-orbit must have zero μ-measure since $H \leqslant L$.

There exists a subgroup of H that is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$, which acts ergodically on X with respect to μ. This follows from the Mautner phenomenon. — Indeed H is by assumption an almost direct product of non-compact simple Lie groups and each of these contains a subgroup that is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$. Now consider a diagonally embedded subgroup that is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$ and that projects non-trivially to each simple almost direct factor. By Theorem 2.11 this subgroup satisfies the Mautner phenomenon for H, and since H acts ergodically so does the subgroup. So we may assume that H is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$. Furthermore, we let $U \leqslant H$ be the subgroup corresponding to the upper unipotent subgroup in $\mathrm{SL}_2(\mathbb{R})$. Once again by the Mautner phenomenon (Proposition 2.9) U acts ergodically with respect to μ.

† This case is interesting as the proof is relatively straightforward, even though there may be a large gap in the dimensions of the acting group and the group that gives rise to the ambient space. Furthermore, due to this gap there may be a large collection of possible intermediate subgroups $H \leqslant L \leqslant G$. However, the use of this special case is limited as the acting group is not amenable and hence it is *a priori* not even clear why we should have any H-invariant probability measure on a given orbit closure $\overline{H\bullet x} \subseteq X$.

By the structure theory of finite-dimensional representations of $\mathrm{SL}_2(\mathbb{R})$ (see Knapp [**?**, Th. 1.64], for example), we see that the H-invariant subspace $\mathfrak{l} \leqslant \mathfrak{g}$ (with respect to the adjoint action) has an H-invariant complement $V < \mathfrak{g}$ (though we have no reason to expect that V is a Lie algebra).

Now let $K \subseteq X$ be a set of μ-measure exceeding 0.99 comprising uniformly generic points for $U < H$. We would like to find points $x_n, y_n \in K$ with

$$y_n = g_n \raisebox{0.5pt}{.} x_n,$$

for some $g_n \neq I$ with $g_n \in \exp(V)$ belonging to the 'transverse' direction for all $n \geqslant 1$, and with $g_n \to I$ as $n \to \infty$. We then may consider the polynomials

$$p_n(r) = \mathrm{Ad}_{u(T_n r)}(\log g_n), \tag{6.9}$$

assume that these converge as $n \to \infty$, and apply Proposition 6.9. By the H-invariance of V all the polynomials p_n would have values in V and so we would then be able to find a polynomial $p : \mathbb{R} \to \mathfrak{g}$ taking values in V and with μ preserved by $\exp p(r)$ for all $r > 0$. The existence of such a polynomial contradicts the definition of $L = \mathrm{Stab}(\mu)^o$.

To find x_n, y_n as above, we can apply a relatively simple Fubini argument as follows (crucially, using the fact that μ is invariant under L).

So let $B_\delta^L = B_\delta^L(I)$ be a small open metric ball in L around the identity, and define

$$Y = \left\{ x \in X \mid \int_{B_\delta^L} \mathbb{1}_K(\ell \raisebox{0.5pt}{.} x) \,\mathrm{d}m_L(\ell) > 0.9 m_L(B_\delta^L) \right\}.$$

We claim first that $\mu(Y) > 0.9$, which may be seen by looking at the complement as follows:

$$\begin{aligned}
\mu(X \smallsetminus Y) &= \mu\left(\left\{ x \in X \mid \int_{B_\delta^L} \mathbb{1}_{X \smallsetminus K}(\ell \raisebox{0.5pt}{.} x) \,\mathrm{d}m_L(\ell) \geqslant 0.1 m_L(B_\delta^L) \right\} \right) \\
&\leqslant \frac{1}{0.1 m_L(B_\delta^L)} \int_X \int_{B_\delta^L} \mathbb{1}_{X \smallsetminus K}(\ell \raisebox{0.5pt}{.} x) \,\mathrm{d}m_L(\ell) \,\mathrm{d}\mu \\
&= \frac{1}{0.1 m_L(B_\delta^L)} \int_{B_\delta^L} \underbrace{\int_X \mathbb{1}_{X \smallsetminus K}(\ell \raisebox{0.5pt}{.} x) \,\mathrm{d}\mu}_{=\mu(X \smallsetminus K)} \,\mathrm{d}m_L \qquad \text{(by Fubini)} \\
&= \frac{\mu(X \smallsetminus K)}{0.1} < \frac{0.01}{0.1} = 0.1,
\end{aligned}$$

since L preserves μ and $\mu(K) > 0.99$.

We now claim that for any nearby points $x, y \in Y$ we can find $\ell_x, \ell_y \in B_\delta^L$ such that

$$x' = \ell_x \raisebox{0.5pt}{.} x \in K, \tag{6.10}$$

$$y' = \ell_y \boldsymbol{\cdot} y \in K, \tag{6.11}$$

and

$$y' = \exp(v) \boldsymbol{\cdot} x' \tag{6.12}$$

with $v \in V$. To see this, notice that if δ is sufficiently small, then (by the inverse mapping theorem) the map

$$\begin{aligned} \psi : B_{2\delta}^L \times B_{2\delta}^V(0) &\longrightarrow G \\ (\ell, v) &\longmapsto \ell \exp(v) \end{aligned}$$

is a diffeomorphism from $B_{2\delta}^L \times B_{2\delta}^V(0)$ onto an open neighborhood O of the identity in G. Let now $g \in B_\kappa^G(I)$ be chosen so that $y = g \boldsymbol{\cdot} x$. Then we would like to find $\ell_x, \ell_y \in B_\delta^L$ with $g\ell_x^{-1} = \ell_y^{-1} \exp(v)$, which will give (6.12). This can be done using the local diffeomorphism above: if κ is sufficiently small, then $g\ell_x^{-1} \in O$ and may define ℓ_y and v by

$$\psi^{-1}(g\ell_x^{-1}) = (\ell_y^{-1}, v). \tag{6.13}$$

However, we still have to worry about the conditions (6.10) and (6.11).

For this, we are going to see that most points $\ell_x \in B_\delta^L$ (and the corresponding ℓ_y) will satisfy this. Indeed, by definition of Y, at least 90% of all $\ell_x \in B_\delta^L$ satisfy $x' = \ell_x \boldsymbol{\cdot} x \in K$, and at least 90% of all $\ell_y \in B_\delta^L$ satisfy $y' = \ell_y \boldsymbol{\cdot} y \in K$. However, we need to do this while ensuring that (6.13) (or equivalently, (6.12)) holds. So define the map

$$\begin{aligned} \phi : B_\delta^L &\longrightarrow B_{2\delta}^L \\ \ell_x &\longmapsto \ell_y \end{aligned}$$

with ℓ_y as in (6.13). This smooth map depends on the parameter $g \in B_\kappa^G$ and is close to the identity in the C^1-topology if κ is sufficiently small (all maps we deal with are analytic and for $g = e$ we have $\phi = I_{B_\delta^L}$). Therefore ϕ does not distort the chosen Haar measure of L much, and sends B_δ^L into a ball around the identity that is not much bigger than B_δ^L (both with respect to the metric structure and with respect to the measure). In other words, if κ is sufficiently small, then

$$\begin{aligned} m_L\left(\phi\left(\left\{\ell_x \in B_\delta^L \mid \ell_x \boldsymbol{\cdot} x \in K\right\}\right) \cap B_\delta^L\right) &> 0.9 m_L\left(\phi\left(\left\{\ell_x \in B_\delta^L \mid \ell_x \boldsymbol{\cdot} x \in K\right\}\right)\right) \\ &> 0.8 m_L\left(\left\{\ell_x \in B_\delta^L \mid \ell_x \boldsymbol{\cdot} x \in K\right\}\right) \\ &> (0.8)(0.9) m_L\left(B_\delta^L\right) > 0.7 m_L\left(B_\delta^L\right). \end{aligned}$$

Together with

$$m_L\left(\{\ell_y \in B_\delta^L \mid \ell_y \boldsymbol{\cdot} y \in K\}\right) > 0.9 m_L\left(B_\delta^L\right),$$

we see that there are many points $\ell_x \in B_\delta^L$ with $\ell_x \cdot x \in K$ for which ℓ_y defined by (6.13) also satisfies $\ell_y \cdot y \in K$.

The theorem now follows relatively quickly as outlined earlier. Recall that we may assume that every $L = \mathrm{Stab}(\mu)^o$-orbit has μ-measure zero. Let

$$z \in \mathrm{Supp}\,(\mu|_Y)\,.$$

Then for every $\kappa = \frac{1}{n}$ there exist $x_n = z, y_n = g_n \cdot x_n \in Y$ with

$$g_n \in B_{1/n}^G(I) \smallsetminus L.$$

Applying the procedure above to x_n, y_n (which we certainly may if n is large) then we get

$$x_n', y_n' = \exp(v_n) \cdot x_n' \in K, v_n \in V, v_n \neq 0, v_n \to 0$$

as $n \to \infty$. There are now two cases to consider.

If v_n is in the eigenspace of Ad_{u_s} for infinitely many n (and so let us assume for all n by passing to that subsequence), then we may apply Lemma 6.7 to each v_n and deduce that $\exp(v_n)$ preserves μ. However, since $v_n \to 0$ as $n \to \infty$ and the unit sphere in V is compact, we may assume that $\frac{v_n}{\|v_n\|} \to w$ as $n \to \infty$ by passing to a subsequence again. We conclude that since $\mathrm{Stab}(\mu)$ is closed, $\exp(tw) \in \mathrm{Stab}(\mu)$ for all t. Since V is a linear complement to the Lie algebra of $L = \mathrm{Stab}(\mu)^o$, this is a contradiction.

So assume that v_n is not in an eigenspace for any $n \geqslant 1$ (by deleting finitely many terms). In this case we may define T_n such that the polynomials in (6.9) have norm one. Use compactness of the set of polynomials with bounded degree and norm one to choose a subsequence (again denoted (p_n)) that converges to a polynomial p, and then apply Proposition 6.9 to see that μ is preserved by $\exp p(t)$ for all $t > 0$. Since p is the limit of $\mathrm{Ad}_{T_n r}(v_n) \in V$, p also takes values in V which again contradicts the definition of V. $\square$

6.6 Transverse Divergence and Entropy for the horocycle flow

†We will reprove (up to a fact regarding entropy which we will assume for the moment) the classification of invariant measures under a weaker assumption. Note that the assumption made in Theorem 6.18 below is weaker than the assumption in Theorem 5.3, as we do not assume that Γ is a lattice. As a result, the proof of the theorem below is a better representation of the

† Even though the result of this section may not go much beyond what we already understand, we take this case as a starting point for a tour of cases ending with the general case of unipotent flows on quotients of $\Gamma\backslash\mathrm{SL}_3(\mathbb{R})$

general measure classification results, and indeed is a result of Ratner (see Theorem 6.1 and the survey [**?**]). The use of entropy below goes back to work of Margulis and Tomanov [**?**].

Theorem 6.18 (Invariant measures for the horocycle flow). *Let Γ be a discrete subgroup of* $\mathrm{SL}_2(\mathbb{R})$*, let* $X = \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$*, and let*

$$U = \left\{ u_s = \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\}.$$

Suppose that μ is a U-invariant and ergodic probability measure on X. Then either

- *μ is supported on a single periodic orbit of U; or*
- *$\mu = m_X$ and Γ is a lattice.*

We want to apply an argument similar to the one in Section 6.3.2. It is easy to check that the argument as it is presented there is not going to be helpful since it would always just imply invariance under $\{u_s\}$. We start by generalizing Section 6.3.1.

Lemma 6.19 (Normalizer lemma). *Let $X = \Gamma\backslash G$ for some closed linear group G and some discrete subgroup $\Gamma < G$, let $U < G$ be a one-parameter subgroup, and let μ be an U-invariant and ergodic probability measure on X. Suppose that $x, y = g \bullet x \in X$ are generic for the U-action (in both directions) and the measure μ and suppose*

$$g \in N_G(U) = \{g \in G \mid gUg^{-1} = U\}.$$

Then g preserves μ.

For $G = \mathrm{SL}_2(\mathbb{R})$ and the horocycle subgroup U we have that $g \in N_G(U)$ implies that

$$g = \begin{pmatrix} \lambda & t \\ & \lambda^{-1} \end{pmatrix}$$

for some $\lambda \neq 0$ and $t \in \mathbb{R}$. Indeed, if

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

normalizes U, then we may calculate

$$\mathrm{Ad}_g \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -c & a-d \\ 0 & c \end{pmatrix}$$

and deduce that $c = 0$. We note that the lemma also implies in this case that

$$a = \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$$

preserves μ. We refer to Figure 6.4 for the geometrical picture of the proof.

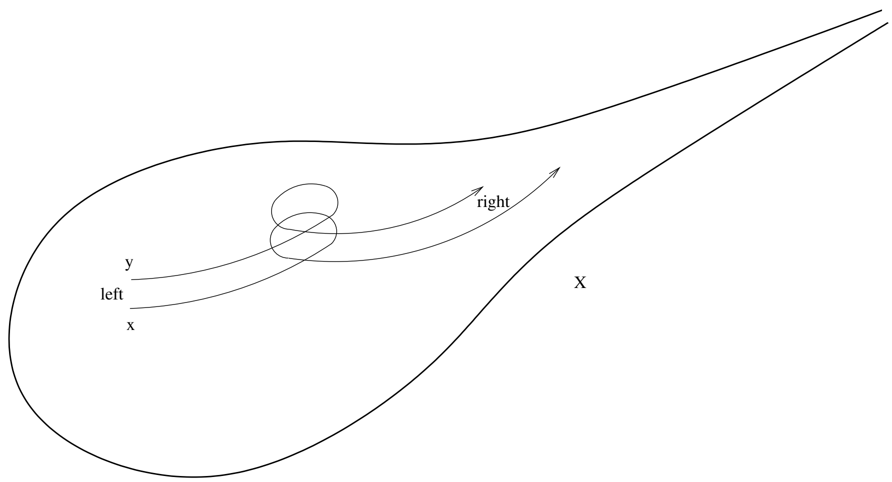

Fig. 6.4 If $y = xh^{-1}$ with $h \in N_G(V)$, then the two orbits are again parallel as in Figure 6.2, but xu_s^{-1} may not be close to yu_s^{-1} but instead be close to yu_r^{-1} for some r.

Proof of Lemma 6.19. The lemma follows from the argument used in Section 6.3.1, taking into account the fact that g conjugates $u_s \in U$ into

$$gu_s g^{-1} = u_{\lambda s}$$

for some fixed $\lambda \in \mathbb{R}^\times$. Let us assume for simplicity of notation that $\lambda > 0$, the case $\lambda < 0$ is very similar. Hence the piece of the orbit

$$u_{[0,T]} \cdot x = \{u_s \cdot x \mid s \in [0, T]\}$$

is mapped under g to

$$gu_{[0,T]} \cdot x = \{u_{\lambda s} g \cdot x \mid s \in [0, T]\} = u_{[0,\lambda T]} \cdot y.$$

As before, the normalized Lebesgue measure on $u_{[0,T]} \cdot x$ and $u_{[0,\lambda T]} \cdot y$ both approximate μ as $T \to \infty$, and we deduce that g preserves μ. $\square$

Just as in the discussion in Section 6.3.1, for the proof of Theorem 6.18 we cannot hope in general for this propitious situation — the requirement that $g \in N_G(U)$ restricts the displacement between the two typical points to a two-dimensional group sitting inside the three-dimensional $\mathrm{SL}_2(\mathbb{R})$. Thus we will be forced in the argument developed below to work with an element

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $ad - bc = 1$ but with no other constraint, and in particular with c permitted to be non-zero. As we will see in the first part of the proof we will be able to use 'transverse divergence' to produce additional invariance.

PROOF OF THEOREM 6.18. Suppose $K \subseteq X$ is a uniformly generic set of measure $\mu(K) > 0.99$. Suppose that $z \in \operatorname{Supp} \mu|_K$. If for some $\delta > 0$ we have

$$B_\delta(z) \cap K \subseteq U \cdot z,$$

then $U \cdot z$ has positive measure. This shows that z is therefore a periodic orbit and μ is its normalized periodic orbit measure.

Otherwise, it follows that we can choose $x_n \in K$ and $y_n \in K$ with

$$y_n = g_n \cdot x_n$$

and $g_n \notin U$ for all $n \geqslant 1$ and $g_n \to e$ as $n \to \infty$. If for some n we have

$$g_n = \begin{pmatrix} \lambda & t \\ 0 & \lambda^{-1} \end{pmatrix} \in N_G(U),$$

then by Lemma 6.19 we know that $a = \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$ with $\lambda \neq \pm 1$ preserves μ. Below we will show how this implies that $\mu = m_X$.

Next we discuss the more general case, so assume that

$$g_n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$$

with $c_n \neq 0$. We may assume that $c_n > 0$ for all $n \geqslant 1$, for if not we can interchange x_n and y_n, thereby replacing g_n by g_n^{-1}. We would like to argue along the lines of Proposition 6.9, but we already learned from the proof of Lemma 6.19 that we might have to use different clocks for the parametrization of the orbits of x and y. We have seen before (see (2.6) in the proof of Proposition 2.9 on p. 58) the calculation that lies behind this:

$$\begin{aligned} u_{s_y} g_n u_{-s_x} &= \begin{pmatrix} 1 & s_y \\ & 1 \end{pmatrix} \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \begin{pmatrix} 1 & -s_x \\ & 1 \end{pmatrix} \\ &= \begin{pmatrix} a_n + c_n s_y & b_n - a_n s_x + s_y(d_n - c_n s_x) \\ c_n & d_n - c_n s_x \end{pmatrix}. \end{aligned} \tag{6.14}$$

If we set $s_y = s_x$ then the upper-right entry, which corresponds to the subgroup U, is a quadratic term and for small $\mathsf{d}(g_n, I)$ this quadratic term is the most significant entry. As this would not lead anywhere, we instead choose

$$s_y = \frac{a_n s_x}{d_n - c_n s_x}.$$

Having made this choice, we obtain the simpler formula

$$u_{s_y} g_n u_{-s_x} = \begin{pmatrix} a_n + c_n s_y & b_n \\ c_n & d_n - c_n s_x \end{pmatrix}.$$

Once again we want to speed up the time parameter $s_x = T_n r$ by defining T_n to be $\frac{1}{c_n}$ and

$$\phi_n(r) = u_{a_n T_n r/(d_n - r)} g_n u_{-T_n r} = \begin{pmatrix} a_n + \frac{a_n r}{d_n - r} & b_n \\ c_n & d_n - r \end{pmatrix},$$

which defines a sequence of rational functions taking values in $\mathrm{SL}_2(\mathbb{R})$. The pole of the nth rational function in this sequence is at $r = d_n$, which is approximately 1. It follows that (ϕ_n) converges uniformly on $[0, \frac{1}{2}]$ to the rational function

$$\phi(r) = \begin{pmatrix} \frac{1}{1-r} & 0 \\ 0 & 1-r \end{pmatrix}.$$

We claim that

$$\mu \text{ is preserved by } \phi\left(\tfrac{1}{2}\right) = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}. \tag{6.15}$$

Once we know this we are at the same stage as in the previous special case.

To prove the claim in (6.15), fix some $\varepsilon > 0$ and $f \in C_c(X)$. Then there exists a $\delta > 0$ with

$$\mathsf{d}(h_1, h_2) < \delta \implies |f(h_1 \centerdot x) - f(h_2 \centerdot x)| < \varepsilon \tag{6.16}$$

for all $x \in X$. Next notice that there exists some $\kappa > 0$ such that

$$\mathsf{d}(\phi(r), \phi(\tfrac{1}{2})) < \tfrac{\delta}{2}$$

for all $r \in [\frac{1}{2} - \kappa, \frac{1}{2}]$, which implies that

$$\mathsf{d}(\phi_n(r), \phi(\tfrac{1}{2})) < \delta \tag{6.17}$$

for all sufficiently large n. Taking a convex combination of two ergodic averages, and keeping κ fixed, we can deduce (just as in Section 6.3.2) that

$$\mathsf{A}_n = \frac{1}{\kappa T_n} \int_{(\frac{1}{2} - \kappa) T_n}^{\frac{1}{2} T_n} f^{\phi(1/2)}(u_s \centerdot x_n) \, \mathrm{d}s \longrightarrow \int_X f^{\phi(1/2)} \, \mathrm{d}\mu.$$

Now

$$f^{\phi(1/2)}(u_s \centerdot x_n) = f\left(\phi(1/2) u_s \centerdot x_n\right)$$

is within ε of

$$f\left(\phi_n(c_n s) u_s \centerdot x_n\right),$$

since $c_n s \in [\frac{1}{2} - \kappa, \frac{1}{2}]$ and because of (6.16) and (6.17). Next we recall the definition of ϕ_n to get

$$\phi_n(\underbrace{c_n s}_{r})u(s) = u_{a_n T_n r/(d_n - r)} g_n u_{-T_n c_n s} u_s = u_{a_n s/(d_n - c_n s)} g_n.$$

Together we deduce that

$$\left| \mathsf{A}_n - \frac{1}{\kappa T_n} \int_{(\frac{1}{2}-\kappa)T_n}^{\frac{1}{2}T_n} f\Big(u_{a_n s/(d_n - c_n s)}\, \underbrace{g_n \bullet x_n}_{y_n}\Big)\, \mathrm{d}s \right| < \varepsilon.$$

The integral in this estimate is almost of the same form for y_n as the ergodic average A_n for x_n — except that the orbit is run through non-linearly. For that reason we now use the substitution $t = \frac{a_n s}{d_n - c_n s}$. Its derivative is given by

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{a_n d_n}{(d_n - c_n s)^2},$$

which for large n and sufficiently small κ satisfies

$$\left| \frac{1}{4}\frac{\mathrm{d}t}{\mathrm{d}s} - 1 \right| < \varepsilon. \tag{6.18}$$

This shows that

$$\left| \mathsf{A}_n - \frac{1}{4\kappa T_n} \int_{(\frac{1}{2}-\kappa)T_n}^{\frac{1}{2}T_n} f\Big(\, u_{\underbrace{a_n s/(d_n - c_n s)}_{t}}\Big) y_n \bullet \frac{\mathrm{d}t}{\mathrm{d}s}\, \mathrm{d}s \right| < \varepsilon\left(1 + \|f\|_\infty\right),$$

and equivalently

$$\left| \mathsf{A}_n - \frac{1}{4\kappa T_n} \int_{(\frac{1}{2}-\kappa)\frac{a_n T_n}{d_n - \frac{1}{2}+\kappa}}^{\frac{1}{2}\frac{a_n T_n}{d_n - \frac{1}{2}}} f\left(u_t \bullet y_n\right)\, \mathrm{d}t \right| < \varepsilon\left(1 + \|f\|_\infty\right).$$

From (6.18) (or alternatively apply the above to the constant function $f = \mathbb{1}$) we also deduce that the length of the interval for the integral is asymptotic to $4\kappa T_n$ as $\varepsilon \to 0$. More precisely we have

$$\left| 1 - \frac{\text{total length}}{4\kappa T_n} \right| < 2\varepsilon.$$

Applying the convex combination argument to the intervals

$$\left[0, \tfrac{1}{2}\tfrac{a_n T_n}{d_n - \frac{1}{2}}\right] \text{ and } \left[0, (\tfrac{1}{2} - \kappa)\tfrac{a_n T_n}{d_n - \frac{1}{2}+\kappa}\right],$$

the initial point y_n and the function f we get

$$\frac{1}{\text{total length}} \int_{(\frac{1}{2}-\kappa)\frac{a_n T_n}{d_n-\frac{1}{2}+\kappa}}^{\frac{1}{2}\frac{a_n T_n}{d_n-\frac{1}{2}}} f\left(u_t \boldsymbol{\cdot} y_n\right) \mathrm{d}t \longrightarrow \int_X f \,\mathrm{d}\mu.$$

Together we see after taking the limits as $n \to \infty$ that

$$\left| \int_X f^{\phi(1/2)} \,\mathrm{d}\mu - \int_X f \,\mathrm{d}\mu \right| < O(\varepsilon) \left(1 + \|f\|_\infty\right),$$

and this holds for any $\varepsilon > 0$ and $f \in C_c(X)$. Hence we have shown that in any case μ is invariant under a non-trivial element $a = \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$ of the geodesic flow.

We now finish the proof using the entropy theory developed[†] in Chapter 10. In particular, by Theorem 10.2(2) we have

$$h_\mu(a) = 2 \log |\lambda|,$$

since μ is U-invariant and U is precisely the stable horospherical subgroup. However, by Theorem 10.5, this implies that μ is also invariant under the opposite unipotent subgroup

$$\left\{ \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\}.$$

Since these unipotent subgroups together generate $\mathrm{SL}_2(\mathbb{R})$, we see that μ must be the Haar measure m_X on X, which also forces Γ to be a lattice. □

Exercises for Section 6.6

Exercise 6.6.1. Let

$$G =_2 (\mathbb{R}) = \mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2 = \left\{ \begin{pmatrix} A & v \\ 0 & 1 \end{pmatrix} \mid A \in \mathrm{SL}_2(\mathbb{R}), v \in \mathbb{R}^2 \right\},$$

$\Gamma =_2 (\mathbb{Z}) = G \cap \mathrm{SL}_3(\mathbb{R})$, and $X = \Gamma \backslash G$.

Consider the following[‡] choices of one-parameter unipotent subgroups and prove for each of them the classification of invariant measures:

(1) We could set $u_s = \begin{pmatrix} 1 & 0 & s \\ & 1 & 0 \\ & & 1 \end{pmatrix}$.

[†] We hope that this and future forward references to Theorem 10.2 and Theorem 10.5 serve as a good motivation for learning the basics of entropy theory from [**?**] and the more refined arguments of Part II.

[‡] The reader may check whether these are all, up to conjugation.

(2) A more interesting choice is given by $u_s = \begin{pmatrix} 1 & s & 0 \\ & 1 & 0 \\ & & 1 \end{pmatrix}$. Here you may use again entropy theory (see the next section on how to avoid this).

(3) Finally we could also set $u_s = \begin{pmatrix} 1 & s & \frac{s^2}{2} \\ & 1 & s \\ & & 1 \end{pmatrix}$. We note that this case is easier to deal with, no entropy theory is needed.

6.7 Joinings of the Horocycle Flow

In this section we consider the group $G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ and its quotient $X = \Gamma\backslash G$ by a lattice. Up to conjugation, G allows three different choices of one-parameter unipotent flows:

- $\left\{ u_s = \left(\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \right) \,\middle|\, s \in \mathbb{R} \right\}$;
- $\left\{ u_s = \left(\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \right) \,\middle|\, s \in \mathbb{R} \right\}$; and
- $\left\{ u_s = \left(\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \right) \,\middle|\, s \in \mathbb{R} \right\}$.

The first two are actually horospherical subgroups so the discussion in Chapter 5 applies to these cases. Thus we will only consider the third (most difficult) case (which of course is a special case of Ratner's measure classification in Theorem 6.1).

Theorem 6.20 (The first non-horospherical case). *Let*

$$G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}),$$

let $\Gamma < G$ be a lattice[†]*, and define the quotient space $X = \Gamma\backslash G$. Let*

$$U = \left\{ u_s = \left(\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \right) \,\middle|\, s \in \mathbb{R} \right\}.$$

Then any U-invariant ergodic probability measure μ on X is algebraic.

We will give two proofs, both of which start in the same way. The first proof will use entropy theory and works in the stated generality. The second proof is due to Ratner from her earlier work on the rigidity properties of the horocycle flow [**?**, **?**, **?**] and will not use entropy theory but only works in a special case (more precisely, only for reducible lattices).

[†] In some portions of the proof it will be convenient to refer to Theorem 5.7 where we assumed that $\Gamma\backslash G$ is isomorphic to an orbit in d for some $d \geqslant 2$. So strictly speaking we should assume this. Alternatively see Exercise 6.7.2.

Start of proof of Theorem 6.20 using transverse divergence. Let $K \subseteq X$ be a set of uniformly generic points of measure $\mu(K) > 0.99$ and suppose that $y = g{\bullet}x$ with $x \in X$ and $g \in G$ close to the identity $I \in G$. As in the last section, we now want to study how the U-orbits of x and of y move apart, where we allow different time parameters for the orbit of x and the orbit of y. To this end we calculate

$$u_{s_y}{\bullet}y = u_{s_y} g u_{-s_x}{\bullet}(u_{s_x}{\bullet}x)$$

and for

$$g = \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \right)$$

obtain (by applying the calculation (6.14) for each component)

$$u_{s_y} g u_{-s_x} = \left(\begin{pmatrix} a + cs_y & b - as_x + s_y(d - cs_x) \\ c & d - cs_x \end{pmatrix}, \begin{pmatrix} a' + c's_y & b' - a's_x + s_y(d' - c's_x) \\ c' & d' - c's_x \end{pmatrix} \right).$$

We again set $s_y = \frac{as_x}{d - cs_x}$, so that the above simplifies to

$$\phi(s_x) = \left(\begin{pmatrix} a + cs_y & b \\ c & d - cs_x \end{pmatrix}, \begin{pmatrix} a' + c's_y & b' - a's_x + s_y(d' - c's_x) \\ c' & d' - c's_x \end{pmatrix} \right), \tag{6.19}$$

and this already ensures that any limit of functions like ϕ (with b approaching 0) does not take values in $U \smallsetminus \{I\}$.

Once again we need to speed up the time parameter s_x by setting $s_x = Tr$ for some $T > 0$ to be defined later. In the last section we defined T to be $\frac{1}{c}$ in order to ensure that the limit of the first matrix in ϕ is interesting. Here we need to be more careful, as with that choice the second matrix defining ϕ could diverge[†].

Clearly if $\phi(s_x)$ is constant, then it will be difficult to make it more interesting by a speeding up. Hence it will be useful to ask when this happens. This could be done by analyzing the concrete function ϕ as above in detail, but it is easier to do this abstractly.

Lemma 6.21. *Let $G \leqslant \mathrm{SL}_d(\mathbb{R})$ be a linear group, let*

$$U = \left\{ u(s) \mid s \in \mathbb{R}^k \right\} \leqslant G$$

be a unipotent subgroup parameterized by some polynomial map[‡] *$u : \mathbb{R}^k \to U$ with $u(0) = I$. Fix some $g \in G$. Suppose also that $s_y = s_y(s_x)$ is defined on an open neighborhood of $0 \in \mathbb{R}^k$ (for example, by a rational function) such*

[†] For example, if $c > 0$ is much smaller than c' this will happen.

[‡] We are not assuming that U is abelian, nor are we assuming that u is a homomorphism.

that $s_y(0)=0$ *and*

$$\phi(s_x)=u(s_y)gu(-s_x)$$

is constant where defined. Then $g\in N_G(U)$.

Proof. Let $s_x\in\mathbb{R}^k$ be close to 0. Then

$$\phi(s_x)=u(s_y)gu(-s_x)=\phi(0)=g$$

is equivalent to

$$gu(s_x)g^{-1}=u(s_y)\in U.$$

As this holds for all s_x near 0 and U is connected, the lemma follows. □

Suppose for a moment that $\phi(s_x)$ as in (6.19) is indeed constant. Then we have $g\in N_G(U)$, and by Lemma 6.19 we also have $g\in\mathrm{Stab}_G(\mu)$. Suppose that $x\in\mathrm{Supp}\,\mu|_K$, and that we are in this case for all $y=g\boldsymbol{\cdot}x\in K$ sufficiently close to x. Then $\mathrm{Stab}_G(\mu)\boldsymbol{\cdot}x$ has positive measure, and hence has full measure and the theorem follows.

It remains to consider the case where there is a sequence (y_n) with

$$y_n=g_n\boldsymbol{\cdot}x$$

for $n\geqslant 1$ with $x\in K$ and $g_n\to I$ as $n\to\infty$ for which the rational map ϕ_n defined as above is not constant. Then

$$\Phi_n(s_x)=(d_n-c_ns_x,(d_n-c_ns_x)\phi_n(s_x))$$

is a tuple of polynomials with not both being constant. We define the speeding-up parameters $T_n>0$ such that

$$\sup_{r\in[0,1]}\|\Phi_n(T_nr)-\Phi_n(0)\|_\infty=1.$$

We may choose a subsequence[†] such that

$$\Phi_n(T_nr)\longrightarrow\Psi(r)$$

converges uniformly as $n\to\infty$ on compact subsets of $\mathbb{R}$. It follows that

$$\Psi(r)=(1-\alpha r,\psi_0(r))$$

for some $\alpha\in[-1,1]$ and some polynomial ψ_0. Moreover,

$$\frac{1}{1-\alpha r}\psi_0(r)=\psi(r)$$

[†] As usual we simplify notation by not introducing a further subscript to denote the subsequence.

is the limit of the sequence of functions $(r \mapsto \phi_n(T_n r))$ uniformly on compact subsets of $\mathbb{R}\smallsetminus\{\frac{1}{\alpha}\}$. Now we can ask for the behavior of the function

$$r \mapsto \psi(r) \in G$$

for $r \neq \frac{1}{\alpha}$, without calculating it explicitly.

Lemma 6.22. *Let $G \leqslant \mathrm{SL}_d(\mathbb{R})$ be a closed linear group. Let*

$$U = \{u(s) \mid s \in \mathbb{R}^k\} < G$$

be a unipotent subgroup with a polynomial parameterization $u : \mathbb{R}^k \to U$ such that $u(0) = I$. Let $\{M_t : U \to U \mid t \in \mathbb{R}\}$ be a one-parameter group of automorphisms of U such that M_1 uniformly expands U. Let (g_n) be a sequence in G with $g_n \to I$ as $n \to \infty$. Suppose further that there exists a sequence (t_n) with $t_n \to \infty$ as $n \to \infty$, and a sequence of rational functions $\left(s_y^{(n)} : \mathbb{R}^k \to \mathbb{R}^k\right)$, well-defined except possibly on a proper subvarieties, such that†

$$\psi_n(r) = M_{t_n}\left(u(s_y^{(n)}(r))\right) g_n M_{t_n}(u(-r))$$

converges uniformly on some open subset $O \subseteq \mathbb{R}^k$ to some function

$$\psi : O \to G.$$

Then $\psi(O) \subseteq N_G(U)$.

We note that in the case we are currently interested we have $U \cong \mathbb{R}$ and we may define $M_t : U \to U$ by multiplication with e^t.

Proof. Let $u \in U$ and $r \in O$. Then

$$M_{t_n}(u(-r))u = M_{t_n}(u(-r)M_{-t_n}(u)) = M_{t_n}(u(-(r+\varepsilon_n))) \tag{6.20}$$

for some sequence (ε_n) with $\varepsilon_n \to 0$ as $n \to \infty$ since $M_{-t_n}(u) \to I$ in U as $n \to \infty$. By uniform convergence this implies that

$$\begin{aligned}\psi(r) &= \lim_{n\to\infty} \psi_n(r+\varepsilon_n)\\ &= \lim_{n\to\infty} M_{t_n}(u(s_y^{(n)}(r+\varepsilon_n)))g_n M_{t_n}(u(-(r+\varepsilon_n)))\\ &= \lim_{n\to\infty} M_{t_n}(u(s_y^{(n)}(r+\varepsilon_n)))g_n M_{t_n}(u(-r))u\end{aligned}$$

for $r \in O$ by (6.20). Comparing this with the definition of $\psi(r)$ we see that

$$\psi(r) = \lim_{n\to\infty} u_n' \psi_n(r) u$$

† As the reader may notice, the automorphism M_{t_n} does the speeding-up in this more general setting.

for some $u_n' \in U$. As

$$\lim_{n\to\infty} \psi_n(r) = \psi(r)$$

we must have $\lim_{n\to\infty} u_n' = u' \in U$ and

$$\psi(r) = u'\psi(r)u,$$

or equivalently

$$\psi(r)u\psi(r)^{-1} = (u')^{-1} \in U.$$

As this holds for all $u \in U$, the lemma follows. □

Using the same argument as in the proof of Theorem 6.18 and Lemma 6.22 it now follows that μ is invariant under all elements $\psi(r)$ for all $r \in (-1,1)$. Analyzing the construction of ψ we also see together with Lemma 6.22 that

$$\psi(r) \in \left\{ \left(\begin{pmatrix} * & \\ & * \end{pmatrix}, \begin{pmatrix} * & * \\ & * \end{pmatrix} \right) \right\} \cap N_G(U).$$

We claim also that ψ is not constant. Indeed if ψ is constant, then $\alpha = 0$ which implies that $\psi_0 = \psi$ is also constant in contradiction to the definition of T_n.

From this it follows quickly that

$$\psi(r) = \left(\begin{pmatrix} \frac{1}{1-\alpha r} & \\ & 1-\alpha r \end{pmatrix}, \begin{pmatrix} \frac{1}{1-\alpha r} & \beta(r) \\ & 1-\alpha r \end{pmatrix} \right)$$

for some rational function $\beta(r)$.
Case I: Assume first that $\alpha = 0$ so that

$$\psi(r) = \left(\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & \beta(r) \\ & 1 \end{pmatrix} \right)$$

for some nonconstant $\beta(r)$. This gives that μ is invariant under the horosphere

$$\left\{ \left(\begin{pmatrix} 1 & * \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \right) \right\},$$

and the result follows from Theorem 5.11.
Case II: Suppose now $\alpha \in [-1,1]$ is nonzero so that

$$\psi(\tfrac{1}{2}) = \left(\begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}, \begin{pmatrix} \lambda & s \\ & \lambda^{-1} \end{pmatrix} \right)$$

for some positive $\lambda \neq 1$. We claim that we may assume in the following that $s = 0$. In fact, replacing μ by

$$\left(\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & s' \\ & 1 \end{pmatrix} \right)_* \mu$$

gives a new measure that is still invariant under U and is also invariant under

$$\left(\begin{pmatrix}1&\\&1\end{pmatrix},\begin{pmatrix}1&s'\\&1\end{pmatrix}\right)\left(\begin{pmatrix}\lambda&\\&\lambda^{-1}\end{pmatrix},\begin{pmatrix}\lambda&s\\&\lambda^{-1}\end{pmatrix}\right)\left(\begin{pmatrix}1&\\&1\end{pmatrix},\begin{pmatrix}1&-s'\\&1\end{pmatrix}\right)$$
$$=\left(\begin{pmatrix}\lambda&\\&\lambda^{-1}\end{pmatrix},\begin{pmatrix}\lambda&s+\lambda^{-1}s'-\lambda s'\\&\lambda^{-1}\end{pmatrix}\right).$$

Since $\lambda \neq \lambda^{-1}$ we can choose s' so that the new measure is invariant under the diagonally embedded element

$$a=\left(\begin{pmatrix}\lambda&\\&\lambda^{-1}\end{pmatrix},\begin{pmatrix}\lambda&\\&\lambda^{-1}\end{pmatrix}\right).$$

We assume without loss of generality $\lambda \in (0,1)$. We note that a acts ergodically with respect to μ. Indeed if $f \in L^2(X,\mu)$ is a-invariant we may apply Proposition 2.13 and obtain that f is also U-invariant and so constant by our assumption on μ. We will continue the proof in two different ways (using entropy, and in a more special case without entropy theory). □

Finishing the proof of Theorem 6.20 using entropy theory. By Theorem 10.2(2) we have

$$h_\mu(a) \geqslant h_\mu(a,U) = 2|\log\lambda|,$$

since U belongs to the stable horospherical subgroup G_a^- of a and μ is U-invariant (which forces the leafwise measures μ_x^U to be the Haar measure on U for a.e. x and makes it easy to calculate $h_\mu(a,U)$, see the easy half of Theorem 10.5). We now consider the opposite horospherical subgroup

$$G_a^+ = \left\{\left(\begin{pmatrix}1&\\s_1&1\end{pmatrix},\begin{pmatrix}1&\\s_2&1\end{pmatrix}\right)\,\middle|\, s_1,s_2\in\mathbb{R}\right\}$$

and the subgroup

$$U^+ = \left\{\left(\begin{pmatrix}1&\\s&1\end{pmatrix},\begin{pmatrix}1&\\s&1\end{pmatrix}\right)\,\middle|\, s\in\mathbb{R}\right\}\subseteq G,$$

which should be considered as opposite to U.

Case A: It could be that the leafwise measure $\mu_x^{G_a^+}$ (which can also be used to describe the entropy since $h_\mu(a)=h_\mu(a^{-1})$) are almost surely supported on U^+. Suppose this is the case. Then

$$2|\log\lambda| = h_\mu(a,U) \leqslant h_\mu(a) = h_\mu(a,G_a^+) = h_\mu(a,U^+),$$

where we used the above inequality, Theorem 10.2(2) applied to G_a^+, and our assumption regarding the leafwise measures. However, together with (the difficult part of) Theorem 10.5 applied to U^+ this shows that μ is invariant under U^+. Since U and U^+ generate the diagonally embedded copy of $\mathrm{SL}_2(\mathbb{R})$, we may now refer to Theorem 6.17 which proves the theorem.

CASE B: Now suppose that the leafwise measures $\mu_x^{G_a^+}$ are not almost surely supported on U^+. Section 9.3.1 shows that there does not exist a set X' of full measure such that $g{\scriptstyle\bullet}x, x \in X'$ with $g \in G_a^+$ implies that $g \in U^+$. We now use this, again in a more general setting, to produce tuples of uniformly generic points with a specific relationship to each other.

Lemma 6.23. *Suppose $X = \Gamma\backslash G$ for some Lie group G and discrete subgroup Γ, let $a \in G$ be such that Ad_a is diagonalizable with positive eigenvalues, and let $U < G_a^-$ be a unipotent subgroup. Suppose μ is a U-invariant and ergodic probability measure on X which is also a-invariant. Let $U^+ \leqslant G_a^+$ be a subgroup and suppose that the leafwise measures $\mu_x^{G_a^+}$ are not almost surely supported on U^+. Then there exists a uniformly generic set of points K for the action of U, two points $y = g{\scriptstyle\bullet}x$ and x with $y \in G_a^+ \smallsetminus U^+$, and infinitely many $n \geqslant 1$ with*

$$a^{-n}{\scriptstyle\bullet}y = a^{-n}ga^n{\scriptstyle\bullet}(a^{-n}{\scriptstyle\bullet}x) \in K \text{ and } a^{-n}{\scriptstyle\bullet}x \in K.$$

PROOF. Let K be a sequence of uniformly generic sets of points for the action of U with $\mu(K) > \frac{9}{10}$. By Proposition 2.13 μ is also ergodic with respect to the action of a. Let

$$X' = \left\{ x \in X \mid \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_K(a^{-n}{\scriptstyle\bullet}x) > \tfrac{9}{10} \right\}, \tag{6.21}$$

so that $\mu(X') = 1$ by the pointwise ergodic theorem applied to μ and to the action of a. Applying Section 9.3.1 to X' and the subgroups $U^+ \leqslant G_a^+$, it follows that there exist points $y = g{\scriptstyle\bullet}x$ and $x \in X'$ with $g \in G_a^+ \smallsetminus U^+$, and the lemma follows. $\square$

Applying Lemma 6.23 in our case, we find two points $y = g{\scriptstyle\bullet}x$ and x with

$$g = \left(\begin{pmatrix} 1 & \\ c & 1 \end{pmatrix}, \begin{pmatrix} 1 & \\ c' & 1 \end{pmatrix} \right) \in G_a^+$$

and $c \neq c'$ (since by assumption $g \notin U^+$). Moreover, we have infinitely many $n \geqslant 1$ with $y_n = a^{-n}{\scriptstyle\bullet}y$, $x_n = a^{-n}{\scriptstyle\bullet}x$ in K. Notice that

$$g_n = a^{-n}ga^n = \left(\begin{pmatrix} 1 & \\ \lambda^{2n}c & 1 \end{pmatrix}, \begin{pmatrix} 1 & \\ \lambda^{2n}c' & 1 \end{pmatrix} \right).$$

We now start the arguments from the very beginning of the proof again, using the points $y_n = g_n{\scriptstyle\bullet}x_n, x_n \in K$. The calculation there now simplifies, since†

† We write $*$ for an entry of the matrices that we do not care about.

$$u\left(\tfrac{s_x}{1-cs_x}\right) g u\left(-s_x\right) = \left(\begin{pmatrix} * & 0 \\ c & 1-cs_x \end{pmatrix}, \begin{pmatrix} * & -s_x + s_y(1-c's_x) \\ c' & 1-c's_x \end{pmatrix}\right),$$

which after conjugation by a^{-n} gives

$$u\left(\lambda^{-2n}\tfrac{s_x}{1-cs_x}\right) g_n u\left(-\lambda^{-2n}s_x\right) = \\ \left(\begin{pmatrix} * & 0 \\ \lambda^{-2n}c & 1-cs_x \end{pmatrix}, \begin{pmatrix} * & \lambda^{-2n}s_x^2\tfrac{c-c'}{1-cs_x} \\ \lambda^{2n}c' & 1-c's_x \end{pmatrix}\right).$$

Notice first that in the first matrix on the right the top right entry is zero (and so equal to the same entry of g_n), so the matrix is $\phi_n(\lambda^{-2n}s_x)$. Then notice that the top right entry of the second matrix is non-zero and diverges to infinity as $n \to \infty$. This shows that λ^{-2n} is too large a speeding-up parameter. More precisely, the speeding-up parameter T_n must satisfy

$$\frac{T_n}{\lambda^{-2n}} \longrightarrow 0$$

as $n \to \infty$. This shows (after choosing a converging subsequence and taking the limit) that

$$\psi(r) = \left(\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & \kappa r^2 \\ & 1 \end{pmatrix}\right)$$

for some $\kappa \neq 0$, and that μ is invariant under $\psi(r)$ for $r \in (-1,1)$. Thus we are now back in Case I of the beginning of the proof, which concludes the argument. □

In the second, entropy-free, argument we will assume that $\Gamma = \Gamma_1 \times \Gamma_2$ for lattices $\Gamma_1, \Gamma_2 \in \mathrm{SL}_2(\mathbb{R})$.

Proof of Theorem 6.20 without using entropy theory. We set

$$X_i = \Gamma_i \backslash \mathrm{SL}_2(\mathbb{R})$$

and consider the projections $\pi_1(x,x') = x$ and $\pi_2(x,x') = x'$ from

$$X = \Gamma_1 \backslash \mathrm{SL}_2(\mathbb{R}) \times \Gamma_2 \backslash \mathrm{SL}_2(\mathbb{R})$$

to X_1 respectively X_2. Let $\mu_i = (\pi_i)_*\mu$, and obtain in this way a horocycle-invariant probability measure on each X_i for $i = 1, 2$. By Theorem 5.3 these measures are therefore known to be algebraic, which leads us to three cases.

[(i)]μ_1 and μ_2 are both periodic orbit measures, which reduces the classification of the possibilities for μ to the classification of invariant measures on $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. One of the two measures is a periodic orbit measure, but the other is Haar measure. Both measures are Haar measures, in which case μ is, by definition, a joining for the horocycle flow.

We consider the case (i) dealt with and show that case (ii) is also quite easy to handle. Suppose without loss of generality that $(\pi_1)_*\mu = \mu_1$ is a periodic orbit measure while $(\pi_2)_*\mu = m_{X_2}$ is the Haar measure. Suppose that $s > 0$ is the period of the horocycle flow on $\operatorname{Supp}\mu_1$. In this case $\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix}$ acts trivially on the first factor and ergodically on the second. Applying the decomposition $\mu = \int \mu^{\mathscr{A}}_{(x,x')}\,\mathrm{d}\mu(x,x')$ into conditional measures for the σ-algebra

$$\mathscr{A} = \mathscr{B}_{X_1} \times \{\varnothing, X_2\},$$

we notice that u_s preserves every element of $\mathscr{A}$ modulo μ (since μ-almost everywhere u_s does not change the first component). By [**?**, Cor. 5.24] this implies a.s. that

$$(u_s)_*\mu^{\mathscr{A}}_{(x,x')} = \mu^{\mathscr{A}}_{u_s\cdot(x,x')} = \mu^{\mathscr{A}}_{(x,x')}.$$

To summarize we have that $\mu^{\mathscr{A}}_{(x,x')}$ does not depend on x', is supported on $\{x\} \times X_2$, and is invariant under the horocycle flow on $\{x\} \times X_2$. Since

$$m_{X_2} = (\pi_2)_*\mu = (\pi_2)_* \int \mu^{\mathscr{A}}_{(x,x')}\,\mathrm{d}\mu(x,x') = \int (\pi_2)_*\mu^{\mathscr{A}}_{(x,x')}\,\mathrm{d}\mu_{(x,x')}$$

expresses m_{X_2} as an integral convex combination of other horocycle-invariant probability measures, it follows by ergodicity that

$$(\pi_2)_*\mu^{\mathscr{A}}_{(x,x')} = m_{X_2},$$

or equivalently

$$\mu^{\mathscr{A}}_{(x,x')} = \delta_x \times m_{X_2}$$

for μ-almost every (x,x'). It follows that $\mu = (\pi_1)_*\mu \times m_{X_2}$ is algebraic.

So we now (and for the rest of the section) consider (iii), the most interesting case, of a joining μ with

$$(\pi_i)_*\mu = m_{X_i}$$

for $i = 1, 2$. By the beginning of the proof on page 207, we can derive additional transverse invariance. Either we are in Case I, in which case we have horospherical invariance and hence the trivial joining $\mu = m_{X_1} \times m_{X_2}$ by an argument very similar to (ii), or we may assume after modifying μ slightly that μ is invariant under a diagonally embedded diagonal element[†] a.

We again set $\mathscr{A} = \mathscr{B}_{X_1} \times \{\varnothing, X_2\}$ and consider the conditional measures $\mu^{\mathscr{A}}_{(x,x')}$ which describes μ on the atom $[(x,x')]_{\mathscr{A}} = \{x\} \times X_2$ for μ-almost every (x,x').

If $\mu^{\mathscr{A}}_{(x,x')}$ is not atomic almost everywhere, then $\mu = m_{X_1} \times m_{X_2}$ is the trivial joining.

[†] This element a will be used later in the proof, but its entropy properties will not be used.

For each $m \geqslant 1$ let K_m be a set of uniformly generic points with $\mu(K_m) > 1 - \frac{1}{m}$. Replacing K_m by $K_1 \cup \cdots \cup K_m$ if necessary, we may assume that

$$K_1 \subseteq K_2 \subseteq \cdots$$

and let

$$X' = \bigcup_{m \geqslant 1} K_m$$

so that $\mu(X') = 1$. Since we assume that $\mu_{(x,x')}^{\mathscr{A}}$ is not atomic a.e., we see that $\mu_{(x,x')}^{\mathscr{A}}|_{X'}$ is not atomic μ-almost everywhere. Therefore there exists some set K_m and some $(x, x') \in X$ such that $\mu_{(x,x')}^{\mathscr{A}}|_{K_m}$ is non-atomic. As $\operatorname{Supp} \mu_{(x,x')}^{\mathscr{A}} \subseteq \{x\} \times X_2$, we see that there exists a sequence of points

$$x_n, y_n = g_n \boldsymbol{\cdot} x_n \in K_m$$

with $x_n \neq y_n$, $\pi_1(x_n) = x = \pi_1(y_n)$ where the displacement satisfies

$$g_n = \left(I, \begin{pmatrix} a_n' & b_n' \\ c_n' & d_n' \end{pmatrix}\right) \neq I$$

with

$$\begin{pmatrix} a_n' & b_n' \\ c_n' & d_n' \end{pmatrix} \longrightarrow I$$

as $n \to \infty$. We now apply the argument from the beginning of the proof on page 207 to see that μ is invariant under the action of

$$\left\{\left(I, \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix}\right)\right\},$$

either because

$$g_n = \left(I, \begin{pmatrix} 1 & b_n' \\ & 1 \end{pmatrix}\right)$$

with $b_n \to 0$ as $n \to \infty$ and so we may apply the centralizer lemma (Lemma 6.7), or because we are back in Case I of the transverse divergence argument.

Thus we may suppose that $\mu_{(x,x')}^{\mathscr{A}}$ is atomic a.e., in which case we make the following claim.

There exists some $m \geqslant 1$ and functions[†] $f_1, \dots, f_m : X_1 \to X_2$ such that the measure $\mu_{(x,x')}^{\mathscr{A}}$ may be expressed in the form[‡]

[†] In some sense it is better to think of $\{f_1, \dots, f_m\}$ as a correspondence or an m-valued function from X_1 to X_2.

[‡] In the following we will write $\mu_{(x,\cdot)}^{\mathscr{A}} = \mu_{(x,x')}^{\mathscr{A}}$ as the conditional measure does not depend on the second coordinate x'.

$$\mu^{\mathscr{A}}_{(x,\cdot)} = \frac{1}{m}\sum_{i=1}^{m} \delta_{(x,f_i(x))}$$

for m_{X_1}-almost every x.

We note that the claim shows in particular that μ is determined by m_{X_1} and the set of functions $\{f_1,\dots,f_m\}$

We define a function

$$f((x,x')) = \mu^{\mathscr{A}}_{(x,\cdot)}\left(\{(x,x')\}\right)$$

which by the previous claim is positive almost surely. Notice that $u_s^{-1}\mathscr{A} = \mathscr{A}$, so that

$$\begin{aligned} f\left(u_s\text{•}(x,x')\right) &= \mu^{\mathscr{A}}_{u_s\text{•}(x,\cdot)}\left(\{u_s\text{•}(x,x')\}\right) \\ &= (u_s)_*\mu^{u_s^{-1}\mathscr{A}}_{(x,\cdot)}\left(\{u_s\text{•}(x,x')\}\right) \\ &= \mu^{\mathscr{A}}_{(x,\cdot)}\left(\{(x,x')\}\right) = f((x,x')) \end{aligned}$$

by [**?**, Cor. 5.24]. This shows that f is a u_s-invariant function†. Therefore, f is constant μ-almost everywhere, so that we also have

$$\mu^{\mathscr{A}}_{(x,\cdot)}\left(\{(x,x')\}\right) = f\left((x,x')\right) = f\left((x,y')\right) = \mu^{\mathscr{A}}_{(x,\cdot)}\left(\{(x,y')\}\right)$$

if both (x,x') and (x,y') belong to this full-measure set and share the same first coordinate. As $\mu^{\mathscr{A}}_{(x,\cdot)}$ is by construction a probability measure, it follows that there is some $m \geqslant 1$ and m points

$$\{f_1(x),\dots,f_m(x)\} \subseteq X_2$$

such that

$$\mu^{\mathscr{A}}_{(x,\cdot)} = \frac{1}{m}\sum_{i=1}^{m} \delta_{(x,f_i(x))},$$

for μ-almost every (x,x') (or equivalently for m_{X_1}-a.e. x).

We may choose the functions $f_1, f_2, \dots, f_m : X_1' \to X_2$ to be measurable on a subset $X_1' \subseteq X_1$ of full measure.

We let X_1' be the set on which $\mu^{\mathscr{A}}_{(x,\cdot)}$ is defined and has the property in the last claim. Using a countable basis of the topology of X_2, we find a sequence of finite or countable partitions $(\mathscr{P}_n)$ such that

$$\mathscr{P}_n \leqslant \sigma\left(\mathscr{P}_{n+1}\right)$$

† This function is also measurable, which the reader may check by exhibiting f as a pointwise limit of a sequence of measurable functions using $\mu^{\mathscr{A}}_{(x,x')}(B)$ for elements B chosen from a refining sequence of partitions of X_2. We skip this proof, but refer the reader to the next step for a similar argument.

and

$$\mathscr{B}_{X_2} = \bigvee_{n=1}^{\infty} \sigma\left(\mathscr{P}_n\right).$$

We also order the elements of

$$\mathscr{P}_n = \{P_{n,1}, \dots\}$$

where we may assume that $P_{n,i}$ has diameter smaller than $\frac{1}{n}$ for $i \geqslant 1$. We will define f_1 as in the claim to be a limit of a sequence of measurable functions $\left(f_1^{(n)}\right)$.

Pick some $y_{1,i} \in P_{1,i}$ for $i \geqslant 1$ and define

$$\begin{aligned} f_1^{(1)}(x) = y_{1,1} \text{ on } B_{1,1} &= \{x \in X_1 \mid \mu_{(x,\cdot)}^{\mathscr{A}}(\{x\} \times P_{n,1}) > 0\} \\ f_1^{(1)}(x) = y_{1,2} \text{ on } B_{1,2} &= \{x \in X_1 \mid \mu_{(x,\cdot)}^{\mathscr{A}}(\{x\} \times P_{n,2}) > 0\} \smallsetminus B_{1,1}, \end{aligned}$$

and so on. In defining $f_1^{(2)}$ we again use some $y_{2,i} \in P_{2,i}$ for $i \geqslant 1$, but we require the property that $f_1^{(2)}(x)$ and $f_1^{(2)}(x)$ belong to the same partition element of $\mathscr{P}_1$. We can ensure this by requiring that each $P_{1,i}$ is split into finitely many partition elements of $\mathscr{P}_2$, and the subsets of $P_{1,i}$ appear before the subsets of $P_{1,j}$ in the enumeration of the elements of $\mathscr{P}_2$ whenever $i < j$. With this allowed assumption we can simply follow the same procedure for the construction of $f_1^{(2)}$. Repeating this for all n we get a sequence of piece-wise constant (and, in particular, measurable) functions $f_1^{(n)}$ with the property that

$$\mathsf{d}(f_1^{(n)}(x), f_1^{(k)}(x)) < \tfrac{1}{k}$$

if $n > k$. Therefore

$$f_1(x) = \lim_{n\to\infty} f_1^{(n)}(x)$$

exists for all $x \in X_1$ and defines a measurable function $f_1 : X_1' \to X_2$. By construction there exists for every n some $Q_n \in \mathscr{P}_n$ with $f_1(x) \in \overline{Q_n}$, $\mu_{(x,\cdot)}^{\mathscr{A}}(\{x\} \times Q_n) > 0$ and so also $\mu_{(x,\cdot)}^{\mathscr{A}}(Q_n) \geqslant 1/m$. Since Q_n has diameter $\leqslant 1/n$ we see that $\bigcap_{n=1}^{\infty} \overline{Q_n} = \{(x, f_1(x))\}$ which gives $\mu_{(x,\cdot)}^{\mathscr{A}}(\{(x, f_1(x))\}) = 1/m$ for all $x \in X_1'$. If $m > 1$ then we remove $(x, f_1(x))$ from $\mu_{(x,\cdot)}^{\mathscr{A}}$ by replacing the measure with

$$\mu_{(x,\cdot)}^{\mathscr{A}} - \tfrac{1}{m}\delta_{(x,f_1(x))}$$

and repeat the procedure as necessary.

As the above arguments already show we will work more and more with points in X_1 and will below use frequently dynamical arguments on X_1 with respect to the factor measure $m_{X_1} = (\pi_1)_*\mu$ to derive additional properties of the functions $f_1, \dots, f_m$. To simplify the notation for these arguments we set

$$u = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} \text{ and } a = \begin{pmatrix} \lambda & \\ & \lambda^{-1} \end{pmatrix}$$

with $\lambda \in (0,1)$ as before. Since we already know that (u,u) and (a,a) preserve μ (which is determined by m_{X_1} and the functions $f_1, \dots, f_m$), we get the "compatibility properties" for the functions $f_1, \dots, f_m$. In fact

$$\{f_1(u \cdot x), \dots, f_m(u \cdot x)\} = u \cdot \{f_1(x), \dots, f_m(x)\}$$

a.s. and similarly with u replaced by a. Indeed, by [**?**, Cor. 5.24] we have

$$\begin{aligned} (u,u) \cdot (\{x\} \times \{f_1(x), \dots, f_m(x)\}) &= (u,u) \cdot \operatorname{Supp} \mu_{(x,\cdot)}^{\mathscr{A}} \\ &= \operatorname{Supp} (u,u)_* \mu_{(x,\cdot)}^{(u,u)^{-1}\mathscr{A}} \\ &= \operatorname{Supp} \mu_{(u\cdot x,\cdot)}^{\mathscr{A}}, \end{aligned}$$

which in turn may be written as

$$\operatorname{Supp} \mu_{(u\cdot x,\cdot)}^{\mathscr{A}} = \{u \cdot x\} \times \{f_1(u \cdot x), \dots, f_m(u \cdot x)\},$$

m_{X_1}-almost everywhere. This is the claimed equivariance property of the set of functions for u, the case of a is identical. We now suppose that these equivariance formulas hold for all $x \in X_1'$ and that X_1' is invariant under both u and a.

Our main aim is to show that for the element

$$v_t = \begin{pmatrix} 1 & \\ t & 1 \end{pmatrix}$$

we have the analogous formula

$$\{f_1(v_t \cdot x), \dots, f_m(v_t \cdot x)\} = v_t \cdot \{f_1(x), \dots, f_m(x)\}, \tag{6.22}$$

which will show that μ (which is determined by m_{X_1} and $f_1, \dots, f_m$) is also (v_t, v_t)-invariant.

Now that we have set the stage and know what we are aiming at, it is time to get to the heart of the matter, namely the following ingenious argument due to Ratner which we first outline in the case $m = 1$ as follows.

The proof resembles in some ways a double Hopf argument (see [**?**, Sec. 9.5]). Consider the points

$$(x, f_1(x)) \text{ and } (v_t \cdot x, f_1(v_t \cdot x)) = (v_t \cdot x, g \cdot f_1(x))$$

(with $g = v_t$ being our goal). Applying the equivariance property for a to f_1 we obtain

$$\begin{aligned} f_1(a^{-n} v_t \cdot x) &= a^{-n} g f_1(x) = a^{-n} g a^n \cdot f_1(a^{-n} \cdot x) \\ &= f_1(v_{\lambda^{2n} t} a^{-n} \cdot x). \end{aligned} \tag{6.23}$$

Using the ergodic theorem for the action of a^{-1}, and the fact that f_1 is nearly continuous by Lusin's theorem, we see that for many $n \geqslant 1$ the point in (6.23) and the point $f_1(a^{-n} \cdot x)$ are close together since $\lambda^{2n} t \to 0$ as $n \to \infty$. Unfortunately this does not imply much about g itself, because we could certainly have† $a^{-n} g a^n \to \infty$ as $n \to \infty$.

Using u^ℓ instead of a^{-n} gives a better situation, as follows. If t is very small, then

$$u^\ell v_t u^{-\ell} = \begin{pmatrix} 1 & \ell \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & \\ t & 1 \end{pmatrix} \begin{pmatrix} 1 & -\ell \\ & 1 \end{pmatrix} = \begin{pmatrix} 1+t\ell & -\ell^2 t \\ t & 1-t\ell \end{pmatrix}$$

will still be small for ℓ smaller than $1/\sqrt{|t|}$‡. Using once again the ergodic theorem for u and the fact that f_1 is nearly continuous by Lusin's theorem, we obtain that for most ℓ in $[0, 1/\sqrt{|t|})$ we have that

$$u^\ell \cdot f_1(v_t \cdot x) = \left(u^\ell g u^{-\ell}\right) u^\ell \cdot f_1(x)$$

is very close to $u^\ell \cdot f_1(x)$. However, this time $u^\ell g u^{-\ell}$ is a polynomial in ℓ (rather than an exponential function) which will allow us to derive constraints on the entries of g. Since ℓ is constrained to an interval $[0, 1/\sqrt{|t|})$, the constraints on the entries of

$$(v_t, g) = \left(\begin{pmatrix} 1 & \\ t & 1 \end{pmatrix}, \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right)$$

will take the form of inequalities

$$|c| \ll |t|, |d-a| \ll \sqrt{|t|}, |b| \ll 1.$$

Since we are aiming to prove that $g = v_t$, this also appears to be a hopeless venture. In the argument below we will be double-dipping in the following sense. By using a^{-n} we will be able to make t smaller and smaller indefinitely (without winning back any information about g). By using u^ℓ for longer and longer intervals as n grows, we will be able to obtain better and better constraints on the entries of g.

In order for this double-dipping to work, we need to define some sets, for which we will return to the general case of $m \geqslant 1$. By Lusin's theorem there exists a compact set $K \subseteq X_1'$ with $\mu(K) > 1 - \frac{1}{30}$ such that $f_i|_K$ is continuous for $i = 1, \ldots, m$. We define

† The geodesic flow has many pairs of orbits that are close for a large percentage of time without being close for a good (meaning algebraic) reason.

‡ It might appear disadvantageous to use u instead of a^{-1}, since $a^{-n} v_t a^n$ actually converges to I as $n \to \infty$, whereas the corresponding expression for u is only small for certain times. The utility of u for the argument will become clear soon.

$$Y_1 = \left\{ x \in X_1' \,\middle|\, \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{1}_K \left(u^\ell \boldsymbol{\cdot} x \right) \geqslant \tfrac{9}{10} \text{ for all } L \geqslant 1 \right\}$$

and

$$Y_2 = \left\{ x \in X_1' \,\middle|\, \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{Y_1 \cap K} \left(a^{-n} \boldsymbol{\cdot} x \right) > \tfrac{1}{2} \right\}.$$

By the maximal ergodic theorem applied to the action of u we have $m_{X_1}(Y_1) \geqslant \frac{2}{3}$, hence $m_{X_1}(Y_1 \cap K) > \frac{1}{2}$, and by the pointwise ergodic theorem applied to the action of a we have $m_{X_1}(Y_2) = 1$. We now derive the promised inequalities.

(3) **Lemma 6.24 (Linearization for the correspondence).** *Depending on K there exists some $\delta > 0$ such that for all*

$$y = v_t \boldsymbol{\cdot} x, x \in Y_1 \cap K$$

with $t \in (-\delta, \delta)$ and all i there exists j such that

$$f_i(y) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \boldsymbol{\cdot} f_j(x)$$

for some $a, b, c, d \in \mathbb{R}$ with $|c| \ll |t|$, $|a-1|, |d-1| \ll \sqrt{|t|}$ and $|b| \ll 1$.

In the proof we will use the fact that $y = v_t \boldsymbol{\cdot} x$ and x satisfy that $u^\ell \boldsymbol{\cdot} x$ and $u^\ell \boldsymbol{\cdot} y$ are close together as long as $\ell^2 t$ is small. Applying $f_1, \dots, f_m$ we have the weaker property that the image points are some fixed percentage of this time window (if $m = 1$ this would be 80%) close in X_2. Here we will need the following lemma.

Lemma 6.25 (Linearization for two orbits). *Let $X = \Gamma \backslash \mathrm{SL}_2(\mathbb{R})$ be a quotient by a lattice. For any $p \in (0,1)$ and any compact subset $K \subseteq X$ there exists some $\kappa \in (0,1]$ with the following property. Suppose that $L > 1$, the points $x \in K$ and $y \in X$ satisfy*

$$\frac{1}{L} \left| \left\{ \ell \in \{0, \dots, L-1\} \mid u^\ell \boldsymbol{\cdot} x \in K \text{ and } \mathsf{d}(u^\ell \boldsymbol{\cdot} x, u^\ell \boldsymbol{\cdot} y) < \kappa \right\} \right| \geqslant p.$$

Then $y = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \boldsymbol{\cdot} x$ with $|c| \ll_p \frac{1}{L^2}, |a-1| \ll_p \frac{1}{L}, |d-1| \ll_p \frac{1}{L}$, and $|b| \ll_p 1$.

PROOF. The main idea of the proof is similar to the proof of the non-divergence for the horocycle flow in 2 in Section 4.1. We let $\rho \in (0,1])$ be chosen so that 2ρ is an injectivity radius on K, and let

$$S = \{\ell \in \{0, \dots, L\} \mid u^\ell \boldsymbol{\cdot} x \in K \text{ and } \mathsf{d}(u^\ell \boldsymbol{\cdot} y, u^\ell \boldsymbol{\cdot} x) < \rho\}.$$

For $\ell \in S$ we let $g_\ell \in \mathrm{SL}_2(\mathbb{R})$ be the unique matrix satisfying $u^\ell \boldsymbol{\cdot} y = g_\ell u^\ell \boldsymbol{\cdot} x$ and

$$\mathsf{d}\left(g_\ell, I\right) = \mathsf{d}\left(u^\ell \boldsymbol{\cdot} y, u^\ell \boldsymbol{\cdot} x\right) < \rho.$$

We say that $\ell, m \in S$ are *equivalent* if the corresponding points $u^\ell \boldsymbol{\cdot} y, u^\ell \boldsymbol{\cdot} x$ respectively $u^m \boldsymbol{\cdot} y, u^m \boldsymbol{\cdot} x$ are close and are so 'for the same reason'. More precisely we define $\ell, m \in S$ to be equivalent if

$$g_m = u^{m-\ell} g_\ell u^{-(m-\ell)}$$

and that $d(u^{k-\ell} g_\ell u^{-(k-\ell)}, I) < \rho$ for all[†] k between ℓ and m.

Suppose for a moment that S consists of one equivalence class. If $0 \in S$ then we already defined g_0. Otherwise we let $g_0 = u^{-\ell} g_\ell u^\ell$ for some $\ell \in S$. In any case we let

$$g_0 = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

so that

$$u^\ell \begin{pmatrix} a & b \\ c & d \end{pmatrix} u^{-\ell} = \begin{pmatrix} a + c\ell & b + (d-a)\ell - c\ell \\ c & d - c\ell^2 \end{pmatrix}$$

has distance no more than ρ from I for at least the fraction p of the points in $\{0, 1, \dots, L\}$. For those choices of ℓ, we also have

$$|b + (d-a)\ell - c\ell^2| \ll \rho$$

for some absolute implied constant, which depends only on the Riemannian metric. By Lemma[‡] 4.6 this implies that

$$|b + (d-a)\ell + c\ell^2| \ll_p \rho \leqslant 1$$

for all $\ell = 0, \dots, L$, potentially with a different implied constant. The estimates in the lemma now follow by using $\ell = 0$ to see that $|b| \ll_p 1$, $\ell = \frac{L}{2}$ and $\ell = L$ to get

$$|(d-a)\tfrac{L}{2} + c\tfrac{L^2}{4}| \ll_p 1 \text{ and } |(d-a)L + cL^2| \ll_p 1,$$

which gives $|(d-a)L| \ll_p 1$ and $|cL^2| \ll_p 1$. This also implies $ad = ad - bc + O(\frac{1}{L^2}) = 1 + O(\frac{1}{L^2})$. Using the diagonal entry of $u^\ell g_0 u^{-\ell}$ we also see that $|a-1| \ll_p \rho$, $|d-1| \ll_p \rho$. If ρ is sufficiently small, then $(d+1) \geqslant 1$ and so $(a-1)(d+1) = ad - 1 + a - d = O(\frac{\rho}{L} + \frac{1}{L^2})$ implies $|a-1| \ll_p \frac{1}{L}$. The estimate $|d-1| \ll_p \frac{1}{L}$ follows in the same way.

To prove that S contains only one equivalence class, we assume the opposite, choose κ sufficiently small and will again use Lemma 4.6 to derive a contradiction. In fact by that lemma we may choose $\kappa < \rho$ so that

[†] Note that possibly not all of these integers k belong to S due to the additional requirement $u^k \boldsymbol{\cdot} x \in K$ in the definition of S.

[‡] Strictly speaking we use a discrete analogue of the lemma. However, we only need the quadratic case and the proof easily extends to the discrete case.

$$\frac{1}{T}\left|\{t \in \{0,\dots,T-1\} \mid |f(t)| \leqslant \frac{\kappa}{\rho}\|f\|_{\infty,T}\}\right| < \frac{p}{3}$$

for any quadratic polynomial f where

$$\|f\|_{\infty,T} = \sup_{0 \leqslant t \leqslant T-1} |f(t)|.$$

Choosing κ possibly even smaller (to accommodate for the Lipschitz constant of switching between the Riemannian metric and the matrix norm near the identity) we also obtain

$$\frac{1}{T}\left|\{t \in \{0,\dots,T-1\} \mid \mathsf{d}(u^t h u^{-t}, I) \geqslant \kappa\}\right| < \frac{p}{3}$$

if $h \in B_G^\rho$ is such that $\mathsf{d}(u^{-1}hu, I) \geqslant \rho$ or $\mathsf{d}(u^T h u^{-T}, I) \geqslant \rho$.

For each equivalence class $[\ell]$ with $\ell \in S$ as a representative, we define the *protecting intervals* $P_{[\ell]}$ to be the maximal subinterval of $\{0,\dots,L\}$ on which $\mathsf{d}(u^{k-\ell}g_\ell u^{-(k-\ell)}, I) \leqslant \rho$ for all $k \in P_{[\ell]}$. By definition $[\ell] \subseteq P_{[\ell]}$. We may also assume that for each equivalence class $[\ell]$ and its interval $P_{[\ell]}$ we have $\mathsf{d}(u^{k-\ell}g_\ell u^{-(k-\ell)}, I) \geqslant \rho$ for k equal to the left end point minus one or equal to the right end point plus one. Indeed, for otherwise by maximality of $P_{[\ell]}$ those endpoints must be 0 and $L-1$ which gives that $P_{[\ell]} = \{0,\dots,L-1\}$ and so the lemma by the first part of the proof. Hence by our choice of κ

$$\frac{1}{|P_{[\ell]}|}|\underbrace{\{k \in P_{[\ell]} \mid \mathsf{d}(u^{k-\ell}g_\ell u^{-(k-\ell)}, I) \leqslant \kappa\}}_{[\ell]}| < \frac{p}{3}.$$

We also note that an element $\ell \in [0, L]$ could belong to two intervals $P_{[\ell_1]}$ and $P_{[\ell_2]}$ for $[\ell_1] \neq [\ell_2]$, but only to two. In fact suppose $\ell_1 < \ell_2 < \ell_3$ with

$$\ell \in P_{[\ell_1]} \cap P_{[\ell_2]} \cap P_{[\ell_3]}$$

and with $[\ell_1], [\ell_2], [\ell_3]$ all different. Since $u^{\ell_2}\cdot x \in K$ by definition of $S \ni \ell_2$, since $P_{[\ell_1]}$ is maximal interval on which $\mathsf{d}(u^{m-\ell_1}g_{\ell_1}u^{-(m-\ell_1)}, I) < \rho$, and since ρ is smaller than the injectivity radius at K, we see that $\ell_1 \notin P_{\ell_2}$. Since $P_{[\ell_1]} \ni \ell_1$ and $P_{[\ell_2]} \ni \ell_2$ are intervals, we must have $\ell_1 < \ell < \ell_2$. The same argument leads to $\ell_2 < \ell < \ell_3$, which is a contradiciton. Hence any integer between 0 and $L-1$ belongs to at most 2 protecting intervals.

We finally set

$$= \{\ell \in \{0,\dots,L-1\} \mid \mathsf{d}(u^\ell\cdot y, u^\ell\cdot x) < \kappa \text{ and } u^\ell\cdot x \in K\} \subseteq \bigcup_{[\ell]} [\ell]$$

and obtain

$$|| \leqslant \sum_{[\ell]} |_{[\ell]}| \leqslant \sum_{[\ell]} \frac{p}{3} |P_{[\ell]}| \leqslant \frac{2}{3} pL.$$

However, this contradicts our assumptions. Hence there can only be one equivalence class and the lemma follows. □

We return to the setting of Theorem 6.20 and apply the lemma above.

Proof of Lemma 6.24. Since K is compact and the functions $f_1, \dots, f_m$ restricted to K are continuous, the set

$$K' = \bigcup_{i=1}^{m} f_i(K)$$

is a compact subset of X_2. We set $p = \frac{8}{10m}$ and apply Lemma 6.25 to

$$X = X_2 = \Gamma_2 \backslash \mathrm{SL}_2(\mathbb{R})$$

and the compact set K'. This defines some $\kappa > 0$. Since $f_i(x) \neq f_j(x)$ for $i \neq j$ and all x in the domain of these functions by construction, we may also suppose that

$$\mathsf{d}(f_i(x), f_j(x)) > 2\kappa$$

for $x \in K$ and $i \neq j$. Again since f_i restricted to K is continuous we see that there exists a $\delta > 0$ such that

$$x, y = g \cdot x \in K, g \in \mathrm{SL}_2(\mathbb{R}) \text{ with } \mathsf{d}(g, I) < \delta \Longrightarrow \mathsf{d}(f_i(y), f_i(x)) < \kappa$$

for $i = 1, \dots, m$.

Suppose now that $t \in (-\delta, \delta)$ and $x, v_t \cdot x \in K \cap Y_1$. We can now find an interval $I_{x,y}$ of length $\gg_\delta 1/\sqrt{|t|}$ such that for $\ell \in I_{x,y}$ we have

$$\mathsf{d}(u^\ell v_t u^{-\ell}, I) = \mathsf{d}\left(\begin{pmatrix} 1 + \ell t & \ell^2 t \\ t & 1 - \ell t \end{pmatrix}, I\right) < \delta,$$

and (by definition of Y_1) for $\frac{8}{10}$ of all $\ell \in I_{x,y}$ we have $u^\ell \cdot x, u^\ell \cdot y \in K$ and so

$$\mathsf{d}\left(f_i\left(u^\ell \cdot y\right), f_i\left(u^\ell \cdot x\right)\right) < \kappa$$

for $i = 1, \dots, m$. By the properties of $\{f_i \mid i = 1, \dots, m\}$ and our choice of κ this also shows that for $\frac{8}{10}$ of all $\ell \in I_{x,y}$ we have that for all i there exists some $j = j(i, \ell) \in \{1, \dots, m\}$ with

$$\mathsf{d}\left(u^\ell \cdot f_i(y), u^\ell \cdot f_j(x)\right) < \kappa. \tag{6.24}$$

Thus for every i there exists a $j = j(i)$ and a fraction of the interval $I_{x,y}$ exceeding $\frac{8}{10m}$ in proportion such that (6.24) holds (with j independent of s). Applying Lemma 6.25 we obtain that

$$f_i(y) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot f_j(x)$$

with $|c| \ll |t|$, $|a-1| \ll \sqrt{|t|}$, $|d-1| \ll \sqrt{|t|}$ and $|b| \ll 1$. $\square$

We continue with the proof of Theorem 6.20. Let $t \in \mathbb{R}$ and

$$y = v_t \cdot x = \begin{pmatrix} 1 & \\ t & 1 \end{pmatrix} \cdot x, x \in Y_2.$$

Then for more than $\frac{1}{2}$ of all $n \geqslant 1$ we have $a^n \cdot x \in Y_1$, and similarly for y. Therefore, there are infinitely many $n \geqslant 1$ for which both $x_n = a^{-n} \cdot x \in Y_1$ and $y_n = a^{-n} \cdot y \in Y_1$. Choose one such n and notice that

$$y_n = \begin{pmatrix} 1 & \\ \lambda^{2n} t & 1 \end{pmatrix} \cdot x_n,$$

so that these points are, for large $n \geqslant 1$, extremely close. We now apply Lemma 6.25 to y_n and x_n. It follows that for every i there exists some j such that †

$$f_i(y_n) = g_n \cdot f_j(x_n) = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \cdot f_i(x_n)$$

with $|c_n| \ll \lambda^{2n}|t|$, $|a_n - 1| \ll \lambda^n \sqrt{|t|}$, $|d_n - 1| \ll \lambda^n \sqrt{|t|}$ and $|b_n| \ll 1$. Going back to x and y by applying the matrix a^n we see that for every i there exists some j with

$$f_i(y) = a^n g_n a^{-n} \cdot f_j(x) = a^n \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} a^{-n} \cdot f_j(x) = \begin{pmatrix} a_n & \lambda^{2n} b_n \\ \lambda^{-2n} c_n & d_n \end{pmatrix} \cdot f_j(x)$$

where $|\lambda^{-2n} c_n| \ll t$, $|a_n - 1| \ll \lambda^n \sqrt{|t|}$, $|d_n - 1| \ll \lambda^n \sqrt{|t|}$, and $|\lambda^{2n} b_n| \ll \lambda^{2n}$. Here it is crucial that the entries of the matrix $a^n g_n a^{-n}$ are uniformly bounded. Hence we may choose a subsequence such that $a^n g_n a^{-n}$ converges and $j = j(n)$ is constant along this subsequence. Hence we have shown that for every i and every pair $y = v_t \cdot x, x \in Y_2$ there exists some $j = j(x,t,i)$ and some $c = c(x,t,i) \in \mathbb{R}$ with

$$f_i(v_t \cdot x) = v_c \cdot f_j(x).$$

If $c = t$ almost surely and for all i we have obtained our objective (see below). So suppose $c \neq t$ for some choice of i and on a set of positive measure. In this case we are essentially in the same situation as in Case B on page 212 and we can conclude the argument as before. We note that the leafwise measures in Case B are just used to produce the situation that we already have: on every set of full measure we find points

† As was mentioned before we do not know at this stage any relationship between these displacement g_n for different n's.

$$(x, x') = (x, f_j(x)), (y, y') = (v_t \cdot x, f_i(v_t \cdot x)) = (v_t, v_c) \cdot (x, x')$$

with $t \neq c$. Applying this to the set in (6.21) and continuing the argument as just after the proof of Lemma 6.23 we conclude that μ is the trivial joining (which actually contradicts our description of the conditional measures $\mu_{(x,\cdot)}^{\mathscr{A}}$).

Since we now may assume $c = t$ for a.e. $x \in Y_1$ and since both sets $\{f_1(v_t \cdot x), \dots, f_m(v_t \cdot x)\}$ and $\{v_t \cdot f_1(x), \dots, v_t \cdot f_m(x)\}$ contain m elements it follows that (6.22) holds almost surely. Let us now show that this implies that μ is preserved by (v_t, v_t) for any $t \in \mathbb{R}$. So let $f \in C_c(X_1 \times X_2)$. Then

$$\begin{aligned}
&\int_{X_1 \times X_2} f((v_t, v_t) \cdot (x, x')) \, \mathrm{d}\mu = \\
&\qquad \int_{X_1} \int_{\{x\} \times X_2} f((v_t, v_t) \cdot (x, x')) \, \mathrm{d}\mu_{(x,\cdot)}(x, x') \, \mathrm{d}m_{X_1}(x) = \\
&\qquad\qquad \int_{X_1} \frac{1}{m} \sum_{i=1}^{m} f((v_t, v_t) \cdot (x, f_i(x))) \, \mathrm{d}m_{X_1}(x) = \\
&\qquad\qquad\quad \int_{X_1} \frac{1}{m} \sum_{i=1}^{m} f(v_t \cdot x, f_i(v_t \cdot x)) \, \mathrm{d}m_{X_1}(x) = \\
&\qquad\qquad\qquad\quad \int_{X_1} \frac{1}{m} \sum_{i=1}^{m} f(x, f_i(x)) \, \mathrm{d}m_{X_1}(x) = \int_{X_1 \times X_2} f \, \mathrm{d}\mu,
\end{aligned}$$

where we used in order the definition of the conditional measures, our description of them, (6.22) for m_{X_1}-a.e. x, and the fact that v_t preserves m_{X_1}.

Now note that U as in Theorem 6.20 together with $\{(v_t, v_t) : t \in \mathbb{R}\}$ generate the diagonal embedded copy H of $\mathrm{SL}_2(\mathbb{R})$. As H contains U, H acts ergodically with respect to μ. Hence Theorem 6.17 applies and shows that μ is algebraic. □

Exercises for Section 6.7

Exercise 6.7.1. Show the classification of the invariant measures for the one-parameter subgroup as in Exercise 6.6.1(2) without referring to entropy theory.

Exercise 6.7.2. Consider all cases of unipotent one-parameter flows and horospherical subgroups on a quotient $\Gamma \backslash \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ by any discrete subgroup Γ (using entropy theory).

Exercise 6.7.3. Consider all cases of unipotent one-parameter flows and horospherical subgroups on a quotient $\Gamma \backslash \mathrm{SL}_2(\mathbb{C})$ by any discrete subgroup Γ (using entropy theory).

6.8 Invariant Measures on Finite Volume Quotients of $\mathrm{SL}_3(\mathbb{R})$

In this section we let $G = \mathrm{SL}_3(\mathbb{R})$, assume that $\Gamma < G$ is any discrete subgroup, and set $X = \Gamma\backslash G$.

There are two different† choices of one-parameter unipotent subgroups, defined by the two possibilities below:

- $u_s = \begin{pmatrix} 1 & 0 & s \\ & 1 & 0 \\ & & 1 \end{pmatrix}$, or
- $u_s = \begin{pmatrix} 1 & s & \frac{s^2}{2} \\ & 1 & s \\ & & 1 \end{pmatrix}$.

In either case we have the following theorem.

This should be the last proof of a special case. It should use the language of quasi-regular maps!

Theorem 6.26. *Let $X = \Gamma\backslash\mathrm{SL}_3(\mathbb{R})$, and let U be either of the one-parameter subgroups as above. Then an U-invariant and ergodic probability measure μ on X is always algebraic‡.*

Exercises for Section 6.8

Exercise 6.8.1. Find the complete list of all possible connected algebraic subgroups $L \leqslant \mathrm{SL}_3(\mathbb{R})$ that may give rise to some U-invariant and ergodic Haar measure $\mu = m_{L\bullet x}$ for some $X = \Gamma\backslash\mathrm{SL}_3(\mathbb{R})$. To rule out the possibility that $L \simeq \mathrm{GL}_2(\mathbb{R})$, you may assume that $\Gamma = \mathbb{G}(\mathbb{Z})$ for some algebraic group $\mathbb{G}$ defined over $\mathbb{Q}$ with $\mathbb{G}(\mathbb{R}) = \mathrm{SL}_3(\mathbb{R})$.

Notes to Chapter 6

(24) (Page 176) This appeared in print in the work of Dani [**?**, Conjecture II].

(25) (Page 180) This is an instance of a more general result due to Weyl [**?**] giving equidistribution modulo one for the values on the natural numbers of any polynomial with an irrational coefficient. Furstenberg [**?**] showed that this followed from a general result extending unique ergodicity from irrational circle rotations to certain maps on tori. We refer to [**?**, Sec. 4.4.3] for a detailed discussion.

What else should we do here? Linearization? Orbit Closure?

† The reader may check that these are all, up to conjugation.

‡ The proof will essentially give all the possible subgroups $L \leqslant \mathrm{SL}_3(\mathbb{R})$ that may give rise to the Haar measure. However, this list is starting to become longer so we refrain from giving it here. Furthermore, to know precisely which of these possible subgroups can arise arithmetic properties of the lattice become important. In particular, it seems that one of the possible subgroups can only be ruled out by using the Margulis arithmeticity theorem. See also Exercise 6.8.1.

Chapter 7
More on Algebraic Groups

†In this chapter we describe some of the finer properties of algebraic groups, and how these fit into the study of the space $d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$. Some of the deeper facts we cannot prove here (but will again describe these in the context of dynamics on $X = \Gamma\backslash G$).

7.1 Three (instead of Two) Fundamentally Different Classes of Algebraic Groups

There are two fundamental classes of Lie groups, *solvable* Lie groups and *semi-simple* Lie groups. The Levi decomposition for Lie groups shows how a Lie group can be decomposed into a normal solvable Lie subgroup and a semi-simple Lie subgroup.

For algebraic subgroups this story can be repeated in the following way. We will assume that $\mathrm{char}(\mathbb{K}) = 0$. Every Zariski connected linear algebraic group $\mathbb{G}$ has a maximal normal solvable Zariski connected algebraic subgroup $(\mathbb{G})$, generated by all the Zariski connected normal solvable algebraic subgroups, called the *solvable radical* of $\mathbb{G}$. If $\mathbb{G}$ is defined over $\mathbb{K}$, then $(\mathbb{G})$ is also defined over $\mathbb{K}$ since $(\mathbb{G})$ is unique and all Galois automorphisms must fix it (see Lemma 3.20).

Then there also exists a semi-simple algebraic subgroup $\mathbb{L} < \mathbb{G}$ (see below for the formal definition) with a Levi decomposition of algebraic groups

$$\mathbb{G} = \mathbb{L} \cdot (\mathbb{G})$$

and with $\mathbb{L} \cap (\mathbb{G})$ finite. If $\mathbb{G}$ is defined over $\mathbb{K}$, then $\mathbb{L}$ can be chosen to also be defined over $\mathbb{K}$. Indeed by Varadarajan [**?**, Th. 3.14.1] there exists a semi-simple Lie subalgebra $\mathfrak{l}$ of the Lie algebra $\mathfrak{g}$ of $\mathbb{G}$ which is defined over $\mathbb{K}$

† This chapter can be skipped at first by the less algebraically inclined reader, who may return to it later as needed.

and gives a transverse of the Lie algebra of the radical $(\mathbb{G})$. Then $\mathfrak{l}$ is the Lie algebra of an algebraic subgroup $\mathbb{L}$ defined over $\mathbb{K}$, see Section 7.8 for a discussion of this last step.

However, in the context of algebraic groups more can be said. Within the solvable radical $(\mathbb{G})$ there exists a maximal normal unipotent algebraic subgroup $_{\mathrm{u}}(\mathbb{G})$, called the *unipotent radical* of $\mathbb{G}$. If $\mathbb{G}$ is defined over $\mathbb{K}$ and $\mathbb{K}$ has characteristic zero, then as before $_{\mathrm{u}}(\mathbb{G})$ is also defined over $\mathbb{K}$.

Moreover, the Levi decomposition can be refined even further. There exists an algebraic *torus* $\mathbb{T} < (\mathbb{G})$ (see below) such that $(\mathbb{G}) = \mathbb{T} \cdot_{\mathrm{u}} (\mathbb{G})$ so that

$$\mathbb{G} = \mathbb{L} \cdot \mathbb{T} \cdot_{\mathrm{u}} (\mathbb{G}).$$

If $\mathbb{G}$ is defined over $\mathbb{K}$, then $\mathbb{T}$ can be chosen to be defined over $\mathbb{K}$ as well, see [**?**, Th. 10.6(4), Th. 18.2(i)].

Thus there are three different types of algebraic groups, and any algebraic groups can be built from these.

- An algebraic group is *semi-simple* if it is Zariski connected and its Lie algebra is semi-simple.
- An algebraic group is an *algebraic torus* if it is Zariski connected and (as a subgroup of SL_d) it can be simultaneously diagonalized over $\overline{\mathbb{K}}$, or equivalently if it is conjugate over $\overline{\mathbb{K}}$ to a subgroup of the full diagonal subgroup in some SL_d.
- An algebraic group is *unipotent* if it can be conjugated to a subgroup of the unipotent upper triangular subgroup

$$\mathbb{U}_{\max} = \left\{ \begin{pmatrix} 1 & * & \dots & * \\ & 1 & \dots & * \\ & & \ddots & \\ & & & 1 \end{pmatrix} \right\} \subseteq \mathrm{SL}_d \,.$$

We list some (unstructured) observations that may help the reader navigate some of these definitions if they are unfamiliar.

- Algebraically, torus groups and semi-simple algebraic groups share an important feature: all their algebraic representations are *semi-simple*, meaning that they can be decomposed into direct sums of irreducible representations. For torus subgroups this is simply the statement that one can simultaneously diagonalize the (commuting) diagonalizable elements obtained from the representation (see Proposition 3.28). For semi-simple algebraic groups this is a consequence of the representation theory of semi-simple Lie algebras.
- A *reductive algebraic group* is an almost direct product of a semi-simple algebraic group and an algebraic torus (classified by $_{\mathrm{u}}(\mathbb{G})$ being trivial). The algebraic representations of a reductive group are also semi-simple. Furthermore, the Levi decomposition for a reductive group is an almost

direct product (the radical equals the connnected component of the center of $\mathbb{G}$).

- From an abstract ergodic theoretic point of view, algebraic torus subgroups, unipotent algebraic groups, and their combinations in the form of solvable algebraic groups, give rise to amenable[†] groups of $\mathbb{R}$-points (or $\mathbb{C}$-points or $\mathbb{Q}_p$-points as appropriate). Thus the study of their actions fits nicely into the classical setting of ergodic theory, with for example the ergodic theorems for amenable group actions.
- In the context of this book are the actions of the $\mathbb{R}$-points $\mathbb{H}(\mathbb{R})$ of a subgroup $\mathbb{H} < \mathbb{G}$ on a homogeneous space $\Gamma\backslash\mathbb{G}(\mathbb{R})$ important. From this point of view the class of unipotent algebraic subgroups $\mathbb{H} < \mathbb{G}$ and the class of semi-simple subgroups $\mathbb{H} < \mathbb{G}$ also have similar behavior. For these subgroups and their combinations (classified by $(\mathbb{H}) =_{\mathrm{u}} (\mathbb{H})$), orbit closures and invariant measures can be completely classified (see Chapters 5 and 6).

An important example of a reductive algebraic group is GL_d. It also fits into our definition of algebraic subgroups being subgroups of SL_D where D is $d+1$, since

$$\mathrm{GL}_d \cong \left\{ \begin{pmatrix} g & \\ & (\det g)^{-1} \end{pmatrix} \in \mathrm{SL}_{d+1} \right\}.$$

Notice that $\mathrm{GL}_d = \mathrm{SL}_d \cdot \mathbb{T}$, where $\mathbb{T} < \mathrm{GL}_d$ is the subgroup which gets mapped under the above isomorphism to

$$\mathbb{T} \cong \left\{ \begin{pmatrix} aI_d & \\ & a^{-d} \end{pmatrix} \mid a \in \mathbb{G}_m \right\},$$

and that this is an almost direct product since $(\mathrm{SL}_d \cap \mathbb{T})(\overline{\mathbb{K}})$ has no more than d elements (and has precisely d elements if $\mathbb{K}$ has characteristic zero). For $d=1$ we only get another notation for $\mathbb{G}_m \cong \mathrm{GL}_1$.

7.2 ($\mathbb{K}$-)Characters

Definition 7.1. Let $\mathbb{G}$ be an algebraic group defined over $\mathbb{K}$. A *($\mathbb{K}$-)character* of $\mathbb{G}$ is a one-dimensional representation ϕ (defined over $\mathbb{K}$); that is, a group homomorphism $\phi : \mathbb{G} \to \mathbb{G}_m$ (defined over $\mathbb{K}$).

Why should we be interested in these polynomial maps? Following the developments above, an answer is easy to find. In Proposition 3.8 we have seen that the stabilizer subgroup of a $\mathbb{Q}$-vector in a $\mathbb{Q}$-representation of SL_d gives rise to a closed orbit. In Chevalley's theorem (Theorem 3.26) we saw that any $\mathbb{Q}$-group is the stabilizer subgroup of a line spanned by a $\mathbb{Q}$-vector

[†] Semi-simple algebraic groups for which $\mathbb{G}(\mathbb{R})$ is compact also have the property that $\mathbb{G}(\mathbb{R})$ is amenable, but their actions are usually not so interesting.

in a $\mathbb{Q}$-representation. This implies the following corollary (which will be strengthened in Section 7.4).

Corollary 7.2. *If an algebraic subgroup $\mathbb{H} \leqslant \mathrm{SL}_d$ is defined over $\mathbb{Q}$ and has no non-trivial $\mathbb{Q}$-characters, then*

$$\mathrm{SL}_d(\mathbb{Z})\mathbb{H}(\mathbb{R}) \subseteq \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$$

is closed.

Proof. Let $\mathbb{H} = \{g \in \mathrm{SL}_d \mid \rho(g)v \sim v\}$ be as in Theorem 3.26. Suppose $v_i \neq 0$ for $i \in \{1, \dots, D\}$. Then

$$\begin{aligned} \phi : \mathbb{H} &\longrightarrow \mathbb{G}_m \\ g &\longmapsto \tfrac{(\rho(g)v)_i}{v_i} \end{aligned}$$

is a $\mathbb{Q}$-character. If $\mathbb{H}$ has no non-trivial $\mathbb{Q}$-characters, then $\phi(\mathbb{H}) = 1$ and so

$$\mathbb{H} = \{g \in \mathrm{SL}_d \mid \rho(g)v = v\}.$$

The corollary follows by Proposition 3.8. □

In the reverse direction we have the following corollary of the Borel density theorem (Theorem 3.30).

Corollary 7.3. *Suppose that $\mathbb{H} \leqslant \mathrm{SL}_d$ is a Zariski connected algebraic subgroup defined over $\mathbb{Q}$ and that*

$$\mathbb{H}(\mathbb{Z}) = \mathbb{H}(\mathbb{R}) \cap \mathrm{SL}_d(\mathbb{Z})$$

is a lattice in $\mathbb{H}(\mathbb{R})$. Then $\mathbb{H}$ has no non-trivial $\mathbb{Q}$-characters.

Proof. Suppose that $\phi : \mathbb{H} \to \mathbb{G}_m$ is a $\mathbb{Q}$-character. Then† $\phi\left(\mathbb{H}(\mathbb{Z})\right) \subseteq \mathbb{Q}^\times$ is a subgroup. Furthermore, the denominators of the elements in $\phi\left(\mathbb{H}(\mathbb{Z})\right)$ are uniformly bounded (since they can only arise from the finitely many coefficients of the polynomial map ϕ). However, the only two subgroups of $\mathbb{Q}^\times$ with bounded denominators are the trivial subgroup $\{1\}$ and the subgroup $\{\pm 1\}$ with two elements. In either case we have $\phi(\gamma)^2 = 1$ for all $\gamma \in \mathbb{H}(\mathbb{Z})$. As this is a polynomial relation, the same property holds for all elements of the Zariski closure $\mathbb{L} \subseteq \mathbb{H}$ of $\mathbb{H}(\mathbb{Z})$. By the Borel density theorem (Theorem 3.30), $\mathbb{L}$ contains the unipotent radical $_{\mathrm{u}}(\mathbb{H})$ of $\mathbb{H}$. Let $\mathbb{T} \leqslant (\mathbb{H})$ be a torus defined over $\mathbb{Q}$ such that $_{\mathrm{u}}(\mathbb{H}) \cdot \mathbb{T} = (\mathbb{H})$. Let t be an element of $\mathbb{T}(\mathbb{R})$ with decomposition $t = t_{\mathrm{pos}} t_{\mathrm{comp}}$ according to Proposition 3.28. By Theorem 3.30 we

† Here we use the assumption that ϕ is defined over $\mathbb{Q}$ (that is, has coefficients in $\mathbb{Q}$) to obtain $\phi(\mathbb{H}(\mathbb{Z})) \subseteq \mathbb{Q}$ and that it is a character to see that $\phi(\mathbb{H}(\mathbb{Z}))$ is a multiplicative subgroup.

again have $t_{\text{pos}} \in \mathbb{L}$ and so $\phi(t_{\text{pos}})^2 = 1$. Furthermore, $t_{\text{comp}} \in \mathbb{H}(\mathbb{R})$ has the property that

$$\overline{\{t_{\text{comp}}^n \mid n \in \mathbb{Z}\}} \subseteq \mathbb{H}(\mathbb{R})$$

is compact. Hence

$$\overline{\{\phi(t_{\text{comp}})^n \mid n \in \mathbb{Z}\}} \subseteq \mathbb{R}^\times$$

is compact, which implies that $\phi(t_{\text{comp}})^2 = 1$ once again. Putting these together (see Section 7.1), we see that $\phi(g)^2 = 1$ for all $g \in (\mathbb{H})$.

Since $\phi(g) = 1$ for all g in the semi-simple group $\mathbb{L} < \mathbb{H}$ (provided by the Levi decomposition in Section 7.1), we see that $\phi(g)^2 = 1$ for all $g \in \mathbb{H}$. Since $\mathbb{H}$ is Zariski connected (and hence has Zariski connected image under ϕ) we have $\phi(\mathbb{H}) = 1$, and so ϕ is the trivial character. □

7.3 Restriction of Scalars

Restriction of scalars is a powerful construction that may be used to obtain many different examples of algebraic groups. Let $\mathbb{K}'|\mathbb{K}$ be a finite field extension, so that we may identify $\mathbb{K}'$ (as a ring) with a subalgebra of $\mathrm{Mat}_d(\mathbb{K})$ (see also Section 3.3). Given this identification, the idea is simple. We can identify elements in $\mathrm{Mat}_D(\mathbb{K}')$ with block matrices in $\mathrm{Mat}_{dD}(\mathbb{K})$; more precisely, let

$$\phi : \mathbb{K}' \longrightarrow \mathrm{Mat}_d(\mathbb{K})$$

be the map identifying elements in $\mathbb{K}'$ with matrices over $\mathbb{K}$ (with respect to a fixed basis of the vector space $\mathbb{K}'$ over $\mathbb{K}$). Then we may extend ϕ to a map

$$\Phi : \mathrm{Mat}_D(\mathbb{K}') \longrightarrow \mathrm{Mat}_{dD}(\mathbb{K})$$
$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{D1} & \cdots & a_{DD} \end{pmatrix} \longmapsto \begin{pmatrix} \phi(a_{11}) & \cdots & \phi(a_{1n}) \\ \vdots & & \vdots \\ \phi(a_{D1}) & \cdots & \phi(a_{DD}) \end{pmatrix}.$$

Notice that the images of ϕ and of Φ are defined by linear equations over $\mathbb{K}$, and even their $\overline{\mathbb{K}}$-points form an algebra. Now suppose that we are given an algebraic subgroup $\mathbb{H} < \mathrm{SL}_D$ defined over $\mathbb{K}$. Then we can define a new algebraic subgroup

$$_{\mathbb{K}'|\mathbb{K}}(\mathbb{H}) \subseteq \mathrm{SL}_{dD}$$

by demanding that the following relations hold:

- if $g \in_{\mathbb{K}'|\mathbb{K}} (\mathbb{H})$ then

$$g = \begin{pmatrix} g_{11} & \cdots & g_{1D} \\ \vdots & & \vdots \\ g_{D1} & \cdots & g_{DD} \end{pmatrix} \in \langle \operatorname{Im} \Phi \rangle_{\overline{\mathbb{K}}},$$

where each $g_{ij} \in \operatorname{Mat}_d(\overline{\mathbb{K}})$ belongs to the $\overline{\mathbb{K}}$-linear hull of the image of ϕ;

- for any polynomial $p \in \mathscr{J}(\mathbb{H})$ we have

$$p \begin{pmatrix} g_{11} & \cdots & g_{1D} \\ \vdots & & \vdots \\ g_{D1} & \cdots & g_{DD} \end{pmatrix} = 0 \in \operatorname{Im} \phi,$$

where the polynomial $p(x_{11}, \dots, x_{1D}, \dots, x_{D1}, \dots, x_{DD})$ is applied to the matrices $g_{ij} \in \operatorname{Im} \phi$.

We will only be interested in the case $(\mathbb{K}) = 0$ (and even in this setting we will mainly be interested in $\mathbb{K} = \mathbb{Q}$ and $\mathbb{K} = \mathbb{R}$). In this case $\operatorname{Im} \phi$ can be simultaneously diagonalized, and so over $\overline{\mathbb{K}}$ (or over any other field that contains all Galois images of $\mathbb{K}'$) all of this becomes the statement

$$\left({}_{\mathbb{K}'|\mathbb{K}}(\mathbb{H})\right)\left(\overline{\mathbb{K}}\right) \cong \left(\mathbb{H}(\overline{\mathbb{K}})\right)^d, \tag{7.1}$$

and so ${}_{\mathbb{K}'|\mathbb{K}}(\mathbb{H})$ is an algebraic group. To see this, simply simultaneously diagonalize all blocks belonging to $\langle \operatorname{Im}(\phi) \rangle_{\overline{\mathbb{K}}}$ by conjugation with the block matrix

$$\begin{pmatrix} h & & \\ & \ddots & \\ & & h \end{pmatrix},$$

where h is such that it diagonalizes $\operatorname{Im}(\phi)$. After this conjugation the elements of ${}_{\mathbb{K}'|\mathbb{K}}(\mathbb{H})$ are diagonal in each block and we can permute rows and columns so that we obtain a subgroup of $(\mathrm{SL}_D)^d$ (embedded as block matrices along the diagonal into SL_{dD}). By the last requirement we see in each block an independent copy of $\mathbb{H}$.

This construction of groups is interesting because the isomorphism in (7.1) does not usually hold on the level of $\mathbb{K}$-points, so that we obtain often new algebraic groups defined over $\mathbb{K}$.

Example 7.4. To help digest restriction of scalars in a more concrete situation, we discuss ${}_{\mathbb{C}|\mathbb{R}}(\mathbb{G}_m)$, where $\mathbb{G}_m$ can be realized as the algebraic subgroup

$$\left\{ \begin{pmatrix} \alpha & \\ & \alpha^{-1} \end{pmatrix} \mid \alpha \in \mathbb{G}_m \right\} \subseteq \mathrm{SL}_2$$

and $\mathbb{C}$ can be realized as the ring of matrices

$$\left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \mid a, b \in \mathbb{R} \right\}.$$

Hence (following the recipe above for defining ${}_{\mathbb{C}|\mathbb{R}}(\mathbb{G}_m)$) we have

$$\begin{aligned} \left({}_{\mathbb{C}|\mathbb{R}}(\mathbb{G}_m)\right)(\mathbb{R}) &= \left\{ \begin{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix} & \\ & \frac{1}{a^2+b^2} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \end{pmatrix} \mid a, b \in \mathbb{R} \right\} \\ &\cong \mathbb{C}^\times \cong \mathbb{R}^\times_{>0} \times \mathbb{S}^1. \end{aligned}$$

The isomorphism in Example 7.4 is not accidental. From the definitions it follows that[†]

$$\left({}_{\mathbb{K}'|\mathbb{K}}(\mathbb{H})\right)(\mathbb{K}) \cong \mathbb{H}(\mathbb{K}')$$

for any algebraic group $\mathbb{H}$ defined over $\mathbb{K}$, for any finite field extension $\mathbb{K}'|\mathbb{K}$. We will see other examples of this construction below.

7.3.1 Unipotent Algebraic Groups

The most familiar non-trivial unipotent algebraic group is

$$\mathbb{G}_a \cong U = \left\{ \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \right\}.$$

If we apply restriction of scalars for $\mathbb{K}'|\mathbb{K}$ of degree d to U then we obtain

$${}_{\mathbb{K}'|\mathbb{K}}(\mathbb{G}_a) \cong_{\mathbb{K}'|\mathbb{K}} (U) = \left\{ \begin{pmatrix} I_d & g \\ & I_d \end{pmatrix} \mid g \in \operatorname{Im} \phi \right\}.$$

However, since the algebraic group

$$\left\{ \begin{pmatrix} I_d & g \\ & I_d \end{pmatrix} \mid g \in \mathrm{Mat}_d \right\} \cong \mathbb{G}_a^{d^2}$$

is simply isomorphic to a d^2-dimensional vector space, and since $\operatorname{Im} \phi$ in that space is a d-dimensional subspace we see that

$${}_{\mathbb{K}'|\mathbb{K}}(\mathbb{G}_a) \cong \mathbb{G}_a^d,$$

and that the isomorphism takes place over $\mathbb{K}$ not just over $\overline{\mathbb{K}}$.

[†] The reader may now wonder what all this is good for, if it just gives back the group that we started with. However, as we will see this construction can be quite powerful, for example for the purpose of constructing lattices.

Apart from $\mathbb{G}_a^2$ (which is not significantly different from $\mathbb{G}_a$), the next natural unipotent algebraic group is the Heisenberg

$$\mathbb{H} = \left\{ \begin{pmatrix} 1 & x & z \\ & 1 & y \\ & & 1 \end{pmatrix} \right\} \subseteq \mathrm{SL}_3 \,.$$

Let us again apply ${}_{\mathbb{K}'|\mathbb{K}}$ to $\mathbb{H}$, where for concreteness we will consider the quadratic field extension

$$\mathbb{K}' = \mathbb{Q}(\sqrt{d}) \cong \left\{ \begin{pmatrix} a_1 & da_2 \\ a_2 & a_1 \end{pmatrix} \right\} : \mathbb{K} = \mathbb{Q}$$

for a non-square d. Then

$${}_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}}(\mathbb{H}) = \left\{ \begin{pmatrix} I_2 & \begin{pmatrix} x_1 & dx_2 \\ x_2 & x_1 \end{pmatrix} & \begin{pmatrix} z_1 & dz_2 \\ z_2 & z_1 \end{pmatrix} \\ & I_2 & \begin{pmatrix} y_1 & dy_2 \\ y_2 & y_1 \end{pmatrix} \\ & & I_2 \end{pmatrix} \right\}.$$

Over $\mathbb{C}$ (or even over $\mathbb{R}$ in the case $d > 0$) we have

$${}_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}}(\mathbb{H}) \cong \mathbb{H} \times \mathbb{H}. \tag{7.2}$$

However, we claim that the isomorphism in (7.2) does not take place over $\mathbb{Q}$. That is, it is not possible to choose the isomorphism in such a way that the coefficients of the polynomials appearing in it all lie in $\mathbb{Q}$. To see this, notice the following properties.

- There is an element of $(\mathbb{H} \times \mathbb{H})(\mathbb{Q})$ whose centralizer in $\mathbb{H} \times \mathbb{H}$ is 5-dimensional. Indeed,

$$\begin{pmatrix} 1 & 1 & 0 \\ & 1 & \\ & & 1 \end{pmatrix} \times I \in (\mathbb{H} \times \mathbb{H})(\mathbb{Q})$$

  has the centralizer

$$\left\{ \begin{pmatrix} 1 & x & y \\ & 1 & \\ & & 1 \end{pmatrix} \right\} \times \mathbb{H}.$$

- All elements of $\left({}_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}}(\mathbb{H})\right)(\mathbb{Q})$ are either in the center or have a 4-dimensional centralizer. Indeed, if we let

$$h = \begin{pmatrix} I_2 & \begin{pmatrix} x_1 & dx_2 \\ x_2 & x_1 \end{pmatrix} & \begin{pmatrix} z_1 & dz_2 \\ z_2 & z_1 \end{pmatrix} \\ & I_2 & \begin{pmatrix} y_1 & dy_2 \\ y_2 & y_1 \end{pmatrix} \\ & & I_2 \end{pmatrix} \in \left({}_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}}(\mathbb{H}) \right)(\mathbb{Q})$$

and assume that h is not in the center, then $(x_1, x_2, y_1, y_2) \neq 0$. Using the isomorphism in (7.2), the element h corresponds to

$$h \cong \underbrace{\begin{pmatrix} 1 & x_1 + x_2\sqrt{d} & z_1 + z_2\sqrt{d} \\ & 1 & y_1 + y_2\sqrt{d} \\ & & 1 \end{pmatrix}}_{h_+} \times \underbrace{\begin{pmatrix} 1 & x_1 - x_2\sqrt{d} & z_1 - z_2\sqrt{d} \\ & 1 & y_1 - y_2\sqrt{d} \\ & & 1 \end{pmatrix}}_{h_-}.$$

Since neither h_+ nor h_- belongs to the center of $\mathbb{H}$, it follows that the centralizer of h is (as the direct product of the centralizers of h_+ resp. h_-) 4-dimensional.

Thus

$$\left({}_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}}(\mathbb{H}) \right)(\mathbb{Q}),$$

is not isomorphic to

$$(\mathbb{H} \times \mathbb{H})(\mathbb{Q}) = \mathbb{H}(\mathbb{Q}) \times \mathbb{H}(\mathbb{Q}).$$

Hence ${}_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}}(\mathbb{H})$ and $\mathbb{H} \times \mathbb{H}$ are two different algebraic groups over $\mathbb{Q}$, even though they are the same over $\mathbb{C}$. Such statements are of interest for us because this also shows, say for $d > 0$, that the image of ${}_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}}(\mathbb{H})(\mathbb{Z})$ and $\mathbb{H}(\mathbb{Z}) \times \mathbb{H}(\mathbb{Z})$ are two different discrete subgroups of $\mathbb{H}(\mathbb{R}) \times \mathbb{H}(\mathbb{R})$. They are in fact different lattices (as the reader may check).

7.3.2 Torus Groups

We have already discussed — albeit implicitly — in some detail the algebraic torus

$${}_{\mathbb{K}|\mathbb{Q}}(\mathbb{G}_m)$$

obtained from the multiplicative group $\mathbb{G}_m$ and a finite field extension $\mathbb{K}|\mathbb{Q}$. The reader should go back to Section 3.3 and compare the construction there to the construction of ${}_{\mathbb{K}|\mathbb{Q}}(\mathbb{G}_m)$. Here we add a few more important definitions.

An algebraic torus $\mathbb{T}$ over a field $\mathbb{K}$ is called *split over* $\mathbb{K}$ if $\mathbb{T} \cong \mathbb{G}_m^d$ and this isomorphism is defined over $\mathbb{K}$. An algebraic torus $\mathbb{T}$ over a field $\mathbb{K}$ is called *anisotropic over* $\mathbb{K}$ if $\mathbb{T}$ has no $\mathbb{K}$-characters.

The example ${}_{\mathbb{K}|\mathbb{Q}}(\mathbb{G}_m)$ above is neither split (unless $\mathbb{K} = \mathbb{Q}$) nor is it anisotropic. The determinant function (which gives rise to the norm ${}_{\mathbb{K}|\mathbb{Q}}$) is a $\mathbb{Q}$-character on ${}_{\mathbb{K}|\mathbb{Q}}(\mathbb{G}_m)$. However, its kernel

$$\mathbb{T} = \ker(\det) \leqslant_{\mathbb{K}|\mathbb{Q}} (\mathbb{G}_m)$$

is anisotropic over $\mathbb{Q}$ by Proposition 3.11 and Corollary 7.3.

We can also strengthen Corollary 7.2 for torus subgroups to get the following result of Ono [**?**] (which will give the remaining third of the proof of the Borel Harish-Chandra theorem in Section 7.4).

Corollary 7.5. *Let $\mathbb{T} \leqslant \mathrm{SL}_d$ be a torus defined over $\mathbb{Q}$ and anisotropic over $\mathbb{Q}$. Then $\mathbb{T}(\mathbb{Z})$ is a uniform lattice in $\mathbb{T}(\mathbb{R})$, or equivalently the orbit*

$$\mathrm{SL}_d(\mathbb{Z})\mathbb{T}(\mathbb{R}) \subseteq \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})$$

is compact.

PROOF. Let us start with the case where $\mathbb{T} \leqslant \mathrm{SL}_d$ is $\mathbb{Q}$-irreducible, meaning that there is no proper $\mathbb{Q}$-subspace of $\mathbb{Q}^d$ that is invariant under $\mathbb{T}$. Notice that in this case the $\mathbb{Q}$-span of $\mathbb{T}(\mathbb{Q})$,

$$K = \mathbb{Q}[\mathbb{T}(\mathbb{Q})] = \left\{ \sum_i a_i t_i \mid a_i \in \mathbb{Q}, t_i \in \mathbb{T}(\mathbb{Q}) \right\} \subseteq \mathrm{Mat}_d(\mathbb{Q})$$

is a field, which should enable us to use Section 3.3. For if K is not a field, then there exists some $b \in K$ that is a zero divisor, and in that case $\ker(b) \subseteq \mathbb{Q}^d$ is a proper $\mathbb{Q}$-subspace that is invariant under $\mathbb{T}$ (since $\mathbb{T}$ is commutative). Unfortunately, we do not know at this point whether $\mathbb{T}(\mathbb{Q})$ is Zariski dense in $\mathbb{T}$ and so we do not know whether K is big enough (*a priori* it might be $\mathbb{Q}$) for our purposes (so that $\mathbb{T}(\overline{\mathbb{Q}}) \subseteq K \otimes_{\mathbb{Q}} \overline{\mathbb{Q}}$). For this reason, we will give a more complicated definition of K. Let

$$M = M(\overline{\mathbb{Q}}) = \overline{\mathbb{Q}}[\mathbb{T}(\overline{\mathbb{Q}})] = \left\{ \sum_i a_i t_i \mid a_i \in \overline{\mathbb{Q}}, t_i \in \mathbb{T}(\overline{\mathbb{Q}}) \right\} \subseteq \mathrm{Mat}_d(\overline{\mathbb{Q}})$$

be the subspace spanned by $\mathbb{T}(\overline{\mathbb{Q}})$ over $\overline{\mathbb{Q}}$. Since $\mathbb{T}(\overline{\mathbb{Q}})$ can be simultaneously diagonalized, we see that M has dimension at most d. Also note that M is closed under all Galois automorphisms of $\overline{\mathbb{Q}}|\mathbb{Q}$ (since $\mathbb{T}$ is defined over $\mathbb{Q}$). Therefore, M is defined by rational equations (see the argument in the proof of Lemma 3.20) and

$$K = M \cap \mathrm{Mat}_d(\mathbb{Q})$$

is a $\leqslant d$-dimensional sub-algebra. In fact it is a field by the argument outlined above: If $b \in K$ is a zero-divisor, then $\ker b$ is a proper K- (and hence M-, and hence also $\mathbb{T}$-invariant) rational subspace. It also follows that K is d-dimensional, for otherwise $(1,0,\dots,0)K \subseteq \mathbb{Q}^d$ would again be a proper K-invariant rational subspace. By applying[†] Proposition 3.11 it follows that $\mathbb{T}_K = \mathrm{SL}_d \cap M$ gives rise to the compact orbit

$$\mathrm{SL}_d(\mathbb{Z})\mathbb{T}_K(\mathbb{R}) \subseteq \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R}).$$

Since $\mathbb{T} \subseteq \mathbb{T}_K$, it now follows that

$$\mathrm{SL}_d(\mathbb{Z})\mathbb{T}(\mathbb{R}) \subseteq \mathrm{SL}_d(\mathbb{Z})\mathbb{T}_K(\mathbb{R})$$

is a closed subset (by Corollary 7.2) of a compact set. This concludes the case of an irreducible torus subgroup $\mathbb{T} \subseteq \mathrm{SL}_d$ defined over $\mathbb{Q}$.

In the general case, we can apply the same argument together with the assumption that $\mathbb{T}$ has no $\mathbb{Q}$-characters as follows. Since $\mathbb{T}$ is a torus, its natural representation on the d-dimensional space is semi-simple. More concretely, if the action of $\mathbb{T}$ is not irreducible over $\mathbb{Q}$ then we can find a smallest $\mathbb{T}$-invariant non-trivial $\mathbb{Q}$-space on which the representation is irreducible over $\mathbb{Q}$, and a $\mathbb{T}$-invariant complement. Repeating this we see that we can conjugate $\mathbb{T}$ by some rational $g \in \mathrm{GL}_n(\mathbb{Q})$ so that $\mathbb{T}' = g\mathbb{T}g^{-1}$ is of block form with blocks of size $d_1,\dots,d_n$ with $\sum_{j=1}^n d_j = d$. We may assume that $g \in \mathrm{SL}_d(\mathbb{Q})$ by multiplying g on the left with an appropriate block matrix. Since $\mathbb{T}$ (and hence $\mathbb{T}'$) has no non-trivial $\mathbb{Q}$-characters, it follows that

$$\mathbb{T}' \subseteq \mathrm{SL}_{d_1} \times \cdots \times \mathrm{SL}_{d_n}.$$

By the same argument as in the irreducible case within each block, there is a torus $\mathbb{T}_{K_j} \subseteq \mathrm{SL}_{d_j}$ for which

$$\mathrm{SL}_{d_j}(\mathbb{Z})\mathbb{T}_{K_j}(\mathbb{R})$$

is compact, and such that

$$\mathbb{T}' < \mathbb{T}_{K_1} \times \cdots \times \mathbb{T}_{K_n}.$$

Once more, this implies that

$$\mathrm{SL}_d(\mathbb{Z})\mathbb{T}'(\mathbb{R}) \subseteq \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R}) \tag{7.3}$$

is compact.

It remains to dispose of the rational matrix g with $\mathbb{T}' = g\mathbb{T}g^{-1}$. Multiplying (7.3) on the right by g we first find that

$$\mathrm{SL}_d(\mathbb{Z})g\mathbb{T}(\mathbb{R}) \subseteq \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})$$

[†] More precisely, by applying the method of the proof of Proposition 3.11: identify K with $\mathbb{Q}^d$ using the map $K \ni b \mapsto (1,0,\dots,0)b$, transport the determinant function to a polynomial function on $\mathbb{Q}^d$, and apply Mahler's compactness criterion (Theorem 1.17) and Proposition 3.8.

is compact, and so the lattices $\mathbb{Z}^d gt$ for $t \in \mathbb{T}(\mathbb{R})$ are uniformly discrete. Since g lies in $\mathrm{SL}_d(\mathbb{Q})$ there exists some $N \in \mathbb{N}$ such that $\mathbb{Z}^d \subseteq \frac{1}{N}\mathbb{Z}^d g$, so that the lattices

$$\mathbb{Z}^d t \leqslant \tfrac{1}{N}\mathbb{Z}^d gt$$

for $t \in \mathbb{T}(\mathbb{R})$ are also uniformly discrete. By Mahler's compactness criterion (Theorem 1.17) and Corollary 7.2, the theorem follows. □

7.3.3 An example for semi-simple groups

We also have seen the lattice resulting from restriction of scalars applied to a simple group before. In fact the irreducible lattice appearing in Exercise 3.6.1 is of that form. More generally, we have the following.

Example 7.6. Let $\mathbb{K}|\mathbb{Q}$ be a number field, and define $\mathbb{G} =_{\mathbb{K}|\mathbb{Q}} (\mathrm{SL}_d)$. Then

$$\mathbb{G}(\mathbb{R}) \cong \underbrace{\mathrm{SL}_d(\mathbb{R}) \times \cdots \times \mathrm{SL}_d(\mathbb{R})}_{r} \times \underbrace{\mathrm{SL}_d(\mathbb{C}) \times \cdots \times \mathrm{SL}_d(\mathbb{C})}_{s},$$

where $\mathbb{K}$ has r real embeddings and s pairs of complex embeddings. Moreover, $\mathbb{G}(\mathbb{Q}) = \mathrm{SL}_d(\mathbb{K})$ which is diagonally embedded into $\mathbb{G}(\mathbb{R})$ using the $r+s$ different Galois embeddings into $\mathbb{R}$ resp. $\mathbb{C}$. Finally $\mathbb{G}(\mathbb{Z})$ gives rise to a discrete subgroup — a lattice by Section 7.4 — in this product of $(r+s)$ simple groups, which intersects each factor trivially. We note that $\mathbb{G}(\mathbb{Z})$ can be identified with $\mathrm{SL}_d(O)$ for an order $O \subseteq \mathbb{K}$.

7.4 The Borel Harish-Chandra theorem

We are now able to formulate and prove the following characterization of a finite volume quotient for a $\mathbb{Q}$-group.

Theorem 7.7 (General Borel Harish-Chandra theorem). *Let $\mathbb{G}$ be a connected algebraic group defined over $\mathbb{Q}$. Then $\mathbb{G}(\mathbb{Z})$ is a lattice in $\mathbb{G}(\mathbb{R})$ if and only if $\mathbb{G}$ admits no non-trivial $\mathbb{Q}$-characters.*

Proof. If $\mathbb{G}(\mathbb{Z})$ is a lattice then we have already seen in Corollary 7.3 that $\mathbb{G}$ has no non-trivial $\mathbb{Q}$-characters.

Needs checking

Suppose now that $\mathbb{G}$ has no non-trivial $\mathbb{Q}$-characters, and let $\mathbb{G} = \mathbb{G}_{ss}\mathbb{T}\mathbb{U}$ be a Levi decomposition[†] over $\mathbb{Q}$. By Theorem 3.9, $\mathbb{U}(\mathbb{Z})$ is a uniform lattice in $\mathbb{U}(\mathbb{R})$.

[†] Thus in particular $\mathbb{G}_{ss}$ and $\mathbb{T}$ commute and $\mathbb{U} \cap (\mathbb{G}_{ss}\mathbb{T}) = \{I\}$.

We claim that the $\mathbb{Q}$-torus $\mathbb{T}$ has no non-trivial $\mathbb{Q}$-characters. To see this, suppose that $\chi : \mathbb{T} \to \mathbb{G}_m$ is a $\mathbb{Q}$-character. Since the unipotent radical $\mathbb{U}$ is a normal subgroup and any connected algebraic group must act trivially on a torus (if the torus is normalized) there is a map $\mathbb{G} \to \mathbb{G}/\mathbb{U} \cong \mathbb{G}_{ss}\mathbb{T}$. Since $\mathbb{G}_{ss} \cap \mathbb{T}$ is finite, there exists some n for which

$$(g_{ss}tu) \longmapsto (g_{ss}t) \longmapsto t^n \longmapsto \chi(t^n) \in \mathbb{G}_m$$

is well-defined. This defines a $\mathbb{Q}$-character on $\mathbb{G}$. If χ is non-trivial then this induced character on $\mathbb{G}$ would also be non-trivial. Therefore $\mathbb{T}$ has no non-trivial $\mathbb{Q}$-characters. By Corollary 7.5 we deduce that $\mathbb{T}(\mathbb{Z}) < \mathbb{T}(\mathbb{R})$ is a uniform lattice.

Finally, we know by Corollary 7.2 that $\mathbb{G}_{ss}(\mathbb{R})$ has a closed orbit through the identity coset in d. By Theorem 4.11 this shows that $\mathbb{G}_{ss}(\mathbb{Z})$ is a lattice in $\mathbb{G}_{ss}(\mathbb{R})$.

The theorem then follows by applying Lemma 7.8 twice. □

Lemma 7.8. *Let G be a σ-compact locally compact unimodular group, equipped with a left-invariant metric. Suppose that $L \lhd G = LM$ for some closed subgroups L and M with $|L \cap M| < \infty$. Suppose also that a discrete subgroup $\Gamma < G$ has the property that $L \cap \Gamma$ is a lattice in L and $M \cap \Gamma$ is a lattice in M. Then Γ is a lattice in G.*

Proof. By Lemma 1.22 we know that m_G is, up to a scalar multiple, the push-forward of $m_L \times m_M$ on $L \times M$. Let $F_L \subseteq L$ be a fundamental domain for $\Gamma \cap L$ in L, and let $F_M \subseteq M$ be a fundamental domain for $\Gamma \cap M$ in M. We claim that $F = F_L F_M$ is a surjective set for Γ. By the assumption and the description of m_G above, this then implies the lemma. So let $g = \ell m \in G$ with $\ell \in L$ and $m \in M$. Then there exists some $\gamma_M \in \Gamma \cap M$ such that

$$\gamma_M g = \ell'(\gamma_M m) \in LF_M.$$

Furthermore, there exists some $\gamma_L \in \Gamma \cap L$ with

$$\gamma_L \gamma_M g = (\gamma_L \gamma')(\gamma_M m) \in F_L F_M$$

as required. □

Corollary 7.9 (Characterization of compactness). *Let $\mathbb{G}$ be a connected algebraic group defined over $\mathbb{Q}$ such that $\mathbb{G}$ admits no non-trivial $\mathbb{Q}$-characters. Then $\mathbb{G}(\mathbb{Z})$ is a uniform lattice in $\mathbb{G}(\mathbb{R})$ if and only if the semisimple part $\mathbb{G}_{ss}$ in the Levi decomposition over $\mathbb{Q}$ satisfies that $\mathbb{G}_{ss}(\mathbb{Z})$ contains no unipotent elements.*

Proof. □ to come

7.5 Orthogonal groups as algebraic groups

still need to streamline this section

We will show in Section 7.4 that $\mathbb{G}(\mathbb{Z})$ is a lattice in $\mathbb{G}(\mathbb{R})$ if $\mathbb{G}$ is a semi-simple algebraic group defined over $\mathbb{Q}$. We will also see precisely when such a lattice is uniform. This result will subsume the case $\mathbb{G} = \mathrm{SL}_d$ considered in Theorem 1.18 and the case $\mathbb{G} = \mathrm{SO}(Q)$ for an integer quadratic form in $d \geqslant 3$ variables considered in Proposition 3.2. To motivate this and the discussion in Section 3.6 we start with a few more examples of semi-simple algebraic groups defined over $\mathbb{Q}$.

Lemma 7.10 (Orthogonal groups over $\mathbb{C}$). *For $d \geqslant 2$ the group $\mathrm{SO}(d)$ is Zariski connected and $\frac{d(d-1)}{2}$-dimensional. If $d = 2$, then it is a torus. If $d = 3$ or $d \geqslant 5$ then it is simple, and if $d = 4$ then it is semi-simple (and $\mathrm{SO}(4)$ is locally isomorphic to $\mathrm{SO}(3) \times \mathrm{SO}(3)$).*

Outline proof. Clearly

$$\mathrm{SO}(d) = \left\{ g \in \mathrm{SL}_d \mid g g^{\mathrm{t}} = I \right\}$$

has the Lie algebra

$$\mathfrak{so}(d) = \left\{ v \in \mathrm{Mat}_d \mid v + v^{\mathrm{t}} = 0 \right\}.$$

In other words, a matrix v lies in $\mathfrak{so}(d)$ if and only if the diagonal entries of v are zero, and the entries in the lower half of v are determined by those in the upper half via the relation $v = -v^{\mathrm{t}}$. Thus $\mathfrak{so}(d)$ has dimension $\frac{d(d-1)}{2}$. Now consider

$$V = \left\{ a \in \mathbb{C}^d \mid a_1^2 + \cdots + a_d^2 = 1 \right\},$$

which is an irreducible hypersurface in the d-dimensional space[†], and let e_d be the last basis vector in V. We claim that

$$V = e_d \, \mathrm{SO}(d)$$

as subsets of $\mathbb{C}^d$. This claim follows from another lemma in algebraic geometry as follows. By Shafarevich [?, Ch. I, Sec. 5, Th. 6] the set $\{e_d g \mid g \in \mathrm{SO}(d)\}$ contains a Zariski open subset of its Zariski closure $W \subseteq V$. From the image

Maybe this lemma should be included?

$$\{e_d g v \mid v \in \mathfrak{so}(d)\}$$

of the Lie algebra it follows that the tangent plane of W at $e_d g$ must contain $(e_d g)^{\perp}$. This implies that $W = V$. Applying the same argument at another point $a \in V$ shows that both $\{e_d g \mid g \in \mathrm{SO}(d)\}$ and $\{a g \mid g \in \mathrm{SO}(d)\}$ contain a Zariski open subset and so must intersect. As both are orbits of a group action, they therefore coincide and hence $V = e_d \, \mathrm{SO}(d)$.

[†] To see this, notice that if $x_1^2 + \cdots + x_d^2 - 1$ factorizes, then it would have to factor into two linear polynomials, and this may easily be shown to be impossible.

If $\mathrm{SO}(d)$ were not Zariski connected, then the argument above would also apply to its connected component of the identity $\mathbb{L} = \mathrm{SO}(d)^o$. Hence for any $g \in \mathrm{SO}(d) \smallsetminus \mathbb{L}$ we would find some $h \in \mathbb{L}$ with $e_d h = e_d g$ and hence with $g_1 = gh^{-1} \in \mathrm{SO}(d-1) \smallsetminus \mathbb{L}$. Clearly

$$\mathrm{SO}(2) \cong \mathrm{SO}(1,1) \cong \mathrm{SO}(x_1 x_2)$$

is a connected torus subgroup. Hence we may assume that $\mathrm{SO}(d-1)$ is Zariski connected. However, this is a contradiction since $\mathrm{SO}(d-1)$ intersects $\mathbb{L}$ and $g_1\mathbb{L}$ non-trivially.

The last claim of the lemma follows from the classification of Lie algebras, for which we refer to Knapp [**?**] (see also Exercise7.5.1). □

Lemma 7.11 ($\mathbb{R}$-points of orthogonal groups). *Fix $d = p + q \geqslant 2$. Then*

$$\mathrm{SO}(p,q)(\mathbb{R})^o \leqslant \mathrm{SO}(p,q)(\mathbb{R})$$

has index one if $d = p, q = 0$ and has index two if $1 \leqslant q \leqslant p \leqslant d$.

Outline proof. For $d = 2$ there are only two cases to consider:

- $\mathrm{SO}(2)(\mathbb{R})^o = \mathrm{SO}(2)(\mathbb{R})$, and
- $[\mathrm{SO}(1,1)(\mathbb{R}) : \mathrm{SO}(1,1)(\mathbb{R})^o] = [\mathbb{R}^\times : \mathbb{R}^\times_{>0}] = 2$.

For $d = 3$ we have:

- $\mathrm{SO}(3)(\mathbb{R})$, which is again connected, and
- $\mathrm{SO}(2,1)(\mathbb{R})$, which contains the subgroup $\mathrm{SO}(2,1)(\mathbb{R})^o$ of index two.

To see the second claim, consider the connected one-sheeted hyperboloid

$$V = \{a \in \mathbb{R}^3 \mid a_1^2 + a_2^a - a_3^2 = 1\}.$$

By a dimension-counting argument we can again show that

$$V = \{e_1 g \mid g \in \mathrm{SO}(2,1)(\mathbb{R})^o\},$$

so that for any $g \in \mathrm{SO}(2,1)(\mathbb{R})$ we can find $h \in \mathrm{SO}(2,1)(\mathbb{R})^o$ with $e_1 g = e_1 h$, and so $e_1(gh^{-1}) = e_1$. The latter we can interpret as the statement

$$gh^{-1} \in \mathrm{SO}(1,1)(\mathbb{R})$$

(where $\mathrm{SO}(1,1)$ is embedded into $\mathrm{SO}(2,1)$ in the lower right corner). As we already know that case, we can extend the claim to $\mathrm{SO}(2,1)(\mathbb{R})$ also.

In general, we have to know when the set

$$V = \{a \in \mathbb{R}^d \mid a_1^2 + \cdots + a_p^2 - a_{p+1}^2 - \cdots - a_d^2 = 1\}$$

is connected. If $p > 1$ it is, but if $p = 1$ it has two connected components since no element of V can have $a_1 = 0$. □

Example 7.12. Let $d \in \mathbb{N}$ be a non-square, and define the quadratic form

$$Q(x, y, z) = x^2 + y^2 + \sqrt{d}z^2.$$

This makes $\mathrm{SO}(Q)$ into an algebraic group defined over $\mathbb{Q}(\sqrt{d})$, and hence

$$\mathbb{G} =_{\mathbb{Q}(\sqrt{d})|\mathbb{Q}} (\mathrm{SO}(Q))$$

is an algebraic group defined over $\mathbb{Q}$. Since d is positive, the field $\mathbb{Q}(\sqrt{d})$ has two real embeddings (that is, the field splits over $\mathbb{R}$) and hence $\operatorname{Im}\phi$ (in the notation of Section 7.3) can be diagonalized over $\mathbb{R}$. It follows that

$$\begin{aligned}\mathbb{G}(\mathbb{R}) &\cong \mathrm{SO}(x^2 + y^2 + \sqrt{d}z^2)(\mathbb{R}) \times \mathrm{SO}(x^2 + y^2 - \sqrt{d}z^2)(\mathbb{R}) \\ &\cong \mathrm{SO}(3)(\mathbb{R}) \times \mathrm{SO}(2, 1)(\mathbb{R}).\end{aligned}$$

Hence $\mathbb{G}(\mathbb{Z})$ is a discrete subgroup (in fact a lattice by Section 7.4) of a product of a compact and a non-compact semi-simple group, that intersects each factor trivially.

Of course other quadratic forms and other fields give rise to similar examples, with various numbers of factors and signatures.

Exercises for Section 7.5

Exercise 7.5.1. Let $\mathbb{S}^3 \subseteq \mathbb{H} = \mathbb{R}[i, j, k]$ be the unit sphere in the Hamiltonian quaternions $\mathbb{H}$. Show that $\phi(z) : t \mapsto ztz^{-1}$ for $t \in \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$ defines an algebraic group homomorphism from an algebraic group whose group of $\mathbb{R}$-points equals $\mathbb{S}^3$ (equipped with multiplication in the Hamiltonian quaternions) to $\mathrm{SO}_3(\mathbb{R})$. Show that the map has finite kernel and is onto.

Show that $\psi(z_1, z_2) : t \mapsto z_1 t z_2^{-1}$ for $t \in \mathbb{H} = \mathbb{R}[i, j, k]$ defines in the same sense an algebraic group homomorphism from $\mathbb{S}^3 \times \mathbb{S}^3$ into $\mathrm{SO}_4(\mathbb{R})$. Show that the map has finite kernel and is onto.

7.6 An example of an isogeny for algebraic groups

Another source of slightly different algebraic groups are given by *isogenies*. We will now discuss SL_d and $_d$ as a special case of this construction.

Proposition 7.13 (Projective general linear group). *For every $d \geqslant 2$ there is a simple algebraic group $_d$ which*

- *is defined over $\mathbb{Q}$,*
- *is locally isomorphic to SL_d, and*
- *has trivial center.*

There also exists a homomorphism $\phi : \mathrm{SL}_d \to_d$ *whose kernel is the center of* SL_d*. If* $\mathbb{K} = \mathbb{R}$ *(or any other field), then*

$$\left[{}_d(\mathbb{K}) : \phi(\mathrm{SL}_d(\mathbb{K}))\right] = \left[\mathbb{K}^\times : (\mathbb{K}^\times)^d\right]. \tag{7.4}$$

For $d = 2$ *we have* ${}_2 \cong \mathrm{SO}(2,1)$ *over* $\mathbb{Q}$*. The map* $\phi : \mathrm{SL}_d \to_d$ *is called an* isogeny*.*

Proof. For $g \in \mathrm{GL}_d$ we define the restriction

$$\phi(g) = \mathrm{Ad}_g |_{\mathfrak{sl}_d}$$

of the adjoint representation of GL_d to the Lie algebra $\mathfrak{sl}_d$ of SL_d. The center of GL_d belongs to the kernel of ϕ. Hence

$$\phi : \mathrm{GL}_d \longrightarrow \mathrm{SL}(\mathfrak{sl}_d)$$

is a representation defined over $\mathbb{Q}$. We define ${}_d$ to be the Zariski closure of the image of ϕ. Since GL_d is Zariski connected, ${}_d$ is a Zariski connected algebraic group defined over $\mathbb{Q}$.

By a dimension argument (similar to the one used in proving Lemma 7.10, and using Shafarevich [**?**, Ch. I, Sec. 5, Th. 6]) we see that over an algebraically closed field $\overline{\mathbb{K}}$ we have

$${}_d(\overline{\mathbb{K}}) = \phi\left(\mathrm{GL}_d(\overline{\mathbb{K}})\right) = \phi\left(\mathrm{SL}_d(\overline{\mathbb{K}})\right).$$

However, we claim that

$$(\mathbb{K}) = \phi\left(\mathrm{GL}_d(\mathbb{K})\right)$$

over any field $\mathbb{K}$. To see this, let $g \in \mathrm{GL}_d(\overline{\mathbb{K}})$ be such that $\phi(g) \in_d (\mathbb{K})$. Modifying g by an element of the center of GL_d, we may assume that

$$g_{i_0,j_0} \in \mathbb{K} \smallsetminus \{0\}$$

for some $i_0, j_0 \in \{1, \dots, d\}$. Since g acts trivially on the span of the identity element and (by assumption) rationally on the hyperplane $\mathfrak{sl}_d \subseteq \mathrm{Mat}_d$ we see that

$$gE_{ij}g^{-1} \in \mathrm{Mat}_d(\mathbb{K})$$

for all $i, j \in \{1, \dots, d\}$. We choose $i = j_0$ and set $h = g^{-1}$ to conclude that

$$g_{i_0 j_0} h_{j\ell} \in \mathbb{K}$$

for $j, \ell = 1, \dots, d$. This shows that $h = g^{-1} \in \mathrm{GL}_d(\mathbb{K})$, and hence that g is in $\mathrm{GL}_d(\mathbb{K})$.

We may modify g only by elements from $\mathbb{C}(\mathbb{K})$, where $\mathbb{C}$ is the center of GL_d. Doing so modifies the determinant by a dth power of an element of $\mathbb{K}^\times$, which implies the index formula in (7.4).

It remains to show that ${}_2 \cong \mathrm{SO}(2,1)$. For this, notice that

$$\operatorname{tr}\left(\phi(g)(v)\phi(g)(w)\right) = \operatorname{tr}\left(gvg^{-1}gwg^{-1}\right) = \operatorname{tr}(vw),$$

so that the bilinear Killing form

$$\operatorname{tr}(vw)$$

for $v, w \in \mathfrak{sl}_2 \subseteq \mathrm{Mat}_2$ is preserved by all elements $\phi(g)$ for $g \in \mathrm{GL}_2$. Now notice that the quadratic form

$$\operatorname{tr}\left(\begin{pmatrix} a & b \\ c & -a \end{pmatrix}\begin{pmatrix} a & b \\ c & -a \end{pmatrix}\right) = \operatorname{tr}\begin{pmatrix} a^2+bc & 0 \\ 0 & bc+a^2 \end{pmatrix} = 2(a^2+bc)$$

is conjugate over $\mathbb{Q}$ to the quadratic form

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

Since ${}_2 \subseteq \mathrm{SO}(2,1)$ and both ${}_2$ and $\mathrm{SO}(2,1)$ are 3-dimensional and Zariski connected, we obtain the proposition. □

mention "adjoint" vs "simply connected" orthogonal groups may be neither, see SO(4), spin group

7.7 Irreducibility of Lattices for $\mathbb{Q}$-groups

7.7.1 A Weird Irreducible Lattice

Let us show briefly how badly Corollary 3.33 can fail for a general lattice in a product of Lie groups in the presence of compact factors. Lattices arising from $\mathbb{Q}$-groups do not show such weird behavior (see Section 7.7.2).

Example 7.14. Let $\Gamma < \mathrm{SL}_2(\mathbb{R})$ be any lattice that is isomorphic as an abstract group to a free group $\mathbb{F}_\ell$ with $\ell \geqslant 2$ generators. Now let $\phi : \Gamma \to K$ be a group homomorphism into a compact Lie group. For concreteness, we list some examples:

- $K = \mathrm{SO}(2)$ and $\ell = 3$, with the image of the first two generators arbitrary elements of K and the third being the identity;
- $K = \mathrm{SO}(3)$ and $\ell = 3$, with the images as before lying inside a copy of $\mathrm{SO}(2) < K$;
- $K = \mathrm{SO}(5)$ and $\ell = 2$, with two arbitrary images.

The graph

$$\{(\gamma, \phi(\gamma)) \mid \gamma \in \Gamma\} \subseteq G \times K$$

is a lattice in $G \times K$ which does not satisfy the conclusions of Corollary 3.33.

7.7.2 Irreducible Lattices and $\mathbb{Q}$-simple Groups

Lemma 7.15. *Suppose $\mathbb{G} = \mathbb{H}_1\mathbb{H}_2$ is the almost direct product of two semi-simple linear algebraic groups defined over $\mathbb{Q}$. Then $\mathbb{G}(\mathbb{Z})$ is a reducible lattice, and is commensurable with $\mathbb{H}_1(\mathbb{Z})\mathbb{H}_2(\mathbb{Z}) \leqslant \mathbb{G}(\mathbb{Z})$.*

PROOF. By Section 7.4, $\mathbb{H}_i(\mathbb{Z}) \leqslant \mathbb{H}_i(\mathbb{R})$ is a lattice for $i = 1, 2$. Since

$$\mathbb{H}_i(\mathbb{Z}) = \mathbb{G}(\mathbb{Z}) \cap \mathbb{H}_i(\mathbb{R})$$

for $i = 1, 2$, it follows that $\mathbb{G}(\mathbb{Z})$ is a reducible lattice. □

In the following we will assume that the algebraic group $\mathbb{G}$ cannot be decomposed over $\mathbb{Q}$ as in Lemma 7.15. More precisely, we say that a semi-simple linear algebraic group $\mathbb{G}$ defined over $\mathbb{Q}$ is $\mathbb{Q}$-*simple* (or $\mathbb{Q}$-*almost simple*) if it is impossible to write $\mathbb{G}$ as a product $\mathbb{H}_1\mathbb{H}_2$ of two proper semi-simple algebraic subgroups for which $\mathbb{H}_1 \cap \mathbb{H}_2$ is finite. Equivalently, $\mathbb{G}$ is $\mathbb{Q}$-simple if there does not exist a connected normal algebraic subgroup $\mathbb{H}_1 \lhd \mathbb{G}$ defined over $\mathbb{Q}$ (for if such a subgroup $\mathbb{H}_1$ existed, then $\mathbb{H}_2 = C_{\mathbb{G}}(\mathbb{H}_1)^o$ would give the decomposition $\mathbb{G} = \mathbb{H}_1\mathbb{H}_2$). By the following lemma, we have already seen examples of such groups.

Lemma 7.16. *Suppose that $\mathbb{G}$ is a simple*[†] *linear algebraic group defined over a number field $\mathbb{K}|\mathbb{Q}$. Then ${}_{\mathbb{K}|\mathbb{Q}}(\mathbb{G})$ is a semi-simple $\mathbb{Q}$-simple linear algebraic group.*

PROOF. Clearly the Galois group of $\overline{\mathbb{Q}}|\mathbb{Q}$ acts transitively on the set of all embeddings of $\mathbb{Q}$ into $\overline{\mathbb{Q}}$. Each such embedding corresponds to one simple factor of ${}_{\mathbb{K}|\mathbb{Q}}(\mathbb{G})$ over $\overline{\mathbb{Q}}$. Therefore, the Galois group of $\overline{\mathbb{Q}}|\mathbb{Q}$ acts transitively on the set of all simple factors of ${}_{\mathbb{K}|\mathbb{Q}}(\mathbb{G})$. Hence the only connected normal algebraic subgroups $\mathbb{H} \lhd_{\mathbb{K}|\mathbb{Q}} (\mathbb{G})$ are $\mathbb{H} = \{I\}$ and $\mathbb{H} =_{\mathbb{K}|\mathbb{Q}} (\mathbb{G})$. □

Corollary 7.17. *Let $\mathbb{G}$ be a $\mathbb{Q}$-simple linear algebraic group such that $\mathbb{G}(\mathbb{R})$ is non-compact. Then $\mathbb{G}(\mathbb{Z})$ is an irreducible lattice and is Zariski dense in $\mathbb{G}$. Moreover, if $\mathbb{G} = \mathbb{H}_1\mathbb{H}_2$ is an almost direct product of the semi-simple algebraic subgroups $\mathbb{H}_1$ and $\mathbb{H}_2$ defined over $\mathbb{R}$, and $\mathbb{H}_2(\mathbb{R})$ is non-compact, then the projection of $\mathbb{G}(\mathbb{Z})$ to*

$$\mathbb{G}(\mathbb{R})/C(\mathbb{G}(\mathbb{R}))\mathbb{H}_2(\mathbb{R}) \geqslant \mathbb{H}_1(\mathbb{R})/C(\mathbb{H}_2(\mathbb{R}))$$

is dense in $\mathbb{H}_1(\mathbb{R})^o/C(\mathbb{H}_1(\mathbb{R}))$.

Notice that in the presence of compact factors the conclusion above improves on Corollary 3.33 in ways that are not possible for general lattices

[†] Since we may identify $\mathbb{G}$ with its $\overline{\mathbb{Q}}$-points, simple here means simple over $\overline{\mathbb{Q}}$. Sometimes this is also referred to an absolutely simple.

(see Section 7.7.1). In other words, the algebraic nature of the lattice allows us to also obtain density statements for projections to compact subgroups if (and this is clearly a necessary assumption) the kernel of the projection map is non-compact.

Proof of Corollary 7.17. Suppose first that $\mathbb{G} = \mathbb{H}_1\mathbb{H}_2$ is an almost direct product defined over $\mathbb{R}$, and

$$\Gamma \cap \mathbb{H}_i(\mathbb{R}) < \mathbb{H}_i(\mathbb{R})$$

is a lattice for $i = 1, 2$. Since $\mathbb{G}(\mathbb{R})$ is non-compact, at least one of the almost direct factors, say $\mathbb{H}_1$, is also non-compact. Now take the Zariski closure of $\mathbb{G}(\mathbb{Z}) \cap \mathbb{H}_1(\mathbb{R})$ to obtain an algebraic subgroup $\mathbb{L} \leqslant \mathbb{G}$ defined over $\mathbb{Q}$ (by Lemma 3.19). By the Borel density theorem (Theorem 3.30) and our assumption, $\mathbb{L} \leqslant \mathbb{H}_1$ contains all non-compact factors of $\mathbb{H}_1(\mathbb{R})$. The group $\mathbb{L}^o$ is defined over $\overline{\mathbb{Q}}$ and is a direct factor of $\mathbb{G}$ contradicting our assumption that $\mathbb{G}$ is $\mathbb{Q}$-simple. However, it is not so clear why $\mathbb{L}^o$ is normal in $\mathbb{G}$. Because of this, we define the *core*

$$\mathbb{F} = \bigcap_{g\in\mathbb{G}} g\mathbb{L}^o g^{-1} = \bigcap_{g\in\mathbb{H}_1} g\mathbb{L}^o g^{-1}$$

which is a normal algebraic subgroup $\mathbb{F} \lhd \mathbb{G}$ contained in $\mathbb{L}^o \leqslant \mathbb{H}_1$, commuting with $\mathbb{H}_2$. Since $\mathbb{L}^o$ is defined over $\mathbb{Q}$ (by Lemma 3.22) and $\mathbb{G}$ is defined over $\mathbb{Q}$, the group $\mathbb{F}$ is invariant under all Galois automorphisms of $\overline{\mathbb{Q}}|\mathbb{Q}$ and therefore is also defined over $\mathbb{Q}$ (by Lemma 3.20). By construction $\mathbb{F}_1 = \mathbb{F}^o \lhd \mathbb{G}$ is an algebraic subgroup defined over $\mathbb{Q}$ which is contained in $\mathbb{H}_1$ and contains all non-compact factors of $\mathbb{H}_1(\mathbb{R})$. This is a contradiction to the assumption that $\mathbb{G}$ is $\mathbb{Q}$-simple if $\mathbb{H}_2$ is non-trivial, and so we have shown that $\mathbb{G}(\mathbb{Z})$ is an irreducible lattice.

We only needed to assume that $\mathbb{H}_2$ is non-trivial in the argument above at the very last step. Applying the argument to the case $\mathbb{H}_1 = \mathbb{G}$ and $\mathbb{H}_2 = \{I\}$, we obtain the Zariski closure $\mathbb{L} \leqslant \mathbb{G}$ of $\mathbb{G}(\mathbb{Z})$. If $\mathbb{L} \neq \mathbb{G}$, then we obtain moreover the subgroup $\mathbb{F} \lhd \mathbb{G}$ defined over $\mathbb{Q}$, and get just as before a contradiction to the assumption that $\mathbb{G}$ is $\mathbb{Q}$-simple. Therefore, we have shown the strengthening of the Borel density theorem that $\mathbb{G}(\mathbb{Z})$ is Zariski dense in $\mathbb{G}$ (regardless of the presence of compact factors).

Using this strengthened version of Borel density, we can argue as in the proof of Corollary 3.33 to obtain the last claim of the corollary. Indeed, suppose that $\mathbb{G} = \mathbb{H}_1\mathbb{H}_2$ is an almost direct product over $\mathbb{R}$, and that $\mathbb{H}_2(\mathbb{R})$ is non-compact. Now project $\mathbb{G}(\mathbb{Z})$ to $\mathbb{G}(\mathbb{R})/C(\mathbb{G})(\mathbb{R})\mathbb{H}_2(\mathbb{R})$ and denote by F the pre-image in $\mathbb{H}_1(\mathbb{R})$ of the closure:

$$F = \pi_1^{-1}\left(\overline{\pi_1(\mathbb{G}(\mathbb{Z}))}\right) \cap \mathbb{H}_1(\mathbb{R})$$

where

$$\pi_1 : \mathbb{G}(\mathbb{R}) \longrightarrow \mathbb{G}(\mathbb{R})/C(\mathbb{G})(\mathbb{R})\mathbb{H}_2(\mathbb{R})$$

is the natural projection. By construction $\mathbb{G}(\mathbb{Z})$ normalizes F, and hence also its Lie algebra. By the Zariski density of $\mathbb{G}(\mathbb{Z})$ in $\mathbb{G}$, it follows that

$$F^o \lhd \mathbb{H}_1(\mathbb{R})$$

is normal. We wish to show that $F^o = \mathbb{H}_1(\mathbb{R})$, so suppose for now that this is not the case. Then $\mathbb{H}_1 = \mathbb{F}_1\mathbb{F}_2$ is an almost direct product over $\mathbb{R}$ with

$$F^o = \mathbb{F}_2(\mathbb{R})^o$$

and $\mathbb{F}_1 \neq \{I\}$. However, this contradicts once again the assumption that $\mathbb{G}$ is $\mathbb{Q}$-simple: as in Corollary 3.33 we see in turn that the projection of $\mathbb{G}(\mathbb{Z})$ to $\mathbb{G}(\mathbb{R})/C(\mathbb{G}(\mathbb{R}))\mathbb{F}_2(\mathbb{R})\mathbb{H}_2(\mathbb{R})$ is discrete, that

$$\Lambda = \mathbb{G}(\mathbb{Z}) \cap (\mathbb{F}_2(\mathbb{R})\mathbb{H}_2(\mathbb{R}))$$

is a lattice in $\mathbb{F}_2(\mathbb{R})\mathbb{H}_2(\mathbb{R})$, and that the Zariski closure $\mathbb{L}$ of Λ contains all non-compact factors of $\mathbb{F}_2(\mathbb{R})\mathbb{H}_2(\mathbb{R})$. We again construct the core of $\mathbb{L}$, which is a normal $\mathbb{Q}$-group containing all non-compact factors of $\mathbb{F}_2(\mathbb{R})$ and of $\mathbb{H}_2(\mathbb{R})$, and is contained in $\mathbb{F}_2\mathbb{H}_2$. This contradicts the assumption that $\mathbb{G}$ is $\mathbb{Q}$-simple, and completes the proof. □

The reader may get the feeling from the arguments above concerning $\mathbb{Q}$-simple groups, and from the fact that we only had groups arising from restriction of scalars of simple groups as examples, that these are the only $\mathbb{Q}$-simple groups. In fact this is almost correct in the following sense. Restriction of scalars ${}_{\mathbb{K}|\mathbb{Q}}\mathbb{G}$ applied to an absolutely simple group $\mathbb{G}$ defined over $\mathbb{K}$ gives rise to a semi-simple group that is the direct product of its simple factors (instead of an almost direct product). If a $\mathbb{Q}$-simple group $\mathbb{G}$ is the direct product of its simple factors (for example, because it is adjoint or because it is simply connected), then it is of the form $\mathbb{G} =_{\mathbb{K}|\mathbb{Q}} \mathbb{H}$ for an absolutely simple group $\mathbb{H}$ defined over a number field. The proof of this result is not very different to the proofs of this section; we only mention here that the number field in question arises as the field corresponding to the subgroup that stabilizes a given simple factor of $\mathbb{G}$.

Exercises for Section 3.6

Exercise 7.7.1. Let $\mathbb{G}$ be a $\mathbb{Q}$-simple algebraic group which is not simple over $\mathbb{R}$, and let $\mathbb{F} \lhd \mathbb{G}$ be a simple factor over $\mathbb{R}$. Show that $\mathbb{G}(\mathbb{Z}) \cap \mathbb{F}(\mathbb{R})$ is finite.

7.8 Almost Algebraicity and Algebraic Lie Subalgebras

Definition 7.18. A Lie subalgebra $\mathfrak{h} \subseteq \mathfrak{sl}_d$ is called *algebraic* if it is the Lie algebra of an algebraic subgroup of SL_d. A connected Lie subgroup $H \leqslant \mathrm{SL}_d(\mathbb{R})$ is called *almost algebraic* if it is the connected component $\mathbb{H}(\mathbb{R})^o$ of the $\mathbb{R}$-points $\mathbb{H}(\mathbb{R})$ of an algebraic subgroup $\mathbb{H}$ defined over $\mathbb{R}$.

Lemma 7.19. *Let $H \leqslant \mathrm{SL}_d(\mathbb{R})$ be a connected Lie subgroup. Then the following are equivalent:*

(1) *H is almost algebraic;*
(2) *the Lie algebra of H is algebraic.*

Proof. If $H = \mathbb{H}(\mathbb{R})^o$ for some algebraic group $\mathbb{H}$ defined over $\mathbb{R}$, then the Lie algebras of $\mathbb{H}$ and of H coincide. This shows that (1) implies (2). If the Lie algebra $\mathfrak{h}$ of H is also the Lie algebra of an algebraic group $\mathbb{L}$ then $\mathfrak{h}$ uniquely determines $\mathbb{L}^o$ as the Zariski closure of $\exp(\mathfrak{h})$, and so in particular $\mathbb{L}^o$ is defined over $\mathbb{R}$ since $\mathfrak{h}$ is. By dimension counting we must therefore have $H = \mathbb{L}^o(\mathbb{R})^o$, and the lemma follows. □

To see that the notion of almost algebraic is necessary, notice that the Zariski closure of the one-dimensional Lie subgroup

$$H = \left\{ \begin{pmatrix} \mathrm{e}^t & & & \\ & \mathrm{e}^{-t} & & \\ & & 1 & t \\ & & & 1 \end{pmatrix} \mid t \in \mathbb{R} \right\} \leqslant \mathrm{SL}_4(\mathbb{R})$$

is equal to the two-dimensional algebraic group

$$\mathbb{L} = \left\{ \begin{pmatrix} \alpha & & & \\ & \alpha^{-1} & & \\ & & 1 & \beta \\ & & & 1 \end{pmatrix} \mid \alpha \neq 0 \right\}.$$

This follows from Proposition 3.28.

Another type of counterexample comes from subgroups of torus groups. For example,

$$H = \left\{ \begin{pmatrix} \mathrm{e}^t & & \\ & \mathrm{e}^{\alpha t} & \\ & & \mathrm{e}^{-(1+\alpha)t} \end{pmatrix} \mid t \in \mathbb{R} \right\}$$

has for almost every choice of $\alpha \in \mathbb{R}$, the full two-dimensional subgroup

$$\mathbb{T} = \left\{ \begin{pmatrix} a_1 & & \\ & a_2 & \\ & & a_3 \end{pmatrix} \mid a_1 a_2 a_2 = 1 \right\}$$

as algebraic closure.

In some sense these two types of example are the only source of non-algebraic Lie subalgebras.

Proposition 7.20. *Any semi-simple Lie subalgebra* $\mathfrak{h} \subseteq \mathrm{SL}_d$ *is algebraic.*

Proof. First define the normalizer

$$N_{\mathrm{SL}_d}(\mathfrak{h}) = \{g \in \mathrm{SL}_d \mid \mathrm{Ad}_g(\mathfrak{h}) = \mathfrak{h}\}$$

of the subalgebra $\mathfrak{h}$. By Knapp [?, Prop. 1.20, 1.21] the Lie algebra of $N_{\mathrm{SL}_d}(\mathfrak{h})$ is $\mathfrak{h} + \mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h})$ where

$$\mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h}) = \{v \in \mathfrak{sl}_d \mid [v, \mathfrak{h}] = 0\}$$

is the centralizer of $\mathfrak{h}$ in $\mathfrak{sl}_d$. Now define

$$\mathbb{M} = N_{\mathrm{SL}_d}(\mathfrak{h}) \cap C_{\mathrm{SL}_d}\left(\mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h})\right),$$

where

$$C_{\mathrm{SL}_d}\left(\mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h})\right) = \{g \in \mathrm{SL}_d \mid \mathrm{Ad}_g|_{\mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h})} = \mathrm{id}\}$$

is the centralizer of $\mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h})$ in SL_d. Notice that $\mathfrak{h}$ still belongs to the Lie algebra of $\mathbb{M}$. Finally define

$$\mathbb{L} = [\mathbb{M}^o, \mathbb{M}^o]$$

as the Zariski closure of the group generated by all commutators of elements in the connected component $\mathbb{M}^o$. The Lie algebra of $\mathbb{M}$ is $\mathfrak{h} \oplus \mathfrak{f}$ where $\mathfrak{f} \subseteq \mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h})$ is the center of $\mathfrak{c}_{\mathfrak{sl}_d}(\mathfrak{h})$. By the Levi decomposition of algebraic groups, $\mathbb{M}^o = \mathbb{L}\mathbb{F}$ for some abelian $\mathbb{F}$, and

$$\mathbb{L} = [\mathbb{M}^o, \mathbb{M}^o]$$

is the desired algebraic group. □

Corollary 7.21. *If* $\mathfrak{h} = \mathfrak{h}_{ss} \ltimes \mathfrak{u} \subseteq \mathfrak{sl}_d$ *is a semi-direct product of a semi-simple Lie subalgebra* $\mathfrak{h}_{ss} \subseteq \mathfrak{sl}_d$ *and a nilpotent Lie subalgebra* $\mathfrak{u} \subseteq \mathfrak{sl}_d$ *such that* $\exp(\mathfrak{u})$ *is a unipotent subgroup, the* $\mathfrak{h}$ *is an algebraic Lie subalgebra.*

Another way of stating essentially the same result is the following.

Corollary 7.22. *Let* $\mathfrak{h} \subseteq \mathfrak{sl}_d$ *be a Lie subalgebra which is not algebraic. Then the Zariski closure of* $\exp(\mathfrak{h})$ *has a nontrivial torus subgroup in its radical.*

Proof.[Proof of Corollaries 7.21–7.22]

□

Problems for Section 7.8

Exercise 7.8.1. Show that $\mathrm{SO}(2,1) \cong_2 (\mathbb{R})$, and show that $\mathrm{PSL}_2(\mathbb{R})$ is a subgroup of index 2 in $_2(\mathbb{R})$. Deduce that the lattice in Example 3.3(4) gives rise to a uniform lattice in $\mathrm{PSL}_2(\mathbb{R})$.

Exercise 7.8.2. Prove that $\mathbb{G}(\mathbb{Z})$ in Example 7.12 is indeed a uniform lattice (this may be done, for example, by generalizing the arguments used in Section 3.1).

Exercise 7.8.3. Prove that $\mathbb{G}(\mathbb{Z})$ as in Example 7.6 is indeed a lattice in some or all cases (follow the proof of Theorem 1.18, use the NAK decomposition of $\mathbb{G}(\mathbb{R})$, the fact that $N(\mathbb{Z})$ is a lattice in $N(\mathbb{R})$, and Dirichlet's unit theorem).

Chapter 8
S-algebraic Groups and Quotients

8.1 S-algebraic and Adelic Extensions of X_d

† In this section we will define restricted direct products and the adele ring(26) in both zero and positive characteristic. This leads to the notion of S-algebraic group. The reader is assumed to be familiar with the p-adic numbers $\mathbb{Q}_p$ viewed as a completion of $\mathbb{Q}$ with respect to the p-adic norm $|x|_p = p^{-\operatorname{ord}_p(x)}$, the maximal compact subring $\mathbb{Z}_p$, and their positive characteristic analogs (we refer to Koblitz [?] or [?] for a friendly introduction to p-adic numbers, and to Weil [?] for a more advanced treatment). The groups we will be concerned with in this section include

$$\mathrm{SL}_d(\mathbb{Q}_p) = \{g \in \mathrm{Mat}_d(\mathbb{Q}_p) \mid \det g = 1\},$$

and the clopen subgroup

$$\mathrm{SL}_d(\mathbb{Z}_p) = \{g \in \mathrm{Mat}_d(\mathbb{Z}_p) \mid \det g = 1\}.$$

Similarly, we may define $\mathrm{SL}_d(\mathbb{R} \times \mathbb{Q}_p)$, but notice that

$$\mathrm{SL}_d(\mathbb{R} \times \mathbb{Q}_p) \cong \mathrm{SL}_d(\mathbb{R}) \times \mathrm{SL}_d(\mathbb{Q}_p).$$

More generally (see below for the precise definitions) we have

$$\mathrm{SL}_d(\mathbb{A}_{\mathbb{Q}}) \cong \prod_{\sigma}{}' \mathrm{SL}_d(\mathbb{Q}_\sigma).$$

Here the indicated isomorphisms are isomorphisms of topological groups.

† If the reader is only interested in actions on a quotient of a Lie group, then this chapter may be skipped.

8.1.1 Zero Characteristic S-algebraic Groups

Let $S \subseteq \{\infty, 2, 3, 5, 7, \dots\}$ be a subset of the set of *places* of $\mathbb{Q}$. Suppose that for every $\sigma \in S$ we are given a closed subgroup[†] $G_\sigma \subseteq \mathrm{SL}_d(\mathbb{Q}_\sigma)$. We define the associated *S-algebraic group*

$$G_S = \prod_{\sigma \in S}{}' G_\sigma$$

to be the restricted direct product of the groups G_σ with respect to the compact subgroups $G_\sigma \cap \mathrm{SL}_d(\mathbb{Z}_\sigma)$ (the prime on the product symbol denotes the restricted product). We quickly review what this construction means in our setting (see Weil [?] for a full treatment).

If S is finite, then the restricted direct product is simply the direct product

$$G_S = \prod_{\sigma \in S} G_\sigma,$$

with the product topology. If S is not finite, then (as an abstract group)

$$G_S = \bigcup_{\substack{S \cap \{\infty\} \subseteq S' \subseteq S, \\ |S'| < \infty}} \prod_{\sigma \in S'} G_\sigma \times \prod_{p \in S \smallsetminus S'} \left(G_p \cap \mathrm{SL}_d(\mathbb{Z}_p) \right).$$

In the case $|S| = \infty$, we can use for any finite set $S' \subseteq S$ the product topology on each set of the form

$$G_{S,S'} = \prod_{\sigma \in S'} G_\sigma \times \prod_{p \in S \smallsetminus S'} \left(G_p \cap \mathrm{SL}_d(\mathbb{Z}_p) \right), \tag{8.1}$$

and notice that $G_{S,S''}$ is an open subgroup of $G_{S,S'}$ for any finite set S'' with

$$S \cap \{\infty\} \subseteq S'' \subseteq S' \subseteq S.$$

We can therefore define the topology on G_S by requiring that G_S be a topological group and that $G_{S,S'}$ be an open subgroup of G_S for any finite set S' with $S \cap \{\infty\} \subseteq S' \subseteq S$. In all of the expressions above the place denoted ∞, corresponding to the completion $\mathbb{R}$ of $\mathbb{Q}$, plays a special role in the following way. If $\infty \in S$, then ∞ also has to belong to the subsets $S', S'' \subseteq S$ arising.

Lemma 8.1 (Left-invariant proper metric). *The topology on G_S described above is given by a proper[‡] left-invariant metric.*

The existence of the metric is generally established in the appendix. Shall we keep the lemma (with a shortened proof) because it gives a proper metric?

[†] In Chapter 3 we will discuss ‘algebraic groups’ in more detail. In this terminology one usually takes $G_\sigma = \mathbb{G}(\mathbb{Q}_\sigma)$ for an algebraic group $\mathbb{G}$ defined over $\mathbb{Q}$.

[‡] That is, the metric has the property that every closed ball of finite radius defined by the metric is compact.

PROOF. We prove this first in the case $S=\{\sigma\}$, which readily gives the case of S finite. So assume that $S=\{\sigma\}$, and notice that $G_\sigma\subseteq\mathrm{SL}_d(\mathbb{Q}_\sigma)$ is a closed subgroup, so that it is sufficient to prove this for $\mathrm{SL}_d(\mathbb{Q}_\sigma)$.

If $\sigma=\infty$, then $\mathrm{SL}_d(\mathbb{R})$ is a connected Lie group (see the discussion after Lemma 1.24), and the lemma follows by considering the Riemannian metric on $\mathrm{SL}_d(\mathbb{R})$ (see [**?**, Ch. 9]).

If $\sigma=p$ is a finite place, then $\mathrm{SL}_d(\mathbb{Z}_p)$ is a compact open subgroup of $\mathrm{SL}_d(\mathbb{Q}_p)$. The topology of $\mathrm{SL}_d(\mathbb{Z}_p)$ may be induced by considering the metric

$$\mathsf{d}((g_{ij}),(h_{ij}))=\max_{i,j}|g_{ij}-h_{ij}|_p. \tag{8.2}$$

Notice that

$$\mathsf{d}((g_{ij}),(h_{ij}))=\mathsf{d}(k_1(g_{ij})k_2,k_1(h_{ij})k_2),$$

for any $k_1,k_2\in\mathrm{SL}_d(\mathbb{Z}_p)$, so that d gives a bi-invariant metric on $\mathrm{SL}_d(\mathbb{Z}_p)$. This metric can then be extended to a left-invariant metric on $G=\mathrm{SL}_d(\mathbb{Q}_p)$ as follows. Choose a sequence (g_n) in $\mathrm{SL}_d(\mathbb{Q}_p)$ with $g_1=e$ for which

$$\mathrm{SL}_d(\mathbb{Q}_p)=\bigsqcup_{n=1}^{\infty}g_n\,\mathrm{SL}_d(\mathbb{Z}_p). \tag{8.3}$$

This is possible since $\mathrm{SL}_d(\mathbb{Z}_p)$ is a non-empty closed and open subgroup of $\mathrm{SL}_d(\mathbb{Q}_p)$. Recall that $\mathsf{d}(\cdot,\cdot)$ is bounded by 1 on $\mathrm{SL}_d(\mathbb{Z}_p)\times\mathrm{SL}_d(\mathbb{Z}_p)$. Now define an intermediate function by

$$f(h)=\begin{cases}\mathsf{d}(h,I) & \text{if } h\in\mathrm{SL}_d(\mathbb{Z}_p), \text{ and}\\ n & \text{if } h\in g_n\,\mathrm{SL}_d(\mathbb{Z}_p) \text{ for } n\geqslant 2.\end{cases}$$

In words, $f(h)$ is a measurement of the distance from h to I if $h\in\mathrm{SL}_d(\mathbb{Z}_p)$ and if not gives the index, which we may think of as an address, in the union (8.3). We now define

$$\mathsf{d}_2(g_1,g_2)=\widetilde{f}(g_1^{-1}g_2)$$

for $g_1,g_2\in\mathrm{SL}_d(\mathbb{Q}_p)$, where for $g\in\mathrm{SL}_d(\mathbb{Q}_p)$ we define

$$\widetilde{f}(g)=\inf\left\{\sum_{i=1}^{k}f(h_i)\mid g=h_1^{\varepsilon_1}\cdots h_k^{\varepsilon_k}\text{ with }\varepsilon_i\in\{\pm1\},h_i\in\mathrm{SL}_d(\mathbb{Q}_p)\right\}.$$

Notice that $\widetilde{f}(g)\leqslant 1$ implies that $g\in\mathrm{SL}_d(\mathbb{Z}_p)$. In particular, if $g\in\mathrm{SL}_d(\mathbb{Z}_p)$ then the infimum defining $\widetilde{f}(g)$ is attained with $k=1$ and $h_1=g$, so d_2 extends the metric d defined above on $\mathrm{SL}_d(\mathbb{Z}_p)$. For the same reason we see that $\mathsf{d}_2(g_1,g_2)=0$ implies that $g_1=g_2$. It is clear that the construction above gives left-invariance, meaning that

$$\mathsf{d}_2(hg_1, hg_2) = \mathsf{d}_2(g_1, g_2)$$

for all $h, g_1, g_2 \in \mathrm{SL}_d(\mathbb{Q}_p)$. Moreover, we have $\widetilde{f}(g) = \widetilde{f}(g^{-1})$, so

$$\mathsf{d}_2(g_1, g_2) = \mathsf{d}_2(g_2, g_1).$$

Since d_2 is left-invariant and agrees with d on the open subgroup $\mathrm{SL}_d(\mathbb{Z}_p)$, it defines the usual topology on $\mathrm{SL}_d(\mathbb{Q}_p)$. It remains to check the triangle inequality and to show that the resulting metric is proper.

For the triangle inequality we will use the structure of the definition of $\widetilde{f}$. Note that due to the left-invariance, it is sufficient to consider the points I, g_1, g_2. In this case

$$\mathsf{d}_2(e, g_1) = \widetilde{f}(g_1),$$

and so

$$\mathsf{d}_2(e, g_1) + \mathsf{d}_2(g_1, g_2) = \widetilde{f}(g_1) + \widetilde{f}(g_1^{-1}g_2),$$

which by definition is no smaller than $\widetilde{f}(g_2) = \mathsf{d}_2(e, g_2)$ as required.

To see properness, notice that $B_R^{\mathrm{SL}_d(\mathbb{Q}_p)}$ is contained in the finite union of all possible products of $\mathrm{SL}_d(\mathbb{Z}_p)$ and at most R sets (counted with multiplicity) of the form $g_n \mathrm{SL}_d(\mathbb{Z}_p)$ and $(g_n \mathrm{SL}_d(\mathbb{Z}_p))^{-1}$ with $2 \leqslant n \leqslant R$. Since this finite union is compact, this shows the lemma in the case $S = \{p\}$ also.

Now let S be an infinite set, with $\infty \notin S$. Let $S' = \varnothing$ and consider the open subgroup $G_{S,\varnothing}$ as in (8.1), on which it is easy to define a left-invariant proper metric using left-invariant proper metrics on $G_p \cap \mathrm{SL}_d(\mathbb{Z}_p)$, which is a compact open subgroup of G_p. Using the same argument as used above, this metric can be extended to G_S. If $\infty \in S$, then we may consider G_∞ and $G_{S\smallsetminus\{\infty\}}$ separately, and taking the product metric again obtain a metric as claimed in the lemma. $\square$

8.1.2 The p-Adic and Adelic Extension of X_d

Recall from Section 1.3 that $d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$ is the space of unimodular lattices in $\mathbb{R}^d$ and that d itself has finite volume (that is, $\mathrm{SL}_d(\mathbb{Z})$ is a lattice in $\mathrm{SL}_d(\mathbb{R})$). In this section we introduce a 'p-adic extension'

$$d, p = \mathrm{SL}_d(\mathbb{Z}[\tfrac{1}{p}])\backslash \underbrace{\mathrm{SL}_d(\mathbb{R}) \times \mathrm{SL}_d(\mathbb{Q}_p)}_{\cong \mathrm{SL}_d(\mathbb{R}\times\mathbb{Q}_p)} \tag{8.4}$$

where $\gamma \in \mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])$ is identified with the *diagonally embedded* element

$$(\gamma, \gamma) \in \mathrm{SL}_d(\mathbb{R}) \times \mathrm{SL}_d(\mathbb{Q}_p),$$

and the 'adelic extension'

$$d, \mathbb{A}_{\mathbb{Q}} = \mathrm{SL}_d(\mathbb{Q})\backslash \mathrm{SL}_d(\mathbb{A}_{\mathbb{Q}}), \tag{8.5}$$

where we use a similar *diagonal embedding* sending $\gamma \in \mathrm{SL}_d(\mathbb{Q})$ to

$$(\gamma, \gamma, \dots) \in \mathrm{SL}_d(\mathbb{R}) \times \prod_{p \in S'} \mathrm{SL}_d(\mathbb{Q}_p) \times \prod_{p \notin S'} \mathrm{SL}_d(\mathbb{Z}_p)$$

with $S' = S'(\gamma)$ chosen so that $\gamma \in \mathrm{SL}_d(\mathbb{Z}_p)$ for all $p \notin S'$. We will show that both $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])$ and $\mathrm{SL}_d(\mathbb{Q})$ are lattices in their respective groups in Theorem 8.2 and 8.3 below, and describe the meaning of the respective quotients. The space d, p has in some sense more structure than d has, because we have in addition to the usual action of $\mathrm{SL}_d(\mathbb{R})$ an action of $\mathrm{SL}_d(\mathbb{Q}_p)$ on d, p.

To motivate this a bit further, we briefly discuss the p-adic and adelic abelian (and additive) quotients

$$\mathbb{R} \times \mathbb{Q}_p / \mathbb{Z}[\tfrac{1}{p}]$$

and

$$\mathbb{A}_{\mathbb{Q}} / \mathbb{Q}$$

as p-adic and adelic extensions of $\mathbb{R}/\mathbb{Z}$.

For the former abelian quotient, notice that $\mathbb{Z}[\frac{1}{p}]$ is not a discrete subset of $\mathbb{R}$ (nor of $\mathbb{Q}_p$), but is a discrete subset of $\mathbb{R} \times \mathbb{Q}_p$ if we use the diagonal embedding

$$\mathbb{Z}[\tfrac{1}{p}] \ni a \overset{\imath}{\longmapsto} (a, a) \in \mathbb{R} \times \mathbb{Q}_p.$$

This is because

$$\imath\left(\mathbb{Z}[\tfrac{1}{p}]\right) \cap ((-1,1) \times \mathbb{Z}_p) = \imath(\mathbb{Z}) \cap ((-1,1) \times \mathbb{Z}_p) = \imath(0,0).$$

Also notice that

$$\mathbb{R} \times \mathbb{Q}_p / \imath(\mathbb{Z}[\tfrac{1}{p}]) \cong \mathbb{R} \times \mathbb{Z}_p / \imath(\mathbb{Z})$$

since every element $(a_\infty, a_p) \in \mathbb{R} \times \mathbb{Q}_p$ can be modified by some $\imath(a)$ with a in $\mathbb{Z}[\frac{1}{p}]$ to cancel the denominator of the element in $\mathbb{Q}_p$, so that

$$(a_\infty, a_p) + \imath(a) \in \mathbb{R} \times \mathbb{Z}_p.$$

This follows at once from the density of $\mathbb{Z}[\frac{1}{p}]$ in $\mathbb{Q}_p$. Moreover, this shows that $F = [0,1) \times \mathbb{Z}_p$ is a fundamental domain for $\imath(\mathbb{Z}[\frac{1}{p}])$. Hence $\imath(\mathbb{Z}[\frac{1}{p}])$ is a co-compact lattice in $\mathbb{R} \times \mathbb{Q}_p$.

The discussion above (specifically, discreteness together with Lemma 1.1) also shows that

$$\mathsf{Y}_p = \mathbb{R} \times \mathbb{Q}_p / \imath(\mathbb{Z}[\tfrac{1}{p}])$$

is locally isomorphic to $\mathbb{R} \times \mathbb{Z}_p$. Now $\mathbb{Z}_p$ is homeomorphic to the standard middle-third Cantor set, so it is surprising that Y_p is nonetheless connected (though it is neither locally connected nor path connected). To see that it is connected, it is enough to notice that the $\mathbb{R}$-orbits (which are connected) are dense in Y_p. We will not need this property, and so leave it as an exercise.

More globally, one can understand the space Y_p as a projective limit,

$$\mathsf{Y}_p = \varprojlim_{n\to\infty} \mathbb{R}/p^n\mathbb{Z}$$

where each quotient in the limit construction is isomorphic to $\mathbb{T}$ and the projective limit uses the canonical projection maps

$$\begin{aligned} \mathbb{R}/p^{n+1}\mathbb{Z} &\longrightarrow \mathbb{R}/p^n\mathbb{Z} \\ x \pmod{p})^{n+1}\mathbb{Z} &\longmapsto x \pmod{p})^n\mathbb{Z} \end{aligned} \tag{8.6}$$

for all $x \in \mathbb{R}$. In less sophisticated language, we may describe Y_p as resembling a path that looks different at different resolutions. On the largest scale, it resembles a circle, but at a finer resolution labeled n it resembles a tightly wrapped circular path of length p^n, which on closer inspection splits into a wrapped circle of length p^{n+1}, and so on. For this reason, spaces like Y_p are referred to as *solenoids*(27).

The construction above generalizes easily to any finite subset S of the set of all primes $\mathbb{P}$, again giving rise to a co-compact lattice

$$\imath\left(\mathbb{Z}[\tfrac{1}{p} \mid p \in S]\right) \subseteq \mathbb{R} \times \prod_{p\in S} \mathbb{Q}_p.$$

However, it also generalizes to the rationals embedded diagonally in the rational adeles as follows. The image

$$\imath(\mathbb{Q}) \subseteq \mathbb{A}_{\mathbb{Q}}$$

is a discrete subgroup since

$$\imath(\mathbb{Q}) \cap \left((-1,1) \cap \prod_{p\in\mathbb{P}} \mathbb{Z}_p\right) = \imath(0),$$

and $\mathsf{Y}_{\mathbb{A}_{\mathbb{Q}}} = \mathbb{A}_{\mathbb{Q}}/\imath(\mathbb{Q})$ is compact, since

$$[0,1) \times \prod_{p\in\mathbb{P}} \mathbb{Z}_p,$$

which is pre-compact, is a fundamental domain for $\imath(\mathbb{Q})$ in $\mathbb{A}_{\mathbb{Q}}$.

We have deliberately avoided discussing the group structure of Y_p and $\mathsf{Y}_{\mathbb{A}_{\mathbb{Q}}}$ (both of which are interesting in their own right; see Exercise 8.1.2) since in the context of locally homogeneous spaces we generally do not have a group structure on the quotient space. Thus far we have used $\imath$ to denote the diagonal embedding of rational numbers (or subrings of the rational numbers). This quickly becomes wearisome, and instead one often drops this letter, on the understanding that it is clear (sometimes after a little thought) when the diagonal embedding is used. We will follow this practice below.

Let us now fix a prime p and discuss the p-adic extension of

$$d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R}).$$

Theorem 8.2 (p-adic cover of d). *Let d, p be defined as in* (8.4)*, with the subgroup $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])$ embedded diagonally. Then d, p is a single orbit of the subgroup $\mathrm{SL}_d(\mathbb{R} \times \mathbb{Z}_p) \subseteq \mathrm{SL}_d(\mathbb{R} \times \mathbb{Q}_p)$. In fact if we use (for example) the point $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}]) \in d, p$ corresponding to the identity element, then its orbit is given by*

$$\mathrm{SL}_d(\mathbb{Z}[\tfrac{1}{p}])\,\mathrm{SL}_d(\mathbb{R} \times \mathbb{Z}_p) = d, p,$$

and is isomorphic to

$$\mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R} \times \mathbb{Z}_p). \tag{8.7}$$

Moreover, there is a canonical projection map

$$\pi : d, p \longrightarrow d$$

defined on (8.7) *by*

$$\pi\left(\mathrm{SL}_d(\mathbb{Z})(g_\infty, g_p) = \mathrm{SL}_d(\mathbb{Z})g_\infty\right)$$

for $g_\infty \in \mathrm{SL}_d(\mathbb{R})$ and $g_p \in \mathrm{SL}_d(\mathbb{Z}_p)$, with the property that each fiber $\pi^{-1}(x)$ for $x \in d$ is isomorphic to the compact group $\mathrm{SL}_d(\mathbb{Z}_p)$. Moreover, $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])$ is a (non-uniform) lattice in $\mathrm{SL}_d(\mathbb{R} \times \mathbb{Q}_p)$.

We note that this result is not an automatic property for p-adic extensions. It is called *strong approximation* and holds more generally for *(Zariski) simply connected algebraic groups.* We will introduce some of these terms in Chapter 3.

Theorem 8.2 shows that d, p is a compact extension[†] of d, by which we mean that the pre-images of compact sets in d are compact subsets of d, p. We can give a global interpretation of d, p in three equivalent ways:

- $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])\backslash \mathrm{SL}_d(\mathbb{R} \times \mathbb{Q}_p)$,
- $\left\{\mathbb{Z}[\frac{1}{p}]^d(g_\infty, g_p) \mid g_\infty \in \mathrm{SL}_d(\mathbb{R}), g_p \in \mathrm{SL}_d(\mathbb{Q}_p)\right\}$, or

[†] Clearly d, p is not compact, since d is non-compact. To see the claim, let $\mathrm{SL}_d(\mathbb{Z})K$ be compact with $K \leqslant \mathrm{SL}_d(\mathbb{R})$ compact, then

$$\pi^{-1}(K) = \mathrm{SL}_d(\mathbb{Z})K \times \mathrm{SL}_d(\mathbb{Z}_p)$$

if d, p is again described as in (8.7). We refer to Corollary 8.4 for the details.

- the set of $\imath(\mathbb{Z}[\frac{1}{p}])$-submodules of $\mathbb{R}^d \times \mathbb{Q}_p^d$ generated by d vectors $(g_{\infty,j}, g_{p,j})$ with $g_{\infty,j} \in \mathbb{R}^d$, $g_{p,j} \in \mathbb{Q}_p^d$ and with

$$\det \begin{pmatrix} g_{\infty,1} \\ \vdots \\ g_{\infty,d} \end{pmatrix} = \det \begin{pmatrix} g_{p,1} \\ \vdots \\ g_{p,d} \end{pmatrix} = 1.$$

This description follows from the fact that an orbit is always isomorphic to the acting group modulo the stabilizer, and the stabilizer of the diagonally embedded copy of $\mathbb{Z}[\frac{1}{p}]^d$ inside $\mathrm{SL}_d(\mathbb{R}\times\mathbb{Q}_p)$ is precisely the diagonally embedded copy of $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])$. Notice that $\mathbb{Z}[\frac{1}{p}]^d(g_\infty, g_p) \subseteq \mathbb{R}^d \times \mathbb{Q}_p^d$ is a co-volume one co-compact lattice. What is less clear, but is contained in Theorem 8.2, is that we can find a different set of generators for any of these lattices so that $g_{p,j} \in \mathbb{Z}_p^d$ while retaining the property that

$$\det \begin{pmatrix} g_{p,1} \\ \vdots \\ g_{p,d} \end{pmatrix} = 1.$$

The projection map in the theorem is then the map that sends the $\mathbb{Z}[\frac{1}{p}]$-module $\mathbb{Z}[\frac{1}{p}]^d(g_\infty, g_p)$ to the $\mathbb{Z}$-module $\mathbb{Z}^d g_\infty \in d$, which is also determined by the formula

$$\pi_\infty\left(\varLambda \cap \left(\mathbb{R}^d \times \mathbb{Z}_p^d\right)\right)$$

where $\varLambda$ is the $\mathbb{Z}[\frac{1}{p}]$-module corresponding to the point in d, p and

$$\pi_\infty : \mathbb{R}^d \times \mathbb{Q}_p^d \to \mathbb{R}^d$$

is the canonical projection onto the first coordinate.

Another global meaning of d, p may be found by mimicking the discussion of Y_p as a projective limit of circles. Write

$$\varGamma_{p^n} = \{\gamma \in \mathrm{SL}_d(\mathbb{Z}) \mid \gamma \equiv I \pmod{p}^n\},$$

so that

$$\pi_n : \varGamma_{p^{n+1}} \backslash \mathrm{SL}_d(\mathbb{R}) \longrightarrow \varGamma_{p^n} \backslash \mathrm{SL}_d(\mathbb{R})$$

is a generalization of (8.6) to the context of SL_d, and one can check that d, p is the projective limit of the congruence quotients $\varGamma_{p^n} \backslash \mathrm{SL}_d(\mathbb{R})$ of 'full level p^n'. In other words, d, p is the appropriate space to use if one is interested in all congruence quotients of $\mathrm{SL}_d(\mathbb{R})$ defined by powers of p. To discuss other congruence quotients one needs to consider all the primes, and as a result the greater flexibility of the 'adelic cover'.

Theorem 8.3 (Adelic cover of d). *Let $d, \mathbb{A}_{\mathbb{Q}}$ be defined by* (8.5). *Then $d, \mathbb{A}_{\mathbb{Q}}$ is a single orbit of the subgroup*

$$\mathrm{SL}_d(\mathbb{R}) \times \prod_{p\in\mathbb{P}} \mathrm{SL}_d(\mathbb{Z}_p),$$

and there is a canonical projection map $\pi : d, \mathbb{A}_{\mathbb{Q}} \to d$ *whose fibres are each isomorphic to*

$$\prod_{p\in\mathbb{P}} \mathrm{SL}_d(\mathbb{Z}_p).$$

Finally, $\mathrm{SL}_d(\mathbb{Q})$ *is a (non-uniform) lattice in* $\mathrm{SL}_d(\mathbb{A}_{\mathbb{Q}})$.

We can again describe the points in $d, \mathbb{A}_{\mathbb{Q}}$ as $\mathbb{Q}$-modules in $\mathbb{A}_{\mathbb{Q}}^d$ that are generated by special basis elements of determinant one at all places. Alternatively, one can describe $d, \mathbb{A}_{\mathbb{Q}}$ as the projective limit over all congruence quotients $\Gamma_n \backslash \mathrm{SL}_d(\mathbb{R})$ with

$$\Gamma_n = \{\gamma \in \mathrm{SL}_d(\mathbb{Z}) \mid \gamma \equiv I \pmod{n}\}.$$

Before we give the proof of Theorems 8.2 and 8.3, let us state and prove a corollary to the above theorems and discussion.

Corollary 8.4 (Mahler's compactness criteria for d, p **and** $d, \mathbb{A}_{\mathbb{Q}}$**).** *A subset* $K \subseteq d, \mathbb{A}_{\mathbb{Q}}$ *(resp.* d, p*) is compact if and only if it is closed and its image in* d *is compact. Equivalently,* K *is compact if* K *is closed and uniformly discrete. That is, if there exists some* $\delta > 0$ *such that the* $\mathbb{Q}$*-modules (resp.* $\mathbb{Z}[\frac{1}{p}]$*-modules)* Λ *corresponding to the elements of* K *satisfy*

$$\Lambda \cap B_\delta(0) = \{0\}.$$

Proof. By Theorem 8.3 (resp. Theorem 8.2, but for brevity we will restrict attention to the adelic case in the rest of this proof) $d, \mathbb{A}_{\mathbb{Q}}$ is the quotient

$$\mathrm{SL}_d(\mathbb{Q})\backslash\mathrm{SL}_d(\mathbb{A}_{\mathbb{Q}}) \cong \mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R} \times \textstyle\prod_{p\in\mathbb{P}} \mathbb{Z}_p)$$

and the projection map mentioned above is the map that forgets the p-adic matrices

$$\pi\left(\mathrm{SL}_d(\mathbb{Z})(g_\infty, g_2, g_3, \dots)\right) = \mathrm{SL}_d(\mathbb{Z})g_\infty$$

(assuming that $g_p \in \mathrm{SL}_d(\mathbb{Z}_p)$ for all primes p). Therefore, if $\mathrm{SL}_d(\mathbb{Z})K$ is compact (see the paragraph before Definition 1.10 on p. 15) for some compact $K \subseteq \mathrm{SL}_d(\mathbb{R})$, then

$$\pi^{-1}\left(\mathrm{SL}_d(\mathbb{Z})K\right) = \mathrm{SL}_d(\mathbb{Z})K \times \prod_{p\in\mathbb{P}} \mathrm{SL}_d(\mathbb{Z}_p)$$

is again compact. This shows one direction of the first equivalence. Conversely, if K is compact, then $\pi(K)$ is compact. Finally, notice that, for d, Mahler's compactness criterion (Theorem 1.17) is precisely of the same form as in the corollary: $\pi(K)$ is compact if and only if it is closed and uniformly discrete. Further, notice that for small enough values of δ we have

$$B_\delta^{\mathbb{A}_\mathbb{Q}^d} \subseteq \mathbb{R}^d \times \prod_{p\in\mathbb{P}} \mathbb{Z}_p^d$$

by the topology of $\mathbb{A}_\mathbb{Q}^d$. We claim that this implies that the condition of uniform discreteness of the $\mathbb{Q}$-modules $\Lambda_\mathbb{Q}$ associated to the elements of K is actually equivalent to the uniform discreteness of the $\mathbb{Z}$-modules

$$\Lambda_\mathbb{Z} = \pi_\infty \left(\Lambda_\mathbb{Q} \cap \left(\mathbb{R}^d \times \prod_{p\in\mathbb{P}} \mathbb{Z}_p^d \right) \right).$$

The claim and the first equivalence in the corollary then imply the second compactness criterion in the corollary.

The above two notions of uniform discreteness only differ in the following sense. In the first we take the intersection with $B_\delta^{\mathbb{A}_\mathbb{Q}^d}(0)$, and in the second with

$$B_\delta^{\mathbb{R}^d}(0) \times \prod_{p\in\mathbb{P}} \mathbb{Z}_p^d.$$

The former is potentially smaller (making the corresponding uniform discreteness formally easier), but there must exist some $\varepsilon > 0$ and $N \geqslant 1$ such that

$$B_\delta^{\mathbb{A}_\mathbb{Q}^d}(0) \supseteq B_\varepsilon^{\mathbb{R}^d}(0) \times \prod_{p\in\mathbb{P}} (N\mathbb{Z}_p)^d$$

where $N\mathbb{Z}_p = \mathbb{Z}_p$ if $p \nmid N$ so that the set on the right-hand side is still open in $\mathbb{A}_\mathbb{Q}$. Now suppose that

$$\Lambda \cap \left(B_{\varepsilon/N}^{\mathbb{R}^d}(0) \times \prod_{p\in\mathbb{P}} \mathbb{Z}_p^d \right) \neq \{0\},$$

and then we may multiply the non-zero element of Λ by N (also embedded diagonally) and see that the intersection

$$\Lambda \cap B_\delta^{\mathbb{A}_\mathbb{Q}^d}(0) \supseteq \Lambda \cap \left(B_\varepsilon^{\mathbb{R}^d}(0) \times \prod_{p\in\mathbb{P}} (N\mathbb{Z}_p)^d \right) \neq \{0\}$$

is also non-trivial. Therefore, the notions of uniform discreteness agree, and the corollary follows. □

Proof of Theorem 8.2. We need to show that d, p is actually only one orbit of the open subgroup

$$\mathrm{SL}_d(\mathbb{R}\times\mathbb{Z}_p)\subseteq \mathrm{SL}_d(\mathbb{R}\times\mathbb{Q}_p).$$

Alternatively, we need to show that the whole group

$$\mathrm{SL}_d(\mathbb{R}\times\mathbb{Q}_p)=\mathrm{SL}_d(\mathbb{Z}[\tfrac{1}{p}])\,\mathrm{SL}_d(\mathbb{R}\times\mathbb{Z}_p) \tag{8.8}$$

is the product of the discrete subgroup $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])$ and the open subgroup

$$\mathrm{SL}_d(\mathbb{R}\times\mathbb{Z}_p).$$

Since the subset on the right-hand side of (8.8) is right-invariant under $\mathrm{SL}_d(\mathbb{R})$, this all boils down to the projection of the set on the right-hand side of (8.8) to $\mathrm{SL}_d(\mathbb{Q}_p)$. In fact (8.8) is equivalent to

$$\mathrm{SL}_d(\mathbb{Q}_p)=\mathrm{SL}(\mathbb{Z}[\tfrac{1}{p}])\,\mathrm{SL}_d(\mathbb{Z}_p), \tag{8.9}$$

where $\mathrm{SL}_d(\mathbb{Z}_p)$ is still open. We claim that $\mathrm{SL}_d(\mathbb{Z}[\frac{1}{p}])$ is dense, which then proves (8.9), and hence (8.8), and so the theorem. To prove the claim notice that by Lemma 1.24, every $g\in\mathrm{SL}_d(\mathbb{Q}_p)$ is a finite product of matrices

$$u_{ij}(t)=I+tE_{ij}$$

for $t\in\mathbb{Q}_p$ and $i\neq j$. Hence it is sufficient to approximate $u_{ij}(t)$ with $t\in\mathbb{Q}_p$. However, this is easy to show since $\mathbb{Z}[\frac{1}{p}]$ is dense in $\mathbb{Q}_p$. □

8.1.3 Positive Characteristic S-algebraic Groups

The construction from Section 8.1.2 can also be carried out without much change to a set S of places of a global field $\mathbb{K}|\mathbb{F}_p(t)$ with positive characteristic. In this case the local fields $\mathbb{K}_\sigma$ are isomorphic to $\mathbb{F}_q((s))$ for some $q=p^f$, $f\geqslant 1$ and some $s\in\mathbb{K}$ (the element s is a *uniformizer* corresponding to the place σ).

Once more there exists a compact open subring $\mathbb{F}_q[[s]]\subseteq\mathbb{F}_q((s))$ so that $\mathrm{SL}_d\left(\mathbb{F}_q[[s]]\right)\subseteq\mathrm{SL}_d\left(\mathbb{F}_q((s))\right)$ is a compact open subgroup and plays the same role as $\mathrm{SL}_d(\mathbb{Z}_p)$ in the definition of a metric and in the construction of restricted products. One difference to the characteristic zero setting (that is, to $\mathbb{R}$ or $\mathbb{Q}_p$) is the complete absence of locally defined exponential and logarithm maps between $\mathrm{SL}_d\left(\mathbb{F}_q((s))\right)$ and what should be its 'Lie algebra' $\mathfrak{sl}_d\left(\mathbb{F}_q((s))\right)$ (or between closed subgroups and Lie sub-algebras). We refer to Section **??** for a partial correspondence between (and the definitions of) algebraic subgroups and their Lie algebras.

Exercises for Section 8.1

Exercise 8.1.1. Let $G < \mathrm{SL}_d(\mathbb{Q}_p)$ be a closed linear group over the field $\mathbb{Q}_p$ of p-adic rational numbers. Show that a lattice $\Gamma < G$ cannot contain a unipotent element.

Exercise 8.1.2. Show that the character group of Y_p (defined on p. 255) is $\mathbb{Z}[\frac{1}{p}]$, and that the character group of $\mathsf{Y}_{\mathbb{A}_\mathbb{Q}}$ is $\mathbb{Q}$.

8.1.4 S-Arithmetic Quotients of Forms of SL_d

We briefly outline in this section how the notion of irreducible lattices and Corollary 3.33 extends to the setting of S-arithmetic quotients. We do this for (forms of) SL_d as in the next result because this is a familiar group, and because it is an example of a *simply connected* linear algebraic group. We will not discuss the definition of this notion in general†, but we will discuss (indeed, have already discussed in part) the main consequence of this property for SL_d.

Corollary 8.5 (A special case of strong approximation). *Let $\mathbb{G}$ be a linear algebraic group defined over $\mathbb{Q}$ such that*

$$\mathbb{G}(\mathbb{R}) \cong \mathrm{SL}_d(\mathbb{R})$$

and

$$\mathbb{G}(\mathbb{Q}_p) \cong \mathrm{SL}_d(\mathbb{Q}_p)$$

for some $d \geqslant 1$ and some prime $p < \infty$. Then $\mathbb{G}(\mathbb{Z}[\frac{1}{p}])$ is dense both in $\mathbb{G}(\mathbb{R})$ and in $\mathbb{G}(\mathbb{Q}_p)$.

As discussed in Section 8.1, the group $\mathbb{G}(\mathbb{Z}[\frac{1}{p}])$, when diagonally embedded via the map $x \mapsto (x, x)$ into $\mathbb{G}(\mathbb{R}) \times \mathbb{G}(\mathbb{Q}_p)$, is a discrete subgroup. Just as in the purely real case $\mathbb{G}(\mathbb{Z}) < \mathbb{G}(\mathbb{R})$ (see Section 7.4), this diagonally embedded subgroup is actually a lattice in $\mathbb{G}(\mathbb{R}) \times \mathbb{G}(\mathbb{Q}_p)$. Corollaries 8.5 and 3.33 are similar results when viewed in this context. In fact for the projection to $\mathbb{G}(\mathbb{R})$ (that is, for $\mathbb{G}(\mathbb{Z}[\frac{1}{p}])$ viewed as a subgroup of $\mathbb{G}(\mathbb{R})$), the arguments of Corollary 3.33 give the density (since $\mathbb{G}(\mathbb{R})^o \cong \mathrm{SL}_d(\mathbb{R})$ is connected as a Lie group). For $\mathbb{G}(\mathbb{Q}_p) \cong \mathrm{SL}_d(\mathbb{Q}_p)$ this is clearly wrong in several different senses and so we need another property of $\mathrm{SL}_d(\mathbb{Q}_p)$ to help us.

Proposition 8.6 (No unbounded open subgroups). *Let $L \subseteq \mathrm{SL}_d(\mathbb{Q}_p)$ be an unbounded open subgroup. Then $L = \mathrm{SL}_d(\mathbb{Q}_p)$.*

† It is expressed in terms of the weight lattice and the root lattice, and is not related to the more familiar topological notion.

Proof of Corollary 8.5. Let us assume Proposition 8.6 for the moment, and see how it helps us to prove the strong approximation property in this case. We only need to consider $\mathbb{G}(\mathbb{Z}[\frac{1}{p}]) < \mathbb{G}(\mathbb{Q}_p)$, since the real case $\mathbb{G}(\mathbb{Z}[\frac{1}{p}]) < \mathbb{G}(\mathbb{R})$ is similar to the proof of Corollary 3.33. Let $L = \overline{\mathbb{G}(\mathbb{Z}[\frac{1}{p}])} \leqslant \mathbb{G}(\mathbb{Q}_p)$ be the closure of the lattice $\mathbb{G}(\mathbb{Z}[\frac{1}{p}]) \leqslant \mathbb{G}(\mathbb{R})\times\mathbb{G}(\mathbb{Q}_p)$ projected to $\mathbb{G}(\mathbb{Q}_p)$. Clearly its Lie algebra $\mathfrak{l}$ is normalized by $\mathbb{G}(\mathbb{Z}[\frac{1}{p}])$. By the Borel density theorem (Theorem 3.30 extended to include S-arithmetic cases) it follows that $\mathfrak{l}$ is normalized by $\mathbb{G}(\mathbb{Q}_p) \cong \mathrm{SL}_d(\mathbb{Q}_p)$, since the latter is generated by one-parameter unipotent subgroups for which the extended Borel density theorem holds. Thus $\mathfrak{l} \lhd \mathfrak{sl}_d(\mathbb{Q}_p)$ is a Lie ideal in the simple Lie algebra $\mathfrak{sl}_d(\mathbb{Q}_p)$ over $\mathbb{Q}_p$, so $\mathfrak{l} = \{0\}$ or $\mathfrak{l} = \mathfrak{sl}_d(\mathbb{Q}_p)$. The former only happens when $\mathbb{G}(\mathbb{Z}[\frac{1}{p}]) \leqslant \mathbb{G}(\mathbb{Q}_p)$ is discrete, which can be ruled out just as in the proof of Corollary 3.33, since the diagonally embedded lattice $\Gamma = \mathbb{G}(\mathbb{Z}[\frac{1}{p}]) < \mathbb{G}(\mathbb{R}) \times \mathbb{G}(\mathbb{Q}_p)$ intersects the non-compact factor $\mathbb{G}(\mathbb{R}) \times \{I\}$ trivially.

Therefore, $\mathfrak{l} = \mathfrak{sl}_d(\mathbb{Q}_p)$ and $L \leqslant \mathbb{G}(\mathbb{Q}_p)$ is an open subgroup. By Poincaré recurrence L is also unbounded. Indeed, if $a \in \mathbb{G}(\mathbb{Q}_p)$ has $a^n \to \infty$ as $n \to \infty$, then for almost every $(g_\infty, g_p) \in \mathbb{G}(\mathbb{R}) \times \mathbb{G}(\mathbb{Q}_p)$ there exists a sequence (n_k) with $n_k \to \infty$ as $k \to \infty$, and for each k a $\gamma_k \in \mathbb{G}(\mathbb{Z}[\frac{1}{p}])$ with

$$(\gamma_k, \gamma_k)(g_\infty, g_p)(I, a^{n_k}) \longrightarrow (g_\infty, g_p)$$

as $k \to \infty$. Any such sequence (γ_k) in L then goes to infinity in $\mathbb{G}(\mathbb{Q}_p)$. Hence L is unbounded, and so $L = \mathbb{G}(\mathbb{Q}_p)$ by Proposition 8.6. □

Proof of Proposition 8.6 in the case needed for Corollary 8.5. We assume that $L \leqslant \mathrm{SL}_d(\mathbb{Q}_p)$ is an open subgroup, and for every $a \in \mathrm{SL}_d(\mathbb{Q}_p)$ there exist sequences (n_k) in $\mathbb{N}$ and (γ_k) in L with $n_k \to \infty$ as $k \to \infty$ and

$$\gamma_k g a^{n_k} = g\varepsilon_k \longrightarrow g$$

as $k \to \infty$. This may be written as

$$(gag^{-1})^{n_k} = ga^{n_k}g^{-1} = \gamma_k^{-1} g\varepsilon_k g^{-1}$$

for all $k \geqslant 1$. For large enough k we have $g\varepsilon_k g^{-1} \in L$, and so also

$$(gag^{-1})^k \in L.$$

For convenience we flip the conjugation over to L and get the equivalent statement

$$a^k \in g^{-1}Lg.$$

Let us choose, for example

$$a = \begin{pmatrix} p^{d-2} & & & & \\ & p^{d-3} & & & \\ & & \ddots & & \\ & & & p^0 & \\ & & & & p^{\ell} \end{pmatrix} \in \mathrm{SL}_d(\mathbb{Q}_p)$$

with ℓ chosen to ensure that $\det(a) = 1$. Since every sufficiently small element of $\mathrm{SL}_d(\mathbb{Q}_p)$ belongs to $g^{-1}Lg$, and since a (respectively a^{-1}) normalizes and expands each of the standard one-parameter unipotent subgroups, it follows that $g^{-1}Lg$ contains these one-parameter unipotent subgroups completely. By Lemma 1.24, we deduce that

$$g^{-1}Lg = \mathrm{SL}_d(\mathbb{Q}_p) = L.$$

□

Although it is not necessary for our discussion, we will now outline a more general proof (see also Exercise **??**).

Sketch proof of Proposition 8.6 for $d = 2$. Suppose that $L \leqslant \mathrm{SL}_2(\mathbb{Q}_p)$ is an open unbounded subgroup. As in the argument above it suffices to find an element in L whose eigenvalues are of norm not equal to 1. For $\mathrm{SL}_2(\mathbb{Q}_p)$ this is a property of the trace: $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has eigenvalues of norm not equal to 1 if and only if $\mathrm{tr}(g) = a + d \notin \mathbb{Z}_p$.

Now let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in L$ be sufficiently large. If $\mathrm{tr}(g) \notin \mathbb{Z}_p$ then we are done. Since $\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \in L$ for sufficiently small s (say for $|s|_p \leqslant \varepsilon$), we also have

$$g' = g \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} = \begin{pmatrix} a & b + as \\ c & d + cs \end{pmatrix} \in L.$$

If $|c|_p > \varepsilon^{-1}p$, then $g' \in L$ has the desired property for some s with $|s|_p < \varepsilon$. If $|a|_p > \varepsilon^{-2}p$ but $|c|_p \leqslant \varepsilon^{-1}p$, then we can use the argument above for

$$\begin{pmatrix} 1 & \\ s & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c + as & d + bs \end{pmatrix} \in L$$

for some s with $|s| < \varepsilon$. The remaining cases are similar. □

Simply connected groups have properties analogous to those of Lemma 1.24 and Proposition 8.6 (see Margulis [**?**] for the details). For such groups strong approximation also holds, and can be shown in the same way once Proposition 8.6 is known.

Exercises for Section 8.1.4

Exercise 8.1.3. Prove Proposition 8.6 for $d \geqslant 2$.

density of G(Z[1/p]) in the connected component of G(R) follows quite generally

Notes to Chapter 8

(26)(Page 251) More advanced arithmetic aspects of these groups may be found in the work of Weil [**?**].
(27)(Page 256) The terminology comes from electro-magnetic helical coils in physics, and seems to have been coined in this mathematical context by van Danzig [**?**].

Part II
Entropy theory on homogeneous spaces

Chapter 9
Leafwise Measures

In this chapter we will introduce the notion of leafwise measures in the setting of $\Gamma\backslash G$ on orbits of a subgroup $H \leqslant G$, and will discuss the relationship between leafwise measures and dynamical properties of the original measure. In Chapter 10 we will relate the leafwise measures on stable leaves to entropy.

9.1 Fiber Measures for Locally Finite Measures

Before discussing the locally finite case (a measure defined on the Borel σ-algebra of a metric space is called locally finite if every point has an open neighborhood of finite measure), we describe a structural property of the conditional measures with respect to restrictions.

Recall from [**?**, Sec. 5.3] that for any finite[†] measure μ on a σ-compact[(28)] metric space (Y, d), and any sub-σ-algebra $\mathscr{A} \subseteq \mathscr{B}_Y$ of the corresponding Borel σ-algebra $\mathscr{B}_Y$ there exists a system of conditional measures $\mu_y^{\mathscr{A}}$, which are defined almost everywhere, and which have (and are characterized by) the following properties:

(1) $y \mapsto \mu_y^{\mathscr{A}}$ is $\mathscr{A}$-measurable, meaning that for any $f \in \mathscr{L}^1(Y, \mathscr{B}_Y, \mu)$ the integral $\int f \,\mathrm{d}\mu_y^{\mathscr{A}}$, defined almost everywhere, is $\mathscr{A}$-measurable as a function of $y \in Y$, and

(2) $\displaystyle\int_A f \,\mathrm{d}\mu = \int_A \int f \,\mathrm{d}\mu_y^{\mathscr{A}} \,\mathrm{d}\mu(y)$ for all $A \in \mathscr{A}$.

Notice that an equivalent formulation of these properties is to require that

[†] This easily extends from probability measures to finite measures as follows. If $\mu(Y) \in (0, \infty)$ then define a new measure by $\mu' = \frac{1}{\mu(Y)}\mu$ and consider the conditional measures for μ' — these are also conditional measures for the original measure μ in the sense described. Notice that the conditional measures $\mu_y^{\mathscr{A}}$ are always probability measures on Y for any finite measure μ.

$$E_\mu\left(f\middle|\mathscr{A}\right)(y)=\int f\,\mathrm{d}\mu_y^{\mathscr{A}}$$

for any $f\in\mathscr{L}^1(Y,\mathscr{B}_Y,\mu)$.

Lemma 9.1. *Let Y' be a Borel subset of a σ-compact metric space Y, and let $\mathscr{A}\subseteq\mathscr{B}_Y$ be a countably-generated sub-σ-algebra. Then the conditional measures for the restriction*[†] *$\mu|_{Y'}$ of a finite measure μ on Y with $\mu(Y')>0$ satisfy*

$$(\mu|_{Y'})_y^{\mathscr{A}}=\frac{1}{\mu_y^{\mathscr{A}}(Y')}\mu_y^{\mathscr{A}}|_{Y'} \tag{9.1}$$

for almost every $y\in Y'$.

Notice that if we set

$$A_0=\{y\in Y\mid\mu_y^{\mathscr{A}}(Y')=0\}\in\mathscr{A}, \tag{9.2}$$

then

$$\mu(A_0\cap Y')=\int_{A_0}\mathbb{1}_{Y'}\,\mathrm{d}\mu=\int_{A_0}\mu_y^{\mathscr{A}}(Y')\,\mathrm{d}\mu=0,$$

by the properties of the conditional measures $\mu_y^{\mathscr{A}}$. This shows that the right-hand side of (9.1) is well-defined on the complement of a null set.

Proof of Lemma 9.1. It is clear that

$$\nu_y=\frac{1}{\mu_y^{\mathscr{A}}(Y')}\mu_y^{\mathscr{A}}|_{Y'}$$

is $\mathscr{A}$-measurable wherever it is defined. For any $f'\in\mathscr{L}^1(Y,\mathscr{B}_Y,\mu|_{Y'})$ define $f\in\mathscr{L}^1(Y,\mathscr{B}_Y,\mu)$ by

$$f(y)=\begin{cases}f'(y) & \text{if } y\in Y',\\ 0 & \text{if } y\in Y\smallsetminus Y'.\end{cases}$$

Notice that f and f' coincide as elements of $L^1(Y,\mathscr{B}_Y,\mu|_{Y'})$ since $Y\smallsetminus Y'$ is a null set with respect to $\mu|_{Y'}$. We have to show that

$$\int f'\,\mathrm{d}\nu_y=E_{\mu|_{Y'}}\left(f'\middle|\mathscr{A}\right)(y)$$

for $y\in Y$. To that end, fix a set $A\in\mathscr{A}$. Since A_0 defined in (9.2) is a null set with respect to $\mu|_{Y'}$, we may assume that $A\subseteq Y\smallsetminus A_0$, which makes the measure ν_y well-defined in all of the calculations below. We have

[†] The restriction is defined by $\mu|_{Y'}(B)=\mu(B\cap Y')$ for all $B\in\mathscr{B}_Y$.

$$
\begin{aligned}
\int_A \int f' \mathrm{d}\nu_y \mathrm{d}\mu|_{Y'}(y) &= \int_A \frac{1}{\mu_y^{\mathscr{A}}(Y')} \int_{Y'} f'(z) \mathrm{d}\mu_y^{\mathscr{A}}(z) \mathrm{d}\mu|_{Y'}(y) \quad \text{(by definition of } \nu_y\text{)} \\
&= \int_A \mathbb{1}_{Y'}(y) \frac{1}{\mu_y^{\mathscr{A}}(Y')} \underbrace{\int f(z)\, \mathrm{d}\mu_y^{\mathscr{A}}(z)}_{=E_\mu(f|\mathscr{A})(y)} \mathrm{d}\mu(y) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(by definition of } f\text{)} \\
&= \int_A \int \mathbb{1}_{Y'}(z) \underbrace{\frac{1}{\mu_z^{\mathscr{A}}(Y')} E_\mu\left(f\middle|\mathscr{A}\right)(z)}_{\mathscr{A}\text{-measurable } \&\ \mu_y^{\mathscr{A}}-\text{a.e. constant}} \mathrm{d}\mu_y^{\mathscr{A}}(z)\, \mathrm{d}\mu(y)
\end{aligned}
$$

where we have used property (2) of the conditional measures from p. 269. As the second two terms (as indicated) are now constant for the integration with respect to $\mu_y^{\mathscr{A}}$, we can now take this constant out of the integral to deduce that

$$
\begin{aligned}
\int_A \int f' \mathrm{d}\nu_y \mathrm{d}\mu|_{Y'}(y) &= \int_A \frac{1}{\mu_y^{\mathscr{A}}(Y')} E_\mu\left(f\middle|\mathscr{A}\right)(y) \int \mathbb{1}_{Y'}(z)\, \mathrm{d}\mu_y^{\mathscr{A}}(z)\, \mathrm{d}\mu(y) \\
&= \int_A \frac{1}{\mu_y^{\mathscr{A}}(Y')} E_\mu\left(f\middle|\mathscr{A}\right)(y) \mu_y^{\mathscr{A}}(Y')\, \mathrm{d}\mu \\
&= \int_A f\, \mathrm{d}\mu = \int_A f'\, \mathrm{d}\mu|_{Y'}, \qquad\qquad \text{(by definition of } f\text{)}
\end{aligned}
$$

which proves the lemma. □

Lemma 9.1 suggests how we may define conditional measures for a locally finite measure.

Proposition 9.2. *Let X be a σ-compact metric space, let μ be a locally finite measure, and let $\mathscr{A} \subseteq \mathscr{B}_X$ be a countably-generated sub-σ-algebra of the Borel σ-algebra $\mathscr{B}_X$ of X. Then there exists a family of locally finite measures $\mu_x^{\mathscr{A}}$, defined for almost every $x \in X$, with the following defining properties.*

(1) *For any measurable set $Y \subseteq X$, we have that $\mu_y^{\mathscr{A}}(Y) > 0$ for almost every $y \in Y$, $\mu_x^{\mathscr{A}}(Y)$ depends measurably on $x \in X$, and if $\overline{Y}$ is compact then*
$$
(\mu|_Y)_y^{\mathscr{A}} = \frac{1}{\mu_y^{\mathscr{A}}(Y)} \mu_y^{\mathscr{A}}|_Y
$$
for almost every $y \in Y$.

(2) *If $[x]_{\mathscr{A}} = [y]_{\mathscr{A}}$ (that is, if x and y belong to the same $\mathscr{A}$-atom) then $\mu_x^{\mathscr{A}}$ is proportional to $\mu_y^{\mathscr{A}}$, whenever both measures are defined.*

(3) *The family of measures $\mu_x^{\mathscr{A}}$ is, up to proportionality, uniquely characterized by property (1). More concretely, if ν_x is a family of locally finite measures satisfying (1) then there is a measurable function $s : X \to \mathbb{R}_{>0}$ with*

$$\nu_x = s(x)\mu_x^{\mathscr{A}}$$

almost everywhere.

(4) *If a continuous proper map $T : X \to X$ satisfies $T^{-1}\mathscr{A} = \mathscr{A}$ and preserves μ (that is, $\mu(T^{-1}B) = \mu(B)$ for all measurable $B \subseteq X$), then $T_*\mu_x^{\mathscr{A}} \propto \mu_{Tx}^{\mathscr{A}}$ for almost every $x \in X$ (that is, $T_*\mu_x^{\mathscr{A}}$ and $\mu_{Tx}^{\mathscr{A}}$ are proportional).*

Even though the family of measures constructed in Proposition 9.2 is a direct generalization of the notion of conditional measures (see [**?**, Ch. 5]), we will refer to them as *fiber measures* in order to emphasize the fact that they have quite different properties when the measure is not finite. The term *conditional measure* is also widely used in this setting. Notice that it is possible to modify the construction of $\mu_x^{\mathscr{A}}$ so as to arrange that (almost everywhere) $\mu_x^{\mathscr{A}} = \mu_y^{\mathscr{A}}$ whenever $[x]_{\mathscr{A}} = [y]_{\mathscr{A}}$; that is, to arrange that the map $x \mapsto \mu_x^{\mathscr{A}}$ is $\mathscr{A}$-measurable. However, this will not be needed here (and will not be true for the leafwise measures defined in Section 9.2).

Before starting the proof of Proposition 9.2, we expand a little on the implicit measurability of the map $x \mapsto \mu_x^{\mathscr{A}}$. We stated that

$$x \mapsto \mu_x^{\mathscr{A}}(Y)$$

is measurable for any compact set $Y \subseteq X$, and implicitly stated that

$$y \mapsto \frac{1}{\mu_y^{\mathscr{A}}(Y)}\mu_y^{\mathscr{A}}|_Y$$

is $\mathscr{A}$-measurable on Y. Therefore, for any continuous function with compact support, and for any non-negative measurable function f, the map

$$x \longmapsto \int f \,\mathrm{d}\mu_x^{\mathscr{A}}$$

is measurable (though it may not be $\mathscr{A}$-measurable).

Proof of Proposition 9.2. We start with the almost everywhere uniqueness property in (3). Suppose that ν_x and ρ_x are two families of locally finite measures satisfying (1). Write X as a countable union of compact sets Y_n with $Y_n \subseteq Y_{n+1}^o$. Then the measures $\frac{1}{\nu_x(Y_n)}\nu_x|_{Y_n}$ and $\frac{1}{\rho_x(Y_n)}\rho_x|_{Y_n}$ are well-defined and equal for almost every $x \in Y_n$ by assumption. Let $N \subseteq X$ be a null set with the property that this equality holds for $x \in Y_n \smallsetminus N$ for all $n \geqslant 1$. It follows that for $x \in Y_n \smallsetminus N$ we have $\nu_x = \frac{\nu_x(Y_n)}{\rho_x(Y_n)}\rho_x$ as required for (3).

We next turn to existence of the family $\mu_x^{\mathscr{A}}$. Let $\mu_n = \mu|_{Y_n}$ with Y_n as in the previous paragraph. Then by [**?**, Th. 5.14] the family of conditional measures $(\mu_n)_y^{\mathscr{A}}$ exists, and by Lemma 9.1 we have the compatibility condition

$$(\mu_n)_y^{\mathscr{A}} = \frac{1}{(\mu_k)_y^{\mathscr{A}}(Y_n)}(\mu_k)_y^{\mathscr{A}}|_{Y_n} \tag{9.3}$$

for almost every $y \in Y_n$, whenever $k > n$ (see Figure 9.1).

Let $X' \subseteq X$ be a set of full measure such that (9.3) holds for all $y \in X'$ and all $k > n$ with $y \in Y_n$. In particular,

$$(\mu_k)_y^{\mathscr{A}}(Y_n) > 0$$

if $y \in X' \cap Y_n$ and $k \geqslant n$.

For $m \geqslant 1$, $y \in X' \cap Y_m \smallsetminus Y_{m-1}$, and $B \subseteq X$ a measurable set, we define

$$\mu_y^{\mathscr{A}}(B) = \lim_{k\to\infty} \frac{1}{(\mu_k)_y^{\mathscr{A}}(Y_m)}(\mu_k)_y^{\mathscr{A}}(B). \tag{9.4}$$

To see that this makes sense and defines a measure, suppose first that $B \subseteq Y_n$ for some $n \geqslant m$. Then for $k \geqslant n$, two applications of (9.3) shows that

$$\begin{aligned}\frac{1}{(\mu_k)_y^{\mathscr{A}}(Y_m)}(\mu_k)_y^{\mathscr{A}}(B) &= \frac{1}{(\mu_k)_y^{\mathscr{A}}(Y_n)(\mu_n)_y^{\mathscr{A}}(Y_m)}(\mu_k)_y^{\mathscr{A}}(Y_n)(\mu_n)_y^{\mathscr{A}}(B)\\ &= \frac{1}{(\mu_n)_y^{\mathscr{A}}(Y_m)}(\mu_n)_y^{\mathscr{A}}(B),\end{aligned}$$

so in this case the sequence in (9.4) is eventually constant.

Now

$$B = B \cap Y_m \sqcup B \cap (Y_{m+1} \smallsetminus Y_m) \sqcup B \cap (Y_{m+2} \smallsetminus Y_{m+1}) \sqcup \cdots,$$

so we may write

$$\begin{aligned}\frac{1}{(\mu_k)_y^{\mathscr{A}}(Y_m)}(\mu_k)_y^{\mathscr{A}}(B) = \frac{1}{(\mu_k)_y^{\mathscr{A}}(Y_m)}\Big((\mu_k)_y^{\mathscr{A}}\big(B \cap Y_m\big)&\\ +(\mu_k)_y^{\mathscr{A}}\big(B \cap (Y_{m+1} \smallsetminus Y_m)\big) + \cdots&\\ +(\mu_k)_y^{\mathscr{A}}\big(B \cap (Y_k \smallsetminus Y_{k-1})\big)\Big)&,\end{aligned}$$

and so

$$\mu_y^{\mathscr{A}}(B) = (\mu_m)_y^{\mathscr{A}}\big(B \cap Y_m\big) + \sum_{k>m} \frac{1}{(\mu_k)_y^{\mathscr{A}}(Y_m)}(\mu_k)_y^{\mathscr{A}}\big(B \cap (Y_k \smallsetminus Y_{k-1})\big)$$

is a well-defined sum (possibly converging to ∞) of non-negative terms. It follows that $\mu_y^{\mathscr{A}}$, defined by (9.4), is a locally finite measure on X.

By construction, the family $\mu_x^{\mathscr{A}}$ satisfies (1) for the compact subsets

$$Y = Y_n$$

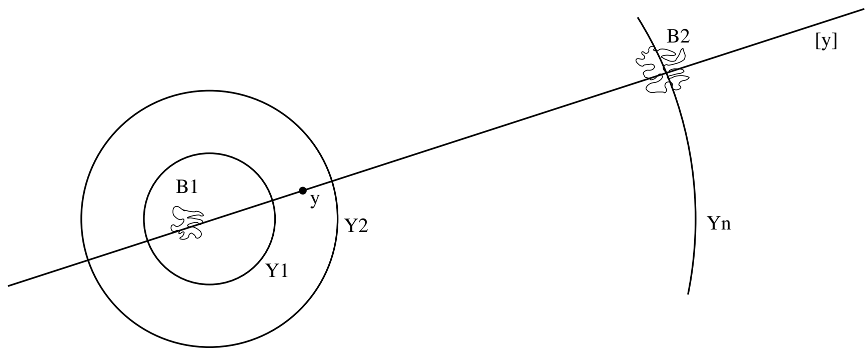

Fig. 9.1 To define $\mu_y^{\mathscr{A}}(B)$ for $y \in Y_2 \smallsetminus Y_1$ and for $B = B_1 \cup B_2$, we need to consider all restrictions of μ_n for $n \geqslant m = 2$, and their conditional measures $(\mu_k)_y^{\mathscr{A}}$.

in the construction. If $Y \subseteq X$ is a subset with $\overline{Y}$ is compact, then

$$Y \subseteq \bigcup_{n=1}^{\infty} Y_n^o$$

by the choice of Y_n and so $Y \subseteq Y_n$ for some n. Thus (1) follows from the special case above together with Lemma 9.1.

Property (2) follows from (1) by the following argument. If

$$[x]_{\mathscr{A}} = [y]_{\mathscr{A}}, x \in Y_n \smallsetminus Y_{n-1}, y \in Y_m \smallsetminus Y_{m-1},$$

then $x, y \in Y_k$ for $k \geqslant \max\{m, n\}$ and by property (1) we have

$$\frac{1}{\mu_x^{\mathscr{A}}(Y_k)} \mu_x^{\mathscr{A}}|_{Y_k} = \frac{1}{\mu_y^{\mathscr{A}}(Y_k)} \mu_y^{\mathscr{A}}|_{Y_k}, \tag{9.5}$$

which implies that

$$\frac{\mu_x^{\mathscr{A}}(Y_k)}{\mu_y^{\mathscr{A}}(Y_k)}$$

is independent of k, and so by (9.5) also that $\mu_x^{\mathscr{A}} \propto \mu_y^{\mathscr{A}}$ as required.

Suppose now that T is as in (4). Then

$$T|_{T^{-1}(Y_n)} : \left(T^{-1}(Y_n), \mathscr{B}(T^{-1}(Y_n)), \mu|_{T^{-1}(Y_n)}\right) \to (Y_n, \mathscr{B}(Y_n), \mu|_{Y_n})$$

is measure-preserving and satisfies $T^{-1}(\mathscr{A}|_{Y_n}) = \mathscr{A}|_{T^{-1}(Y_n)}$. Hence by Exercise 9.1.4 (see [**?**, Cor. 5.24]) we have

$$T_*\left(\left(\mu|_{T^{-1}(Y_n)}\right)_y^{\mathscr{A}|_{T^{-1}(Y_n)}}\right) = (\mu|_{Y_n})_{T(y)}^{\mathscr{A}|_{Y_n}}$$

for μ-almost every $y \in T^{-1}(Y_n)$. Applying (1) to $T^{-1}(Y_n)$ and to Y_n we see that

$$T_*\left(\mu_y^{\mathscr{A}}|_{T^{-1}(Y_n)}\right) \propto \mu_{T(y)}^{\mathscr{A}}|_{Y_n}$$

for μ-almost every $y \in T^{-1}(Y_n)$. As this holds for all $n \geqslant 1$, property (4) follows. □

Exercises for Section 9.1

Exercise 9.1.1. Let X be a σ-compact metric space. Prove that if $(\nu_n)_{n\geqslant 1}$ is a sequence of measures on X, then

$$\nu(B) = \sum_{n=1}^{\infty} \nu_n(B)$$

defines a measure on X as well. Show that if ν_n is locally finite, and for every compact set Y, there is some n_Y such that $\nu_n(Y) = 0$ for $n \geqslant n_Y$, then ν is locally finite as well.

Exercise 9.1.2. Use the following outline for an alternative construction of the fiber measures $\mu_x^{\mathscr{A}}$ as in Proposition 9.2. Let $f \in \mathscr{L}^1(X, \mathscr{B}, \mu)$ be a function with $f > 0$ everywhere. Define a finite measure ν by $\mathrm{d}\nu = f\,\mathrm{d}\mu$, so that $\nu_x^{\mathscr{A}}$ can be defined. Now define $\mu_x^{\mathscr{A}}$ by $\mathrm{d}\mu_x^{\mathscr{A}} = \frac{1}{f}\,\mathrm{d}\nu_x^{\mathscr{A}}$. Prove that this agrees up to proportionality (as in Proposition 9.2(3)) with the definition in Proposition 9.2.

Exercise 9.1.3. Two countably generated σ-algebras $\mathscr{A}$ and $\mathscr{C}$ are countably equivalent if any atom of $\mathscr{A}$ can be covered by at most countably many atoms of $\mathscr{C}$, and *vice versa*. Prove that countable equivalence (as defined on p. 336) is an equivalence relation on the space of σ-algebras on a set.

Exercise 9.1.4. Let $(X, \mathscr{B}_X, \mu)$ and $(Y, \mathscr{B}_Y, \nu)$ be finite measure spaces. Let $\mathscr{A} \subseteq \mathscr{B}_Y$ be a sub-σ-algebra, and let $T : X \to Y$ be a measure-preserving map. Prove that

$$T_*\left(\mu_x^{T^{-1}\mathscr{A}}\right) = \nu_{Tx}^{\mathscr{A}}$$

for μ-almost every $x \in X$.

9.2 Leafwise Measures for Orbit Foliations of $\Gamma\backslash G$

In this section we apply the abstract results from Section 9.1 to obtain, in the context[(29)] of $X = \Gamma\backslash G$, a notion of measures restricted to (conditional on) the orbit of the action of some group H on X, with respect to some measure μ on X. We will have almost no assumption regarding the behavior of μ with respect to the H-action. As we will see later, these measures, which we will call *leafwise measures*, are not very interesting if the measure μ is actually invariant under H (although this property is often desirable) since this is precisely the case in which the leafwise measures will coincide almost everywhere with a Haar measure on H. Instead we will use the leafwise measures in contexts where we know very little about the relationship between the action of H and the measure μ, and we will see that various properties of the H-action with respect to μ can be characterized by the leafwise measures.

9.2.1 The Setting

We begin with a familiar example, where both the algebra and the geometry are very simple.

Example 9.3. Let $G=\mathbb{R}^2$, $\Gamma=\mathbb{Z}^2$, $X=\mathbb{T}^2$, and let $H\subseteq\mathbb{R}^2$ be an irrational line (that is, a line through the origin with irrational slope). We saw in [**?**, Sec. **??**] how the structure of a probability measure — invariant under an automorphism of $\mathbb{T}^2$ — along such an irrational line can affect the entropy of a toral automorphism.

Let μ be a probability measure on $\mathbb{T}^2$. Since the unit square $[0,1)^2$ is a fundamental domain for $\mathbb{Z}^2$ in $\mathbb{R}^2$ we may also view μ as a measure $\mu_{[0,1)^2}$ on $[0,1)^2$. We extend this to all of $\mathbb{R}^2$ by tiling $\mathbb{R}^2$ with translates of the unit square and adding up the translated measures to get

$$\mu_{\mathbb{R}^2}=\sum_{n\in\mathbb{Z}^2}\left(\mu_{[0,1)^2}+n\right)$$

where $\mu_{[0,1)^2}+n$ is the push-forward of $\mu_{[0,1)^2}$ under translation by n.

Finally, notice that $G/H\cong\mathbb{R}$ and so the σ-algebra

$$\mathscr{B}_{G/H}=\{B\in\mathscr{B}_{\mathbb{R}^2}\mid B+H=B\}$$

is easily seen to be countably-generated. Moreover, the atoms of $\mathscr{B}_{G/H}$ are cosets of H in $\mathbb{R}^2$ and so the construction of the fiber measures in Section 9.1 applies, and defines for $\mu_{\mathbb{R}^2}$-almost every x a fiber measure $(\mu_{\mathbb{R}^2})_x^{\mathscr{B}_{G/H}}$ and the leafwise measure can be[†] defined (up to a proportionality constant) by

$$\mu_x^H=(\mu_{\mathbb{R}^2})_x^{\mathscr{B}_{G/H}}-x,$$

by which we mean the push-forward of $(\mu_{\mathbb{R}^2})_x^{\mathscr{B}_{G/H}}$ under translation by $-x$, so that μ_x^H is a locally finite measure on H.

We now generalize Example 9.3. Let G be a σ-compact group with a left-invariant metric d defining its topology, and let $\Gamma\leqslant G$ be a discrete subgroup. We write $I\in G$ for the identity element, and we will also assume that d_G is a *proper metric*, meaning that

need to check the proof: it probably suffices to assume that H has a proper metric — this would be a much weaker assumption as H is abelian or nilpotent most of the time — For H the assumption is definitely used

[†] Our definition below will be $x-(\mu_{\mathbb{R}^2})_x^{\mathscr{B}_{G/H}}$ since in general only $h:x\mapsto xh^{-1}$ will be an action of the group H on X.

$$\overline{B_r^G(I)} = \{g \in G \mid \mathsf{d}_G(g, I) \leqslant r\}$$

is compact for any $r > 0$. The typical examples we have in mind include the following:

- $\Gamma = \mathbb{Z}^n \leqslant G = \mathbb{R}^n$ (with quotient space a torus);
- $\Gamma = \mathbb{Q}^n \leqslant G = \mathbb{A}_{\mathbb{Q}}^n$ (with quotient space a solenoid);
- $G = \mathrm{SL}_n(\mathbb{R})$ (or any closed linear group) and $\Gamma = \mathrm{SL}_n(\mathbb{Z})$ (or a discrete subgroup in a closed linear group).

Define $\pi : G \to \Gamma\backslash G$ to be the natural quotient map $g \mapsto \Gamma g$. Let $H \leqslant G$ be a closed subgroup, and as before we define the action of H on $\Gamma\backslash G$ by $h{\scriptstyle\bullet}x = xh^{-1}$ for $x \in X$. By the results of Chapter 2 this *right action* of H on X is often ergodic unless there are obvious obstructions (for instance, in Example 9.3, an obvious obstruction would be if H were a rational line); this means there are very few Borel subsets of X which are H-invariant. Equivalently, there are very few Borel subsets of X which are unions of H-orbits.

We define the *foliation into H-orbits* to be the partition of X into the H-orbits

$$\mathscr{F}_H = \{xH \mid x \in X\}.$$

We will sometimes refer to the elements of $\mathscr{F}_H$ as *leaves of the foliation*†. In a typical situation, there is no countably generated σ-algebra $\mathscr{A}$ for which the $\mathscr{A}$-atoms are precisely the leaves of $\mathscr{F}_H$ (that is, are the H-orbits; see Exercise 9.2.1).

Nonetheless, we would like to have a description of a probability measure μ on X along the leaves of $\mathscr{F}_H$, and these are what we shall refer to as *leafwise measures*. There are various ways to construct these measures; we will use Section 9.1 after lifting the measure μ to a locally finite measure μ_G on G (as in Example 9.3). The advantage of this transition is that in the lifted situation the H-orbits gH (that is, the left cosets of H) on G are the atoms of a countably-generated σ-algebra $\mathscr{B}_{G/H}$.

We now proceed to the more general framework, describing how to do the two steps in Example 9.3 in a more general setting.

Lemma 9.4. *Let G be a σ-compact, locally compact group with a left-invariant proper metric d_G defining its topology, let $\Gamma < G$ be discrete and define $X = \Gamma\backslash G$. Now let μ be a locally finite measure on‡ X. Then there is a Γ-left-invariant locally finite measure μ_{G} on G with the property that*

$$\int f \,\mathrm{d}\mu_{\mathsf{G}} = \int \sum_{\mathsf{g} \in \mathsf{X}} f(\mathsf{g}) \,\mathrm{d}\mu(\mathsf{x})$$

† We will use this terminology for any group H, even if H is, for example, a p-adic group.

‡ For now ignore the change of notation, or see Remark 9.11.

for any continuous function f with compact support, and for any non-negative measurable function f.

The sum $\sum_{\mathsf{g}\in\mathsf{x}} f(\mathsf{g})$ on the right-hand side in Lemma 9.4 is a function on X, so that the integral makes sense.

PROOF. This is clear from the Riesz representation theorem, since the right-hand side defines a positive linear functional on $C_c(\mathsf{G})$; Γ-invariance of μ_{G} follows from the (manifest) Γ-invariance of this functional. The extension to non-negative measurable functions follows by the monotone convergence theorem. □

An alternative description of the measure μ_{G} in Lemma 9.4 comes from emulating Example 9.3 directly. If $F \subseteq \mathsf{G}$ is a fundamental domain of X (that is, a Borel set for which π restricted to F is a Borel isomorphism[†] between F and $\mathsf{X} = \Gamma\backslash\mathsf{G}$), we can identify μ with a measure μ_F on F and then define

$$\mu_{\mathsf{G}} = \sum_{\gamma\in\Gamma} \gamma_*\mu_F.$$

Applying Lemma 9.4 to $f = \mathbb{1}_B$ for a Borel subset $B \subseteq F$ shows that this is the same construction.

Lemma 9.5. *Let G be a σ-compact, locally compact group with a left-invariant proper metric d, and let H be a closed subgroup of G. There is a countably-generated σ-algebra $\mathscr{B}_{G/H} \subseteq \mathscr{B}_G$ whose atoms are the cosets gH for $g \in G$.*

PROOF. Since G has a left-invariant metric d, it also has a right-invariant metric defined by

$$\mathsf{d}_r(g_1, g_2) = \mathsf{d}(g_1^{-1}, g_2^{-1})$$

for $g_1, g_2 \in G$. Just as for X we can make the left coset space G/H into a metric space by letting

$$\mathsf{d}_{G/H}(g_1H, g_2H) = \inf_{h\in H} \mathsf{d}_r(g_1, g_2h).$$

The canonical quotient map $G \to G/H$ is continuous and open, so G/H is a σ-compact metric space. The Borel σ-algebra on G/H is countably-generated (by the countable set of metric balls with rational radius centered at points in a dense countable set, for example), and we identify this σ-algebra with a sub-σ-algebra $\mathscr{B}_{G/H}$ of $\mathscr{B}_G$ whose atoms will be the cosets gH as required. □

Using the argument above, and the results of Section 9.1, we have for almost every $\mathsf{g} \in \mathsf{G}$ a locally finite fiber measure $(\mu_G)_{\mathsf{g}}^{\mathscr{B}_{\mathsf{G}/H}}$ supported

[†] A map $\phi : Y_1 \to Y_2$ is called a Borel isomorphism if ϕ is bijective, measurable with respect to the Borel σ-algebra, and ϕ^{-1} is also measurable with respect to the Borel σ-algebra.

on $\mathsf{g}H$. By moving these fiber measures to H (using the inverse of the orbit map $\phi_{\mathsf{g}}(h) = h\bullet\mathsf{g}$ for $h \in H$) we will obtain the definition and basic properties of the leafwise measures.

9.2.2 Main Properties of the Leafwise Measures

Theorem 9.6. *Let $G, \Gamma, H, \mathscr{B}_{\mathsf{G}/H}$, and*[†] $\mathsf{X} = \Gamma\backslash\mathsf{G}$ *be as above. Let μ be a locally finite measure on X. Then there exists a set $X' \subseteq \mathsf{X}$ of full μ-measure, and for every $\mathsf{x} \in X'$ a leafwise measure μ_{x}^H which has the following properties.*

(1) *μ_{x}^H is a locally finite measure on H.*
(2) *The identity I lies in $\operatorname{Supp}\mu_{\mathsf{x}}^H$, that is $\mu_{\mathsf{x}}^H(B_\delta^H) > 0$ for any $\delta > 0$.*

By (1) and (2) we may normalize μ_{x}^H so that $\mu_{\mathsf{x}}^H(B_1^H) = 1$. Furthermore,

(3) *The map $\mathsf{x} \mapsto \mu_{\mathsf{x}}^H$ is measurable (that is, for any function $f \in C_c(H)$, or for any non-negative measurable function f on H, the map $\mathsf{x} \mapsto \int_H f \,\mathrm{d}\mu_{\mathsf{x}}^H$ is Borel measurable).*
(4) *Moreover, for $\mathsf{x}, h\bullet\mathsf{x} \in X'$ with $h \in H$ we have*
$$\mu_{h\bullet\mathsf{x}}^H h \propto \mu_{\mathsf{x}}^H,$$
where $\mu_{h\bullet\mathsf{x}}^H h$ is the push-forward of $\mu_{h\bullet\mathsf{x}}^H$ under right-translation by h.
(5) *Write μ_{G} for the induced Γ-left-invariant measure on G. Then for μ_{G}-almost every $\mathsf{g} \in \mathsf{G}$ we have $\mathsf{x} = \pi(\mathsf{g}) \in X'$ and*
$$(\mu_{\mathsf{G}})_{\mathsf{g}}^{\mathscr{B}_{\mathsf{G}/H}} \propto \mu_{\mathsf{x}}^H\bullet\mathsf{g},$$
where we simply write $\mu_{\mathsf{x}}^H\bullet\mathsf{g}$ for the push-forward $(\phi_{\mathsf{g}})_*\mu_{\mathsf{x}}^H$ under the map $\phi_{\mathsf{g}} : h \in H \mapsto h\bullet\mathsf{g}$ obtained by letting the varying $h \in H$ act on a fixed $\mathsf{g} \in \mathsf{G}$.

The formula $\mu_{h\bullet\mathsf{x}}^H h \propto \mu_{\mathsf{x}}^H$ may be interpreted as saying that the leafwise measures define measures
$$\mu_{h\bullet\mathsf{x}}^H\bullet(h\bullet\mathsf{x}) = \left(\mu_{h\bullet\mathsf{x}}^H h\right)\bullet\mathsf{x} \propto \mu_{\mathsf{x}}^H\bullet\mathsf{x}$$
on the H-orbit $H\bullet\mathsf{x}$ which are independent of the chosen starting point
$$\mathsf{x} \in (H\bullet\mathsf{x}) \cap X'$$
apart from a factor of proportionality which may depend on x. Note, however, that $\mu_{\mathsf{x}}^H\bullet\mathsf{x}$ may not itself be a locally finite measure on X. For this reason we

[†] Again ignore the strange fonts for now, or if you have to then look at Remark 9.11 right away.

avoid working directly with $\mu_{\mathsf{x}}^{H}\bullet\mathsf{x}$ and instead work with the orbits in G. An overview of the situation is given in Figure 9.2.

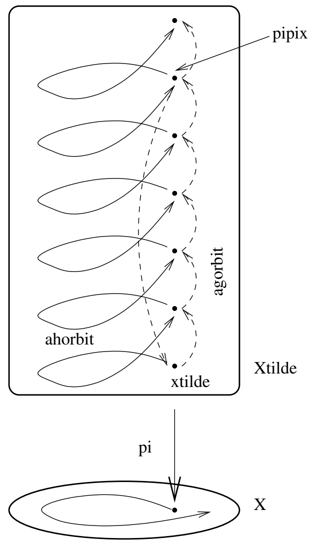

Fig. 9.2 A close return of the H-orbit through x to x is lifted to a close visit of the H-orbit through g to some $\gamma\mathsf{g}$. Hence a locally finite measure on the H-orbit of g may not correspond to a locally finite measure on the H-orbit of x

.

maybe the tower in the Figure doesn't have to be as tall?

Before we turn to the proof of Theorem 9.6 we describe the sense in which the leafwise measures μ_{x}^{H} describe μ along the leaves of $\mathscr{F}_H$ (that is, along the H-orbits).

Definition 9.7. A countably-generated σ-algebra $\mathscr{A} \subseteq \mathscr{B}_Y$ for some measurable set $Y \subseteq \mathsf{X}$ is said to be *H-subordinate* if

$$[\mathsf{x}]_{\mathscr{A}} = V_{\mathsf{x}}\bullet\mathsf{x} = \{\mathsf{x}h^{-1} \mid h \in V_{\mathsf{x}}\}$$

for some open bounded set $V_{\mathsf{x}} \subseteq H$ and every $\mathsf{x} \in Y$.

Definition 9.8. We say a set $P \subseteq \mathsf{X}$ is an *open H-plaque* if $P = V\bullet\mathsf{x}$ for some $\mathsf{x} \in \mathsf{X}$ and some open bounded set $V \subseteq H$.

Corollary 9.9. *In the notation and with the hypotheses of Theorem* 9.6*, assume in addition that for almost every $\mathsf{x} \in \mathsf{X}$ the map $H \ni h \mapsto h\bullet\mathsf{x}$ is injective, let $Y \subseteq \mathsf{X}$ be a measurable set of finite measure, and let $\mathscr{A} \subseteq \mathscr{B}_Y$ be an H-subordinate σ-algebra on Y. Then the conditional measures for $\mathscr{A}$ are*

$$(\mu|_Y)_{\mathsf{y}}^{\mathscr{A}} = \left(\frac{1}{\mu_{\mathsf{y}}^{H}(V_{\mathsf{y}})}\mu_{\mathsf{y}}^{H}|_{V_{\mathsf{y}}}\right)\bullet\mathsf{y} \tag{9.6}$$

for almost every $\mathsf{y} \in Y$ *with* $[\mathsf{y}]_{\mathscr{A}} = V_{\mathsf{y}} \bullet \mathsf{y}$ *and* $V_{\mathsf{y}} \subseteq H$*; that is they are obtained from the corresponding restrictions of the leafwise measures.*

Corollary 9.9 will be proved on p. 291. We will show in Section 9.2.4 the existence of special H-subordinate σ-algebras, which we will call (H, R)-*flowers*. In fact one can use such σ-algebras to define the leafwise measures$^{(30)}$.

We will also be interested in the behavior of null sets with respect to the leafwise measures as in the next corollary.

Corollary 9.10. *Let* $\mathsf{X} = \Gamma\backslash\mathsf{G}$*,* H*, and* μ *be as in Theorem* 9.6*, and let* N *be a null subset of* X*. Then*

$$\mu_{\mathsf{x}}^{H}\left(\{h \in H \mid h\bullet\mathsf{x} \in N\}\right) = 0$$

for almost every† $\mathsf{x} \in \mathsf{X}$.

Remark 9.11. The reader may have noticed that we use the notation G and X instead of G and X in many of the above results. The reason for that is that the above results are sometimes used in the following slightly stronger form.

Let Ω be a compact metric space, let us refer to it as the auxiliary space. Then we define $\mathsf{X} = X \times \Omega$ and $\mathsf{G} = G \times \Omega$ (so that if Ω is a singleton we may identify G with G and X with X). Both actions, the action of Γ on the left and the action of G on the right extend to G by letting the action be trivially on Ω. More precisely, if $\mathsf{g} = (g, \omega) \in \mathsf{G}$, $\gamma \in \Gamma$, and $h \in G$, then we define $\gamma\mathsf{g} = \gamma(g, \omega) = (\gamma g, \omega)$ and $h\bullet\mathsf{g} = \mathsf{g}h^{-1} = (gh^{-1}, \omega)$. In that sense we still have $\mathsf{X} = \Gamma\backslash\mathsf{G}$ and G (and so also its subgroup $H < G$) still has a well-defined right action on X.

There are a few reasons why we simply hide the extension of the results in the notation instead of putting the generalization full-fledged into the above statements. First, the extension is only used in some applications of the results here and comes basically for free, i.e. the proofs are not getting more complicated unless we would allow the notation to become more complicated. Second, as all standard Borel spaces are Borel isomorphic ([]) we could even assume that the auxiliary space Ω is a compact metric group (for example, setting it equal to $\mathbb{T}$) and replace G by the group $G \times \Omega$ but as we do not want to add this theorem to our prerequisites and in our application of the extension Ω is by definition usually not a group we have chosen the compromise of hiding the extension in the notation.

need a good ref

For the discussion in this section we also note some corollaries of this extended definition. If F_G is a fundamental domain for Γ in G then $F_{\mathsf{G}} = F_G \times \Omega$ is a fundamental domain for Γ in G. Moreover, if $\mathscr{B}_{G/H}$ is the countably generated σ-algebra in G which has the atoms gH for $g \in G$ then $\mathscr{B}_{\mathsf{G}/H} = \mathscr{B}_{G/H} \times \mathscr{B}_{\Omega}$ is the countably generated σ-algebra whose atoms are once more the H-orbits $H\bullet\mathsf{g} = \mathsf{g}H = gH \times \{\omega\}$.

† Strictly speaking, we should say here for almost every $\mathsf{x} \in X'$ as μ_{x}^{H} is not defined for $\mathsf{x} \in \mathsf{X} \smallsetminus X'$, but as X' has full measure itself this amounts to the same.

9.2.3 Proof of Theorem 9.6 — *Construction of Leafwise Measures*

We start by using Proposition 9.2 to construct μ_{x}^H for $\mathsf{x} \in X'$.

Proof of Theorem 9.6. Given the measure μ on $\mathsf{X} = \Gamma\backslash\mathsf{G}$ as in the theorem, construct $\nu = \mu_{\mathsf{G}}$ using Lemma 9.4. Now let $\mathscr{A} = \mathscr{B}_{\mathsf{G}/H}$ be the σ-algebra on G whose atoms are the cosets $\mathsf{g}H$ for $\mathsf{g} \in \mathsf{G}$ (as in Lemma 9.5). Applying Proposition 9.2, we obtain the fiber measures $\nu_{\mathsf{g}}^{\mathscr{A}}$ for all $\mathsf{g} \in G'$, where $G' \subseteq \mathsf{G}$ is a set of full μ_{G}-measure. Since Γ is countable and preserves the measure μ_{G} under its left-action, we may assume that $\Gamma G' = G'$. Moreover, $\gamma\mathscr{A} = \mathscr{A}$ for all $\gamma \in \Gamma$ (indeed, for all $\gamma \in \mathsf{G}$), so by Proposition 9.2(4) we may assume that

$$\gamma\left(\nu_{\mathsf{g}}^{\mathscr{A}}\right) \propto \nu_{\gamma\mathsf{g}}^{\mathscr{A}} \tag{9.7}$$

for all $\gamma \in \Gamma$ and $\mathsf{g} \in G'$, where we simply write $\gamma(\nu_{\mathsf{g}}^{\mathscr{A}})$ for the push-forward of $\nu_g^{\mathscr{A}}$ under left-multiplication by γ. Now let $F \subseteq \mathsf{G}$ be a fundamental domain for the action of Γ on G, so that we can define

$$\mu_{\mathsf{x}}^H = \left(\phi_{\mathsf{g}}^{-1}\right)_* \left(\nu_{\mathsf{g}}^{\mathscr{A}}\right)$$

for $\mathsf{x} \in X' = \pi(G')$ where $\mathsf{g} \in F$ is the unique element with $\pi(\mathsf{g}) = \mathsf{x}$ and $\phi_{\mathsf{g}} : H \to \mathsf{G}$ is defined by $\phi_{\mathsf{g}}(h) = h{\scriptstyle\bullet}\mathsf{g}$ as in the statement of the theorem. Notice that $\nu_{\mathsf{g}}^{\mathscr{A}}$ is a locally finite measure on G that is supported on $\mathsf{g}H$, which is precisely the image of ϕ_{g}. Hence, taking the push-forward under ϕ_{g}^{-1} makes sense and we see that μ_{x}^H is a locally finite measure on H. Therefore (5) holds by definition for $\mathsf{g} \in G' \cap F$. If now $\mathsf{g} \in G'$ and $\gamma \in \Gamma$ has $\gamma\mathsf{g} \in G' \cap F$ then, by (9.7),

$$\mu_{\mathsf{x}}^H = \left(\phi_{\gamma\mathsf{g}}^{-1}\right)_* \nu_{\gamma\mathsf{g}}^{\mathscr{A}} \propto \left(\phi_{\gamma\mathsf{g}}^{-1}\right)_* \left(\gamma\nu_{\mathsf{g}}^{\mathscr{A}}\right) = \left(\phi_{\mathsf{g}}^{-1}\right)_* \nu_{\mathsf{g}}^{\mathscr{A}}$$

since

$$\phi_{\gamma\mathsf{g}}^{-1}\left(\gamma\mathsf{g}h^{-1}\right) = h = \phi_{\mathsf{g}}^{-1}(\mathsf{g}h^{-1})$$

for all $h \in H$. Therefore, (5) holds for any $\mathsf{x} \in G'$. It is clear from Proposition 9.2 that (1) holds.

In order to prove (2) it is enough to prove the translated statement

$$\mathsf{g} \in \operatorname{Supp} \nu_{\mathsf{g}}^{\mathscr{A}}$$

for ν-almost every $g \in \mathsf{G}$. To see this, choose a countable dense subset $\{\mathsf{g}_1, \dots, \mathsf{g}_k, \dots\} \subseteq \mathsf{G}$. Now apply Proposition 9.2(1) and find a null set N so that

$$\nu_{\mathsf{g}}^{\mathscr{A}}\left(B_{1/n}^{\mathsf{G}}(\mathsf{g}_k)\right) > 0$$

for all $\mathsf{g} \in G' \cap B_{1/n}^{\mathsf{G}}(\mathsf{g}_k) \smallsetminus N$, $k \geqslant 1$ and $n \geqslant 1$, which implies by density that

$$\nu_{\mathsf{g}}^{\mathscr{A}}\left(B_{2/n}^{\mathsf{G}}(g)\right)>0$$

for all $\mathsf{g}\in G'\smallsetminus N$, which means that $\mathsf{g}\in\operatorname{Supp}\nu_{\mathsf{g}}^{\mathscr{A}}$. Recall now that

$$\phi_{\mathsf{g}}:H\to H{\scriptstyle\bullet}\mathsf{g}=\mathsf{g}H$$

is a homeomorphism, and hence $I\in\operatorname{Supp}\left(\phi_{\mathsf{g}}^{-1}\right)_{*}\nu_{\mathsf{g}}^{\mathscr{A}}$. This proves (2), and also shows that we may modify our definition of μ_{x}^{H} by a multiplicative scalar as indicated in the theorem.

Assume now that $\mathsf{x}_1=\pi(\mathsf{g}_1)$, $\mathsf{x}_2=\pi(\mathsf{g}_2)$, and that $\mathsf{x}_2=h{\scriptstyle\bullet}\mathsf{x}_1$ as assumed in (4). We may choose g_1 and g_2 such that $\mathsf{g}_2=h{\scriptstyle\bullet}\mathsf{g}_1$. Then

$$\phi_{\mathsf{g}_2}(h')=h'{\scriptstyle\bullet}\mathsf{g}_2=h'h{\scriptstyle\bullet}\mathsf{g}_1=\phi_{\mathsf{g}_1}(h'h)$$

and so by property (5) of the theorem, Proposition 9.2(2) and since $\phi_{\mathsf{g}_1}^{-1}$ coincides with $\phi_{\mathsf{g}_2}^{-1}$ followed by right multiplication with h, we have

$$\mu_{\mathsf{x}_1}^{H}\propto\left(\phi_{\mathsf{g}_1}^{-1}\right)_{*}\nu_{\mathsf{g}_1}^{\mathscr{A}}\propto\left(\phi_{\mathsf{g}_1}^{-1}\right)_{*}\nu_{\mathsf{g}_2}^{\mathscr{A}}=\left(\left(\phi_{\mathsf{g}_2}^{-1}\right)_{*}\nu_{\mathsf{g}_2}^{\mathscr{A}}\right)h,$$

as required for (4).

It remains to show (3). Fix a function $f\in C_c(H)$. We claim that

$$\psi:\mathsf{g}\longmapsto\int f(h)\,\mathrm{d}\nu_{\mathsf{g}}^{\mathscr{A}}(h{\scriptstyle\bullet}\mathsf{g})$$

is measurable. Recall that $\nu_{\mathsf{g}}^{\mathscr{A}}$ is supported on $H{\scriptstyle\bullet}g$, and that $H{\scriptstyle\bullet}\mathsf{g}$ is homeomorphic to H, so writing the variable in the integral as $h{\scriptstyle\bullet}\mathsf{g}$ makes sense. This then implies the same property for any non-negative measurable function f on H, and in particular for $f=\mathbb{1}_{B_1^H}$ (the indicator function of the open unit ball in H which is used to normalize μ_{x}^{H}), so (3) will follow.

To see that $\mathsf{g}\longmapsto\psi(\mathsf{g})$ is indeed measurable, we recall that by Proposition 9.2 the integral

$$\Psi(\mathsf{g})=\int F(\mathsf{g}')\,\mathrm{d}\nu_{\mathsf{g}}^{\mathscr{A}}(\mathsf{g}')\tag{9.8}$$

is a measurable function of $\mathsf{g}\in\mathsf{G}$ whenever $F\geqslant 0$ is a measurable function on G, or alternatively whenever $F\in C_c(\mathsf{G})$. Note that the definition of ψ differs from the definition of Ψ in that for ψ the integrand $f(h)$ depends also on g (since h is the offset between $\mathsf{g}'=h{\scriptstyle\bullet}\mathsf{g}$ and g), whereas in Ψ only the measure depends on g. Nonetheless the measurability extends from Ψ to ψ by the following continuity argument.

By the Tietze–Urysohn extension theorem, there exists some $F_0\in C_c(G)$ with $F_0|_H=f$. As G is separable there exists a sequence $(g_k)_{k\geqslant 1}$ in G with

$$G=\bigcup_{k=1}^{\infty}B_{1/n}^{G}(g_k)$$

for every $n \geqslant 1$. We now define[†]

$$\Psi_n(\mathsf{g}) = \int F_0\left((g')^{-1} g_k\right)\,\mathrm{d}\nu_{\mathsf{g}}^{\mathscr{A}}(g')$$

where we choose $k = k_n(g) \geqslant 1$ minimally with the property that

$$g \in B_{1/n}^{\mathsf{G}}(g_k).$$

In other words, $g \mapsto \Psi_n(g)$ is piecewise equal to functions defined as in (9.8). Hence $g \mapsto \Psi_n(g)$ is measurable. We claim that

$$\lim_{n\to\infty} \Psi_n(g) = \psi(g)$$

for all $g \in \mathsf{G}$. Indeed, for $g \in \mathsf{G}$ and $k_n(g) \geqslant 1$ minimal as in the definition of $\Psi_n(\mathsf{g})$ we have

$$\lim_{n\to\infty} g_{k_n(g)} = g,$$

and so

$$\lim_{n\to\infty} F_0\left((g')^{-1} g_{k_n(g)}\right) = F_0\left((g')^{-1} g\right),$$

where the convergence takes place in $C_c(G)$. Since $\nu_{\mathsf{g}}^{\mathscr{A}}$ is locally finite, we can apply dominated convergence. Moreover, for $g' = h{\bullet}g = gh^{-1} \in gH$,

$$\lim_{n\to\infty} F_0\left(hg^{-1}\mathsf{g}_{k_n(g)}\right) = F_0(h) = f(h)$$

so that $\lim_{n\to\infty} \Psi_n(g) = \psi(g)$ as claimed. It follows that ψ is measurable, and (3) follows as discussed earlier. □

Proof of Corollary 9.10. Let $N \subseteq \mathsf{X}$ be a null set, and let $N_{\mathsf{G}} = \pi^{-1}(N)$ be its pre-image in G. Then $\mu_{\mathsf{G}}(N_{\mathsf{G}}) = 0$ by Lemma 9.4. Therefore,

$$\mu_{\mathsf{G}}|_{B_R^{\mathsf{G}}}(N_{\mathsf{G}}) = 0$$

and so

$$(\mu_{\mathsf{G}})_{\mathsf{g}}^{\mathscr{B}_{\mathsf{G}/H}}\left(N_{\mathsf{G}} \cap B_R^{\mathsf{G}}\right) = 0$$

for all $R \geqslant 1$ and μ_{G}-almost every g, so we also have

$$(\mu_{\mathsf{G}})_{\mathsf{g}}^{\mathscr{B}_{\mathsf{G}/H}}(N_{\mathsf{G}}) = 0.$$

The corollary follows from Theorem 9.6(5). □

[†] We write g for the group component of the tuple $\mathsf{g}(g,\omega) =\in \mathsf{G}$.

9.2.4 H-subordinate σ-algebras

The following notion is useful for the construction of H-subordinate σ-algebras. For convenience, we introduce the notation B_r^H for $B_r^H(I)$, the metric open ball of radius r around the identity in H.

Definition 9.12. Let $H \leqslant G$ be a closed subgroup, and let $\Gamma \leqslant G$ be discrete. Let $\delta \in (0,1]$ and $R \geqslant 1$. A subset $T \subseteq \mathsf{X} = \Gamma\backslash\mathsf{G}$ is an (R,δ)-*cross-section for* H at $\mathsf{x}_0 \in X$ if

(1) $B_\delta^X(\mathsf{x}_0) \subseteq B_1^H \cdot T$, and
(2) the natural map $B_R^H \times T \to B_R^H \cdot T$ is injective and bi-measurable†.

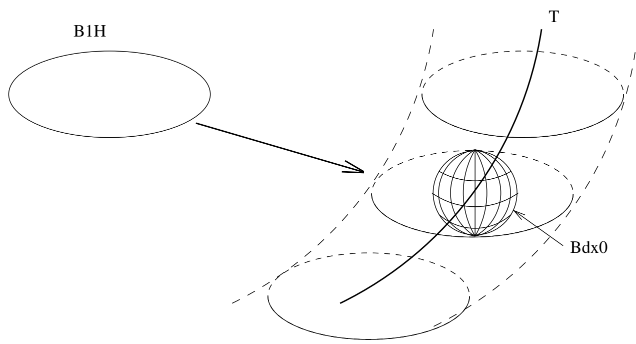

Fig. 9.3 As we will see later, in the case of a Lie group the cross-section T may be chosen as a line transverse to the H-orbit through the center of the δ-ball $B_\delta^{\mathsf{X}}(\mathsf{x})$.

Proposition 9.13. *With the assumptions of Definition 9.12, let x_0 be a point in $\mathsf{X} = \Gamma\backslash\mathsf{G}$ and let $R \geqslant 1$ be chosen so that*

$$\overline{B_R^H} \ni h \longmapsto h{\bullet}\mathsf{x}_0$$

is injective. Then there exists some $\delta > 0$ and some $T \subseteq X = \Gamma\backslash\mathsf{G}$ which is an (R,δ)-cross-section for H at x_0. In particular, there exists an H-subordinate σ-algebra $\mathscr{A} = \mathscr{A}(\mathsf{x}_0, R, \delta)$ on $B_R^H{\bullet}T$ for which

$$[h{\bullet}t]_{\mathscr{A}} = B_R^H{\bullet}t$$

for all $h \in B_R^H$ and $t \in T$.

† This property is actually automatic under the assumed injectivity by Parthasarathy [?], but it is often quite easy to show the measurability of the inverse map directly without using general properties of standard Borel spaces.

Definition 9.14. Let $T \subseteq \mathsf{X} = \Gamma\backslash\mathsf{G}$ be an (R,δ)-cross-section for H at the point $\mathsf{x}_0 \in \mathsf{X}$. Then the σ-algebra $\mathscr{A}(\mathsf{x}_0, R, \delta)$ on $B_R^H{\boldsymbol\cdot}T$ with atoms $B_R^H{\boldsymbol\cdot}t$ for $t \in T$ is called an *(H,R)-flower with base $B_\delta(\mathsf{x}_0)$*.

We will give a proof of Proposition 9.13 for the case where G and H are Lie groups or S-algebraic groups (see Section 8.1), and for these cases we will show that T can be chosen to be σ-compact. For the more general setting we refer to the notes of Einsiedler and Lindenstrauss [**?**]. Also see Exercise 9.2.2 for the case where G is a finite product of real and p-adic Lie groups and H is an arbitrary closed subgroup.

Beginning of proof of Proposition 9.13. Assume that R and x_0 are as in the statement of the proposition.[†] We claim that there exists some $\eta > 0$ so that

$$\overline{B_R^H B_\eta^G} \ni g \longmapsto g{\boldsymbol\cdot}x_0 \tag{9.9}$$

is also injective.

Suppose the claim is not true. Then we can find, for every $\eta = \frac{1}{n}$, distinct elements $g_1^{(n)}$ and $g_2^{(n)}$ with

$$g_1^{(n)} = h_1^{(n)}\varepsilon_1^{(n)}, g_2^{(n)} = h_2^{(n)}\varepsilon_2^{(n)} \in \overline{B_R^H B_\eta^G}$$

such that $g_1^{(n)}{\boldsymbol\cdot}x_0 = g_2^{(n)}{\boldsymbol\cdot}x_0$. Passing to a subsequence, we may assume that

$$\lim_{k\to\infty} h_i^{(n_k)} = h_i$$

for $i = 1, 2$. Then $h_1, h_2 \in \overline{B_R^H}$ and $h_1{\boldsymbol\cdot}x_0 = h_2{\boldsymbol\cdot}x_0$, which implies that

$$h_1 = h_2$$

by the assumed injectivity. However, this forces

$$\lim_{n\to\infty} g_1^{(n)} = \lim_{n\to\infty} g_2^{(n)} = h_1,$$

which together with the initial assumptions $g_1^{(n)} \neq g_2^{(n)}$ and $g_1^{(n)}{\boldsymbol\cdot}x_0 = g_2^{(n)}{\boldsymbol\cdot}x_0$ gives a contradiction to the injectivity radius at $h_1{\boldsymbol\cdot}x_0$. This proves the claim in (9.9).

Proof of Proposition 9.13: construction for a Lie group. Now let $\mathfrak{h} = \operatorname{Lie} H \subseteq \mathfrak{g} = \operatorname{Lie}(G)$ be the Lie algebras of H and of G. Let $V \subseteq \mathfrak{g}$ be a linear complement of $\mathfrak{h} \subseteq \mathfrak{g}$. We let $T_G \subseteq B_\eta^G$ be the image of a sufficiently small neighborhood U_V of $0 \in V$ under the exponential map. We will show

---

[†] As the auxiliary space Ω plays no role in the construction we construct the cross-section T in X and then may simply use $T \times \Omega$ as the cross-section for X.

that if T_G is sufficiently small, then[†] $T = T_G \raisebox{0.1ex}{\textbullet} x_0$ satisfies property (1) of Definition 9.12 and that it also satisfies a local version of property (2).

First notice that the map

$$\begin{aligned} H \times V &\longrightarrow G \\ (h, v) &\longmapsto h \exp(v) \end{aligned}$$

is differentiable, and that its derivative at the point $(I, 0)$ is given by

$$\begin{aligned} \mathfrak{h} \times V &\longrightarrow \mathfrak{g} \\ (w, v) &\longmapsto w + v, \end{aligned}$$

which is invertible by our choice of V. It follows that the map above is a local diffeomorphism, and so by choosing $T_G = \exp(U_V)$ small enough we may ensure that

$$B_\varepsilon^H \times T_G \longrightarrow B_\varepsilon^H T_G$$

is a homeomorphism. Moreover, there exists some $\delta > 0$ with

$$B_\delta^G \subseteq B_1^H T_G.$$

it follows that $T = T_G \cdot x_0$ satisfies Definition 9.12(1), and

$$B_\varepsilon^H \times T \longrightarrow B_\varepsilon^H \cdot T$$

is a homeomorphism.

Finishing the proof of Proposition 9.13. Suppose now that the map

$$B_R^H \times T \to B_R^H \cdot T$$

in (2) from Definition 9.12 is not injective. Choose $(h_1, t_1) \neq (h_2, t_2)$ with

$$h_1 \cdot t_1 = h_2 \cdot t_2$$

and $t_1', t_2' \in T_G$ with $t_i = t_i' \cdot x_0$ for $i = 1, 2$ so that $(h_1 t_1') \cdot x_0 = (h_2 t_2') \cdot x_0$. By the choice of η, the map in (9.9) is injective, and so $h_1 t_1' = h_2 t_2'$. This implies that $(h_2^{-1} h_1) t_1' = t_2'$, and so $\mathsf{d}(h_2^{-1} h_1, I) \geqslant \varepsilon$.

Moreover, $h_2^{-1} h_1 = t_2' (t_1')^{-1}$, which implies that

$$\varepsilon \leqslant \mathsf{d}(t_2'(t_1')^{-1}, I) \leqslant \operatorname{diam}(T_G T_G^{-1}) \leqslant 4\eta.$$

However, if we shrink η then we can keep ε unchanged, we may replace T_G and T by some smaller subsets and obtain a contradiction. This shows the injectivity of the map in (2).

[†] In the case the auxiliary space Ω is not a singleton, we set $T = T_G \cdot x_0 \times \Omega$. The argument trivially generalizes to this case.

Clearly, $B_R^H \times T \to B_R^H{\bullet}T$ is measurable, since it is continuous. Since both $B_R^H = \bigcup_{n=1}^\infty K_n$ and $T = \bigcup_{m=1}^\infty T_m$ are countable unions of compact sets, this shows the same property for $B_R^H{\bullet}T_m = \bigcup_{n,m} K_n{\bullet}T_m$. Moreover, the inverse of the map $K_n \times T_m \to K_n T_m$ is also continuous and hence measurable. This implies that the map in (2) is bi-measurable.

Finally, let $\mathscr{A}$ be the image of the σ-algebra

$$\{\varnothing, B_R^H\} \times \mathscr{B}(T) \subseteq \mathscr{B}\left(B_R^H \times T\right)$$

under the map from $B_R^H \times T$ to $B_R^H{\bullet}T$. Then the $\mathscr{A}$-atom of $h{\bullet}t$ is given by $B_R^H{\bullet}t$ for any $h \in B_R^H$ and $t \in T$. □

Notice that the proof above made use of the assumption that G is a Lie group only in the construction part of the proof. In Section 9.2.5 we will generalize this construction, and the remainder of the proof will also apply to the more general setting considered later.

9.2.5 Proposition 9.13 *for S-algebraic groups*

Now consider a discrete subgroup $\Gamma \leqslant G$ and a subgroup $H \leqslant G_S$ defined by $H = H_{S'}$ with $|S'| < \infty$ and with $\mathbb{H}_\sigma \leqslant \mathbb{G}_\sigma$ for $\sigma \in S'$. We will now prove Proposition 9.13 in this context (once again producing a σ-compact section T).

Manfred needs to continue checking things from here

Proof of Proposition 9.13 for S-algebraic groups. As in the beginning of the proof on p. 286, the map from (9.9) is injective for sufficiently small $\eta > 0$. We now replace the construction for a Lie group on p. 286 by a more algebraic argument.

For every $\sigma \in S'$ we let $\mathfrak{h}_\sigma = \mathrm{Lie}(\mathbb{H}_\sigma) \subseteq \mathrm{Lie}(\mathbb{G}_\sigma) \subseteq \mathfrak{gl}_r(\mathbb{Q}_\sigma) = \mathrm{Mat}_{rr}(\mathbb{Q}_\sigma)$ be the Lie algebras of $\mathbb{H}_\sigma$, $\mathbb{G}_\sigma$ and $\mathrm{GL}_r(\mathbb{Q}_\sigma)$. We set $\mathfrak{h}_\sigma = \{0\}$ if $\sigma \in S\smallsetminus S'$, and let $V_\sigma \subseteq \mathfrak{gl}_r(\mathbb{Q}_\sigma)$ be a linear complement of $\mathfrak{h}_\sigma \subseteq \mathfrak{gl}_r(\mathbb{Q}_\sigma)$ for all $\sigma \in S$. We now define

$$T_{G_S} = \overline{B_\eta^{G_S}} \cap \prod_{\sigma\in S}(I + V_\sigma),$$

where as usual I denotes the $r \times r$ identity matrix. We claim that if $\eta > 0$ and $\varepsilon > 0$ are sufficiently small, then the map

$$\overline{B_\varepsilon^H} \times T_{G_S} \longrightarrow \overline{B_\varepsilon^H} \cdot T_{G_S} \tag{9.10}$$

is a homeomorphism, and the image $\overline{B_\varepsilon^H} \cdot T_{G_S}$ contains $B_\delta^{G_S}$ for some $\delta > 0$. Since the map is continuous with compact domain, in order to prove the claim it is enough to show injectivity and the inclusion $\overline{B_\varepsilon^H} \cdot T_{G_S} \supseteq B_\delta^{G_S}$. As S' is finite, for both of these statements it is enough to consider a single place $\sigma \in S'$.

So let $h_0 \in H_\sigma$ be a non-identity point close to the identity, so that

$$h_0 = I + w + \mathrm{O}(|h_0 - I|_\sigma^2) \tag{9.11}$$

for some $w \in \mathfrak{h}_\sigma$, while on the other hand if $w_0 \in \mathfrak{h}_\sigma$ is close to zero then

$$h = I + w_0 + \mathrm{O}(|w_0|_\sigma^2) \tag{9.12}$$

for some $h \in H_\sigma$. In fact, if h_0 is given, then we can define w as a power series,

$$w = \log h_0 = (h_0 - I) - \tfrac{1}{2}(h_0 - I)^2 + \tfrac{1}{3}(h_0 - I)^3 - \cdots,$$

and if w_0 is given then we can define h by the power series

$$h = \exp w_0 = I + w_0 + \tfrac{1}{2!}w_0^2 + \tfrac{1}{3!}w_0^3 + \cdots,$$

both of which are absolutely convergent if $h_0 - I$ (resp. w_0) are sufficiently small.

We now show that (9.11) and (9.12) together imply the injectivity of the map in (9.10), and that

$$\overline{B_\varepsilon^{H_\sigma}} \cdot \underbrace{(\overline{B_\eta^{G_\sigma}} \cap (I + V_\sigma))}_{T_\sigma} \supseteq B_\delta^{G_\sigma}$$

for some $\delta > 0$. Here we may assume for convenience that $\sigma \neq \infty$ as the case $\sigma = \infty$ was discussed earlier.

Suppose therefore that $h_1, h_2 \in B_\varepsilon^{H_\sigma}$ and $t_1, t_2 \in T_\sigma$ with $h_1 \cdot t_1 = h_2 \cdot t_2$. Set $h = h_2^{-1} h_1 \in \overline{B_{2\varepsilon}^{H_\sigma}}$ and $t_i = I + v_i$ with $v_i \in V_\sigma$ for $i = 1, 2$. Then by (9.12) we have

$$I + w + \mathrm{O}(|w|_\sigma^2) = h = (I + v_2)^{-1}(I + v_2 + v_1 - v_2) = \\ I + (I + v_2)^{-1}(v_1 - v_2) = I + v$$

Notice that we have $w \in \mathfrak{h}_0$ and that the the vector v on the right is in a complementary linear subspace (which is close to the subspace V_σ since v_2 is close to zero). We may use this to finish the proof in the following way. Let ϕ be the linear map from $\mathfrak{gl}_r(\mathbb{Q}_\sigma)$ to V_σ with the kernel $\mathfrak{h}_\sigma$ and whose restriction to V is the identity. Then for sufficiently small v_2 we have that $|v| \ll |\phi(v)| = |\phi(w + \mathrm{O}(|w|_\sigma^2))| = \mathrm{O}(|w|_\sigma^2)$. This implies now with the above identity that $w = \mathrm{O}(|w|^2)$. Hence if we work with sufficiently small elements we see that $w = 0$ and also $v = 0$ which gives $v_2 = v_1$, $h = I$ and $h_1 = h_2$ as required to show injectivity.

Assume now that $g \in B_\delta^{G_\sigma}$ for some sufficiently small $\delta > 0$. Then we can write

$$g = I + w_0 + v_0$$

with $w_0 \in \mathfrak{h}_\sigma$, $v_0 \in V_\sigma$ of norm not exceeding $c\delta$ for some absolute constant c. We may now apply (9.12) to $-w_0$ to find some h_0, and for this h_0 we have

$$\begin{aligned} gh_0 &= (I + w_0 + v_0)\left(I - w_0 + \mathrm{O}(|w_0|_\sigma^2\right) \\ &= I + v_0 + \mathrm{O}\left(\max\{|w_0|_\sigma^2, |w_0|_\sigma |v_0|_\sigma\}\right) \\ &= I + w_1 = v_1, \end{aligned}$$

with $w_1 \in \mathfrak{h}_\sigma$ of norm $\mathrm{O}\left(\delta^2\right)$ and $v_1 \in \mathfrak{h}_\sigma$ satisfying $|v_1 - v_0|_\sigma = \mathrm{O}\left(\delta^2\right)$. Now apply (9.12) to $v_0 - w_1$ to find some h_1 with

$$\begin{aligned} gh_0h_1 &= I + v_1 + \mathrm{O}\left(\max\{|w_1|_\sigma^2, |w_1|_\sigma |v_1|_\sigma\}\right) \\ &= I + w_2 + v_2 \end{aligned}$$

with $w_2 \in \mathfrak{h}_\sigma$ of norm $\mathrm{O}\left(\delta^3\right)$ and $v_2 \in \mathfrak{h}_\sigma$ with $|v_2 - v_1|_\sigma = \mathrm{O}\left(\delta^3\right)$. We wish to iterate this, so we have to take care of the implicit constant in the O expressions. However, as we have assumed that the place σ is finite, the corresponding norm is non-Archimedean. Analyzing the argument above again gives $|w_1|_\sigma \leqslant (c\delta)^2$, $|v_1 - v_0|_\sigma \leqslant (c\delta)^2$, and (assuming that $c\delta < 1$) $|v_1|_\sigma \leqslant c\delta$, $|w_2|_\sigma \leqslant (c\delta)^3$, and $|v_2 - v_1|_\sigma \leqslant (c\delta)^3$. The same argument produces $h_2, h_3, \dots, h_n$ in H_σ for which

$$gh_0h_1 \cdots h_n \to I + v$$

as $n \to \infty$ for some $v \in V_\sigma$. Moreover, if δ is sufficiently small (how small depending on ε) then the elements h_n belong to some small compact subgroup of H_σ, so that we obtain

$$h_0h_1 \cdots h_n \to h \in \overline{B_\varepsilon^{H_\sigma}}.$$

This implies, for sufficiently small $\delta > 0$ (depending on η), that $gh \in T_\sigma$ and so $g \in \overline{B_\varepsilon^{H_\sigma}} \cdot T_\sigma$. Thus the map in (9.10) is a homeomorphism onto a neighborhood of $I \in G_S$. Arguing now as we did in finishing the proof of Proposition 9.13 on p. 287 completes the proof of Proposition 9.13 in the case of S-algebraic groups. □

Proof of Proposition 9.13 for S-algebraic groups of positive characteristic. As mentioned in the argument above, Lemma 14.8 proves the two claims (9.11) and (9.12). These claims were shown on p. 289 using the assumption of zero characteristic, but the remainder of the argument (starting on p. 288) did not use this assumption. Thus Proposition 9.13 also follows in the case where G and H are defined as S-algebraic groups of positive characteristic. □

9.2.6 Proof of the corollaries

We now proceed to the proof of Corollary 9.9, for which we will need to combine Proposition 9.13 with the additional assumption of the corollary.

PROOF OF COROLLARY 9.9. Let $\mathscr{A}$ be an H-subordinate σ-algebra on Y, and let $\mathsf{x}_0 \in \mathsf{X}$ be chosen so that $H \ni h \mapsto h{\boldsymbol\cdot}\mathsf{x}_0$ is injective. Let $R \geqslant 1$ be arbitrary. Then there exists a $\delta > 0$ as in Proposition 9.13, and we let $Y_1 = B_R^H{\boldsymbol\cdot}T$ and write $\mathscr{A}(\mathsf{x}_0, R, \delta)$ for the σ-algebra on Y_1 as in that proposition. We will now compare, for each $y \in Y_1$, the leafwise measures μ_y^H with the conditional measures $(\mu|_{Y_1})_y^{\mathscr{A}(\mathsf{x}_0,R,\delta)}$ and then (for $y \in Y \cap Y_1$) the latter to $(\mu|_Y)_y^{\mathscr{A}}$.

Assume that $y = h{\boldsymbol\cdot}t \in Y_1$ with $h \in B_R^H$ and $t \in T$. We claim that

$$\mu_y^{\mathscr{A}(\mathsf{x}_0,R,\delta)} = \frac{1}{\mu_y^H(B_R^H h^{-1})} \left(\mu_y^H|_{B_R^H h^{-1}}{\boldsymbol\cdot}y\right).$$

almost surely. For this, notice that $(Y_1, \mathscr{A}(\mathsf{x}_0, R, \delta), \mu|_{Y_1})$ is isomorphic to

$$(B_R^H{\boldsymbol\cdot}T_G, \mathscr{B}_{G/H}, \mu_G|_{B_R^H{\boldsymbol\cdot}T_G}),$$

where $T_G \subseteq G$ is bounded and has the property that the natural projections $\pi|_{T_G}$ to T and $\pi|_{B_R^H{\boldsymbol\cdot}T_G}$ to $B_R^H{\boldsymbol\cdot}T$ are bi-measurable†. Therefore, if $y = h{\boldsymbol\cdot}t \in Y_1$ and $t_G \in T_G$ with $\pi(t_G) = t$, then by the argument above and Proposition 9.2(1), we have

$$\mu_{h{\boldsymbol\cdot}t}^{\mathscr{A}(\mathsf{x}_0,R,\delta)} = \pi_*\Big(\Big(\mu_G|_{B_R^H{\boldsymbol\cdot}T_G}\Big)_{h{\boldsymbol\cdot}t_G}^{\mathscr{B}_{G/H}}\Big) \propto \pi_*\Big((\mu_G)_{h{\boldsymbol\cdot}t_G}^{\mathscr{B}_{G/H}}|_{B_R^H{\boldsymbol\cdot}T_G}\Big)$$

almost surely. By the construction of $\mu_{h{\boldsymbol\cdot}t}^H$ from Theorem 9.6(5), we have moreover

$$(\mu_G)_{h{\boldsymbol\cdot}t_G}^{\mathscr{B}_{G/H}} \propto \mu_{h{\boldsymbol\cdot}t}^H{\boldsymbol\cdot}(h{\boldsymbol\cdot}t_G),$$

which implies for $y = h{\boldsymbol\cdot}t \in Y_1$ that

$$\mu_y^{\mathscr{A}(\mathsf{x}_0,R,\delta)} = \frac{1}{\mu_y^H(B_R^H h^{-1})} \left(\mu_y^H|_{B_R^H h^{-1}}{\boldsymbol\cdot}y\right) \tag{9.13}$$

almost surely.

We now need to describe the relationship between the conditional measures of $\mathscr{A}|_{Y_1}$ and of $\mathscr{A}(\mathsf{x}_0, R, \delta)|_Y$. Define $\mathscr{C} = \mathscr{A}|_{Y_1} \vee \mathscr{A}(\mathsf{x}_0, R, \delta)|_Y \subseteq \mathscr{B}(Y_2)$ for $Y_2 = Y \cap Y_1$. We will need the following lemma.

Lemma 9.15. *Let $(Y, \mathscr{B}, \mu)$ be a finite measure space with Y a Borel subset of a compact metric space $\overline{Y}$. Let $\mathscr{A} \subseteq \mathscr{C} \subseteq \mathscr{B}$ be countably-generated σ-algebras. Suppose that for any $x \in Y$ the $\mathscr{A}$-atom $[x]_{\mathscr{A}}$ is a countable union*

† Strictly speaking we should write $\mathscr{B}_{G/H}|_{B_R^T{\boldsymbol\cdot}T_G}$ instead of just $\mathscr{B}_{G/H}$, but the context should make this clear.

of $\mathscr{C}$-atoms $[y]_{\mathscr{C}}$ (with $y \in [x]_{\mathscr{A}}$). Then

$$\mu_x^{\mathscr{C}} = \frac{1}{\mu_x^{\mathscr{A}}([x]_{\mathscr{C}})} \mu_x^{\mathscr{A}}|_{[x]_{\mathscr{C}}}$$

for μ-almost every $x \in Y$.

Proof. This is a special case of the formula $\left(\mu_x^{\mathscr{A}}\right)_x^{\mathscr{C}} = \mu_x^{\mathscr{C}}$ for $\mathscr{A} \subseteq \mathscr{C} \subseteq \mathscr{B}$ proved in [**?**, Prop. 5.20]. Alternatively, one can check the claim directly (just as in the proof of Lemma 9.1) in the case where $\mathscr{C}$ coincides with the refinement of $\mathscr{A}$ with a finite partition. Then one can choose a countable generating set $\{C_1, C_2, \dots\}$ for $\mathscr{C}$ and define

$$\mathscr{C}_n = \mathscr{A} \vee \sigma(\{C_1, \dots, C_n\}).$$

By induction, the lemma holds for any $\mathscr{C}_n$. Hence if $f \in \mathscr{L}^\infty$, then

$$E(f \big| \mathscr{C}_n)(x) = \frac{1}{\mu_x^{\mathscr{A}}([x]_{\mathscr{C}_n})} \int_{[x]_{\mathscr{C}_n}} f(y) \,\mathrm{d}\mu_x^{\mathscr{A}}(y)$$

converges to $E(f \big| \mathscr{C})(x)$ by the increasing martingale theorem[†], respectively to

$$\frac{1}{\mu_x^{\mathscr{A}}([x]_{\mathscr{C}})} \int_{[x]_{\mathscr{C}}} f(y) \,\mathrm{d}\mu_x^{\mathscr{A}}(y)$$

if $\mu_x^{\mathscr{A}}([x]_{\mathscr{C}}) \neq 0$. We claim that

$$N = \{x \mid \mu_x^{\mathscr{A}}([x]_{\mathscr{C}}) = 0\}$$

is a null set. To see that N is measurable, notice that

$$\mu_x^{\mathscr{A}}([x]_{\mathscr{C}}) = \lim_{n\to\infty} \mu_x^{\mathscr{A}}([x]_{\mathscr{C}_n})$$

is a limit of measurable functions. Now

$$\mu(N) = \int_Y \mu_y^{\mathscr{A}}(N) \,\mathrm{d}\mu(y)$$

and recall that, by assumption, $[x]_{\mathscr{A}} = \bigcup_{y \in [x]_{\mathscr{A}}} [y]_{\mathscr{C}}$ is a countable union. This implies that $\mu_y^{\mathscr{A}}(N) = 0$ for all $y \in Y$, and so $\mu(N) = 0$. Using a countable dense set of functions in $C(\overline{Y})$, the lemma follows. □

We return to the proof of Corollary 9.9. We wish to apply Lemma 9.15 to $\mathscr{C}$ and to $\mathscr{A}|_{Y_1}$. Thus we need to show that the $\mathscr{A}|_{Y_1}$-atoms are countable unions of $\mathscr{C}$-atoms. By assumption, every $\mathscr{A}$-atom $[y]_{\mathscr{A}}$ for $y \in Y$ is an open H-plaque, that is a set of the form $[y]_{\mathscr{A}} = V_y \cdot y$ for $V_y \subseteq H$ an open

[†] See [**?**, Th. 5.5]; this generalizes immediately from the case of a probability space to a finite measure space (see also the footnote on p. 269).

neighborhood of $I \in H$. If $y \in Y$, then the set $\{h \in H \mid h \centerdot y \in Y_1\}$ is also open. Hence $[y]_{\mathscr{A}|_{Y_1}}$ are open H-plaques as well. Similarly, it follows that the atoms of $\mathscr{A}(\mathsf{x}_0, R, \delta)|_Y$ and of $\mathscr{C}$ are open H-plaques. However, since by assumption H is second-countable, we see that $[\mathsf{x}]_{\mathscr{A}|_{Y_1}}$ (resp. $[\mathsf{x}]_{\mathscr{A}(\mathsf{x}_0,R,\delta)|_Y}$) must be a countable union of $\mathscr{C}$-atoms. Therefore by Lemmas 9.1 and 9.15 applied twice, we see that for μ-almost every $\mathsf{x} \in Y_2$,

$$\mu_{\mathsf{x}}^{\mathscr{A}}|_{[\mathsf{x}]_{\mathscr{C}}} \propto \mu_{\mathsf{x}}^{\mathscr{A}|_{Y_1}}|_{[\mathsf{x}]_{\mathscr{C}}} \propto \mu_{\mathsf{x}}^{\mathscr{C}} \propto \mu_{\mathsf{x}}^{\mathscr{A}(\mathsf{x}_0,R,\delta)|_Y}|_{[\mathsf{x}]_{\mathscr{C}}} \propto \mu_{\mathsf{x}}^{\mathscr{A}(\mathsf{x}_0,R,\delta)}|_{[\mathsf{x}]_{\mathscr{C}}}.$$

Now let $\mathsf{x} \in Y \cap B_\delta^{\mathsf{X}}(\mathsf{x}_0)$ and $V_{\mathsf{x}} \subseteq H$ be chosen with $[\mathsf{x}]_{\mathscr{A}} = V_{\mathsf{x}} \centerdot \mathsf{x}$. Then

$$B_{R-1}^H \centerdot \mathsf{x} \subseteq [\mathsf{x}]_{\mathscr{A}(\mathsf{x}_0,R,\delta)} \subseteq Y_1$$

and hence $\left(V_{\mathsf{x}} \cap B_{R-1}^H\right) \centerdot \mathsf{x} \subseteq [\mathsf{x}]_{\mathscr{C}} = [\mathsf{x}]_{\mathscr{A}} \cap [\mathsf{x}]_{\mathscr{A}(\mathsf{x}_0,R,\delta)}$.

Together with (9.13) this shows that

$$\mu_{\mathsf{x}}^{\mathscr{A}}|_{B_{R-1}^H \centerdot \mathsf{x}} \propto \mu_{\mathsf{x}}^{\mathscr{A}(\mathsf{x}_0,R,\delta)}|_{(V_{\mathsf{x}} \cap B_{R-1}^H)\centerdot \mathsf{x}} \propto \mu_{\mathsf{x}}^H|_{V_{\mathsf{x}} \cap B_{R-1}^H} \centerdot \mathsf{x}$$

for μ-almost every $\mathsf{x} \in Y \cap B_\delta(\mathsf{x}_0)$. If $V_{\mathsf{x}} \subseteq B_{R-1}^H$ then the statement above is precisely (9.6) from Corollary 9.9.

It remains to show that, for every $R \geqslant 1$ there exists a countable collection of (H,R)-flowers with base $B_{\delta_n}(\mathsf{x}_n)$ such that

$$\mu\left(Y \cap \bigcup_{n=1}^{\infty} B_{\delta_n}(\mathsf{x}_n)\right) = \mu(Y).$$

Then the argument above gives the corollary. □

Proposition 9.16. *Under the assumptions of Corollary* 9.9 *there exists (for every $R \geqslant 1$) a countable collection $\{\mathsf{x}_n\} \subseteq \mathsf{X}$ and a sequence (δ_n) in $(0,1]$ for which the (H,R)-flowers $\mathscr{A}(\mathsf{x}_n, R, \delta_n)$ with base $B_{\delta_n}(\mathsf{x}_n)$ exist and satisfy*

$$\mu\left(Y \cap \bigcup_{n=1}^{\infty} B_{\delta_n}^X(\mathsf{x}_n)\right) = \mu(Y).$$

Proof. By assumption, μ-almost every $\mathsf{x} \in \mathsf{X}$ has the property that the map $\theta_{\mathsf{x}} : H \to H \centerdot \mathsf{x}$ is injective. Let

$$O = \bigcup_{(\mathsf{x},\delta_{\mathsf{x}})} B_{\delta_{\mathsf{x}}}(\mathsf{x}),$$

where the union is over all δ_{x}-balls around points $\mathsf{x} \in Y$ for which the σ-algebra $\mathscr{A}(\mathsf{x}, R, \delta_{\mathsf{x}})$ exists by Proposition 9.13. However, as X is second countable, Y is also, and hence O can be written as a countable union $B_{\delta_{\mathsf{x}_n}}(\mathsf{x}_n)$ of a subcollection of these balls. Therefore $\mu(O \cap Y) = \mu(Y)$. □

9.2.7 Growth Rate of the Leafwise Measures†

We have seen and used in several settings the properties of the space $\mathscr{M}(X)$ of Borel probability measures on a compact metric space X, together with the weak*-topology. This topology is given by a metric (see [**?**, Th. B.11] for the details) and in this topology $\mathscr{M}(X)$ is compact. If X is a compact metric space and μ is a Borel probability measure on the Borel σ-algebra $\mathscr{B}_X$, then for any sub-σ-algebra $\mathscr{A}$ we may construct the conditional measures $\mu_{\mathsf{x}}^{\mathscr{A}}$ for almost every $\mathsf{x} \in X$ (see [**?**, Sec. 5.3]). In this case we may understand something of the geometry of these constructed measures by noting that they all lie in the compact metric space $\mathscr{M}(X)$. This allows us to ask if, for example, the map $\mathsf{x} \mapsto \mu_{\mathsf{x}}^{\mathscr{A}}$ has regularity properties, is measurable, and so on.

It would be useful to find a similar compact metric space in which the leafwise measures constructed in Section 9.2 live. The discussion below is taken from the lecture notes of Einsiedler and Lindenstrauss [**?**, Para. 6.29-30], where the proof of the theorem below is given.

Recall that we assume that $\mu_{\mathsf{x}}^{H}(B_1^H) = 1$ almost surely. Unfortunately this conceals any of the real geometric properties of the leafwise measures. For example, this normalization does not give any insight into the size of larger metric balls B_n^H with respect to the measure μ_{x}^{H}. We also have to confront the fact that the measures being constructed are locally finite rather than finite. In particular, because we do not know a bound on the size of $\mu_{\mathsf{x}}^{H}(B_2^H)$, after the normalization above, the leafwise measures do not belong to a compact subset of the space of Radon[(31)] measures, which may be topologized using the weak*-topology induced by the functions in $C_c(H)$.

An alternative approach is therefore to measure the growth in $\mu_{\mathsf{x}}^{H}(B_n^H)$, and use this to make an adapted normalization resulting in values in a compact metric space. We will not prove or use this theorem.

Theorem. Assume, in addition to the assumptions of Theorem 9.6 and Corollary 9.9, that μ is a Borel probability measure on X, and that the group H is unimodular with bi-invariant Haar measure λ_H. Fix a sequence of positive weights (b_n) with $\sum_{n=1}^{\infty} b_n^{-1} < \infty$, and an increasing sequence (r_n) with $r_n \to \infty$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \frac{\mu_{\mathsf{x}}^{H}\left(B_{r_n}^H\right)}{b_n \lambda_H\left(B_{r_n+5}^H\right)} = 0$$

for μ-almost every $\mathsf{x} \in X$.

That is, the leafwise measure of a sequence of large metric open balls $B_{r_n}^H$ cannot grow asymptotically much faster than the Haar measure of a sequence

† The reader may skip this section as the approach will not be used in the book.

of slightly larger metric open balls. This allows us to define[†] — in many different ways — a function

$$f : H \to \mathbb{R}_{>0}$$

which is integrable with respect to the measure μ_x^H for almost every $x \in X$. In particular, we can normalize μ_x^H so as to guarantee that

$$\int_H f(h)\,\mathrm{d}\mu_x^H(h) = 1,$$

and with this normalization all of the leafwise measures μ_x^H lie in the compact metric space of Radon measures on H with $\int_H f\,\mathrm{d}\nu \leqslant 1$, equipped with the weak*-topology induced by the space of continuous functions with compact support. With this understanding, it makes sense to ask if $x \longmapsto \mu_x^H$ is measurable, and this follows from the measurability referred to in Theorem 9.6(3).

We will not need the theorem above, and refer to the notes of Einsiedler and Lindenstrauss [**?**] for the proof. Instead we will use the set-up of the next subsection to interpret the measurability of the map $x \mapsto \mu_x^H$.

9.2.8 Measurability of Leafwise Measures as a Map

true ? where is this used first?

[‡] Just as in Section 9.2.7, our aim is to provide a compact metric space $\mathscr{M}$ so that we may interpret the leafwise measures μ_x^H as elements of $\mathscr{M}$, and thereby interpret Theorem 9.6(3) by saying that the map from X to $\mathscr{M}$ is measurable with respect to the Borel σ-algebras. In this section we explain a less informative but easier approach to this.

Recall that for any compact metric space Y the weak*-topology on $\mathscr{M}(Y)$ makes $\mathscr{M}(Y)$ into a compact metric space (see [**?**, Sec. B.5]). It follows that

$$\mathscr{M} = \prod_{n=1}^{\infty} \mathscr{M}(Q_n) \tag{9.14}$$

is also a compact metric space, whenever (Q_n) is a sequence of compact subsets of H. Below we will also assume that $I \in Q_n^o$ for all n, and

$$H = \bigcup_{n=1}^{\infty} Q_n^o.$$

[†] For example, we could define $f(x) = 1/\left(b_n^2 \lambda_H\left(B_{r_n+5}^T\right)\right)$ for $x \in B_{r_n}^T \smallsetminus B_{r_{n-1}}^T$, for each $n \geqslant 1$.

[‡] The reader may postpone reading this section, as it will be used first in Chapter 10.

For example, we could use $Q_n = \overline{B_n^H}$ for all $n \geqslant 1$.

Corollary 9.17. *Under the assumptions of Theorem* 9.6 *we have the following.*

(1) *The map*

$$X' \ni x \longmapsto \left(\frac{\mu_x^H|_{Q_1}}{\mu_x^H(Q_1)}, \dots, \frac{\mu_x^H|_{Q_n}}{\mu_x^H(Q_n)}, \dots \right) \tag{9.15}$$

is measurable as a map from X' to the compact metric space $\mathscr{M}$ defined in (9.14).

(2) *(Lusin's theorem for leafwise measures) For any $\varepsilon > 0$ there exists a compact set $K \subseteq X$ of measure $\mu(K) > 1 - \varepsilon$ such that the map in* (9.15) *is continuous when restricted to K.*

Proof. We start with the second statement. Let $n \geqslant 1$ and $f \in C(Q_n)$. Then

$$x \mapsto \frac{1}{\mu_x^H(Q_n)} \int f \mathbb{1}_{Q_n} \, \mathrm{d}\mu_x^H \tag{9.16}$$

defines a function $X' \to \mathbb{R}$ which depends measurably on $x \in X'$. Thus by Lusin's theorem (see [**?**, Th. A.20]) for every $\varepsilon' > 0$ there exists a compact set $K = K(\varepsilon') \subseteq X'$ of measure at least $1 - \varepsilon'$ for which the restriction of the map in (9.16) depends continuously on $x \in K$.

If now $D_n = \{f_{n,k} \mid k \geqslant 1\} \subseteq C(Q_n)$ is a dense countable set, then one can apply the argument above for any $f = f_{n,k}$ with $\varepsilon' = \frac{\varepsilon}{2^{n+k}}$ to obtain sets $K_{n,k}$. Let

$$K = \bigcap_{n,k=1}^{\infty} K_{n,k},$$

so that $\mu(K) > 1-\varepsilon$. Then the function (9.16) depends continuously on $x \in K$ for any $f_{n,k} \in D_n$. However, notice that (9.16) for $f_{n,k}$ converges uniformly to (9.16) for f whenever f_{n,k_i} converges uniformly to $f \in C(Q_n)$. Hence we conclude that (9.16) is continuous for any $f \in C(Q_n)$ when restricted to K. As this holds for any $n \geqslant 1$, we see that the pre-image of any set of the form

$$O = \left\{ (\nu_n) \in \mathscr{M} \mid \left| \int f \, \mathrm{d}\nu_n - r \right| < \varepsilon \right\} \tag{9.17}$$

for $f \in C(Q_n)$, $r \in \mathbb{R}$ and $\varepsilon > 0$ intersected with K is open. The sets of the shape (9.17) generate the topology of $\mathscr{M}$ by taking finite intersections and countable unions, so property (2) follows by taking unions and intersections of such sets.

For property (1), it is sufficient to consider countable unions of intersections of sets of the form (9.17) since $\mathscr{M}$ is compact and metric, and hence second countable. Thus it is enough to show that the pre-image of O under the map in (9.15) is measurable. This pre-image is given by

$$\left\{ \mathsf{x} \in X' \mid \left| \frac{1}{\mu_{\mathsf{x}}^{H}(Q_n)} \int f \mathbb{1}_{Q_n} \, \mathrm{d}\mu_{\mathsf{x}}^{H} - r \right| < \varepsilon \right\},$$

which is measurable by Theorem 9.6(3). □

Exercises for Section 9.2

Exercise 9.2.1. Let H be a σ-compact, locally compact, metric, group acting continuously on a σ-compact locally compact metric space X. Assume that there is an H-invariant ergodic probability measure on X which does not give measure one to any single H-orbit. Show that there is no countably-generated σ-algebra $\mathscr{A}$ with the property that the $\mathscr{A}$-atoms are the H-orbits.

Exercise 9.2.2. Let $G = \prod_{\sigma \in S} G_\sigma$ be a product of real (corresponding as usual to $\sigma = \infty$) and p-adic Lie groups[†]. Let H be a closed subgroup of G. Show that there exists an open subgroup $H_0 \leqslant H$ which is a product $\prod_{\sigma \in S} H_\sigma$ of closed subgroups $H_\sigma \leqslant G_\sigma$. Prove Proposition 9.13 (once again with a σ-compact T) in this case.

Exercise 9.2.3. Let $X = \Gamma\backslash G$, μ, and H be as in the statement of Theorem 9.6. For a measurable set $B \subseteq X$, show that

$$(\mu|_B)_{\mathsf{x}}^{H} \propto \mu_{\mathsf{x}}^{H}|_{\{h \in H \mid h \boldsymbol{\cdot} \mathsf{x} \in B\}}.$$

Exercise 9.2.4. Let G, Γ, X and H be as in Example 9.3, and let μ be the one-dimensional Lebesgue measure on $\mathbb{T} \times \{0\}$. Describe the leafwise measures μ_{x}^{H} for $\mathsf{x} \in \mathbb{T}^2$.

9.3 Characterizing Properties of μ Along H-Orbits

We continue working under the assumptions of Theorem 9.6 and, where appropriate, Corollary 9.9.

9.3.1 Triviality and Alignment

Definition 9.18 (H-trivial measures). Let $H \leqslant G$ be a closed subgroup. A locally finite measure μ on $\mathsf{X} = \Gamma\backslash\mathsf{G}$ is called *H-trivial* if there is a measurable set $X' \subseteq \mathsf{X}$ of full measure for which $\mathsf{x}, h\boldsymbol{\cdot}\mathsf{x} \in X'$ for some $h \in H$ implies that $h = I$, the identity of H.

Definition 9.19 (Trivial conditional measures). The leafwise measures μ_{x}^{H} associated to a locally finite measure μ on X and a closed subgroup $H \leqslant G$

[†] We define a p-adic Lie group as a topological group which has an open subgroup that is isomorphic as a topological group to a closed linear subgroup of $\mathrm{SL}_n(\mathbb{Q}_p)$ for some $n \geqslant 1$.

are said to be *trivial* if $\mu_{\mathsf{x}}^H = \delta_I$, the point mass at the identity, for almost every $\mathsf{x} \in \mathsf{X}$.

Lemma 9.20 (Characterizing H-trivial measures). *A locally finite measure μ on X is H-trivial if and only if the associated leafwise measures are trivial.*

PROOF. Suppose the leafwise measures are trivial, and let $X' \subseteq \mathsf{X}$ be a set of full measure such that $\mu_{\mathsf{x}}^H = \delta_I$ for $\mathsf{x} \in X'$ and on which Theorem 9.6(4) holds. Then, if $\mathsf{x}_1, \mathsf{x}_2 \in X'$ satisfy $\mathsf{x}_2 = h{\bullet}\mathsf{x}_1$, we have

$$\delta_I = \mu_{\mathsf{x}_2}^H \propto \mu_{\mathsf{x}_1}^H h^{-1} = \delta_I h^{-1},$$

so $h = I$ and μ is H-trivial.

If μ is H-trivial, let X' be the set in Definition 9.18 so that $N = X \smallsetminus X'$ is a null set. Then by Corollary 9.10 we have $\mu_{\mathsf{x}}^H(H \smallsetminus \{I\}) = 0$ for almost every x, so $\mu_{\mathsf{x}}^H = \delta_I$ almost everywhere. □

As we have seen the following property is useful in the proof of the measure classification for unipotent flows in Chapter 6.

Definition 9.21 (Aligned measures). Let $L \leqslant H \leqslant G$ be closed subgroups. A locally finite measure μ on X is called (H, L)-*aligned* if there is a measurable set $X' \subseteq \mathsf{X}$ of full measure for which $\mathsf{x}, h{\bullet}\mathsf{x} \in X'$ for some $h \in H$ implies that $h \in L$.

Lemma 9.22 (Characterizing (H, L)-aligned measures). *Let $L \leqslant H \leqslant G$ be closed subgroups. A locally finite measure μ on X is (H, L)-aligned if and only if the leafwise measure for H satisfy $\mu_x^H(L \smallsetminus H) = 0$ for μ-a.e. $x \in X$.*

The proof of Lemma 9.20 extends to also give a proof of the above (see Exercise 9.3.1).

9.3.2 Recurrence and Transience

Another interesting property of μ with respect to the action of H (of intermediate strength) that can be characterized using the leafwise measures is *recurrence*. The results in this section go back to the work of Lindenstrauss on arithmetic quantum unique ergodicity [**?**].

Definition 9.23. Let X be a σ-compact, locally compact metric space, let μ be a locally finite measure on X, and let H be a σ-compact, locally compact metric group that acts continuously on X. Then μ is said to be H-*recurrent* if for every measurable set $B \subseteq X$ of positive measure, and for almost every $x \in B$, the set $\{h \in H \mid h{\bullet}x \in B\}$ is unbounded† in H.

† By assumption on the metric, a subset of H is unbounded if and only if its closure is not compact.

Theorem 9.24. *Let $X = \Gamma\backslash G$ be as in Theorem 9.6, and let μ be a finite measure on X with the property that $H \to H\bullet x$ is injective for μ-almost every x where $H \leqslant G$ is a closed subgroup. Then μ is recurrent with respect to the H-action on X if and only if $\mu_x^H(H) = \infty$ for μ-almost every $x \in X$.*

Proof. Assume that μ is recurrent with respect to the right H-action. Let $Y = \{x : \mu_x^H(H) < \infty\}$, and suppose that $\mu(Y) > 0$. Then, if n is sufficiently large, we may assume that the set

$$Y' = \{x \in Y : \mu_x^H(B_n^H) > \tfrac{9}{10}\mu_x^H(H)\}$$

also has $\mu(Y') > 0$. We claim that for $y \in Y'$, the set of return times to Y',

$$\{h \mid h\bullet y \in Y'\},$$

is almost surely a subset of $\subseteq B_{2n}^H$. Since $\mu(Y') > 0$, this then shows that μ is not H-recurrent. To see the claim, pick any return time h. By the definition of the set Y', we know that $\mu_y^H(B_n^H) > \frac{9}{10}\mu_y^H(H)$ and $\mu_{h\bullet y}^H(B_n^H) > \frac{9}{10}\mu_{h\bullet y}^H(H)$. On the other hand, by Theorem 9.6(4) we know that $\mu_{h\bullet y}^H h \propto \mu_y^H$ almost surely, so

$$\mu_y^H(B_n^H h) > \tfrac{9}{10}\mu_y^H(Hh) = \tfrac{9}{10}\mu_y^H(H).$$

Thus the sets B_n^H and $B_n^H h$ both contain more than $\frac{9}{10}$ of the μ_y^H measure, so $B_n^H \cap B_n^H h \neq \varnothing$. Thus $h \in (B_n^H)^{-1}B_n^H$, as claimed.

Now assume (for the purposes of a contradiction) that the leafwise measures satisfy $\mu_x^H(H) = \infty$ for μ-almost every $x \in X$, but that μ is not H-recurrent. Then there exists some measurable set $B \subseteq X$ of positive measure for which it is not true that $\{h \in H \mid h\bullet x \in B\}$ is unbounded for almost every x. We will replace the set B by subsets of B through the proof, always retaining this property and positive measure, by adding additional properties.

If (K_n) with $K_n \subseteq B$ is a sequence of compact sets with

$$\mu\left(B \smallsetminus \bigcup_{n\geqslant 1} K_n\right) = 0,$$

then it follows that there is some compact set $K = K_n \subseteq B$ for which it is also not true that $\{h \in H \mid h\bullet x \in K\}$ is unbounded for almost every $x \in K$.

Given any $R \in \mathbb{N}$, the set

$$L_R = K \smallsetminus \bigl((H \smallsetminus B_R^H)\bullet K\bigr)$$

is measurable (since K is compact and $H \smallsetminus B_R^H$ is σ-compact) and comprises the set of points $x \in K$ for which $\{h \in H \mid h\bullet x \in K\} \subseteq B_R^H$. By assumption,

$$\mu\left(\bigcup_{R\geqslant 1} L_R\right) > 0$$

and so there exists some $L = L_R$ with $\mu(L) > 0$ and $\{h \in H \mid h{\scriptstyle\bullet}x \in L\} \subseteq B_R^H$ for all $x \in L$. Applying Proposition 9.16 to $R+2$, we may find some $x_0 \in X$ and some $\delta > 0$ such that $\mu\left(L \cap B_\delta(x_0)\right) > 0$, and the $(H, R+2)$-flower

$$\mathscr{A} = \mathscr{A}(x_0, R+2, \delta)$$

with base $B_\delta(x_0)$ exists. Changing B if necessary, we may now assume that $B \subseteq L \cap B_\delta(x_0)$ is a compact subset with $\mu(B) > 0$.

The construction above shows that $x \in B$, $h{\scriptstyle\bullet}x \in B$ implies $h \in B_R^H$. This is the final set B for which we will show a contradiction of the assumptions, by showing that B is a μ-null set.

Let $T \subseteq X$ be the $(R+2, \delta)$-cross-section for the σ-algebra $\mathscr{A}$, so that the elements $t \in T$ are in one-to-one correspondence with the atoms

$$B_{R+2}^H{\scriptstyle\bullet}t = [t]_{\mathscr{A}}$$

of the σ-algebra $\mathscr{A}$. As $B \subseteq B_1^H{\scriptstyle\bullet}T$, we can define the subset

$$D = T \cap \left(B_1^H{\scriptstyle\bullet}B\right)$$

of all elements of T corresponding to those $\mathscr{A}$-atoms that intersect B non-trivially. This also implies that $B \subseteq B_1^H{\scriptstyle\bullet}D$. Indeed, if $x \in B$, then $x = h{\scriptstyle\bullet}t$ with $h \in B_1^H$ and $t \in T$ so that $h^{-1}{\scriptstyle\bullet}x = t \in D$ and so $x \in B_1^H{\scriptstyle\bullet}D$.

As a first step towards the contradiction we seek, we claim that the map

$$\begin{aligned} H \times D &\longrightarrow H{\scriptstyle\bullet}D \\ (h, d) &\longmapsto h{\scriptstyle\bullet}d \end{aligned}$$

is injective. This is a significant strengthening of the cross-section property in Definition 9.12, since we may use the same set D for all balls $B_n^H \subseteq H$ with $n \geqslant 1$. This property, together with the fact that $\mu_x^H(H) = \infty$ almost everywhere, will imply that $\mu(B) = 0$.

Suppose therefore that $h{\scriptstyle\bullet}z = h'{\scriptstyle\bullet}z'$ for some $h, h' \in H$ and $z, z' \in D$. By the construction of the set D and the inclusion $B \subseteq B_1^H{\scriptstyle\bullet}D$, there exist points $h_x, h_{x'} \in B_1^H$ with $z = h_x{\scriptstyle\bullet}x$, $z' = h_{x'}{\scriptstyle\bullet}x'$ and $x, x' \in B$. Therefore

$$hh_x{\scriptstyle\bullet}x = h'h_{x'}{\scriptstyle\bullet}x',$$

which implies that $(h_{x'})^{-1}(h')^{-1}hh_x \in B_R^H$ by the assumed properties of B. This shows that $(h')^{-1}h \in B_{R+2}^H$, which implies that $h' = h$ and $z = z'$ since $T \supseteq D$ is an $(R+2, \delta)$-cross-section. This is the injectivity claim above.

Thus we can choose any $n > 1$ and, since $B_n^H \times D \to B_n^H{\scriptstyle\bullet}D$ is injective, we always obtain (just as in Proposition 9.13) an H-subordinate σ-algebra $\mathscr{A}_n$

defined on $B_n^H \bullet D$ for which

$$[h \bullet z]_{\mathscr{A}_n} = B_n^H \bullet z$$

for $h \in B_n^H$ and $z \in D$. If T is a countable union of compact sets (which we may assume if G is a Lie group or an S-algebraic group of any characteristic), then so is D, by definition. In this case bi-measurability of the map

$$B_n^H \times D \to B_n^H \bullet D$$

is clear. The general case follows from general properties of standard Borel σ-algebras (see Parthasarathy [**?**]).

Given the σ-algebra $\mathscr{A}_n$ defined on $B_n^H \bullet D$, Corollary 9.9 implies that

$$\mu_x^{\mathscr{A}_n}(B) = \frac{\mu_x^H\left(\{h \in V_x \mid h \bullet x \in B\}\right)}{\mu_x^H(V_x)}$$

for μ-almost every $x \in B_n^H \bullet D$. Here

$$V_{h \bullet d} = B_n^H h^{-1}$$

is the shape of the atom $[h \bullet d]_{\mathscr{A}_n} = B_n^H \bullet d$ viewed from the point $x = h \bullet d$ for $h \in B_n^H, d \in D$. Since $B \subseteq B_1^H \bullet D$, we have $B_{n-1}^H \subseteq V_x$ for all $x \in B$, and so

$$\mu_x^{\mathscr{A}_n}(B) \leqslant \frac{\mu_x^H\left(B_R^H\right)}{\mu_x^H\left(B_{n-1}^H\right)},$$

which approaches zero as $n \to \infty$ for almost every $x \in B$ by our assumption on the leafwise measures μ_x^H. Let

$$B' = \{x \in B \mid \mu_x^{\mathscr{A}_n}(B) \to 0 \text{ as } n \to \infty\},$$

so that $\mu(B \smallsetminus B') = 0$.

We now define, for each $n \geqslant 1$ a function f_n by

$$f_n(y) = \begin{cases} \mu_y^{\mathscr{A}_n}(B) & \text{if } y \in B_n^H \bullet D, \\ 0 & \text{if } y \notin B_n^H \bullet D. \end{cases}$$

We claim that

$$f_n(y) \to 0 \tag{9.18}$$

as $n \to \infty$ for almost every y. Using this claim and dominated convergence (which we may apply since $f_n \leqslant 1$ for all $n \geqslant 1$ and $\mu(X) < \infty$) we obtain

$$\mu(B) = \int f_n \,\mathrm{d}\mu \to 0$$

as $n \to \infty$, giving the desired contradiction.

The claim in (9.18) holds trivially for any $y \notin H\boldsymbol{\cdot}D$. If $y \in B_{n_0}^H\boldsymbol{\cdot}D$ for some $n_0 \geqslant 1$ and $\mu_y^{\mathscr{A}_{n_0}}(B) > 0$, then $\mu_y^{\mathscr{A}_{n_0}}(B') > 0$ almost surely, since $\mu\left(B\smallsetminus B'\right) = 0$. Therefore, there exists some $x \in B'$ with $[x]_{\mathscr{A}_n} = [y]_{\mathscr{A}_n}$, so we also have

$$f_n(y) = f_n(x) = \mu_x^{\mathscr{A}_n}(B)$$

for any $n \geqslant n_0$. Since $\mu_x^{\mathscr{A}_n}(B) \to 0$ as $n \to \infty$ by definition of B', we obtain the claim (9.18). $\square$

A complementary property to recurrence is *transience*, defined below. Notice that transience is not the negation of recurrence (see Exercise 9.3.4.)

Definition 9.25. Let X be a σ-compact, locally compact metric space, let μ be a probability measure on X, and let H be a σ-compact, locally compact metric group that acts continuously on X. Then μ is said to be H*-transient* if for every $\varepsilon > 0$ there exists some set $X' \subseteq X$ of measure $\mu(X') > 1 - \varepsilon$ such that for every $x \in X'$ the set $\{h \in H \mid h\boldsymbol{\cdot}x \in B\}$ is bounded in H.

Just as in the first part of the proof of Theorem 9.24, it is easy to show that $\mu_x^H(H) < \infty$ almost surely implies that μ is H-transient. The converse also holds quite generally (see Exercise 9.3.4).

9.3.3 Invariance

The final property of μ with respect to the action of H that we wish to characterize using the leafwise measures is *invariance*.

Theorem 9.26. *Let $X = \Gamma\backslash G$, μ and H be as in Theorem 9.6. Then μ is H-invariant if and only if μ_x^H is a left Haar measure on H for almost every $x \in X$.*

For the proof the following lemma will be helpful.

Lemma 9.27. *Let $X = \Gamma\backslash G$, with Γ discrete in G. Then the operator*

$$\begin{aligned} S : C_c(G) &\longrightarrow C_c(X) \\ f(\,\cdot\,) &\longmapsto \sum_{\gamma\in\Gamma} f(\gamma\,\cdot\,) \end{aligned}$$

is surjective.

Notice that the operator in Lemma 9.27 appeared implicitly in Lemma 9.4.

Proof of Lemma 9.27. Let $f \in C_c(X)$ and let $r > 0$ be an injectivity radius on $K = \operatorname{Supp}(f)$. Cover K with finitely many metric r-balls, which we write as

$$K \subseteq \bigcup_{i=1}^{n} B_r^X(x_i),$$

and let $\phi_1, \dots, \phi_n \in C_c(X)$ be a continuous partition of unity[†] on K with respect to $B_r^X(x_i)$. Then it is clear that $f \cdot \phi_i$ is in the image of S for $1 \leqslant i \leqslant n$, and therefore f is also. □

Proof of Theorem 9.26. We first reduce the theorem to the case of G. Indeed, we claim that μ is H-invariant if and only if μ_G is H-invariant, and for brevity we write $\nu = \mu_G$. If μ is H-invariant, then for $f \in C_c(G)$ and $h \in H$ we have

$$\begin{aligned}\int f(gh)\,\mathrm{d}\nu &= \int \sum_{\gamma \in \Gamma} f(\gamma g h)\,\mathrm{d}\mu \\ &= \int \sum_{\gamma \in \Gamma} f(\gamma g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\nu,\end{aligned}$$

by the relationship between μ and ν and the assumed invariance of μ. This implies that ν is H-invariant.

Conversely, if ν is H-invariant, then we reverse the argument as follows. If $F \in C_c(X)$ and $h \in H$, then there exists some $f \in C_c(G)$ with $F = S(f)$ by Lemma 9.27. Then

$$\begin{aligned}\int F(xh)\,\mathrm{d}\mu(x) &= \int \sum_{\gamma \in \Gamma} f(\gamma g h)\,\mathrm{d}\mu(\Gamma g) \\ &= \int f(gh)\,\mathrm{d}\nu(g) \\ &= \int f(g)\,\mathrm{d}\nu(g) = \int F\,\mathrm{d}\mu,\end{aligned}$$

which implies the invariance of μ under the action of H.

Now assume that μ_x^H is a left Haar measure on H for μ-almost every x. Then by Theorem 9.6(5) the measure $\nu_{\mathsf{g}}^{\mathscr{B}_{G/H}}$ is a right Haar measure on the coset gH for ν-almost every $g \in G$. If now $B \subseteq G$ is a bounded subset, $h \in H$ and $Y = B \cup Bh$ then, by Proposition 9.2(1),

$$(\nu|_Y)_{\mathsf{g}}^{\mathscr{B}_{G/H}} = \frac{1}{\nu_{\mathsf{g}}^{\mathscr{B}_{G/H}}(Y)} \nu_{\mathsf{g}}^{\mathscr{B}_{G/H}}|_Y,$$

so that

$$(\nu|_Y)_{\mathsf{g}}^{\mathscr{B}_{G/H}}(B) = (\nu|_Y)_{\mathsf{g}}^{\mathscr{B}_{G/H}}(Bh)$$

for almost every $g \in Y$. Therefore,

[†] That is, the functions ϕ_i for $i = 1, \dots, n$ are continuous with $0 \leqslant \phi_i \leqslant 1$ and $\sum_{i=1}^{n} \phi_i(x) = 1$ for all $x \in K$.

$$\begin{aligned}\nu(B) &= \int_Y (\nu|_Y)_{\mathtt{g}}^{\mathscr{B}_{G/H}}(B)\,\mathrm{d}\mu\\ &= \int_Y (\nu|_Y)_{\mathtt{g}}^{\mathscr{B}_{G/H}}(Bh)\,\mathrm{d}\mu = \nu(Bh),\end{aligned}$$

showing the first implication.

Assume now that μ and (hence also) $\nu = \mu_G$ are H-invariant, and let $h \in H$. Then we may apply Proposition 9.2(4) to the measure-preserving map T defined by $T(g) = gh$ to obtain

$$\nu_{\mathtt{g}}^{\mathscr{B}_{G/H}} h = T_* \nu_{\mathtt{g}}^{\mathscr{B}_{G/H}} \propto \nu_{T\mathtt{g}}^{\mathscr{B}_{G/H}} = \nu_{gh}^{\mathscr{B}_{G/H}} \propto \nu_{\mathtt{g}}^{\mathscr{B}_{G/H}}$$

almost surely. We need to show that the proportionality constant relating $\nu_{\mathtt{g}}^{\mathscr{B}_{G/H}} h$ to $\nu_{\mathtt{g}}^{\mathscr{B}_{G/H}}$ is equal to one almost surely. So let

$$B_{>1} = \left\{ g \mid \nu_{\mathtt{g}}^{\mathscr{B}_{G/H}} h = c_g \nu_{\mathtt{g}}^{\mathscr{B}_{G/H}} \text{ for some } c_g > 1 \right\}.$$

Assuming that $\nu(B_{>1}) > 0$, we can find some bounded subset $B \subseteq B_{>1}$ with $\nu(B) > 0$. Let $Y = B \cup Bh^{-1}$. Then every atom $[y]_{\mathscr{B}_{G/H}} = yH$ intersects $B \subseteq B_{>1}$ non-trivially for $y \in Y$, and so $\nu(Y \smallsetminus B_{>1}) = 0$ by Proposition 9.2(2). Therefore,

$$\begin{aligned}\nu(B) &= \int_Y \frac{\nu_y^{\mathscr{B}_{G/H}}(B)}{\nu_y^{\mathscr{B}_{G/H}}(Y)}\,\mathrm{d}\mu && \text{(by Proposition 9.2(1))}\\ &= \int \frac{1}{c_g}\frac{\nu_y^{\mathscr{B}_{G/H}}(Bh^{-1})}{\nu_y^{\mathscr{B}_{G/H}}(Y)}\,\mathrm{d}\mu\\ &< \int \frac{\nu_y^{\mathscr{B}_{G/H}}(Bh^{-1})}{\nu_y^{\mathscr{B}_{G/H}}(Y)}\,\mathrm{d}\mu = \nu(Bh^{-1}).\end{aligned}$$

However, this inequality contradicts the assumed invariance of μ under right-multiplication by h. It follows that $\mu(B_{>1}) = 0$. Applying the same argument to h^{-1} we see that

$$(\mu_G)_{\mathtt{g}}^{\mathscr{B}_{G/H}} h = (\mu_G)_{\mathtt{g}}^{\mathscr{B}_{G/H}} \tag{9.19}$$

for almost every $g \in G$. Choosing a dense countable set $D \subseteq H$ and applying the argument above to each $h \in D$ shows that (9.19) holds for all $h \in D$. Since a locally finite Borel measure on gH which is right-invariant under the action of D must be right-invariant under the action of all of H, it follows that $(\mu_G)_{\mathtt{g}}^{\mathscr{B}_{G/H}}$ is a right Haar measure on gH for almost every $g \in G$. This implies the theorem by Theorem 9.6(5). $\square$

Exercises for Section 9.3

Exercise 9.3.1. Let μ be as in the statement of Theorem 9.6, and let

$$L \leqslant H \leqslant G$$

be closed subgroups. Prove that μ_{x}^{H} is supported on L if and only if μ is (H, L)-aligned for a.e. x.

Exercise 9.3.2. Let X, H, μ be as in Theorem 9.24. Assume that μ is H-recurrent. Prove that for any measurable set $B \subseteq X$ of positive measure we have

$$\mu_x^H\left(\{h \in H \mid h{\bullet}x \in B\}\right) = \infty$$

for μ-almost every $x \in B$.

Exercise 9.3.3. Let $L \leqslant H \leqslant G$ be closed subgroups of a Lie group G, and let $\Gamma \leqslant G$ be a discrete subgroup. Let μ be a locally finite measure on $X = \Gamma\backslash G$. Prove that μ is L-right-invariant if and only if μ_x^H is almost surely invariant under left-multiplication by L.

Exercise 9.3.4. Let $X = \Gamma\backslash G$, μ, and H be as in Theorem 9.24. Prove the following statements.

(a) μ is H-transient if and only if $\mu_x^H(H) < \infty$ for μ-almost every x.
(b) μ is a combination $\mu = \mu_1 + \mu_2$ of finite mutually singular measures such that μ_1 is H-recurrent and μ_2 is H-transient.

9.4 Invariant Foliations in Various Settings

Our primary motivation for studying leafwise measures is to better understand entropy in terms of leafwise measures for the action of certain subgroups. For this purpose, the material in this section will be useful. In Sections 9.4.2–9.4.4 we will introduce various cases within the general framework of $X = \Gamma\backslash G$ which we will allow in our later discussions of entropy. The reader may skip some of these, and restrict attention to the cases of direct interest.

9.4.1 Behavior of Leafwise Measures

We start with a consequence of Theorem 9.6; in the statement of Corollary 9.28 we write $a\mu_x^H a^{-1}$ for the push-forward of μ_x^H under the conjugation by a map $h \mapsto aha^{-1}$.

Corollary 9.28. *Suppose that $X = \Gamma\backslash G$, H, and μ are as in Theorem* 9.6. *Let*

$$a \in N_G(H) = \{g \in G \mid gHg^{-1} = H\}$$

be an element of the normalizer of H, and suppose that $T : x \mapsto xa^{-1}$ for x in X preserves μ. Then

$$\mu^H_{xa^{-1}} \propto a\mu^H_x a^{-1}$$

for μ-almost every $x \in X$.

Notice in particular that if

$$a \in C_G(H) = \{g \in G \mid gh = hg \text{ for all } h \in H\}$$

lies in the centralizer of H, then $\mu^H_{a\cdot x} = \mu^H_x$ for μ-almost every $x \in X$.

PROOF. First notice that the assumed invariance of μ implies that the measure μ_G on G defined as in Lemma 9.4 is also invariant under $T : g \mapsto a\cdot g$. Therefore, Proposition 9.2(4) implies that

$$(\mu_G)^{\mathscr{B}_{G/H}}_g a^{-1} = T_* (\mu_G)^{\mathscr{B}_{G/H}}_g \propto (\mu_G)^{\mathscr{B}_{G/H}}_{ga^{-1}}$$

almost surely, which gives by Theorem 9.6(5) that

$$\left(\mu^H_x \cdot g\right) a^{-1} \propto \mu^H_{a\cdot x} \cdot ga^{-1} = \left(\mu^H_{a\cdot x} a\right) \cdot g$$

for μ-almost every $x \in \Gamma g$. This is equivalent to

$$\left(a\mu^H_x\right) \cdot g \propto \left(\mu^H_{a\cdot x} a\right) \cdot g,$$

which gives the corollary. □

9.4.2 Invariant Foliations for the n-Torus

We describe here how the dynamics of a toral automorphism $T_A : \mathbb{T}^r \to \mathbb{T}^r$ associated to a matrix $A \in \mathrm{GL}_r(\mathbb{Z})$ fits into the framework of Corollary 9.28.

Example 9.29. Let $A \in \mathrm{GL}_r(\mathbb{Z})$ be an invertible integer matrix giving rise to an automorphism $T_A : \mathbb{T}^r \to \mathbb{T}^r$, and let $H \subseteq \mathbb{R}^r$ be an A-invariant subspace. Thus H is a direct sum of eigenspaces or generalized eigenspaces for A restricted to H. In the study of entropy we will always assume that the eigenvalues of A restricted to H have absolute value not equal to one. The sum of a generalized eigenspace with eigenvalues λ such that $|\lambda| < 1$ is called stable. We define below a horospherical subgroup (denoted by G^-_a) whose orbits are called stable manifolds. We define a group G to be the semi-direct product $G = \mathbb{R}^r \rtimes \mathbb{Z}$ with the group operation

$$(v, n)\cdot(w, m) = (v + A^n(w), n + m),$$

and define a discrete subgroup by $\Gamma = \mathbb{Z}^r \rtimes \mathbb{Z} \leqslant G$. This gives for the structure of the quotient space $\Gamma\backslash G$

$$\Gamma(w,m) = \Gamma(0,-m)\bullet(w,m) = \Gamma(A^{-m}w,0)$$

for all $(w,m) \in G$, and hence

$$\Gamma\backslash G \cong \mathbb{Z}^r\backslash\mathbb{R}^r \cong \mathbb{T}^r.$$

Now define $a = (0,1)$. Then for $w \in \mathbb{R}^r$ we have

$$\begin{aligned}\Gamma(w,0)a^{-1} &= \Gamma(w,0)(0,-1)\\ &= \Gamma(w,-1)\\ &= \Gamma(0,1)(w,-1)\\ &= \Gamma(A(w),0),\end{aligned}$$

which shows that right multiplication by a^{-1} corresponds to applying A on $\mathbb{T}^r$. Similarly, $a(w,0)a^{-1} = (A(w),0)$.

Thus in this case Corollary 9.28 means that

$$\mu^H_{T_A x} \propto A\left(\mu^H_x\right)$$

for μ-almost every x.

9.4.3 Automorphisms Arising From Algebraic Numbers

Example 9.29 can be generalized to solenoids. In this section we discuss a particular case of an automorphism of a solenoid constructed from an algebraic number[(32)].

Example 9.30. Let K be a number field, and let $a \in K\smallsetminus\{0\}$. Let $\mathscr{O}_S$ be an order in K localized at finitely many places with the property that $a \in \mathscr{O}_S^\times$ is a unit of $\mathscr{O}_S$. Here $S = S_\infty \cup S_f$, where S_∞ is the (finite) set of Archimedean places and S_f is a finite subset of the set of non-Archimedean places of K. We will show that $X = \widehat{\mathscr{O}_S}$ (or, more generally, $X = \widehat{J}$ for any $\mathscr{O}_S$-ideal J), with T the dual of multiplication by a and H being defined by some subset $S' \subseteq S$ of the places, fits into the framework of Corollary 9.28. In the study of entropy properties, we will always assume that $|a|_\sigma < 1$ for all $\sigma \in S'$.

Let K_σ denote the completion with respect to the norm $|\cdot|_\sigma$ associated to the place $\sigma \in S$. We refer to Weil [?] for background on the valuation theory of algebraic number fields. If the place σ is Archimedean, then either $K_\sigma \cong \mathbb{R}$ or $K_\sigma \cong \mathbb{C}$, and the isomorphism to $\mathbb{R}$ or $\mathbb{C}$ corresponds to one of the Galois embeddings $K \hookrightarrow \mathbb{C}$. If the place σ is non-Archimedean then $K_\sigma|\mathbb{Q}_p$ is a finite field extension of the field of p-adic numbers for some rational prime p, and in this case we say that σ is a place of K above the place p of $\mathbb{Q}$. In either case, the character group $\widehat{K_\sigma}$ of (the additive structure of) K_σ is isomorphic[(33)] as a topological group to K_σ, so the dual of the group K_S, which we define to be $\prod_{\sigma\in S} K_\sigma$, is once again isomorphic to K_S. The order $\mathscr{O}_S$ embeds into K_S via the *diagonal embedding*

$$\mathscr{O}_S \ni \lambda \hookrightarrow (\lambda, \dots, \lambda) \in K_S,$$

and the image, which we identify with $\mathscr{O}_S$, is a uniform lattice in K_S. It follows that

$$X = \widehat{\mathscr{O}_S} \cong K_S/\mathscr{O}_S^{\perp},$$

and that the annihilator $\mathscr{O}_S^{\perp}$ is a uniform lattice as well. We define a map by

$$T(\lambda) = a\lambda$$

for $\lambda \in \mathscr{O}_S$, and

$$T\left((w_\sigma)_{\sigma\in S}\right) = (aw_\sigma)_{\sigma\in S}$$

for $(w_\sigma)_{\sigma\in S} \in \prod_{\sigma\in S} K_\sigma$. The isomorphism $\widehat{K_S} \cong K_S$ can be chosen so that the dual map to T on $\widehat{K_S} \cong K_S$ is again given by T. If $S' \subseteq S$ is any subset, then

$$H = \prod_{\sigma\in S'} K_\sigma$$

can be considered as a subgroup of K_S with $T(H) = H$. Since we are assuming that $a \in \mathscr{O}_S^{\times}$ is a unit, we have $a\mathscr{O}_S = \mathscr{O}_S$ and (by choice of the isomorphisms between $\widehat{K_\sigma}$ and K_σ), $T\left(\mathscr{O}_S^{\perp}\right) = \widehat{\mathscr{O}_S}$. Thus

$$T(x) = T(w + \mathscr{O}_S^{\perp}) = T(w) + \mathscr{O}_S^{\perp}$$

is a well-defined map on $X = K_S/\mathscr{O}_S^{\perp}$, and the leaves of the foliation

$$\{x + H \mid x \in X\}$$

are sent via T to leaves of the same foliation. If $S' = \{\sigma \in S \mid |a|_\sigma < 1\}$ then we will call the corresponding group G_a^- the *stable horospherical subgroup* and its orbits the *stable manifolds*. Using the arguments from Example 9.29 we see once more that Corollary 9.28 also applies to any T-invariant measure μ on X.

We record two special cases of the construction above.

Example 9.31. If $K = \mathbb{Q}$ and $\mathscr{O}_S = \mathbb{Z}[\frac{1}{p} \mid p \in S_f]$ then

$$X = \widehat{\mathscr{O}_S} = \mathbb{R} \times \prod_{p\in S_f} \mathbb{Q}_p/\mathbb{Z}[\tfrac{1}{p} \mid p \in S_f]$$

where the group being quotiented out is embedded diagonally via the map

$$\mathbb{Z}[\tfrac{1}{p} \mid p \in S_f] \ni r \longmapsto (\underbrace{r, r, \dots, r, r}_{1+|S_f| \text{ terms}}) \in \mathbb{R} \times \prod_{p \in S_f} \mathbb{Q}_p$$

For example, if $S_f = \{2, 3\}$ then[(34)]

$$X = \mathbb{R} \times \mathbb{Q}_2 \times \mathbb{Q}_3 / \mathbb{Z}[\tfrac{1}{6}] \tag{9.20}$$

and we can use the constructions above with $H = \mathbb{R}$, $\mathbb{Q}_2$, or $\mathbb{Q}_3$, or we may take for H the subgroup generated by any two of these subgroups, $\mathbb{R} \times \mathbb{Q}_2$, $\mathbb{R} \times \mathbb{Q}_3$, or $\mathbb{Q}_2 \times \mathbb{Q}_3$. Moreover, we can take a to be any element of the (multiplicative) group $2^{\mathbb{Z}}3^{\mathbb{Z}}$. This geometrical structure was already used implicitly in the proof of Rudolph's Theorem in [**?**].

Example 9.32. If $K|\mathbb{Q}$ is a non-trivial finite field extension and $S_f = \varnothing$, then any unit $a \in \mathscr{O}^\times$ of K defines an $\mathbb{R}$-linear map[(35)]

$$A : \prod_{\sigma \in S} K_\sigma \cong \mathbb{R}^r \times \mathbb{C}^s \longrightarrow \mathbb{R}^r \times \mathbb{C}^s,$$

given by multiplication by (the corresponding Galois image of) a in each coordinate. This case represents a special case of Example 9.29, and the various choices of subsets $S' \subseteq S$ correspond (unless $a \in \mathscr{O}^\times$ belongs to a proper subfield of K) to the possible choices of the sums of the eigenspaces of the linear map.

The setup from Example 9.30 can also be extended to allow the function field case, that is to a finite extension $K|\mathbb{F}_p(t)$ for some prime p. We refer to the paper of Einsiedler and Lind [**?**] for the details, and will not pursue this further here.

9.4.4 Horospherical Subgroups in General

Example 9.33. Let $G \leqslant \mathrm{SL}_r(\mathbb{R})$ be a closed real linear group, and let $\Gamma \leqslant G$ be a discrete subgroup. We endow the Lie-algebra $\mathrm{Lie}(G) = \mathfrak{g}$ of G with an inner product, which induces a left-invariant Riemannian metric on G and, via Definition **??**, a metric on $X = \Gamma\backslash G$. Fix $a \in G$ and define[†] the *stable horospherical subgroup*

$$G_a^- = \{g \in G \mid a^n g a^{-n} \to e \text{ as } n \to \infty\} \tag{9.21}$$

and the *unstable horospherical subgroup*

$$G_a^+ = \{g \in G \mid a^n g a^{-n} \to e \text{ as } n \to -\infty\}. \tag{9.22}$$

[†] These subgroups are generalizations of the horocyclic subgroups of $\mathrm{SL}_2(\mathbb{R})$ to higher dimensions, hence the name.

The sign in the exponent of the horospherical groups indicates the sign of $\log|\lambda|$ for any of the eigenvalues λ of Ad_a restricted to the Lie algebra $\mathfrak{g}_a^{\pm}$ of $G_a^{\pm}$.

We claim that $G_a^{\pm}$ are closed subgroups of G. For this, we transform a and G via conjugation by some element of $\mathrm{SL}_r(\mathbb{R})$ into Jordan normal form (allowing 2×2 blocks along the diagonal for complex eigenvalues). We may assume that the eigenvalues $\lambda_1,\dots,\lambda_s$ of a are arranged so that

$$|\lambda_1| \leqslant |\lambda_2| \leqslant \cdots \leqslant |\lambda_s|.$$

We claim that $g=(g_{ij})\in G$ belongs to G_a^- if and only if the matrix entries g_{ij} satisfy $g_{ij}=\delta_{ij}$ for all pairs of indices (i,j), where i (resp. j) corresponds to an eigenvalue λ_{k_i} (resp. λ_{k_j}) with $|\lambda_{k_i}|\geqslant|\lambda_{k_j}|$. This implies that $G_a^{\pm}$ are closed subgroups.

Indeed, assume that

$$a=\begin{pmatrix}\lambda_1 I_1+U_1 & & & \\ & \lambda_2 I_2+U_2 & & \\ & & \ddots & \\ & & & \lambda_s I_s+U_s\end{pmatrix}$$

where $I_1,\dots,I_s$ are identity matrices of the appropriate dimensions (if λ_i is not real, then the corresponding $\lambda_i I_i$ is a block matrix of the appropriate dimension) and $U_1,\dots,U_s$ are upper unipotent matrices. Then we may write g as a block matrix of the same shape,

$$g=\begin{pmatrix} g_{(1,1)} & \cdots & g_{(1,s)}\\ \vdots & & \vdots\\ g_{(s,1)} & \cdots & g_{(s,s)}\end{pmatrix}$$

and so $a^n g a^{-n}$ takes the form

$$\begin{pmatrix}(\lambda_1 I_1+U_1)^n\, g_{(1,1)}\,(\lambda_1 I_1+U_1)^{-n} & \cdots & (\lambda_1 I_1+U_1)^n\, g_{(1,s)}\,(\lambda_s I_s+U_s)^{-n}\\ \vdots & & \vdots\\ (\lambda_s I_s+U_s)^n\, g_{(s,1)}\,(\lambda_1 I_1+U_1)^{-n} & \cdots & (\lambda_s I_s+U_1)^n\, g_{(s,s)}\,(\lambda_s I_s+U_s)^{-n}\end{pmatrix}.$$

Therefore $a^n g a^{-n}$ converges to the identity as $n\to\infty$ if and only if

$$(\lambda_i I_i+U_i)^n\, g_{(i,j)}\,(\lambda_j I_j+U_j)^{-n} \longrightarrow \begin{cases} I_i & \text{if } i=j, \text{ and}\\ 0 & \text{otherwise}\end{cases}$$

as $n\to\infty$.

If $|\lambda_i|=|\lambda_j|$ then

$$(\lambda_i I_i+U_i)^n\, g_{(i,j)}\,(\lambda_j I_j+U_j)^{-n}-\delta_{ij}I_i$$

$$= \left(I_i + \lambda_i^{-1} U_i\right)^n \left(g_{(i,j)} - \delta_{ij} I_i\right) \left(I_j + \lambda_j^{-1} U_j\right)^{-n} \left(\frac{\lambda_i}{\lambda_j}\right)^n,$$

which converges to zero if and only if

$$\left(I_i + \lambda_i^{-1} U_i\right)^n \left(g_{(i,j)} - \delta_{ij} I_i\right) \left(I_j + \lambda_j^{-1} U_j\right)^{-n}$$

converges to zero. However, the latter expression is a polynomial in n, and so the stated convergence can therefore only happen if this polynomial is constant and equal to zero, hence $g_{(i,j)} = \delta_{ij} I_i$.

If $|\lambda_j| > |\lambda_i|$ then the convergence is guaranteed. Finally, if $|\lambda_j| < |\lambda_i|$, the convergence can only happen if $g_{(i,j)} = 0$.

Now consider the adjoint Ad_a, which on $\mathfrak{sl}_r(\mathbb{R})$ has the eigenvalues $\frac{\lambda_i}{\lambda_j}$ for $1 \leqslant i, j \leqslant s$. Then the eigenvalues of Ad_a restricted to $\mathfrak{g}_a$ are contained in the set

$$\{\tfrac{\lambda_i}{\lambda_j} \mid |\lambda_i| < |\lambda_j|\},$$

depending on $\mathfrak{g} \subseteq \mathfrak{sl}_r(\mathbb{R})$.

Since we have shown that $G_a^- \leqslant G$ is closed, we may choose $H = G_a^-$ (or any closed a-normalized subgroup $H \leqslant G_a^-$) in Corollary 9.28.

There are several possible extensions of Example 9.33, a few of which we list here (cf. Section 8.1).

(1) We can take G to be a connected, simply connected real (or complex) Lie group. Here the horospherical subgroups $G_a^\pm$ defined as in (9.21) and (9.22) are closed (this may be seen in several ways; for example since G is a cover of a closed real or complex linear group).
(2) We can take $G = G_\infty \times G_f$ where G_∞ is a closed real linear group, G_f is any other locally compact σ-compact group, and $a \in G_\infty$. Here G_f could be a finite product of p-adic Lie groups (for example, resulting in $\mathrm{SL}_r(\mathbb{R}) \times \mathrm{SL}_r(\mathbb{Q}_{p_1}) \times \cdots \times \mathrm{SL}_r(\mathbb{Q}_{p_s})$) or the group comprising the $\mathbb{A}_f$-points of an algebraic group (for example, $G = \mathrm{SL}_r(\mathbb{A}_f)$).
(3) We can drop the assumption $a \in G_\infty$ in (2), and allow any $a \in G_\infty \times G_f$, where G_∞ is a connected, simply connected real or complex Lie group and G_f is either a finite product of closed p-adic linear groups or the group of $\mathbb{A}_f$-points of a linear algebraic group. Here $G_a^\pm$ is again a closed subgroup.
(4) Finally, we can take $G = \mathbb{G}(\mathbb{F}_p((t)))$ to be the group of $\mathbb{F}_p((t))$-points of a linear algebraic group $\mathbb{G}$ defined over $\mathbb{F}_p((t))$ (or a finite product of such groups, or the adelic points of a linear algebraic group defined over $\mathbb{F}_p(t)$), and $a \in G$. In this setting the subgroups $G_a^\pm$ are once again closed subgroups of G.

In any of these settings we will refer to the orbit $G_a^- \cdot x$ of $x \in X$ under the action of G_a^- as the *stable manifold* through x.

Exercises for Section 9.4

Add ref to entropy volume

Exercise 9.4.1. Show that the quotient group in (9.20) is indeed isomorphic to the invertible extension appearing in the proof of Rudolph's Theorem.

Exercise 9.4.2. Let G be as in Example 9.33, and suppose that $\Gamma \leqslant G$ is a uniform lattice so that $X = \Gamma\backslash G$ is compact. Show that for any $x \in X$,

$$G_a^{-}\centerdot x = \{y \in X \mid \mathsf{d}(a^n\centerdot x, a^n\centerdot y) \to 0 \text{ as } n \to \infty\}.$$

Exercise 9.4.3. Let $X = \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ and $a = \begin{pmatrix} \mathrm{e} & \\ & \mathrm{e}^{-1} \end{pmatrix}$. Show that

$$G_a^{-}\centerdot x \subseteq \{y \in X \mid \mathsf{d}(a^n\centerdot x, a^n\centerdot y) \to 0 \text{ as } n \to \infty\},$$

with equality if the forward orbit $\{a^n\centerdot x \mid n \in \mathbb{N}\}$ is bounded. Find an example of an $x \in X$ for which equality does not hold.

Notes to Chapter 9

(28)(Page 269) A metric space (X, d) is σ-compact if every point of X has a compact neighborhood and X is a countable union of compact sets. Similarly, a measure space is σ-finite if it is a countable union of measurable sets of finite measure. In [**?**, Chap. **??**] we considered compact metric spaces; the one-point compactification of a σ-compact metric space is a compact metric space, so the results there apply in the σ-compact case also.
(29)(Page 275) We restrict attention in this section to a manageable degree of generality. Most of the results can be formulated more generally; see, for example, Lindenstrauss [**?**], where a general axiomatic treatment is given.
(30)(Page 281) See the work of Lindenstrauss [**?**] on quantum unique ergodicity or the notes of Einsiedler and Lindenstrauss [**?**] for details of how this can be done.
(31)(Page 294) A Radon measure is a measure on the Borel σ-algebra of a Hausdorff space that is locally finite and inner regular (that is, the measure of any measurable set can be approximated by the measure of compact subsets). The space of Radon measures was introduced by Radon [**?**] in the context of Borel measures on $\mathbb{R}^n$ as a way to deal with some of the problems arising in infinite measure spaces, using the correspondence between measures and positive linear functions on the space of compactly supported continuous functions. We refer to Hewitt and Stromberg [**?**] or Schwartz [**?**] for complete treatments.
(32)(Page 307) We recall from Definition **??** that a solenoid is a compact metric abelian group of finite topological dimension $d \in \mathbb{N}$; equivalently a locally compact abelian group whose character group is a subgroup of $\mathbb{Q}^d$ for some $d \geqslant 1$. For $d = 1$ solenoids are easy to classify — a straightforward description may be found in a short paper of Beaumont and Zuckerman [**?**], but the results are significantly older. These solenoids are with few exceptions and up to isomorphism, in one-to-one correspondence with subsets of the set of rational primes. For $d = 2$ (and hence for any $d \geqslant 2$) it is well-known that there is no effective way to describe the lattice of subgroups of $\mathbb{Q}^d$, by work of Kechris [**?**]. Thus two natural classes of d-dimensional solenoids which do have a reasonable classification are products $\prod_{i=1}^{d} \Sigma_i$ of 1-dimensional solenoids, and those corresponding to subgroups of $\mathbb{Q}^d$ that are isomorphic to an algebraic number field K. Despite this, the entropy of automorphisms of a solenoid is readily computed (see Section **??** or the paper of Lind and Ward [**?**]) because the distinction between the (unknown) subgroup of $\mathbb{Q}^d$ and $\mathbb{Q}^d$ itself does not matter for the entropy calculation; more surprisingly work of Miles [**?**]

shows that calculating the number of periodic points of an endomorphism of a solenoid (where the distinction certainly does matter) can always be reduced to the direct product of the number field cases. A solenoid corresponding to an algebraic number field K has a convenient description as a quotient of the adele ring of K, $K_{\mathbb{A}}$ (see Weil [**?**, Sec. IV.2] for an elegant and sophisticated treatment, or Ramakrishnan and Valenza [**?**] for an accessible account) much as in the case of the rationals. Viewed as measure-theoretic dynamical systems, automorphisms of solenoids are readily described: as discussed in Example **??**(3) a compact group automorphism with respect to Haar measure is measurably isomorphic to a Bernoulli shift if it is ergodic. Despite this, the description of the quotient space of ergodic group automorphisms by the equivalence of measurable isomorphism is not known — in particular the question of whether this space is countable or uncountable remains open (this is an arithmetic question; see Lind [**?**] or Everest and Ward [**?**] for an overview). Viewed as topological dynamical systems, a rich structure emerges (see for example work of Chothi, Everest, Miles, Stevens and Ward [**?**], [**?**] and [**?**]). Perhaps the most natural notion of equivalence — finitary isomorphism, topological conjugacy after removal of a null set — remains mysterious for group automorphisms.

(33) (Page 307) This is an instance of a more general phenomena; convenient sources include the monograph of Weil [**?**] and Tate's thesis [**?**]. If k is a locally compact non-discrete field and $\chi : k \to \mathbb{S}^1$ is any non-trivial character, then the map sending a to the character $x \mapsto \chi(ax)$ is an isomorphism of topological groups $k \mapsto \widehat{k}$.

(34) (Page 309) An extremely explicit description of this quotient group and the geometry of the $\times 2, \times 3$-action of $\mathbb{Z}^2$ on it, may be found in the paper of Ward and Yayama [**?**].

(35) (Page 309) If the field K has r embeddings into $\mathbb{R}$ and $2s$ embeddings into $\mathbb{C}$, then Dirichlet's unit theorem (Theorem 3.10) states that the multiplicative group of units $O_K^{\times}$ is isomorphic to the direct product $\mu_K \times \mathbb{Z}^{r+s-1}$, where μ_K is the group of roots of unity in K.

Chapter 10
Leafwise Measures and Entropy

In this chapter we will use the leafwise measures constructed in Chapter 9 to obtain a better understanding of the entropy theory of algebraic maps defined on quotients of Lie groups. This will also lead to more general and flexible results on the uniqueness of maximal measures for such maps. Once again the rather intricate geometric and measure-theoretic arguments ultimately reduce to the strict convexity of the map $t \mapsto -\log t$ (see p. 337 for example) and to understanding the way in which local geometric dilation controls the creation of entropy. Furthermore, we will use the connection between entropy and leafwise measures in subsequent chapters leading to partial classifications of invariant measures for higher-rank actions.

10.1 Entropy and Leafwise Measures: the Theorems

In this section we present the main theorems concerning the relationship between entropy and leafwise measures for the horospherical subgroup G_a^- or its closed subgroups. We allow here any of the settings described in Section 9.4, so in particular $G_a^- \leqslant G$ is a closed subgroup. The proof of these theorems will occupy the rest of the chapter. The ideas behind the material in this section come from work of Pesin [**?**], Furstenberg [**?**], Ledrappier and Young [**?**], [**?**], Katok, Ledrappier and Strelcyn [**?**], [**?**], Margulis and Tomanov [**?**] and others, and we follow the treatment of Einsiedler and Lindenstrauss [**?**, Sec. 7] closely.

10.1.1 Entropy and the Horospherical Subgroups

We start by relating the volume growth of the growing sequence of sets

$$\left(a^{-n} B_1^U a^n\right)_{n \geqslant 1}$$

for a closed subgroup U of G_a^- to the entropy of the map $x \mapsto a \boldsymbol{\cdot} x$ (which we abbreviate simply as a) with respect to the leafwise measures. By definition, G_a^- is a-invariant and is contracted by the action of a.

Definition 10.1. Let U be a closed subgroup of G_a^- that is normalized by a, and let μ be an a-invariant Borel probability measure on $X = \Gamma\backslash G$. Then the *entropy contribution of U at $x \in X$* is defined to be the asymptotic volume growth

$$\mathsf{vol}_\mu^U(a)(x) = \lim_{n\to\infty} \frac{1}{n} \log \mu_x^U \left(a^{-n} B_1^U a^n\right)$$

wherever the limit exists.

Recall that we write $\mathscr{E}$ for the σ-algebra of measurable sets invariant under the map a.

Theorem 10.2. *Let G be a connected, simply connected Lie group, or an S-algebraic group as described in Section* 8.1*, and let $\Gamma \leqslant G$ be a discrete subgroup. Let μ be an a-invariant probability measure on $\Gamma\backslash G$, and let U be a closed subgroup of G_a^- normalized by a. Then*

(1) *The sequence defining the entropy contribution $\mathsf{vol}_\mu^U(a)(x)$ of U at x converges for almost every $x \in X$, and the function $x \mapsto \mathsf{vol}_\mu^U(a)(x)$ is a-invariant. Moreover, $\mathsf{vol}_\mu^U(a)(x)$ also measures the volume decay at I in the sense that*

$$\mathsf{vol}_\mu^U(a)(x) = \lim_{n\to\infty} -\frac{1}{n} \log \mu_x^U (a^n B_1^U a^{-n}).$$

(2) *For almost every $x \in X$,*

$$\mathsf{vol}_\mu^U(a)(x) \leqslant h_{\mu_x^{\mathscr{E}}}(a),$$

with equality if $U = G_a^-$.

(3) *For almost every $x \in X$ the following statements are equivalent:*
- $\mathsf{vol}_\mu^U(a)(x) = 0$*;*
- μ_x^U *is a finite measure;*
- μ_x^U *is a trivial measure (in the sense of Definition* 9.19*).*

need to add some remark about the ergodic case

Combining Theorem 10.2(2) and (3) with Theorem 9.24 gives the following corollary.

Corollary 10.3. *The measure μ is G_a^--recurrent if and only if $h_{\mu_x^{\mathscr{E}}}(a) > 0$ for almost every $x \in X$. If μ is also assumed to be ergodic for the map a, then μ is G_a^--recurrent if and only if $h_\mu(a) > 0$.*

Some remarks are in order.

(1) If the stable horospherical subgroup G_a^- is trivial, then Theorem 10.2 shows that the entropy must vanish for all a-invariant measures. For example, the horocycle flow (see [**?**, Sec. 9.2]), or any flow under a unipotent element, has vanishing entropy.

(2) The opposite extreme to (1) occurs when $a \in G$ is diagonalizable. For example, the time-one map of the geodesic flow corresponding to the action of

$$a = \begin{pmatrix} \mathrm{e} & \\ & \mathrm{e}^{-1} \end{pmatrix}$$

on $\mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$ (see [**?**, Ch. 9] for the details) has stable horospherical subgroup given by the horocyclic subgroup

$$U^- = \left\{ \begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\},$$

so Theorem 10.2 shows that the entropy of the geodesic flow is determined by the leafwise measures for the horocyclic subgroup.

(3) Since $h_\mu(a) = h_\mu(a^{-1})$, Theorem 10.2 also holds for closed a-normalized subgroups $U < G_a^+$.

10.1.2 Entropy and G_a^--Invariance

Before stating the second main theorem concerning entropy and leafwise measures, notice that in the case $X = \Gamma\backslash G$, where Γ is a lattice in G, we can describe the leafwise measures of the measure m_X inherited from the Haar measure on G. Since m_X is certainly invariant under the action of G_a^-, the leafwise measures must be the Haar measure on G_a^-. It follows that the entropy contribution of Definition 10.1 can be computed easily, giving

$$\mathsf{vol}_{m_X}^{G_a^-}(a) = -\log \operatorname{mod}(a, G_a^-),$$

where $\operatorname{mod}(a, G_a^-)$ is the scalar determining how the action of a on G_a^- scales the Haar measure on G_a^-:

$$m_{G_a^-}(aBa^{-1}) = \operatorname{mod}(a, G_a^-) m_{G_a^-}(B)$$

for all measurable $B \subseteq G$ (see [**?**, Cor. 8.8]).

We record two special cases to illustrate how these quantities arise in simple situations (see [**?**] for the relationship between the quantities and entropy).

(1) If $G = \mathbb{R}^r$ and $\Gamma = \mathbb{Z}^r$ and $a \in \mathrm{GL}_r(\mathbb{Z})$ is a toral automorphism, then

$$\mathrm{mod}(a, G_a^-) = \prod_{|\lambda|<1} |\lambda|$$

is the product of all the eigenvalues with absolute value less than one, counted with their algebraic multiplicity.

(2) If G is a connected, simply-connected Lie group, then

$$\mathrm{mod}(a, G_a^-) = \left|\det\left(\mathrm{Ad}_a|_{\mathfrak{g}_a^-}\right)\right|,$$

where $\mathfrak{g}_a^-$ is the Lie algebra of G_a^- and Ad_a is the adjoint action of a on the Lie algebra of G.

The second main theorem of this section gives a characterization in terms of the entropy $h_\mu(a)$ (or in terms of the entropy contribution of a closed subgroup $U \leqslant G_a^-$) for a measure μ to be invariant under G_a^- (resp. under the subgroup U). To formalize this, we extend the pointwise Definition 10.1 by averaging across the space as follows.

Definition 10.4. Let U be a closed subgroup of G_a^- that is normalized by a, and let μ be an a-invariant Borel probability measure on $\Gamma\backslash G$. Then the *entropy contribution of* U is

$$h_\mu(a, U) = \int \mathsf{vol}_\mu^U(a)(x)\,\mathrm{d}\mu.$$

In particular, applying this with $U = G_a^-$ we have $h_\mu(a, G_a^-) = h_\mu(a)$ (by Theorem 10.2).

Theorem 10.5. *Let $X = \Gamma\backslash G$ be as in Theorem* 10.2*. Let U be a closed subgroup of G_a^- that is normalized by a, and let μ be an a-invariant Borel probability measure on $\Gamma\backslash G$. Then the entropy contribution of U satisfies*

$$h_\mu(a, U) \leqslant -\log \mathrm{mod}(a, U),$$

and equality holds if and only if μ is U-invariant.

From one point of view this result — and in particular the last part — is a continuation of one of our major themes, characterizing measures via their entropy properties. If Γ is a uniform lattice and Ad_a is diagonalizable, then we have already seen in [**?**, Th. **??**] that the Haar measure is the unique measure of maximal entropy. Theorem 10.5 contains many significant generalizations of this result, as shown in the next example.

Example 10.6. Let $X = \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ and $a \in \mathrm{SL}_2(\mathbb{R})$. Since the stable horospherical subgroup is the upper unipotent subgroup

$$U^- = \left\{\begin{pmatrix} 1 & s \\ & 1 \end{pmatrix} \mid s \in \mathbb{R}\right\} \subseteq \mathrm{SL}_2(\mathbb{R}),$$

it follows by Theorem 10.5 that an a-invariant measure μ with

$$h_\mu(a) = h_{m_X}(a)$$

must be invariant under the upper unipotent subgroup. Notice that this does not follow from [**?**, Th. **??**], since X is not compact.

On the other hand $h_\mu(a) = h_\mu(a^{-1})$, so the same argument shows that an a-invariant measure μ with $h_\mu(a) = h_{m_X}(a)$ must be invariant under the lower unipotent subgroup

$$U^+ = \left\{ \begin{pmatrix} 1 & \\ s & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\} \subseteq \mathrm{SL}_2(\mathbb{R}).$$

However, the upper and lower unipotent subgroups together generate all of $\mathrm{SL}_2(\mathbb{R})$, so that $h_\mu(a) = h_{m_X}(a)$ implies that μ is invariant under all of $\mathrm{SL}_2(\mathbb{R})$, and therefore $\mu = m_X$. Thus the Haar measure m_X is the unique measure of maximal entropy for a.

The argument in Example 10.6 generalizes to more general situations (subject to the condition that the stable and unstable horospherical groups together generate G, or at least generate a subgroup whose action is uniquely ergodic). Recall that a single continuous map on a compact metric space is said to be uniquely ergodic if there is only one invariant measure (see [**?**, Sec. 4.3]); we extend this to a group action on a σ-compact space by saying that the group action is uniquely ergodic if there is only one invariant probability measure for the whole action.

Corollary 10.7. *Let G be a Lie group with $\Gamma \leqslant G$ a lattice, and let $X = \Gamma\backslash G$. If $a \in G$ has the property that the action of the subgroup generated by G_a^- and G_a^+ is uniquely ergodic on X, then m_X is the unique measure of maximal entropy for the action of a on X.*

The proofs of Theorems 10.2 and 10.5 will take up several sections.

10.1.3 Some Initial Observations

We begin by proving the technical assumption in Corollary 9.9.

Lemma 10.8. *Let μ be an a-invariant probability measure on $X = \Gamma\backslash G$. Then for μ-almost every $x \in X$ the map $G_a^- \ni u \longmapsto u \cdot x$ is injective.*

Proof. Suppose that $x = u \cdot x$ for some non-trivial element $u \in G_a^-$. Then

$$0 = \mathsf{d}(a^n \cdot x, a^n u \cdot x) = \mathsf{d}(a^n \cdot x, a^n u a^{-n} \cdot (a^n \cdot x)) \tag{10.1}$$

for all $n \geqslant 1$. However, by definition of the stable horospherical subgroup, $a^n u a^{-n} \to e$ as $n \to \infty$, so the injectivity radius at $a^n \cdot x$ must converge to 0 as $n \to \infty$. This shows that the point x does not satisfy Poincaré recurrence. By the Poincaré recurrence theorem (Theorem 1.8), it follows that there is some null set that contains all such points x, giving the lemma. □

As we will now show, Corollary 9.28 already gives the first part of Theorem 10.2.

Proof of Theorem 10.2(1). Consider the measurable function

$$f(x) = \log \mu_x^U \left(a^{-1} B_1^U a\right) \geqslant 0,$$

which is non-negative since $a^{-1} B_1^U a \supseteq B_1^U$ and $\mu_x^U \left(B_1^U\right) = 1$ by the normalization used in Theorem 9.6. By Corollary 9.28 we have, for every $k \in \mathbb{Z}$,

$$\begin{aligned} f(a^k \cdot x) &= \log \mu_{a^k \cdot x} \left(a^{-1} B_1^U a\right) \\ &= \log \frac{\mu_{a^k \cdot x} \left(a^{-1} B_1^U a\right)}{\mu_{a^k \cdot x} \left(B_1^U\right)} \\ &= \log \mu_x^U \left(a^{-(k+1)} B_1^U a^{k+1}\right) - \log \mu_x^U \left(a^{-k} B_1^U a^k\right) \end{aligned} \tag{10.2}$$

for almost every x. This implies that

$$\sum_{k=0}^{n-1} f(a^k \cdot x) = \log \mu_x^U \left(a^{-n} B_1^U a^n\right)$$

for every $n > 0$ since the sum is telescoping, and hence

$$\frac{1}{n} \log \mu_x^U \left(a^{-n} B_1^U a^n\right) \to E\left(f \middle| \mathscr{E}\right)(x) = \mathsf{vol}_\mu^U(a)(x)$$

almost surely as $n \to \infty$ by the pointwise ergodic theorem, where $\mathscr{E}$ is the σ-algebra of a-invariant Borel sets in X.

It remains to consider the limit as $n \to -\infty$. In this case we may use (10.2) again to see that

$$\begin{aligned} \sum_{k=0}^{n} f(a^{-k} \cdot x) &= \sum_{k=0}^{n} \left(\log \mu_x^U \left(a^{(k-1)} B_1^U a^{-(k-1)}\right) - \log \mu_x^U \left(a^k B_1^U a^{-k}\right)\right) \\ &= \log \mu_x^U \left(a^{-1} B_1^U a^1\right) - \log \mu_x^U \left(a^n B_1^U a^{-n}\right). \end{aligned}$$

Dividing by n and applying the pointwise ergodic theorem again gives

$$\frac{-\log \mu_x^U \left(a^n B_1^U a^{-n}\right)}{n} \longrightarrow E\left(f \middle| \mathscr{E}\right)$$

almost surely, completing the proof of Theorem 10.2(1). □

Lemma 10.9. *Under the assumptions of Theorem* 10.2*, we have*

$$\mathsf{vol}_\mu^U(a)(x) = \lim_{n\to\infty} \frac{\log \mu_x^U\left(a^{-n} B_r^U a^n\right)}{n}$$

for any $r > 0$.

Proof. Given $r > 0$, there exists some n_0 such that

$$a^{n_0} B_1^U a^{-n_0} \subseteq B_r^U \subseteq a^{-n_0} B_1^U a^{n_0},$$

which implies that

$$\mu_x^U\left(a^{n_0-n} B_1^U a^{-(n_0-n)}\right) \leqslant \mu_x^U\left(a^{-n} B_r^U a^n\right) \leqslant \mu_x^U\left(a^{-(n+n_0)} B_1^U a^{n+n_0}\right),$$

and the lemma follows since $\frac{n\pm n_0}{n} \to 1$ as $n \to \infty$. $\square$

Exercises for Section 10.1

Exercise 10.1.1. Show that for any Lie group G and any $a \in G$, the Lie algebra generated by $\mathfrak{g}_a^-$ and $\mathfrak{g}_a^+$ is a Lie ideal in $\mathfrak{g}$. Deduce that the assumption regarding a in Corollary 10.7 is satisfied whenever G is a simple real Lie group and $\mathfrak{g}_a^-$ is non-trivial.

10.2 Reduction to the Ergodic Cases

In this section we will show how the general statements in Theorems 10.2 and 10.5 follow from their counterparts with the additional hypothesis that μ is assumed to be ergodic for the action of a. The reader who is willing to assume this ergodicity† of the action of a, may continue reading with the next section.

Recall from [**?**, Th. 6.2] that for any a-invariant measure μ there is the ergodic decomposition

$$\mu = \int \mu_x^{\mathscr{E}} \,\mathrm{d}\mu(x).$$

Also recall from [**?**, Th. **??**] that the entropy $h_\mu(a)$ may be obtained by integrating the entropies $h_{\mu_x^{\mathscr{E}}}(a)$ over the ergodic decomposition.

An important observation (the Hopf argument(36)) is that we can choose the elements of $\mathscr{E}$ to be not only a-invariant, but in fact invariant under the larger group $\langle G_a^-, a\rangle$. We will write

† This assumption should not be confused with A-ergodicity, which we will assume in later chapters, but which in general does not imply ergodicity of the action of a.

$$\mathsf{A}_n^f(x) = \frac{1}{n}\sum_{k=0}^{n-1} f(a^k \bullet x)$$

for the ergodic average of f of length n at x.

Proposition 10.10. *If C is an a-invariant measurable subset of X, then there is a $\langle G_a^-, a\rangle$-invariant set $\widetilde{C}$ such that $\mu\left(C \triangle \widetilde{C}\right) = 0$. Therefore, the σ-algebra $\mathscr{E}$ comprising the a-invariant measurable subsets of X is equivalent modulo μ to a countably generated σ-algebra $\widetilde{\mathscr{E}}$ that consists of $\langle G_a^-, a\rangle$-invariant measurable subsets of X.*

Proof (using the Hopf argument). Fix $\varepsilon > 0$ and choose a continuous function with compact support $f \in C_c(X)$ for which

$$\|f - \mathbb{1}_C\|_1 < \varepsilon. \tag{10.3}$$

By the pointwise ergodic theorem we have

$$\frac{1}{n}\sum_{k=0}^{n-1} f(a^k \bullet x) \longrightarrow g(x)$$

for all $x \in X \smallsetminus N$, where N is a μ-null set and g is some a-invariant function. As C is a-invariant, we have

$$\|g - \mathbb{1}_C\|_1 = \lim_{n\to\infty} \|\mathsf{A}_n^f - \mathbb{1}_C\|_1 \leqslant \|f - \mathbb{1}_C\|_1 < \varepsilon.$$

We define $C_\varepsilon = g^{-1}\left((\frac{1}{2}, \infty)\right)$ and obtain

$$C \triangle C_\varepsilon \subseteq N \cup \underbrace{\{x \in C \mid g(x) \leqslant \tfrac{1}{2}\} \cup \{x \notin C \mid g(x) > \tfrac{1}{2}\}}_{E}.$$

Clearly

$$E \subseteq \{x \mid |g(x) - \mathbb{1}_C(x)| \geqslant \tfrac{1}{2}\}$$

and so

$$\tfrac{1}{2}\mu(E) \leqslant \|g - \mathbb{1}_C\|_1 < \varepsilon.$$

For any $h \in G_a^-$, we have

$$a^n \bullet (h \bullet x) = a^n h a^{-n} \bullet (a^n \bullet x),$$

so that $a^n \bullet (h \bullet x)$ and $a^n \bullet x$ are asymptotic to one another as $n \to \infty$ (see (10.1) on p. 319). Since we have chosen f to have compact support, it is uniformly continuous. Hence

$$\frac{1}{n}\sum_{i=0}^{n-1} \left(f(a^i \bullet x) - f(a^i \bullet (h \bullet x))\right) \longrightarrow 0$$

as $n \to \infty$, uniformly in x. It follows that C_ε is G_a^--invariant.

To finish the proof we choose $\varepsilon_n = 2^{-n}$ and run through the construction above for each $n \geqslant 1$. Then

$$\widetilde{C} = \limsup_{n\to\infty} C_{2^{-n}} = \bigcap_n \bigcup_{k\geqslant n} C_{2^{-k}}$$

is the set claimed in the proposition. To see the final claim, notice that $\mathscr{E}$ is equivalent modulo μ to a countably generated σ-algebra $\widetilde{\mathscr{E}}$ (since $L^1(X,\mathscr{E},\mu)$ is a separable subspace of $L^1(X,\mathscr{B},\mu)$), which by the argument above we may assume to be generated by $\langle G_a^-, a\rangle$-invariant measurable sets. □

Corollary 10.11. *Let μ be an a-invariant Borel probability measure, and let $U \leqslant G_a^-$ be a closed subgroup. Then for μ-almost every x, and for $\mu_x^{\mathscr{E}}$-almost every y, we have $\mu_y^U = (\mu_x^{\mathscr{E}})_y^U$.*

In other words, by changing the leafwise measures for $\mu_x^{\mathscr{E}}$ at most on a $\mu_x^{\mathscr{E}}$-null set, we may consistently define $(\mu_x^{\mathscr{E}})_y^U$ to be equal to μ_y^U. With this definition in place, we also have $(\mu_x^{\mathscr{E}})_x^U = \mu_x^U$. In the statement of Corollary 10.11 we did not state this formula because $\{x\}$ is a null set for $\mu_x^{\mathscr{E}}$, so making claims for the leafwise measure at x would seem irrelevant in the almost-everywhere statement.

Proof of Corollary 10.11. By Proposition 10.10, we may replace $\mathscr{E}$ by a countably generated σ-algebra $\widetilde{\mathscr{E}}$ consisting of $\langle U, a\rangle$-invariant sets. Let

$$\mathscr{A} = \mathscr{A}(x_0, R, \delta)$$

be a (U,R)-flower with base $B_\delta(x_0)$ for some $x_0 \in X$, $\delta \in (0,1]$, and $R \geqslant 1$. Let T be the corresponding cross-section and define $Y = B_R^U \boldsymbol{\cdot} T$. We now study the relationship between $\widetilde{\mathscr{E}}|_Y$ and $\mathscr{A}$. Notice first that the atoms of $\widetilde{\mathscr{E}}$ are unions of atoms of $\mathscr{A}$ (since the sets in $\widetilde{\mathscr{E}}$ are U-invariant and the atoms of $\mathscr{A}$ are, by definition, open U-plaques). However, using conditional measures for $\mathscr{A}$ it is easy to see that a measurable function that is constant on $\mathscr{A}$-atoms is in fact $\mathscr{A}$-measurable modulo μ (see [**?**, Sec. 5.3]). Therefore, we have $\mathscr{E}|_Y \subseteq \mathscr{A}$ modulo μ. However, this inclusion of σ-algebras implies that

$$E\Big(E(f|\mathscr{A})\big|\widetilde{\mathscr{E}}|_Y\Big) = E\left(f\big|\widetilde{\mathscr{E}}|_Y\right)$$

for any $f \in L^1(Y)$. In turn, using the defining properties of conditional measures (in terms of conditional expectations) this gives the following relation between the conditional measures: for μ-almost every $x \in Y$ we have, for $\mu_x^{\widetilde{\mathscr{E}}}|_Y$-almost every y,

$$\left(\mu_x^{\widetilde{\mathscr{E}}}|_Y\right)_y^{\mathscr{A}} = (\mu|_Y)_y^{\mathscr{A}}$$

(see [**?**, Prop. 5.20] and Proposition 9.2(1)). We are now going to translate this to a property of leafwise measures. In fact, by (in order) Corollary 9.9 for μ, the argument above, and Corollary 9.9 for $\mu_x^{\widetilde{\mathscr{E}}}$, we have for $y = h{\boldsymbol\cdot}t \in Y$,

$$\mu_y^U|_{B_R^U h^{-1}{\boldsymbol\cdot}y} \propto (\mu|_Y)_y^{\mathscr{A}} = \left(\mu_x^{\widetilde{\mathscr{E}}}|_Y\right)_y^{\mathscr{A}} \propto \left(\mu_x^{\widetilde{\mathscr{E}}}\right)_y^U|_{B_R^U h^{-1}{\boldsymbol\cdot}y}$$

for μ-almost every x and $\mu_x^{\widetilde{\mathscr{E}}}$-almost every $y \in Y$. For $y \in B_\delta^X(x_0)$ this implies that

$$\mu_y^U|_{B_{R-1}^U} \propto \left(\mu_x^{\widetilde{\mathscr{E}}}\right)_y^U|_{B_{R-1}^U}.$$

Applying Proposition 9.16 for $R \in \mathbb{N}$, the corollary follows. □

Proof of reduction to ergodic case in Theorems 10.2 and 10.5. Working with double conditional measures as in Corollary 10.11 may be confusing, but is useful for the following reason. In the proofs of Theorems 10.2 and 10.5 we compare the entropy of the ergodic components with the entropy contribution arising from the subgroup $U \leqslant G_a^-$. By [**?**, entropyoverergodic-components] we know that

$$h_\mu(a) = \int h_{\mu_x^{\mathscr{E}}}(a)\,\mathrm{d}\mu.$$

We would like to make use of a similar relationship between $\mathsf{vol}_\mu^U(a)(x)$ and $\mathsf{vol}_{\mu_x^{\mathscr{E}}}^U(a)(x)$. Using the identity $(\mu_x^{\mathscr{E}})_x^U = \mu_x^U$ as in the discussion immediately after Corollary 10.11, we get

$$\mathsf{vol}_{\mu_x^{\mathscr{E}}}^U(a)(x) = \mathsf{vol}_\mu^U(a)(x).$$

Since $\mu_x^{\mathscr{E}}$ is a-invariant and ergodic for μ-almost every x, if we assume the statements of Theorem 10.2 in the ergodic case, the general case follows quickly. Indeed, we have already shown Theorem 10.2(1) in general, and Theorem 10.2(2) and (3) state relationships between $\mathsf{vol}_\mu^U(a)(x)$, $h_{\mu_x^{\mathscr{E}}}(a)$, and μ_x^U, which with the argument above generalizes from the ergodic to the general case. To see that the ergodic case of Theorem 10.5 also implies its general case, notice that

$$\mathsf{vol}_\mu^U(a)(x) = h_{\mu_x^{\mathscr{E}}}(a,U)$$

so that

$$\mathsf{vol}_\mu^U(a)(x) \leqslant -\log|\operatorname{mod}(a,U)|,$$

which implies that

$$h_\mu(a,U) \leqslant -\log|\operatorname{mod}(a,U)|.$$

If we have equality here, then we have

$$\mathsf{vol}_\mu^U(a)(x) = h_{\mu_x^{\mathscr{E}}}(a,U) = -\log|\operatorname{mod}(a,U)|$$

almost everywhere with respect to μ. By the assumed ergodic case, this implies that $\mu_x^{\mathscr{E}}$ is U-invariant for almost every x, which implies that μ is U-invariant (by Theorem 9.26, for example). If, on the other hand, μ is U-invariant, then μ_x^U is almost surely a Haar measure on U, and it follows from the definition that

$$\mathsf{vol}_\mu^U(a)(x) = -\log|\operatorname{mod}(a,U)|,$$

which implies equality. □

10.3 Entropy Contribution and the Proof of Theorem 10.2

In this section we will prove Theorem 10.2 under the assumption of Proposition 10.21, which will be proved in Section 10.5.

10.3.1 Descending Subordinate σ-algebras

Let $X = \Gamma\backslash G$, $a \in G$ and μ be as in Theorem 10.2. Then by the Kolmogorov–Sinaĭ Theorem we have

$$h_\mu(a) = h_\mu(a,\xi) = H_\mu\left(\xi\middle|\xi_1^\infty\right)$$

if ξ is a generator for the map a, that is if

$$\xi_{-\infty}^\infty = \bigvee_{n\in\mathbb{Z}} a^{-n}\bullet\xi \underset{\mu}{=} \mathscr{B}_X.$$

For the proof of Theorem 10.2, we will need a generator with good geometrical properties with respect to the horospherical subgroup G_a^-.

Definition 10.12. Let $U \leqslant G_a^-$ be a closed subgroup, let $Y \subseteq X$ be measurable, and let $\mathscr{A} \subseteq \mathscr{B}_X$ be a countably-generated σ-algebra. Then $\mathscr{A}$ is *subordinate to U on Y modulo μ* if for μ-almost every $x \in Y$ there exists some $\delta > 0$ with

$$B_\delta^U\bullet x \subseteq [x]_{\mathscr{A}} \subseteq B_{1/\delta}^U\bullet x,$$

and $\mathscr{A}$ is *subordinate to U modulo μ* if $\mathscr{A}$ is subordinate to U on X modulo μ. Finally we say that $\mathscr{A}$ is *a-descending* if $a^{-1}\bullet\mathscr{A} \subseteq \mathscr{A}$.

While we have already seen some σ-algebras that are subordinate to U (see Proposition 9.13) and it is easy to find a-descending σ-algebras (of the form $\bigvee_{n=1}^\infty a^{-n}\bullet\mathscr{A}$ for any given σ-algebra $\mathscr{A}$), it is *a priori* not clear how to find a-descending σ-algebras that are also subordinate to U modulo μ. Assuming ergodicity of μ, we will construct such a generator in Proposition 10.21 This result, whose lengthy proof will occupy Section 10.5, will be needed for the proof of Theorem 10.2(2) linking the entropy contribution for G_a^- to the entropy.

The construction of σ-algebras ξ^U for which

$$\mathscr{A} = \bigvee_{n=0}^{\infty} a^{-n} \bullet \xi^U$$

is U-subordinate is somewhat easier.

[MLE] used to start at 1, now at 0 — decide once and for all

Proposition 10.13. *Let $X = \Gamma\backslash G$, $a \in G$, $U < G_a^-$ and μ be as in the statement of Theorem 10.2. Then there exists, for every $\varepsilon > 0$, a countably-generated σ-algebra $\mathscr{A}$ which is both a-descending and subordinate to U on a set of measure exceeding $1 - \varepsilon$ modulo μ. If X is compact or μ is ergodic, then $\mathscr{A}$ can be chosen to be subordinate to U modulo μ on the whole space.*

In order to prove Proposition 10.13, we will need some preparatory material. In any metric space (X, d) we define the δ-*boundary of* B to be the set

$$\partial_\delta B = \{y \in X \mid \inf_{z \in B} \mathsf{d}(y, z) + \inf_{z \notin B} \mathsf{d}(y, z) < \delta\}$$

for any subset $B \subseteq X$. Notice that $\partial_\delta B$ is not the δ-neighborhood of the boundary ∂B of B: it may very well happen that ∂B is empty while $\partial_\delta B$ is non-empty for $\delta > 0$.

Lemma 10.14. *Let X be a locally compact metric space, and let μ be a Radon measure on X. Then, for every $x \in X$ and Lebesgue almost every $r > 0$, there exists a constant $c = c_{x,r}$ such that $\mu(\partial_\delta B_r(x)) \leqslant c\delta$ for all sufficiently small $\delta > 0$.*

Proof. Let $f(r) = \mu(B_r(x))$. Then $f : [0, \infty) \to \mathbb{R}$ is monotone, and hence differentiable Lebesgue almost everywhere[(37)]. Suppose that f is differentiable at r, and let $c = 2(f'(r) + 1)$. Then we have

$$\partial_\delta B_r(x) \subseteq B_{r+\delta}(x) \smallsetminus B_{r-\delta}(x)$$

and so

$$\begin{aligned} \frac{\mu(\partial_\delta B_r(x))}{\delta} &\leqslant \frac{f(r+\delta) - f(r-\delta)}{\delta} \\ &= \frac{f(r+\delta) - f(r)}{\delta} + \frac{f(r) - f(r-\delta)}{\delta} < c \end{aligned}$$

for sufficiently small $\delta > 0$. □

Definition 10.15. We say that a set B in a metric space X equipped with a measure μ has *thin boundary* if there is some constant $c > 0$ with $\mu(\partial_\delta B) \leqslant c\delta$ for all $\delta > 0$. Furthermore, we say that a partition ξ of X has *μ-thin boundary* if there is a constant c with

$$\mu\left(\partial_\delta \xi\right) \leqslant c\delta \tag{10.4}$$

for all $\delta > 0$, where

$$\partial_\delta \xi = \bigcup_{P \in \xi} \partial_\delta P.$$

Notice that if $\mu(X) < \infty$ then Lemma 10.14 provides many sets with thin boundaries.

Lemma 10.16. *Let $X = \Gamma\backslash G$ and let μ be a probability measure on X. Then for any $\varepsilon > 0$ and $\rho > 0$, there exists a finite partition ξ of X with μ-thin boundary containing only one unbounded set P_0 with $\mu(P_0) < \varepsilon$ and with the property that all other elements $P \in \xi\smallsetminus\{P_0\}$ have diameter less than $\rho r_{X\smallsetminus P_0}$, where $r_{X\smallsetminus P_0}$ is an injectivity radius on the bounded set $X\smallsetminus P_0$.*

Proof. First apply Lemma 10.14 to find some set

$$\Omega = B_r(x_0)$$

with thin boundary and with $\mu(\Omega) > 1 - \varepsilon$. Since μ is a finite measure, we can adjust the constant c to remove the assumption that δ should be sufficiently small in Lemma 10.14. Now cover $\overline{\Omega}$ with balls of the form $B_r(x)$ with $x \in \overline{\Omega}$ and $r \leqslant \rho r_\Omega$ as in Lemma 10.16. Choose a finite subcover $B_{r_1}(x_1), \dots, B_{r_n}(x_n)$ and then define

$$\begin{aligned}
P_0 &= X\smallsetminus\Omega,\\
P_1 &= B_{r_1}(x_1) \cap \Omega,\\
P_2 &= B_{r_2}(x_2) \cap \Omega\smallsetminus P_1,\\
&\;\;\vdots\\
P_n &= B_{r_n}(x_n) \cap \Omega\smallsetminus\left(P_1 \cup \dots \cup P_{n-1}\right),
\end{aligned}$$

which implies the lemma. □

Lemma 10.17. *There exists some $\alpha > 0$ and $d > 0$ depending on a and G such that for every $r \in (0, 1]$ we have*

$$a^n(B_r^{G_a^-})a^{-n} \subseteq B^G_{d\mathrm{e}^{-n\alpha}r} \tag{10.5}$$

for all $n \geqslant 1$.

Proof. First consider the case where G is a real connected Lie group ($\mathrm{SL}_r(\mathbb{R})$, for example). In this case we are using a left-invariant Riemannian metric on G, and we have

$$\mathsf{d}\left(g_1a^{-n}, g_2a^{-n}\right) \leqslant \|\operatorname{Ad}_a^{-n}\|\mathsf{d}(g_1,g_2).$$

In that case the real Jordan normal form (cf. Example 9.33) of the adjoint $\operatorname{Ad}_a : \mathfrak{g} = \operatorname{Lie} G \to \mathfrak{g}$ proves the lemma. If G is not connected but is linear then we may argue as in the proof of Lemma 8.1 and use the metric arising from the inclusion $G \leqslant \mathrm{SL}_r(\mathbb{R})$ for some $r \geqslant 1$, and the lemma follows in that case also. If G is a real Lie group which is not connected or linear, then we may define the metric

$$\mathsf{d}(g_1,g_2) = \begin{cases} 2\frac{\mathsf{d}^0(e,h)}{1+\mathsf{d}^0(e,h)} & \text{if } g_2 = g_1h \text{ and } h \in G^0;\\ 2 & \text{if } g_2 \notin g_1G^0,\end{cases}$$

where d^0 is the left-invariant Riemannian metric on the connected component $G^0 \leqslant G$. In this case $B_1^G(I)$ and $B_1^{G^0}(I)$ together with their metrics d and d^0 respectively, are Lipschitz equivalent, and the case of the lemma for connected Lie groups also implies the general case of any Lie group as a result.

Now assume that $G \leqslant \mathrm{SL}_r(\Bbbk)$ is a linear group over the local field $\Bbbk = \mathbb{Q}_p$ or $=\mathbb{F}_q((s))$ (as in Section 8.1.3). In these cases we defined the left-invariant metric on G using (8.2). After applying a conjugation (which is a bi-Lipschitz map) we may assume that a is in Jordan normal form over $\Bbbk$ (cf. Example 9.33). Once again, the lemma follows since the non-zero off-diagonal matrix entries of a^nga^{-n} are of the form

$$\left(\frac{\lambda_i}{\lambda_j}\right)^n p_{ij}(n)g_{ij}$$

with $|\lambda_i|_\Bbbk < |\lambda_j|_\Bbbk$ for some polynomial $p_{ij} \in \Bbbk[t]$.

If G is a restricted product, then G_a^- is a finite product, and the cases considered above put together prove this case also. □

We now show how to combine Lemmas 10.16 and 10.17 to give a lower bound for the size of atoms for ξ_0^∞.

Lemma 10.18. *Suppose that ξ is a finite partition of $X = \Gamma\backslash G$ with μ-thin boundary. Then for almost every $x \in X$ there is some $\delta > 0$ such that*

$$B_\delta^{G_a^-}\!\centerdot x \subseteq [x]_{\xi_0^\infty}. \tag{10.6}$$

Proof. Let c be a constant with (10.4), and let α and d be constants satisfying (10.5), which we will be applying with $r = 1$. Fix some $\delta > 0$, and write

$$E_n = a^{-n}\bullet\partial_{d\mathrm{e}^{-n\alpha}\delta}\xi$$

for $n \geqslant 0$. By construction,

$$\mu\Big(\bigcup_{n\geqslant 0} E_n\Big) \leqslant cd\Big(\sum_{n\geqslant 0}\mathrm{e}^{-n\alpha}\Big)\delta,$$

which shows that for almost every $x \in X$ there is some δ with $x \notin \bigcup_{n\geqslant 0} E_n$. We claim that (10.6) holds for any such x and the corresponding δ.

To see this, let $h \in B_\delta^{G_a^-}$ and suppose that $h\bullet x \notin [x]_{\xi_0^\infty}$. Then there is some $n \geqslant 0$ for which $a^n\bullet x$ and $a^n\bullet(h\bullet x)$ belong to different elements of ξ. However, conjugation by a contracts the horospherical subgroup G_a^-, and indeed

$$\mathsf{d}(a^n h a^{-n}, I) < d\mathrm{e}^{-n\alpha}\delta$$

by Lemma 10.17. Thus $a^n\bullet(h\bullet x) = a^n h a^{-n}\bullet(a^n\bullet x)$ and $a^n\bullet x$ are no more than $d\mathrm{e}^{-n\alpha}\delta$ apart and belong to different elements of ξ, showing that both belong to $\partial_{d\mathrm{e}^{-n\alpha}\delta}\xi$. This contradicts the definition of E_n and the choice of x and δ above, showing (10.6). □

Example 10.19. If $X = \Gamma\backslash G$ is a torus or a solenoid and a is a hyperbolic automorphism of X, then Lemmas 10.16 and (10.18) together already imply Proposition 10.5. For this we also refer to the argument in [**?**, Th. **??**]. In order to remove the hyperbolicity hypothesis one has to refine the partition further.

Proof of Proposition 10.13. Let $\xi = \{P_0, P_1, \dots, P_f\}$ be a partition chosen as in Lemma 10.16 for $\rho = 1$, and let r be an injectivity radius on $X\smallsetminus P_0$. Then $P_n = \pi(B_n)$ for a set $B_n \subseteq G$ with $\operatorname{diam}(B_n) < r$ for $n = 1, \dots, f$. We define

$$\xi^U = \sigma\left(\{P_0, \pi(B_n \cap A) \mid n = 1, \dots, f, A \in \mathscr{B}_{G/U}\}\right),$$

which is clearly subordinate to U on $X\smallsetminus P_0$.

By construction the σ-algebra $\mathscr{A} = \left(\xi^U\right)_0^\infty$ is a-descending. Moreover, it is clear that

$$[x]_{\mathscr{A}} \subseteq [x]_{\xi^U} \subseteq B_r^U\bullet x$$

for $x \in X\smallsetminus P_0$, by the construction of ξ^U. We now apply Lemma 10.18 and assume that

$$B_\delta^{G_a^-}\bullet x \subseteq [x]_{\xi_0^\infty} \tag{10.7}$$

for some $x \in X$ and $\delta \in (0, d^{-1}r)$ with d as in Lemma 10.17. We claim we must then also have

$$B_\delta^U\bullet x \subseteq [x]_{\mathscr{A}}, \tag{10.8}$$

which with the observations above will imply the first part of the proposition. Assume therefore that $y = u\bullet x$ with $u \in B_\delta^U$. Then, for $n \geqslant 0$, $a^n\bullet y$ and $a^n\bullet x$

belong to the same partition element $P_k \in \xi$ by (10.7). If $k = 0$, then it is clear from the definition of ξ^U that $a^n \cdot x$ and $a^n \cdot y$ lie in the same atom of ξ^U. If $k \geqslant 1$, then notice that a normalizes and contracts U, so we have

$$a^n \cdot x, a^n \cdot y = u_n \cdot (a^n \cdot x) \in P_k = \pi(B_k)$$

with $u_n = a^n u a^{-n} \in B_r^U$ by Lemma 10.17. This once more implies that $a^n \cdot x$ and $a^n \cdot y$ lie in the same atom of ξ^U, and so (10.8) follows.

If X is compact, then we may choose $P_0 = \varnothing$ and obtain the final claim of the proposition in this case. Assume now that μ is ergodic for the action of a. In this case we claim that $\mathscr{A}$ is subordinate to U modulo μ on all of X. The proof of the lower bound (10.8) was general, so we only have to consider the upper bound

$$[x]_{\mathscr{A}} \subseteq B_R^U \cdot x \tag{10.9}$$

for $x \in P_0$. We claim that (10.9) holds for every $x \in P_0$ (for some R depending on n) for which there exists some $n \geqslant 1$ with $a^n \cdot x \notin P_0$. In fact, $a^{-n} \cdot \mathscr{A} \subseteq \mathscr{A}$, and so

$$[x]_{\mathscr{A}} \subseteq [x]_{a^{-n} \cdot \mathscr{A}} = a^{-n} \cdot [a^n \cdot x]_{\mathscr{A}} \subseteq \left(a^{-n} B_r^U a^n\right) \cdot x.$$

By ergodicity, (10.9) holds for almost every $x \in P_0$ and the final claim of the proposition follows. □

10.3.2 Proof of Theorem 10.2

Proposition 10.20. *Let $X = \Gamma\backslash G$ and $a \in G$ be as in Theorem* 10.2*. Assume in addition that μ is an a-invariant ergodic probability measure on X. Let*

$$U' < U < G_a^-$$

be closed a-normalized subgroups, and let $\mathscr{A}$ be an a-descending U-subordinate σ-algebra. Then

$$H_\mu\left(\mathscr{A} \middle| a^{-1} \cdot \mathscr{A}\right) = h_\mu(a, U) = \mathsf{vol}_\mu^U(a)(x)$$

for μ-almost every $x \in X$. Moreover,

$$h_\mu(a, U') \leqslant h_\mu(a, U).$$

Proof. We start by showing that

$$-\frac{1}{n} \log \mu_x^{a^{-n} \cdot \mathscr{A}}([x]_{\mathscr{A}}) \to H_\mu(\mathscr{A} | a^{-1} \cdot \mathscr{A}).$$

Notice first that by Lemma 9.15 we have

$$\mu_x^{a^{-1}\bullet\mathscr{A}}\big|_{[x]_{\mathscr{A}}} = \mu_x^{a^{-1}\bullet\mathscr{A}}([x]_{\mathscr{A}})\mu_x^{\mathscr{A}}$$

for almost every x, since (after removing a null set from X) $[x]_{a^{-1}\bullet\mathscr{A}}$ is a countable union of $\mathscr{A}$-atoms. More generally, by repeating the same argument we obtain

$$\mu_x^{a^{-n}\bullet\mathscr{A}}([x]_{\mathscr{A}}) = \prod_{i=1}^{n} \mu_x^{a^{-i}\bullet\mathscr{A}}([x]_{a^{-(i-1)}\bullet\mathscr{A}}).$$

Also notice that $\mu_{a\bullet x}^{\mathscr{A}} = a_*\mu_x^{a^{-1}\bullet\mathscr{A}}$ (see Exercise **??**). Combining these and taking logarithms gives

$$\begin{aligned} -\frac{1}{n}\log\mu_x^{a^{-n}\bullet\mathscr{A}}([x]_{\mathscr{A}}) &= \sum_{i=1}^{n}\frac{-\log\mu_x^{a^{-i}\bullet\mathscr{A}}([x]_{a^{-(i-1)}\bullet\mathscr{A}})}{n} \\ &= \frac{1}{n}\sum_{i=0}^{n-1} I_\mu(\mathscr{A}\big|a^{-1}\bullet\mathscr{A})(a^i\bullet x) \\ &\to H_\mu(\mathscr{A}\big|a^{-1}\bullet\mathscr{A}) \end{aligned} \tag{10.10}$$

by the pointwise ergodic theorem (since μ is assumed to be ergodic for the action of a).

We may also obtain in a similar manner that

$$\frac{\log\mu_x^U(a^{-n}B_1^U a^n)}{n} \to \int \log\mu_x^U(a^{-1}B_1^U a)\,\mathrm{d}\mu(x), \tag{10.11}$$

where we assume the normalization $\mu_x^U(B_1^U) = 1$, by applying the argument in the proof of Theorem 10.2(1) on p. 320.

We outline the remainder of the proof. The limits (10.10) and (10.11) measure the growth rate of a dynamically expanded set in relation to a fixed set. By Corollary 9.9, the fact that in (10.10) we are using the conditional measure $\mu_x^{a^{-n}\bullet\mathscr{A}}$ while in (10.11) the leafwise measure μ_x^U seems irrelevant. What is unclear is the precise relationship between the shape $V_{n,x} \subseteq U$ of the atoms $[x]_{a^{-n}\bullet\mathscr{A}} = V_{n,x}\bullet x$ and the set $a^{-n}B_1^U a^n$. We will show below that the influence of the shape is negligible as $n \to \infty$, thereby completing the proof.

Fix $\delta > 0$ such that

$$Y := \{x \in X \mid B_\delta^U\bullet x \subseteq [x]_{\mathscr{A}} \subseteq B_{\delta^{-1}}^U\bullet x\} \tag{10.12}$$

has positive measure. By Lemma 10.9,

$$\lim_{n\to\infty}\frac{\log\mu_x^U(a^{-n}B_r^U a^n)}{n}$$

is independent of r for almost every x. Moreover, for almost every $x \in X$ there exists a sequence $(n_j)_{j\geqslant 1}$ of integers for which $a^{n_j}\bullet x \in Y$. For any j we therefore have

$$[x]_{a^{-n_j}\bullet\mathscr{A}} = a^{-n_j}\bullet[a^{n_j}\bullet x]_{\mathscr{A}} \subseteq a^{-n_j}B^U_{\delta^{-1}}a^{n_j}\bullet x$$

and similarly,

$$[x]_{a^{-n_j}\bullet\mathscr{A}} \supseteq a^{-n_j}B^U_\delta a^{n_j}\bullet x.$$

Thus $a^{-n_j}B^U_\delta a^{n_j} \subseteq V_{n_j,x} \subseteq a^{-n_j}B^U_{\delta^{-1}}a^{n_j}$. Also recall that

$$\mu_x^{a^{-n_j}\bullet\mathscr{A}} \propto \mu_x^U|_{V_{n_j,x}}\bullet x$$

by Corollary 9.9. Hence

$$\mu_x^{a^{-n_j}\bullet\mathscr{A}}([x]_{\mathscr{A}}) = \frac{c(x)}{\mu_x^U(V_{n_j,x})}$$

where $c(x) = \mu_x^U(V_{0,x})$. With this notation, the inclusions above imply that

$$\mu_x^U(a^{-n_j}B^U_\delta a^{n_j}) \leqslant \mu_x^U(V_{n_j,x}) = c(x)\mu_x^{a^{-n_j}\bullet\mathscr{A}}([x]_{\mathscr{A}})^{-1} \leqslant \mu_x^U(a^{-n_j}B^U_{\delta^{-1}}a^{n_j})$$

for almost every x. Taking logarithms, letting $j \to \infty$, and using the independence of the limit in Lemma 10.9 completes the proof.

Assume now that U and U' are closed a-normalized subgroups of G_a^- with $U < U'$ as in the statement of the proposition. From the construction of the σ-algebra in Proposition 10.13, we see that there exist two σ-algebras $\mathscr{A}$ and $\mathscr{A}'$ which are a-descending and subordinate to U and to U' respectively, such that additionally $\mathscr{A} \supseteq \mathscr{A}'$. In order to obtain these, one may use the same finite partition ξ and then carry the construction through with both groups.

We claim that

$$\mathscr{A}' \vee a^{-1}\bullet\mathscr{A} \underset{\mu}{=} \mathscr{A}.$$

We already know one inclusion. To see the other, we describe the atoms for the σ-algebra $\mathscr{C} = \mathscr{A}' \vee a^{-1}\bullet\mathscr{A}$. Suppose that y and x are equivalent with respect to $\mathscr{C}$. Then almost surely there exists some $u \in U$ with $y = u\bullet x$. Notice that we have no initial control over the size of u, since the $a^{-1}\bullet\mathscr{A}$-atoms are in general bigger than the $\mathscr{A}$-atoms. To make this more precise, assume that y, x belong to the set Y which was used in the constructions of the σ-algebras. We cannot assume that $\mathsf{d}(e, u)$ is smaller than the injectivity radius of Y. However, we do know that $y = u'\bullet x$ for some $u' \in U'$ since the points x and y are by assumption $\mathscr{A}'$-equivalent, and that $\mathsf{d}(e, u')$ is less than the injectivity radius. Since for almost every x the G_a^--leaf (that is, the set $G_a^-\bullet x$) is embedded by Lemma 10.8, we must have $u = u'$. This implies that x and $y = u\bullet x$ are $\mathscr{A}$-equivalent, first under the assumption that $x, y \in Y$ and then the general case follows by the same argument and ergodicity after using the minimal $n \geqslant 1$ for which $a^n\bullet x, a^n\bullet y \in Y$, which must exist almost surely by Poincaré recurrence. As the atoms of the σ-algebra determine the σ-algebra modulo μ, the claim follows.

The claim implies the inequality sought, since

$$h_\mu(a,U) = H_\mu\left(\mathscr{A}|a^{-1}\bullet\mathscr{A}\right) = H_\mu\left(\mathscr{A}'|a^{-1}\bullet\mathscr{A}\right) \leqslant H_\mu\left(\mathscr{A}'|a^{-1}\bullet\mathscr{A}'\right) = h_\mu(a,U')$$

by monotonicity of the entropy function with respect to the σ-algebra conditioned on (see [**?**, Prop. **??**]). □

Proof of Theorem 10.2(2). By the argument in Section 10.2, we may assume that μ is ergodic for the action of a. By Proposition 10.21 (to be proved in Section 10.5) there exists a generator ξ for which $\mathscr{A} = \xi_0^\infty$ is G_a^--subordinate. By the Kolmogorov–Sinaĭ Theorem we have

$$h_\mu(a) = h_\mu(a,\xi) = H_\mu\left(\xi_0^\infty \middle| a^{-1}\bullet\xi_0^\infty\right) = \mathsf{vol}_\mu^U(a)(x)$$

for almost every $x \in X$.

If $U < G_a^-$ is a closed a-normalized subgroup, then we also have

$$\mathsf{vol}_\mu^U(a)(x) \leqslant \mathsf{vol}_\mu^{G_a^-}(a)(x) = h_\mu(a)$$

by Proposition 10.20. □

Proof of Theorem 10.2(3). Recall that we wish to show the equivalence of the three statements

- $\mathsf{vol}_\mu^U(a)(x) = 0$;
- μ_x^U is a finite measure;
- μ_x^U is a trivial measure (in the sense of Definition 9.19).

It is clear that triviality implies finiteness. Also, if μ_x^U is finite, then the entropy contribution $\mathsf{vol}_\mu^U(a)$ vanishes because it measures a growth rate (see Lemma 10.9).

Now assume that $\mathsf{vol}_\mu^U(a) = 0$. By Proposition 10.13 (and the assumed ergodicity) there exists a σ-algebra $\mathscr{A}$ which is a-descending and subordinate to U modulo μ. Then

$$H_\mu\left(\mathscr{A}\middle|a^{-1}\bullet\mathscr{A}\right) = \int \left(-\log \mu_x^{a^{-1}\bullet\mathscr{A}}([x]_{\mathscr{A}})\right) \mathrm{d}\mu = 0,$$

which implies that $\mu_x^{a^{-1}\bullet\mathscr{A}}([x]_{\mathscr{A}}) = 1$ almost everywhere, so

$$\mathscr{A} \underset{\mu}{=} a^{-1}\bullet\mathscr{A}.$$

Iterating the same argument shows that

$$a^m\bullet\mathscr{A} \underset{\mu}{=} a^{-m}\bullet\mathscr{A}$$

and therefore $\mu_x^{a^{-m}\bullet\mathscr{A}}\left([x]_{a^m\bullet\mathscr{A}}\right) = 1$ almost everywhere, for all $m \geqslant 1$. By Proposition 9.2, this implies that $\mu_x^U\left(V_{-m,x} \smallsetminus V_{m,x}\right) = 0$ almost everywhere,

where $V_{m,x}$ denotes the shape of the $a^m\bullet\mathscr{A}$-atom of x (cf. p. 331). Using again the set Y in (10.12), we see that the precise shapes do not matter since $V_{-m,x}\nearrow U$ and $V_{m,x}\searrow\{I\}$ as $m\to\infty$ for almost every x. It follows that $\mu_x^U\propto\delta_I$. □

Exercises for Section 10.3

Exercise 10.3.1. (This will be a construction of a good generator for a specific? quasihyperbolic toral automorphism)

10.4 Establishing Invariance using Entropy, Proof of Theorem 10.5

We will see that the proof of Theorem 10.5 rests on convexity of the map $t\mapsto-\log t$ for $t\in\mathbb{R}$. However, we will have to use this convexity on every atom $[x]_{a^{-1}\bullet\mathscr{A}}$ for an a-descending σ-algebra which is subordinate to U. We refer to the second proof of [**?**, theorem:maximalityfortoralgoldenmean] on [**?**, page:secondproofofmaximalityfortoralgoldenmean] for a simpler case of this kind of argument.

Proof of Theorem 10.5. Let $U<G_a^-$ be a closed a-normalized subgroup, and let μ be an a-invariant ergodic probability measure on $X=\Gamma\backslash G$. By Proposition 10.13, there exists an a-descending σ-algebra $\mathscr{A}$ which is subordinate to U modulo μ. By Proposition 10.20, the entropy contribution of U is given by

$$h_\mu(a,U)=H_\mu\left(\mathscr{A}\middle|a^{-1}\bullet\mathscr{A}\right).$$

We wish to show that $h_\mu(a,U)\leqslant J$, where $J=-\log\bigl|\mathrm{mod}(a,U)\bigr|$ is the negative logarithm of the modular character of a restricted to U.

We fix a Haar measure m_U on U, and note that

$$m_U(a^{-1}Ba)=\mathrm{e}^J m_U(B) \tag{10.13}$$

for any measurable $B\subseteq U$. For $x\in X$ recall that we write $V_x\subseteq U$ for the shape of the $\mathscr{A}$-atom so $V_x\bullet x=[x]_{\mathscr{A}}$ almost everywhere. Recall that $\mu_x^{a^{-1}\bullet\mathscr{A}}$ is a probability measure on

$$[x]_{a^{-1}\bullet\mathscr{A}}=a^{-1}[a\bullet x]_{\mathscr{A}}=a^{-1}V_{a\bullet x}a\bullet x$$

satisfying

$$H_\mu(\mathscr{A}\big|a^{-1}\bullet\mathscr{A})=-\int\log\mu_x^{a^{-1}\bullet\mathscr{A}}([x]_{\mathscr{A}})\,\mathrm{d}\mu(x). \tag{10.14}$$

We wish to compare this to a similar expression where we will use (in a careful manner) the Haar measure m_U on U instead of the conditional measures. Notice however that we will always work with the given measure μ on X, and in particular the notion of almost everywhere will always mean with respect to μ. Define τ_x^{Haar} to be the normalized push forward of $m_U|_{a^{-1}V_{a\centerdot x}a}$ under the orbit map, so

$$\tau_x^{\mathrm{Haar}} = \frac{1}{m_U(a^{-1}V_{a\centerdot x}a)} m_U\big|_{a^{-1}V_{a\centerdot x}a}\centerdot x,$$

which once again is a probability measure on $[x]_{a^{-1}\mathscr{A}}$.

We define

$$p(x) = \mu_x^{a^{-1}\centerdot\mathscr{A}}([x]_{\mathscr{A}}),$$

the measure used in (10.14) to define $H_\mu(\mathscr{A}\big|a^{-1}\centerdot\mathscr{A})$. We also define the analogous function

$$p^{\mathrm{Haar}}(x) = \tau_x^{\mathrm{Haar}}([x]_{\mathscr{A}}) = \frac{m_U(V_x)}{m_U(a^{-1}V_{a\centerdot x}a)} = \frac{m_U(V_x)}{m_U(V_{a\centerdot x})}\mathrm{e}^{-J}$$

where we used (10.13). We now take logarithms and average over the orbit to obtain the telescoping sum

$$\frac{1}{n}\sum_{k=0}^{n-1}\log p^{\mathrm{Haar}}(a^k\centerdot x) = \underbrace{\frac{1}{n}\left(\log m_U(V_x) - \log m_U(V_{a^n\centerdot x})\right)}_{\Sigma_n} - J. \tag{10.15}$$

We claim that this implies the identity

$$-\int \log p^{\mathrm{Haar}}(x)\,\mathrm{d}\mu(x) = J. \tag{10.16}$$

For this we need to show that $x \mapsto p^{\mathrm{Haar}}(x)$ is measurable. Since $\mathscr{A}$ is countably generated, the equivalence relation

$$R_{\mathscr{A}} = \{(x,y)\in X\times X \mid [x]_{\mathscr{A}} = [y]_{\mathscr{A}}\} = \bigcap_{n=1}^{\infty}\bigcup_{P\in\xi_n} P\times P \subseteq X\times X,$$

where the intersection on the right-hand side is taken over countable partitions $\xi_n|_{\mathscr{A}}$ of X, and is a measurable subset of $X\times X$. This shows that

$$\begin{aligned} m_U\left(V_x\right) &= m_U\left(\{u\in U \mid (x,u\centerdot x)\in R_{\mathscr{A}}\}\right)\\ &= \lim_{n\to\infty} m_U\left(\{u\in U \mid (x,u\centerdot x)\in R_{\xi_n}\}\right) \end{aligned}$$

is a measurable function of x. Together with the the same argument for $R_{a^{-1}\centerdot\mathscr{A}}$, this shows that $x\mapsto p^{\mathrm{Haar}}(x)$ is measurable.

Next notice that *a priori* we do not know whether $\log p^{\mathrm{Haar}}(x)$ is integrable, but we do have

$$\log p^{\mathrm{Haar}}(x) \leqslant 0$$

almost surely, so we may still apply the pointwise ergodic theorem to this function. Moreover, for almost every x there is a sequence $(n_k)_{k\geqslant 1}$ for which

$$\log m_U(V_x) - \log m_U(V_{a^n\boldsymbol{\cdot}x})$$

is bounded as a function of n. It follows that $\Sigma_{n_k} \to 0$ as $k \to \infty$ in (10.15), giving (10.16).

Recall that the σ-algebras $\mathscr{A}$ and $a^{-1}\boldsymbol{\cdot}\mathscr{A}$ are both subordinate to U modulo μ, which implies (after removing a null set) that the $a^{-1}\boldsymbol{\cdot}\mathscr{A}$-atoms are countable unions of $\mathscr{A}$-atoms.

Thus there is a null set N such that for $x \notin N$ the $\mathscr{A}$-atom of x contains an open neighborhood of x in the U-orbit. We may also assume that for $x \notin N$ there are infinitely many positive and negative integers n for which $a^n\boldsymbol{\cdot}x \in Y$, where Y is the set defined in (10.12). Since U is second countable, this implies that

$$[x]_{a^{-1}\boldsymbol{\cdot}\mathscr{A}} \smallsetminus N = \bigsqcup_{i=1}^{\infty} [x_i]_{\mathscr{A}} \smallsetminus N,$$

where the union is disjoint. For almost every x it is safe to ignore the potentially rather subtle distinction between the set N being a null set for the original measure μ and it being a null set for the conditional measure $\mu_x^{a^{-1}\boldsymbol{\cdot}\mathscr{A}}$, since it will almost surely also be a null set for the conditional measure. However, it may not be a null set for the measure τ_x^{Haar}. To deal with this possibility we write

$$[x]_{a^{-1}\boldsymbol{\cdot}\mathscr{A}} = \bigcup_{i=1}^{\infty} [x_i]_{\mathscr{A}} \cup N_x,$$

where N_x is a null set for $\mu_x^{a^{-1}\boldsymbol{\cdot}\mathscr{A}}$ (but is possibly not a null set for τ_x^{Haar}) for each $x \in X$. We may assume that $\mu_x^{a^{-1}\boldsymbol{\cdot}\mathscr{A}}([x_i]_{\mathscr{A}}) > 0$, because if $[x_i]_{\mathscr{A}}$ is a null set for $\mu_x^{a^{-1}\boldsymbol{\cdot}\mathscr{A}}$ then we could simply remove it from the list and increase N_x to compensate. Thus

$$\sum_{i=1}^{\infty} \mu_x^{a^{-1}\boldsymbol{\cdot}\mathscr{A}}([x_i]_{\mathscr{A}}) = 1,$$

but *a priori* we only know that

$$\sum_{i=1}^{\infty} \tau_x^{\mathrm{Haar}}([x_i]_{\mathscr{A}}) \leqslant 1.$$

We now integrate $\log p^{\mathrm{Haar}} - \log p$ over the atom $[x]_{a^{-1}\boldsymbol{\cdot}\mathscr{A}}$ to get

$$K(x)=\int \log p^{\mathrm{Haar}}\,\mathrm{d}\mu_x^{a^{-1}\bullet\mathscr{A}}-\int \log p\,\mathrm{d}\mu_x^{a^{-1}\bullet\mathscr{A}}, \tag{10.17}$$

but since both functions are constant on the $\mathscr{A}$-atoms and N_x is a null set for the measure $\mu_x^{a^{-1}\bullet\mathscr{A}}$, this integral is simply the sum

$$\sum_{i=1}^{\infty}\left(\log\frac{\tau_x^{\mathrm{Haar}}([x_i]_{\mathscr{A}})}{\mu_x^{a^{-1}\bullet\mathscr{A}}([x_i]_{\mathscr{A}})}\right)\mu_x^{a^{-1}\bullet\mathscr{A}}([x_i]_{\mathscr{A}}).$$

By convexity of the map $t\mapsto-\log t$ for $t\in\mathbb{R}$, with $\mu_x^{a^{-1}\bullet\mathscr{A}}([x_i]_{\mathscr{A}})$ as the weights at the points $t_i=\frac{\tau_x^{\mathrm{Haar}}([x_i]_{\mathscr{A}})}{\mu_x^{a^{-1}\bullet\mathscr{A}}([x_i]_{\mathscr{A}})}$, we get

$$\begin{aligned}K(x)&=\sum_{i=1}^{\infty}\log(t_i)\mu_x^{a^{-1}\bullet\mathscr{A}}([x_i]_{\mathscr{A}})\\&\leqslant\log\left(\sum_{i=1}^{\infty}t_i\mu_x^{a^{-1}\bullet\mathscr{A}}([x_i]_{\mathscr{A}})\right)\\&=\log\left(\sum_{i=1}^{\infty}\tau_x^{\mathrm{Haar}}([x_i]_{\mathscr{A}})\right)\\&=\log\tau_x^{\mathrm{Haar}}\left(\bigcup_{i=1}^{\infty}[x_i]_{\mathscr{A}}\right)\leqslant 0.\end{aligned}\tag{10.18}$$

Integrating this inequality over all of X and recalling the relation between the function p and the entropy contribution $h_\mu(a,U)=H_\mu(\mathscr{A}\big|a^{-1}\bullet\mathscr{A})$, and the relation between the function p^{Haar} and J, gives the inequality sought.

In the case of equality, we use strict convexity of the map $t\mapsto-\log t$ as follows. If $h_\mu(a,U)=J$, then the integral of the function K in (10.17), which is non-positive by (10.18), vanishes. It follows that for almost every $x\in X$, both sides of (10.18) vanish. Thus $\tau_x^{\mathrm{Haar}}(N_x)=0$ for almost every x, and moreover $t_i=1$ for all i by strict convexity of the map $t\mapsto-\log t$. Notice that the condition $t_i=1$ for all i means that the conditional measure $\mu_x^{a^{-1}\bullet\mathscr{A}}$ gives the same weight to the $\mathscr{A}$-atoms $[x_i]_{\mathscr{A}}$ as does the normalized Haar measure τ_x^{Haar} on the $a^{-1}\bullet\mathscr{A}$-atoms†. Rephrase footnote later

Now $H_\mu(a^k\bullet\mathscr{A}\big|a^{-\ell}\bullet\mathscr{A})=(k+\ell)h_\mu(a,U)=(k+\ell)J$ for any $k,\ell\geqslant 0$, so we may apply the same argument to deduce that the conditional measure $\mu_x^{a^{-\ell}\bullet\mathscr{A}}$ gives the same weight to the $a^k\bullet\mathscr{A}$-atoms as does the normalized Haar measure on the $a^{-\ell}\bullet\mathscr{A}$-atoms.

For almost every x the $a^{-\ell}\bullet\mathscr{A}$-atom can be made arbitrarily large, as we may find a sequence $\ell_n\to\infty$ for which $a^{\ell_n}\bullet x\in Y$. For each ℓ the various $a^k\bullet\mathscr{A}$-atoms for $k\geqslant 0$ generate the Borel σ-algebra on the $a^{-\ell}\bullet\mathscr{A}$-atoms,

† A similar argument is used in [**?**, Sec. 2.3] in rather heavy disguise. The fundamental phenomena goes back to [**?**, Prop. **??**], and we have now seen it in several contexts and in several guises (see, for example, [**?**, Sec. **??**; Ch. **??**]).

at least on the complement of N which is a null set for $\mu_x^{a^{-\ell}\cdot\mathscr{A}}$ and for the normalized Haar measure on the atom. This may be seen as follows. For $\mu_x^{a^{-\ell}\cdot\mathscr{A}}$-almost every y, the $a^k\cdot\mathscr{A}$-atom of y can be made to have arbitrarily small diameter, since for $y \notin N$ there is a sequence $k_n \to \infty$ with $a^{-k_n}\cdot y \in Y$. This shows that $\mu_x^{a^{-\ell}\cdot\mathscr{A}}$ coincides with the normalized Haar measure on the atom $[x]_{a^{-\ell}\cdot\mathscr{A}}$. Using this for all $\ell \geqslant 1$, we see that the leafwise measure μ_x^U is the Haar measure on U, and so μ is U-invariant by Theorem 9.26. This concludes the proof of Theorem 10.5. □

Exercises for Section 10.4

Exercise 10.4.1. Find the places in the proof of Theorem 10.5 where the fact that τ_x^{Haar} is indeed defined using the Haar measure m_U on U (and not, for example, by the Haar measure on a subgroup of U) is used.

10.5 Construction of a Generator

In proving Theorem 10.2 we made essential use of the existence of a generator with good properties adapted to the geometry of the action of a.

Proposition 10.21. *Let $X = \Gamma\backslash G$ be a quotient of a simply-connected Lie group or S-algebraic group by a discrete subgroup, let $a \in G$, and let μ be an a-invariant ergodic probability measure on X. Then there exists a generator ξ for the action of a with finite entropy, for which $\mathscr{A} = \xi_0^\infty$ is a-descending and G_a^--subordinate.*

In addition to the existence of sets and partitions with thin boundary as in Lemma 10.16, we will also need the following lemma which is more closely adapted to the setup of Theorem 10.2, and which relies on the Lipschitz property of the a-action together with the finite-dimensionality of G.

Lemma 10.22. *Let G be a real Lie group or an S-algebraic group, and let a be an element of G. Then, for any $r \in (0,1)$, there exists some $\kappa = \kappa(G,a,r)$ such that for any compact set $\Omega \subseteq G$ there exists a constant $c = c(\Omega,a,r)$ with the property that for any $n \geqslant 1$ the set Ω can be covered by $c\mathrm{e}^{\kappa n}$ sets of the form*

$$g\bigcap_{k=0}^{n} a^{-k}B_r^G a^k$$

with $g \in \Omega$.

Proof of Lemma 10.22 for a Lie group. Let G be a real Lie group equipped with a left-invariant Riemannian metric. Then by [**?**, Sec. 9.3.2] we

have
$$\mathsf{d}(g_1a^{-1},g_2a^{-1})\leqslant\|\operatorname{Ad}_a\|\mathsf{d}(g_1,g_2),$$
which implies that
$$a^kB^G_{\|\operatorname{Ad}_a\|^{-n}r}a^{-k}\subseteq B^G_r$$
for all $n\geqslant 0$ and $k=0,\dots,n$. It follows that
$$B^G_{\|\operatorname{Ad}_a\|^{-n}r}\subseteq\bigcap_{k=0}^n a^{-k}B^G_ra^k,$$
so it is sufficient to show that Ω can be covered by no more than $c\mathrm{e}^{\kappa n}$ balls with radius $\|\operatorname{Ad}_a\|^{-n}r$.

To see this, notice first that it is enough to consider the case where $\Omega=B^G_\varepsilon$ is a small ball around the identity. This case will imply the general case by covering Ω with a bounded number of left-translates of B^G_ε.

Finally, if $\varepsilon>0$ is sufficiently small, then the logarithm map $\log:B^G_\varepsilon\to\mathfrak{g}$ is bi-Lipschitz (see, for example [**?**, Sec. 9.3.2]). However, this linearizes the problem, reducing it to the claim that a compact set in $\mathbb{R}^d$ can be covered by $\ll\mathrm{e}^{\kappa n}$ balls of radius $\ll\|\operatorname{Ad}_a\|^{-n}$, which is clear. □

Proof of Lemma 10.22 for an S-algebraic group. We first consider the case $S=\{\sigma\}$. If $\sigma=\infty$, then we may apply the case of a real Lie group considered above to $\mathrm{SL}_m(\mathbb{R})$, which implies the lemma for $G_\infty\subseteq\mathrm{SL}_m(\mathbb{R})$.

Suppose now that $\sigma=p$. As in the real case, it is sufficient to prove the lemma for a fixed neighborhood, for example $\Omega=B^{G_p}_1(I)=G_p\cap\mathrm{SL}_m(\mathbb{Z}_p)$, of the identity. Notice that $B^{G_p}_1$ is an open compact subgroup of G_p, and hence that
$$B^{G_p}_1\cap a^{-1}B^{G_p}_1a\leqslant B^{G_p}_1$$
is of finite index which we denote by I. Then
$$a^{-1}B^{G_p}_1a\cap a^{-2}B^{G_p}_1a^2\leqslant a^{-1}B^{G_p}_1a$$
also has index I, which implies that
$$B^{G_p}_1\cap a^{-1}B^{G_p}_1a\cap a^{-2}B^{G_p}_1a^2\leqslant B^{G_p}_1$$
has index I^2. Repeating this, we conclude by induction that
$$B^{G_p}_1\cap a^{-1}B^{G_p}_1a\cap\cdots\cap a^{-n}B^{G_p}_1a^n\leqslant B^{G_p}_1$$
has index I^n. This implies the lemma for G_p and $r=1$. For a general $r\in(0,1)$, notice that
$$B^{G_p}_r\leqslant B^{G_p}_1$$
has finite index, denoted $c=c(r)$. This implies that

$$\bigcap_{k=0}^{n} a^{-k} B_r^{G_p} a^k \leqslant B_1^{G_p}$$

has index no more than $c^n I^n$ as required.

Now assume that $|S| < \infty$. As the lemma already holds by the argument above for each individual $\sigma \in S$, the definition of the metric on G_S in Lemma 8.1 shows that the lemma also holds for a finite set of places.

To extend the argument above to an arbitrary set S, notice first that if $S = \{p\}$ and $a \in G \cap \mathrm{SL}_m(\mathbb{Z}_p)$, then right-multiplication by a is an isometry on $G \cap \mathrm{SL}_m(\mathbb{Z}_p)$ for the metric constructed in Lemma 8.1. By left-invariance of the metric, this extends to all of G.

Assume now that $|S| = \infty$ and $a \in G$. Then $a_p \in G_p \cap \mathrm{SL}_m(\mathbb{Z}_p)$ for all but finitely many $p \in S$. Let

$$S_0 = (\{\infty\} \cap S) \cup \left\{ p \in S \mid a_p \notin G_p \cap \mathrm{SL}_m(\mathbb{Z}_p) \text{ or } G_p \cap \mathrm{SL}_m(\mathbb{Z}_p) \nsubseteq B_r^G \right\}.$$

Then $|S_0| < \infty$, and

$$\bigcap_{k=0}^{n} a^{-k} B_r^G a^k \supseteq \prod_{\sigma \in S_0} \bigcap_{i=0}^{n} \left(a_\sigma^{-k} B_r^{G_\sigma} a_\sigma^k \right) \times \prod_{p \in S \smallsetminus S_0} G_p \cap \mathrm{SL}_m(\mathbb{Z}_p),$$

by the definition of the metric. On the other hand, G_{S,S_0} is open in G_S, so that Ω can be covered by finitely many G_{S,S_0}-orbits. Therefore, the lemma for S_0 discussed above implies the lemma for S. □

We are now ready to start the proof of Proposition 10.21. However, before we do this we would like to point out the main difficulty that will arise in the argument.

For some small parameter ρ, apply Lemma 10.16 to find a partition ξ of X. If now

$$x, a{\scriptstyle\bullet}x, \dots, a^k{\scriptstyle\bullet}x \in \Omega$$

and $x, y = g{\scriptstyle\bullet}x$ are equivalent to each other with respect to ξ_0^∞ (that is, $a^\ell{\scriptstyle\bullet}x$ and $a^\ell{\scriptstyle\bullet}y$ lie in the same partition element of ξ for all $\ell \geqslant 0$), then we readily see that

$$\mathsf{d}(a^i g a^{-i}, I) < \rho$$

for $i = 0, \dots, k$. However, if $x \in \Omega$ but $a{\scriptstyle\bullet}x, a^2{\scriptstyle\bullet}x, \dots, a^{k-1}{\scriptstyle\bullet}x \notin \Omega$, while finally $a^k{\scriptstyle\bullet}x \in \Omega$ then this argument will fail unless we choose at the outset a value of ρ adapted to k. However, since the return time k is in general not known at the outset, we cannot choose ρ small enough to cope with all eventualities in this sense. Thus we will have to control the partition more carefully, taking into account all possible return times to Ω.

Proof of Proposition 10.21. In this proof we will construct four partitions, $\eta, \widetilde{\eta}, \zeta$, and ξ, where at every step we refine the previous partitions in some particular way. We will be concerned in particular with their entropies $H_\mu(\eta), H_\mu(\widetilde{\eta}), H_\mu(\zeta)$ and $H_\mu(\xi)$ and with lower bounds on the sizes of the atoms of the σ-algebras $\eta_0^\infty, \widetilde{\eta}_0^\infty, \zeta_0^\infty$ and ξ_0^∞. For the last two partitions ζ and ξ we will also prove various upper bounds on the atoms, eventually showing that ξ_0^∞ is subordinate to G_a^-.

For clarity the proof will be broken into paragraphs.

Construction of the finite partition η. Let $x_0 \in \operatorname{Supp}\mu \subseteq X$ be an arbitrary starting point, and let r_{x_0} be the injectivity radius of $X = \Gamma\backslash G$ at x_0. Applying Lemma 10.16, we see that there exists some positive $r < \frac{1}{16} r_{x_0}$ such that $\Omega = B_r^X(x_0)$ has μ-thin boundary. Below we will impose an additional requirement on how small r should be. We define

$$\eta = \{\Omega, X \smallsetminus \Omega\}.$$

Clearly $H_\mu(\eta) \leqslant \log 2 < \infty$.

Lower bound on atoms for η_0^∞. By Lemma 10.18, for almost every x there exists some $\delta > 0$ with

$$B_\delta^{G_a^-} \cdot x \subseteq [x]_{\eta_0^\infty}.$$

Construction of the partition $\widetilde{\eta}$. We define

$$\widetilde{\eta} = \{X \smallsetminus \Omega, \Omega_1, \Omega_2, \dots\},$$

where

$$\begin{aligned}
\Omega_1 &= \Omega \cap a^{-1} \cdot \Omega \\
\Omega_2 &= \left(\Omega \smallsetminus a^{-1} \cdot \Omega\right) \cap a^{-2} \cdot \Omega \\
&\;\;\vdots \\
\Omega_n &= \left(\Omega \smallsetminus \left(a^{-1} \cdot \Omega \cup a^{-2} \cdot \Omega \cup \cdots \cup a^{-n+1} \cdot \Omega\right)\right) \cap a^{-n} \cdot \Omega
\end{aligned}$$

for any $n \geqslant 1$. In other words, in constructing the partition $\widetilde{\eta}$ we split Ω into countably many sets according to when a given point in Ω next visits Ω (under forward iteration of the action of a). Notice that a consequence of Poincaré recurrence is that

$$\mu\left(\{x \in \Omega \mid a^n \cdot x \notin \Omega \text{ for all } n \geqslant 0\}\right) = 0.$$

It follows that $\widetilde{\eta}$ is a partition of X modulo μ.

Lower bound on atoms of $\widetilde{\eta}_0^\infty$. Notice that $\widetilde{\eta} \leqslant \eta_0^\infty$, which implies once more that for almost every $x \in X$ there exists some $\delta > 0$ with

$$B_\delta^{G_a^-} \cdot x \subseteq [x]_{\widetilde{\eta}_0^\infty}. \tag{10.19}$$

Finite entropy of $\widetilde{\eta}$. We claim that the countable partition $\widetilde{\eta}$ has finite entropy,

$$H_\mu(\widetilde{\eta}) < \infty,$$

where we will use a convenient decomposition of the infinite sum. Notice first that the sets

$$\begin{aligned}
&\Omega_1\\
&\Omega_2, a\bullet\Omega_2\\
&\Omega_3, a\bullet\Omega_3, a^2\bullet\Omega_3\\
&\quad\vdots\\
&\Omega_n, a\bullet\Omega_n, \dots, a^{n-1}\Omega_n
\end{aligned}$$

are all pairwise disjoint, and that (by ergodicity) their union is a co-null set in X, so

$$\sum_{n=1}^{\infty} n\mu(\Omega_n) = 1. \tag{10.20}$$

It follows that

$$\begin{aligned}
H_\mu(\widetilde{\eta}) &= -\mu(X\smallsetminus\Omega)\log\mu(X\smallsetminus\Omega) - \sum_{n=1}^{\infty}\mu(\Omega_n)\log\mu(\Omega_n)\\
&\leqslant -\mu(X\smallsetminus\Omega)\log\mu(X\smallsetminus\Omega) + \sum_{n:\mu(\Omega_n)>\mathrm{e}^{-n}} n\mu(\Omega_n) + \sum_{n:\mu(\Omega_n)\leqslant\mathrm{e}^{-n}} n\mathrm{e}^{-n} + C_1\\
&< \infty,
\end{aligned}$$

where we have used monotonicity of the function $t \mapsto -\log t$ for those n with $\mu(\Omega_n) > \mathrm{e}^{-n}$, and monotonicity of the function $t \mapsto -t\log t$ for small values of t for those n with $\mu(\Omega_n) \leqslant \mathrm{e}^{-n}$. The constant C_1 accommodates the finitely many cases for which the monotonicity of $t \mapsto -t\log t$ does not apply.

Construction of the partition ζ. We now apply Lemma 10.22 with the assumption that $r < \frac{1}{16}r_{x_0}$ to the set Ω. Thus, each Ω_n may be covered by $f_n \leqslant c\mathrm{e}^{\kappa n}$ sets of the form

$$x_i^{(n)}\bigcap_{k=0}^{n} a^{-k}B_r^G a^k$$

with $x_i^{(n)} \in \Omega_n$ for $i = 1, \dots, f_n$. We could use these sets to refine Ω_n, but in order to preserve the same lower bound we instead define sets that are larger in the direction of G_a^-, and then use these to construct a refinement of Ω_n. Define sets by

$$\begin{aligned}
B_1^{(n)} &= \Omega_n \cap x_1^{(n)} \left(\bigcap_{k=0}^{n} a^{-k} B_r^G a^k \right) B_{4r}^{G_a^-}, \\
B_2^{(n)} &= \Omega_n \smallsetminus B_1^{(n)} \cap x_2^{(n)} \left(\bigcap_{k=0}^{n} a^{-k} B_r^G a^k \right) B_{4r}^{G_a^-}, \\
&\ \ \vdots \\
B_{f_n}^{(n)} &= \Omega_n \smallsetminus \left(B_1^{(n)} \cup \dots \cup B_{f_n-1}^{(n)} \right) \cap x_{f_n}^{(n)} \left(\bigcap_{k=0}^{n} a^{-k} B_r^G a^k \right) B_{4r}^{G_a^-}.
\end{aligned}$$

With these sets we define the partition

$$\zeta = \{X \smallsetminus \Omega, B_i^{(n)} \mid n \geqslant 1, i = 1, \dots, f_n\}.$$

Notice that, by construction, ζ refines $\widetilde{\eta}$. Roughly speaking, since Ω has diameter less than $2r < \frac{1}{8} r_{x_0}$, and we apply $B_{4r}^{G_a^-}$ to all the sets used above, we should think of the partition of Ω_n into the sets above as being *transverse* to the G_a^--orbits, and this observation will be used below to prove the lower bound.

Lower bound for the atoms of ζ_0^∞. We will show that

$$B_\delta^{G_a^-} \cdot x \subseteq [x]_{\zeta_0^\infty} \tag{10.21}$$

whenever $\delta d \leqslant r$ and (10.19) holds for $x \in X$, where d is chosen as in Lemma 10.17 so that

$$a^n B_\delta^{G_a^-} a^{-n} \subseteq B_r^{G_a^-} \tag{10.22}$$

for all $n \geqslant 0$. Suppose therefore that $y = u \cdot x$ with $u \in B_\delta^{G_a^-}$ and with

$$[x]_{\widetilde{\eta}_0^\infty} = [y]_{\widetilde{\eta}_0^\infty}.$$

Let $m \geqslant 0$ be fixed, and notice that $a^m \cdot x$ and $a^m \cdot y = (a^m u a^{-m}) \cdot (a^m \cdot x)$ belong to the same element of $\widetilde{\eta}$. If this element is $X \smallsetminus \Omega$, then $a^m \cdot x$ and $a^m \cdot y$ also belong to the same element of ζ (specifically, to $X \smallsetminus \Omega$). So suppose now that

$$a^m \cdot x, a^m \cdot y \in \Omega_n$$

for some $n \geqslant 1$, and that

$$x' = a^m \cdot x \in \Omega_n \cap x_\ell^{(n)} \left(\bigcap_{k=0}^{n} a^{-k} B_r^G a^k \right) B_{4r}^{G_a^-} \tag{10.23}$$

for some $\ell \in \{1, \dots, f_n\}$. We claim that in this case

$$y' = a^m \mathbin{\raisebox{0.1ex}{$\scriptstyle\bullet$}} y \in \Omega_n \cap x_\ell^{(n)} \left(\bigcap_{k=0}^{n} a^{-k} B_r^G a^k \right) B_{4r}^{G_a^-} \tag{10.24}$$

also. By the symmetry between x and y and the construction of the sets, this then implies that $x', y' \in B_\ell^{(n)} \in \zeta$ for the same ℓ. Applying this for all $n \geqslant 1$, we obtain the lower bound in (10.21).

So let $x' = x_\ell^{(n)} g u_x$ with

$$g \in \bigcap_{k=0}^{n} a^{-k} B_r^G a^k$$

and $u_x \in B_{4r}^{G_a^-}$ as in (10.23). This implies that $g u_x \in B_{5r}^G$ and $x', x_\ell^{(n)} \in \Omega = B_r(x_0)$, while the injectivity radius at x_0 is $r_{x_0} > 16r$ by choice of r. Therefore, $g u_x \in B_{2r}^G$ and so $u_x \in B_{3r}^G$. However, $y' = a^m u a^{-m} \bullet x'$ with

$$a^m u a^{-m} \in B_r^{G_a^-}$$

by (10.22). Together this gives

$$y' = x_\ell^{(n)} g \Big(\underbrace{u_x a^m u^{-1} a^{-m}}_{\in B_{4r}^{G_a^-}} \Big)$$

as claimed in (10.24), completing the proof of (10.21).

Proof of finite entropy for ζ. For each $n \geqslant 1$ we define

$$\mu_n = \frac{1}{\mu(\Omega_n)} \mu|_{\Omega_n}$$

to be the normalized restriction of μ to Ω_n. Then

$$H_{\mu_n}(\zeta) \leqslant \log f_n \leqslant \log c + \kappa n,$$

since the partition ζ when restricted to Ω_n contains only f_n elements modulo μ_n. Moreover

$$H_\mu(\zeta) = H_\mu(\widetilde{\eta}) + H_\mu(\zeta \big| \widetilde{\eta})$$

by [**?**, Prop. **??**], and so by (10.20)

$$\begin{aligned} H_\mu(\zeta \big| \widetilde{\eta}) &= \sum_{n=1}^{\infty} \mu(\Omega_n) H_{\mu_n}(\zeta) \\ &\leqslant \log c + \kappa \sum_{n=1}^{\infty} n \mu(\Omega_n) < \infty, \end{aligned}$$

which together with the fact that $H_\mu(\widetilde{\eta}) < \infty$ implies that $H_\mu(\zeta) < \infty$.

Upper bound for the atoms of ζ_0^∞. We claim that for almost every x in Ω we have

$$[x]_{\zeta_0^\infty} \subseteq \left(\bigcap_{k=0}^{\infty} a^{-k} B^G_{4(1+d)r} a^k \right) \cdot x, \tag{10.25}$$

which implies for almost every $x \in X \smallsetminus \Omega$ that

$$[x]_{\zeta_0^\infty} \subseteq \left(\bigcap_{k=n}^{\infty} a^{-k} B^G_{4(1+d)r} a^k \right) \cdot x,$$

where $n \geqslant 1$ is chosen so that $a^n \cdot x \in \Omega$. To see (10.25), note first that if $x \in \Omega_n$ then

$$[x]_\zeta \subseteq x_\ell^{(n)} \left(\bigcap_{k=0}^{n} a^{-k} B^G_r a^k \right) B^{G_a^-}_{4r}.$$

By Lemma 10.17 we also have

$$B^{G_a^-}_{6r} \subseteq \bigcap_{k=0}^{n} a^{-k} B^G_{4dr} a^k,$$

so that

$$[x]_\zeta \subseteq x \bigcap_{k=0}^{n} a^{-k} B^G_{4(1+d)r} a^k.$$

[MLE] need to verify the $6r$

On the other hand, we have

$$[a^n \cdot x]_\zeta \subseteq (a^n \cdot x) \bigcap_{k=0}^{n_1} a^{-k} B^G_{4(1+d)r} a^k$$

almost surely, where $n_1 \geqslant 1$ is such that $a^n \cdot x \in \Omega_{n_1}$. This implies that

$$[x]_{\zeta_0^\infty} \subseteq x \bigcap_{k=0}^{n_1+n} a^{-k} B^G_{4(1+d)r} a^k.$$

Repeating this argument successively, starting with

$$a^{n+n_1} \cdot x \in \Omega_{n_2},$$

gives (10.25).

Understanding the upper bound. We now examine the intersection

$$D = \bigcap_{k=0}^{\infty} a^{-k} B^G_{4(1+d)r} a^k$$

more carefully. Suppose that $g \in D$. We claim that if r is sufficiently small then either $g \in G_a^-$ (the desired case) or there exists some $s = s(g) > 0$ for which

$$\mathsf{d}(a^n g a^{-n}, G_a^-) \geqslant s \tag{10.26}$$

for all $n \geqslant 0$.

Proof of (10.26) for a Lie group. If r is sufficiently small, then we may write

$$w = \log g \in \mathfrak{g},$$

which by assumption has the property that

$$\{\mathrm{Ad}_a^n(w) \mid n \geqslant 0\}$$

is bounded. This implies that (in the complexification of the Lie algebra) w is a sum of generalized eigenvectors of Ad_a for eigenvalues of absolute value less than one and of (non-generalized) eigenvectors of Ad_a with eigenvalues of absolute value equal to one. If the latter eigenvalues appear in w (that is, if the projection of w onto the corresponding sum of eigenspaces is non-zero) then $\|\mathrm{Ad}_a^n(w)\|$ is bounded from below. Moreover, we also see that the distance from $\mathrm{Ad}_a^n(w)$ to the Lie algebra of G_a^- is bounded from below. Since the logarithm and exponential maps are locally Lipschitz, the claim (10.26) follows.

Proof of (10.26) for an S-algebraic group. As above, we know by assumption that

$$w = g - I \in \prod_{\sigma \in S} \mathrm{Mat}_d(\mathbb{Q}_\sigma)$$

has the property that

$$\{a^n w a^{-n} \mid n \geqslant 0\}$$

is bounded. Arguing as before using linear algebra over $\mathbb{Q}_\sigma$ for all $\sigma \in S$, we deduce that either w is a sum of generalized eigenvectors for the conjugation by a with eigenvalues all of absolute value less than one, or that for some σ the distance from $a_\sigma^n w a_\sigma^{-n}$ to $(G_\sigma)_{a_\sigma}^- - I$ has a lower bound. This gives (10.26) once again.

Construction of the partition ξ. To obtain the generator ξ we need to refine the partition ζ so as to ensure that for every $\delta > 0$ there exists an element $P_\delta \in \xi$ which satisfies

$$P_\delta \subseteq y_\delta B_\delta^G B_{4r}^{G_a^-} \tag{10.27}$$

for some $y_\delta \in \mathit{\Omega}$. This will allow us to improve the upper bound for ζ_0^∞ to the statement that ξ_0^∞ is subordinate to G_a^- modulo μ.

So let $Q \in \zeta$ be any of the partition elements with $Q \subseteq \mathit{\Omega}$ and $\mu(Q) > 0$. If there exists some $y_0 \in Q$ with

$$\mu\left(Q\cap y_0B_{4r}^{G_a^-}\right)>0,$$

then we simply replace the set Q by the two sets $Q\cap y_0B_{4r}^{G_a^-}$ and $Q\smallsetminus y_0B_{4r}^{G_a^-}$, obtaining a partition ξ satisfying (10.27). So we may assume that

$$\mu\left(Q\cap y_0B_{4r}^{G_a^-}\right)=0$$

for all $y_0\in Q$. Then we can inductively find points $y_k\in Q$ and values δ_k in $(0,\frac{1}{k})$ with

$$\begin{aligned}\mu\left(Q\cap y_1B_{\delta_1}^{G}B_{4r}^{G_a^-}\right)&\in\left(0,\tfrac12\mu(Q)\right),\\ \mu\left(Q\smallsetminus y_1B_{\delta_1}^{G}B_{4r}^{G_a^-}\cap y_2B_{\delta_2}^{G}B_{4r}^{G_a^-}\right)&\in\left(0,\tfrac14\mu(Q)\right),\\ &\vdots\\ \mu\Bigl(\underbrace{Q\smallsetminus\left(y_1B_{\delta_1}^{G}B_{4r}^{G_a^-}\cup\cdots\cup y_{k-1}B_{\delta_{k-1}}^{G}B_{4r}^{G_a^-}\right)\cap y_kB_{\delta_k}^{G}B_{4r}^{G_a^-}}_{Q_k}\Bigr)&\in\left(0,\tfrac{1}{2^k}\mu(Q)\right)\end{aligned}$$

for $k\geqslant 1$. We define

$$\xi=\zeta\vee\left\{X\smallsetminus\bigcup_{\ell=1}^{\infty}Q_\ell,Q_k\mid k\in\mathbb{N}\right\}$$

Finite entropy of ξ. Monotonicity of the map $t\mapsto -t\log t$ for small values of t implies that

$$\begin{aligned}H_\mu\left(\xi\middle|\zeta\right)&=\mu(Q)\left(-\sum_{n=1}^{\infty}\frac{\mu(Q_n)}{\mu(Q)}\log\frac{\mu(Q_n)}{\mu(Q)}\right.\\ &\qquad\left.-\frac{\mu\left(Q\smallsetminus\bigcup_{k=1}^{\infty}Q_k\right)}{\mu(Q)}\log\frac{\mu\left(Q\smallsetminus\bigcup_{k=1}^{\infty}Q_k\right)}{\mu(Q)}\right)\\ &\leqslant\mu(Q)\left(-\sum_{n=1}^{\infty}\frac{1}{2^n}\log\frac{1}{2^n}+C\right)<\infty,\end{aligned}$$

so $H_\mu(\xi)<\infty$. Here C is some constant added to handle the finitely many terms for which the monotonicity cannot be used.

Lower bound for the atoms of ξ_0^∞. Our definition of ξ as a refinement of ζ shares an essential feature of the definition of ζ as a refinement of $\widetilde{\eta}$. Once again we applied $B_{4r}^{G_a^-}$ on the right to each of the sets that we used in the refinement. Using this feature, the proof of the lower bound for the atoms of ζ_0^∞ also goes through for the atoms of ξ_0^∞.

Upper bound for the atoms of ξ_0^∞. By the upper bound for the atoms of ζ_0^∞, we already know that

$$[x]_{\xi_0^\infty} \subseteq \left(\bigcap_{k=0}^{\infty} a^{-k} B^G_{4(1+d)r} a^k \right) \centerdot x$$

for almost every $x \in \Omega$. Now suppose that $y = g\centerdot x \in [x]_{\xi_0^\infty}$ for some

$$g \in \bigcap_{k=0}^{\infty} a^{-k} B^G_{4(1+d)r} a^k \cap B^G_{2r}.$$

We claim that this implies almost surely that $g \in B_{2r}^{G_a^-}$. Suppose this is not so, in which case

$$\mathsf{d}\left(a^k g a^{-k}, G_a^-\right) \geqslant s \tag{10.28}$$

for some $s > 0$ by (10.26). However, ξ contains a set P_δ of positive measure with

$$P_\delta \subseteq y_s B^G_\delta B^{G_a^-}_{4r}$$

for some $y_s \in \Omega$. By ergodicity and the assumption regarding y, there exists some $n \geqslant 1$ with $a^n\centerdot x, a^n\centerdot y \in P_\delta$ almost surely. We have $a^n\centerdot x = y_s g_x u_x$ and $a^n\centerdot y = y_s g_y u_y$, or equivalently $a^n\centerdot y = (g_y u_y)^{-1}\centerdot z_s$ and $z_s = (g_x u_x)\centerdot(a^n\centerdot x)$ with $g_x, g_y \in B^G_\delta$ and $u_x, u_y \in B^{G_a^-}_{4r}$. It follows that

$$a^n\centerdot y = (a^n g)\centerdot x = \left[(g_y u_y)^{-1}(g_x u_x)a^n\right]\centerdot x.$$

However, since

$$\mathsf{d}\big(a^n g a^{-n}, e\big) < 4(1+d)r$$

and

$$\mathsf{d}\big((g_y u_y)^{-1} g_x u_x, e\big) < 2\delta,$$

we have

$$a^n g a^{-n} = (g_y u_y)^{-1} g_x u_x$$

for sufficiently small δ and r. This shows that

$$\begin{aligned}
\mathsf{d}\big(a^n g a^{-n}, u_y^{-1} u_x\big) &= \mathsf{d}\big(u_y a^n g a^{-n}, u_x\big)\\
&= \mathsf{d}\big(g_y^{-1} g_x u_x, u_x\big)\\
&\leqslant \|\operatorname{Ad}_{u_x}^{-1}\| \mathsf{d}\big(g_y^{-1} g_x, e\big)\\
&\leqslant \|\operatorname{Ad}_{u_x}^{-1}\| 2\delta,
\end{aligned}$$

which is smaller than s for sufficiently small $\delta > 0$. This contradiction of (10.28) shows that for almost every $x \in \Omega$ we must have

$$[x]_{\xi_0^\infty} \subseteq B^{G_a^-}_{2r}\centerdot x.$$

If $x \in X \smallsetminus \Omega$ and $a^n \cdot x \in \Omega$ then we have

$$[x]_{\xi_0^\infty} \subseteq \left(a^{-n} B_{2r}^{G_a^-} a^n\right) \cdot x.$$

In other words, ξ is a finite entropy partition for which ξ_0^∞ is subordinate to G_a^- modulo μ.

ξ IS A GENERATOR. It remains to show that ξ is a generator with respect to a. Let $n \leqslant 0$ be chosen with $a^n \cdot x \in \Omega$. Then by the argument above we have

$$[a^n \cdot x]_{\xi_0^\infty} \subseteq B_{2r}^{G_a^-} \cdot (a^n \cdot x)\,,$$

or equivalently that the atom $[x]_{\xi_n^\infty}$ has a much smaller upper bound of the form

$$[x]_{\xi_n^\infty} \subseteq \left(a^{-n} B_{2r}^{G_a^-} a^n\right) \cdot x.$$

This shows that

$$[x]_{\xi_{-\infty}^\infty} = \{x\}$$

almost surely, which implies that ξ is a generator under the action of a as required. □

Exercises for Section 10.5

Exercise 10.5.1. Show that in the case of a non-Archimedean group G_σ, the constant $\kappa(G, a, r)$ from Lemma 10.22 can be chosen to be independent of r.

Notes to Chapter 10

(36)(Page 321) This argument is due to Hopf [**?**]; see [**?**, Sec. 9.5] for an account of how this idea may be used to prove ergodicity of the geodesic flow on the modular surface.

(37)(Page 326) This important result was one of the crowning achievements of Lebesgue's development of measure theory, and appears at the end of his monograph [**?**] of 1904. Lebesgue also required the function to be continuous, and showed that the set of points where the derivative does not exist has the property that for any $\varepsilon > 0$ it can be covered by a countable sequence of intervals of total length no more than ε. Faber [**?**] showed that the theory of Lebesgue integration is not really required for this result. We refer to Hewitt and Stromberg [**?**, Th. 17.12] for a modern proof.

Chapter 11
Measure Rigidity for Higher-Rank Torus Actions

In Chapters 5 and 6 we introduced unipotent dynamics and showed some cases of the striking rigidity results of the following shape: every invariant probability measure if algebraic, every orbit closure is algebraic, and so on. We also mentioned in passing that the action of a one-parameter split torus subgroup (that is, a one-parameter $\mathbb{R}$-diagonalizable flow), much like the geodesic flow, cannot have any such rigidity properties (see p. 156 and [**?**, Ch. 9]). However, Furstenberg, Margulis, Katok and Spatzier conjectured in various forms that the action of a split two-parameter torus subgroup on a homogeneous space should behave differently, and in particular should exhibit rigidity. The first glimpse of this transition from the extreme flexibility, with a vast diversity of possible invariant measures and invariant subsets for rank one actions to the highly structured situation for rank two actions, with severe restrictions on the possible invariant measures and invariant subsets, was found by Furstenberg in 1967 (see his paper [**?**], and [**?**] for a discussion and additional references).

We write $T_r : \mathbb{T} \to \mathbb{T}$ for the map $x \mapsto rx \pmod 1$ on the circle.

Theorem. If p and q are distinct prime numbers, then the only closed infinite subset of $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ invariant under the maps T_p and T_q is $\mathbb{T}$ itself.

Furstenberg also raised the question of whether this topological rigidity had a measure-theoretic analog.

[Furstenberg] Is the Lebesgue measure $m_{\mathbb{T}}$ on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ the only non-atomic T_p, T_q-invariant probability measure if p and q are multiplicatively independent[†] integers?

This question remains open. The strongest partial result towards it was obtained by Rudolph [**?**] for p and q relatively prime, and this was later generalized by Johnson [**?**] to the multiplicatively independent case (also see [**?**] for a treatment and additional references).

[†] That is, with the property that $p^m = q^n$ for integers n, m requires $m = n = 0$.

Theorem 11.1 (Rudolph–Johnson). *Let p and q be multiplicatively independent integers. A probability measure μ on $\mathbb{T}$ that is invariant and ergodic under the action of the semigroup generated by T_p and T_q and has $h_\mu(T_p) > 0$ must be Lebesgue measure.*

This result motivated Margulis, Katok, and Spatzier to consider higher-dimensional (that is, toral) and homogeneous space analogs. For toral automorphisms the topological case was dealt with by Berend [**?**], and the Rudolph–Johnson theorem was generalized, following some more restricted results by Katok and Spatzier [**?**] resp. Kalinin and Katok [**?**], by Einsiedler and Lindenstrauss [**?**].

In the homogeneous setting we have the following important conjecture.

Conjecture 11.2 (Margulis). Let

$$A = \left\{ \begin{pmatrix} e^{t_1} & & \\ & e^{t_2} & \\ & & e^{t_3} \end{pmatrix} \mid t_1 + t_2 + t_3 = 0 \right\}$$

be the positive diagonal subgroup[†] in $\mathrm{SL}_3(\mathbb{R})$, and let $3 = \mathrm{SL}_3(\mathbb{Z}) \backslash \mathrm{SL}_3(\mathbb{R})$ be the homogeneous space of unimodular lattices in $\mathbb{R}^3$. Then every bounded A-orbit in 3 is a compact periodic orbit.

This conjecture can also be phrased in other similar homogeneous spaces (which we will discuss later), but it remains open in all of them. Somewhat surprisingly, for the analogous question regarding invariant measures our understanding of T_p, T_q-invariant measures on $\mathbb{T}$ and of A-invariant measures on 3 is more or less the same. After some more restricted results by Katok, Spatzier, Kalinin and Einsiedler the following result was proved by Einsiedler, Katok and Lindenstrauss in 2006 in [**?**].

need to sort out references

Theorem 11.3 (Einsiedler–Katok–Lindenstrauss). *Let μ be an A-invariant ergodic probability measure on* 3*. If $h_\mu(a) > 0$ for some $a \in A$, then $\mu = m_3$ is the Haar measure.*

We will prove this and some more general results of the same flavor in Section **??**, but will not prove the most general form obtained by Einsiedler and Lindenstrauss, referring instead to their papers [**?**] and [**?**].

As we will discuss in Section **??** in greater detail, the partial measure classification above makes use of the special nature of the lattice $\mathrm{SL}_3(\mathbb{Z}) < \mathrm{SL}_3(\mathbb{R})$. It also holds for some other natural lattices, but not for all lattices. This is quite different to the case of unipotent dynamics, where the lattice (or discrete subgroup) does not play any role in the statement.

[†] Strictly speaking A is not the complete group of $\mathbb{R}$-points of the torus given by the diagonal subgroup but the compact group of possibilities for the signs of the diagonal entries that is missing does not make any difference for the problems considered here, so we will still describe this as a torus action.

11.1 Product Structure of Leafwise Measures; First Version

We start our discussion of higher-rank torus actions with the following result on leafwise measures. We note that even though this only needs a single semi-simple element, the theorem is most useful in the discussion of the dynamics of higher-rank torus subgroups. The theorem goes back to work of Einsiedler and Katok [**?**] and was generalized by Lindenstrauss [**?**]. We will return to this phenomenon in the higher-rank setting in the next section.

Theorem 11.4 (Product structure of leafwise measures). *Let G be an S-algebraic group, $\Gamma < G$ a discrete subgroup, and let $\mathsf{X} = \Gamma\backslash G$. Let $a \in G$ be a $\mathbb{Q}_S$-diagonalizable element, let $U \leqslant G_a^-$ and $M \leqslant C_G(a)$ be two closed a-normalized subgroups with the property that M also normalizes U. Suppose that μ is an a-invariant probability measure on X such that for μ-almost every $\mathsf{x} \in \mathsf{X}$ the orbit map $M \ni m \mapsto m \centerdot \mathsf{x}$ is injective. Then the same is true for the subgroup $MU \cong M \ltimes U$, and the leafwise measures satisfy*

$$\mu_{\mathsf{x}}^{MU} \propto \mu_{\mathsf{x}}^{M} \times \mu_{\mathsf{x}}^{U}$$

for μ-almost every $\mathsf{x} \in \mathsf{X}$.

As we will see, the proof combines two ideas, both of which we have used before. First notice that a probability measure ν on a product space $Y \times Z$ is a product measure if and only if the conditional measures for the algebra $\mathscr{A} = \{Y\} \times \mathscr{B}_Z$ almost surely have the form $\nu_{(y,z)}^{\mathscr{A}} = \nu_Y \times \delta_z$ for some measure ν_Y that can almost surely be chosen independently of z. If this holds then $\nu = \nu_Y \times \pi_Z(\nu)$. Second, we will use Hopf's argument to show that $\mu_{\mathsf{x}}^M = \mu_{u\centerdot\mathsf{x}}^M$ for μ-almost every $\mathsf{x}, u\centerdot\mathsf{x} \in \mathsf{X}$ with $u \in U$, and more generally $\mu_{\mathsf{x}}^M \propto \left(\mu_{mu\centerdot\mathsf{x}}^M\right) m$ for μ-almost every $\mathsf{x}, mu\centerdot\mathsf{x} \in \mathsf{X}$ with $u \in U$ and $m \in M$. Due to Proposition 9.13, this gives precisely the analog of the characterization above of product measures for leafwise measures.

Proposition 11.5 (Key invariance property). *Let X, a, M, U and μ be as in Theorem 11.4. Then there exists a set of full measure $X' \subseteq \mathsf{X}$ such that $\mathsf{x}, (mu)\centerdot\mathsf{x} \in X'$ implies that $\mu_{\mathsf{x}}^M \propto \mu_{(mu)\centerdot\mathsf{x}}^M m$.*

Proof. By Corollary 9.17, $\mathsf{x} \mapsto \mu_{\mathsf{x}}^M$ is a measurable map. Fix some $\varepsilon > 0$, and let $X_0 \subseteq \mathsf{X}$ be a set of full measure chosen so that all the almost sure properties of the leafwise measure from Chapter 9 hold for $\mathsf{x} \in X_0$. By Lusin's theorem (Corollary 9.17) there exists a compact set $K_\varepsilon \subseteq X_0$ with $\mu(K_\varepsilon) > 1 - \varepsilon$ such that $K_\varepsilon \ni \mathsf{x} \mapsto \mu_{\mathsf{x}}^M$ is continuous. Using the maximal ergodic theorem for the set

$$X_\varepsilon = \left\{ x \in X_0 \mid \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{K_\varepsilon}(a^n \centerdot \mathsf{x}) > \frac{1}{2} \text{ for all } n \geqslant 1 \right\}$$

we see that $\mu(X_\varepsilon) \geqslant 1-2\varepsilon$. Now suppose that $\mathsf{x}, mu\text{•}\mathsf{x} \in X_\varepsilon$. Then there exists a sequence (n_k) with $n_k \nearrow \infty$ as $k\to\infty$ with $a^{n_k}\text{•}\mathsf{x}, a^{n_k}mu\text{•}\mathsf{x} \in K_\varepsilon$ and $a^{n_k}\text{•}\mathsf{x} \to \mathsf{z} \in K_\varepsilon$ as $k\to\infty$. This implies that

$$a^{n_k}mu\text{•}\mathsf{x} = ma^{n_k}ua^{-n_k}\text{•}(a^{n_k}\text{•}x) \longrightarrow mI\text{•}\mathsf{z} = m\mathsf{z}$$

as $k\to\infty$. Using Proposition 9.2, this shows that

$$\mu^M_{a^{n_k}\text{•}\mathsf{x}} = \mu^M_{\mathsf{x}},$$

since conjugation by a acts trivially on M. Similarly,

$$\mu^M_{a^{n_k}mu\text{•}\mathsf{x}} = \mu^M_{mu\text{•}\mathsf{x}}.$$

By continuity of the leafwise measure on K_ε, we also get

$$\mu^M_{\mathsf{x}} = \mu^M_{a^{n_k}\text{•}\mathsf{x}} \longrightarrow \mu^M_{\mathsf{z}}$$

and

$$\mu^M_{mu\text{•}\mathsf{x}} = \mu^M_{a^{n_k}mu\text{•}\mathsf{x}} \longrightarrow \mu^M_{m\text{•}\mathsf{z}}$$

as $k\to\infty$. Using Proposition 9.13 we also have

$$\left(\mu^M_{mu\text{•}\mathsf{z}}\right) m \propto \mu^M_{\mathsf{z}},$$

which gives

$$\mu^M_{mu\text{•}\mathsf{x}} m \propto \mu^M_{\mathsf{x}}.$$

Now choose $\varepsilon = \frac{1}{n}$, assume that

$$K_1 \subseteq K_{1/2} \subseteq K_{1/3} \subseteq \cdots$$

so that

$$X' = \bigcup_{n\geqslant 1} X_{1/n}$$

is an increasing union of sets with $\mu(X')=1$. If now $\mathsf{x}, mu\text{•}\mathsf{x} \in X'$ then there exists some n with $\mathsf{x}, mu\text{•}\mathsf{x} \in X_{1/n}$, and the proposition follows. □

Proof of Theorem 11.4. (to come) □

11.2 The High Entropy Method

The main argument of [**?**] has subsequently been described as the *high entropy method*, as it allows one to prove a quite general measure classification for invariant measures that have high entropy in the sense that it is close to the maximal entropy possible (see Einsiedler and Katok [**?**]). We will follow the argument as it is presented in the notes of Einsiedler and Lindenstrauss [**?**], and will prove the following result.

Theorem 11.6 (High entropy theorem). *Let $X = \Gamma\backslash G$ be the quotient of an S-algebraic semi-simple group by a discrete subgroup. Let $\sigma \in S$ and let A_σ be a $\mathbb{Q}_\sigma$-diagonalizable subgroup contained in a simple subgroup H of G_σ with $\dim A_\sigma \geqslant 2$. Then there exists some $c \in (0,1)$ such that if μ is an A_σ-invariant ergodic probability measure on X, and $a \in A_\sigma \smallsetminus \{I\}$ satisfies*

$$h_\mu(a) > c|\log \operatorname{mod}(a, G_a^-)|,$$

then μ is an H-invariant algebraic measure.

Notice that by Section 10.1.2, if μ was known to be invariant under G_a^- then $|\log \operatorname{mod}(a, G_a^-)|$ would be the entropy of a with respect to μ. In the case $X = \Gamma\backslash \operatorname{SL}_3(\mathbb{R})$ and the action of the full diagonal subgroup A, we can take $c = \frac{1}{2}$ (see Exercise 11.2.1).

11.2.1 The Supporting and the Invariance Subgroup

Our goal is once more to start with the given a-invariance and to deduce additional invariance for μ. More precisely, we wish to show that invariance of μ under some subgroup of G_a^- for some a that preserves μ. To this end we will study the leafwise measures $\mu_x^{G_a^-}$ or μ_x^U for an a-normalized $U < G_a^-$ more carefully. With an eye on future applications, we will do this in greater generality[†] than is needed for the proof of Theorem 11.6.

Definition 11.7. Let G be an S-algebraic group. An element $a \in G$ is called *p-semi-simple*[‡] if $a = (a_\sigma)_{\sigma\in S}$ with $a_\sigma = I$ for all $\sigma \in S \smallsetminus \{\sigma_0\}$ for some $\sigma_0 \in S$ and a_{σ_0} is $\mathbb{R}$-diagonalizable with positive eigenvalues if $\sigma_0 = \infty$ and $\mathbb{Q}_p$-diagonalizable with eigenvalues in $p^{\mathbb{Z}}$ if $\sigma_0 = p < \infty$.

Lemma 11.8 (Supporting Subgroup). *Let $\mathsf{X} = \Gamma\backslash\mathsf{G}$ be the quotient of an S-algebraic by a discrete subgroup. Let $a \in G$ be a p-semi-simple element. Suppose that μ is an a-invariant probability measure, and let P_x^U be the closed subgroup generated by $\operatorname{Supp} \mu_x^U$. Then for μ-almost every x, P_x^U is an a-normalized Zariski closed and Zariski connected subgroup of U. If $\sigma_0 = \infty$, then P_x^U is also connected as a Lie group.*

Proof of Lemma 11.8 for $\sigma_0 = \infty$. Let $X' \subseteq X$ be a subset of full measure on which all the almost sure properties of Chapter 9 hold. Fix $\varepsilon > 0$

[†] As usual, we will not however attempt maximal generality in order to avoid unnecessary complications.

[‡] Here we are using p to denote an arbitrary place of $\mathbb{Q}$ with $p = \infty$ corresponding to the usual absolute value.

and let $K \subseteq X'$ be a Lusin set with measure exceeding $(1-\varepsilon)$ for the map $x \longmapsto \mu_x^U$ (that is, a set on which the map is continuous).

Suppose now that $x \in K$ satisfies Poincaré recurrence for a^{-1}, meaning that there is a sequence $n_k \to -\infty$ as $k \to \infty$ such that $a^{n_k}\boldsymbol{\cdot} x \in K$ for all $k \geqslant 1$ and $a^{n_k}\boldsymbol{\cdot} x \to x$ as $k \to \infty$. Recall from Theorem 9.6 that

$$\mu^U_{a^{n_k}\boldsymbol{\cdot} x} \propto \left(\Theta_a^{n_k}\right)_* \mu_x^U, \tag{11.1}$$

where $\Theta_a : U \to U$ denotes conjugation by a on U. It follows that

$$P^U_{a^{n_k}\boldsymbol{\cdot} x} = \Theta_a^{n_k}\left(P_x^U\right).$$

By Lemma 3.31, we know that the Lie algebra of $P^U_{a^{n_k}\boldsymbol{\cdot} x}$ converges to an a-normalized Lie algebra. Moreover, if P_x^U is not connected then $P^U_{a^{n_k}\boldsymbol{\cdot} x}$ has all other connected components for large k far from the identity component. Since $\operatorname{Supp}\mu^U_{a^{n_k}\boldsymbol{\cdot} x} \subseteq P^U_{a^{n_k}\boldsymbol{\cdot} x}$ for all $k \geqslant 1$ and $\mu^U_{a^{n_k}\boldsymbol{\cdot} x} \to \mu_x^U$ as $k \to \infty$, it follows that $\operatorname{Supp}\mu_x^U$ is contained in a connected a-normalized Lie subgroup of U whose dimension equals the dimension of the Lie algebra of P_x^U. This implies that P_x^U is a connected a-normalized Lie subgroup of U. This shows the lemma for almost every $x \in K$. Since $\mu(K) > 1-\varepsilon$ and $\varepsilon > 0$ was arbitrary, the lemma follows. $\square$

Proof of Lemma 11.8 for $\sigma_0 = p < \infty$. (to come) $\square$

Lemma 11.9 (Invariance subgroup). *Let $X = \Gamma\backslash G$ be the quotient of an S-algebraic group by a discrete subgroup. Let $a \in G$ be a p-semi-simple element, and suppose that μ is an a-invariant probability measure on X. Let $I_x^U = \{u \in U \mid u_*\mu_x^U \propto \mu_x^U\}$. Then, for μ-almost every x, I_x^U is an a-normalized Zariski closed and Zariski connected subgroup of U, and $I_x^U = \{u \in U \mid u_*\mu_x^U = \mu_x^U\}$. If $\sigma_0 = \infty$, then I_x^U is also connected as a Lie group.*

Proof for $\sigma_0 = \infty$. Let $X' \subseteq X$, $\varepsilon > 0$, and $K \subseteq X'$ be as in the beginning of the proof of Lemma 11.8. Suppose that $x \in K$ satisfies Poincaré recurrence for a, so there is a sequence $n_k \to \infty$ as $k \to \infty$ such that $a^{n_k}\boldsymbol{\cdot} x \in K$ for all $k \geqslant 1$ and $a^{n_k}\boldsymbol{\cdot} x \to x$ as $k \to \infty$. Suppose in addition that $u_*\mu_x^U = c\mu_x^U$ for some $u \in U$ and $c > 0$. Applying (11.1) we see that

$$\left(a^{n_k} u a^{-n_k}\right)_* \mu^U_{a^{n_k}\boldsymbol{\cdot} x} = c\mu^U_{a^{n_k}\boldsymbol{\cdot} x}.$$

Notice that since $K \ni u \longmapsto \mu_x^U$ is continuous, there exists some $M > 0$ such that $\mu_x^U\left(B_2^U\right) \leqslant M$ for all $x \in K$. If now $\ell \in \mathbb{Z}$ is arbitrary, then there exists some $k \geqslant 1$ such that $a^{n_k} u^\ell a^{-n_k} \in B_1^U$, and since $a^{n_k}\boldsymbol{\cdot} x \in K$, we have

$$\begin{aligned} c^\ell = \left(a^{n_k} u^\ell a^{-n_k}\right)_* \mu^U_{a^{n_k}\boldsymbol{\cdot} x}\left(B_1^U\right) &= \mu^U_{a^{n_k}\boldsymbol{\cdot} x}\left(\left(a^{n_k} u^\ell a^{-n_k}\right)^{-1} B_1^U\right) \\ &\leqslant \mu^U_{a^{n_k}\boldsymbol{\cdot} x}\left(B_2^U\right) \leqslant M. \end{aligned}$$

Since $\ell \in \mathbb{Z}$ is arbitrary, we deduce that $c = 1$. Thus

$$I_x^U = \{u \in U \mid u_*\mu_x^U \propto \mu_x^U\} = \{u \in U \mid u_*\mu_x^U = \mu_x^U\}$$

for μ-almost every $x \in K$. It remains to show that I_x^U is a-invariant and connected.

Let $\mathfrak{h}_x^U$ be the Lie algebra of I_x^U. By (11.1) we have

$$\mathrm{Ad}_a^{n_k}(\mathfrak{h}_x^U) = \mathfrak{h}_{a^{n_k}\cdot x}^U,$$

and by Lemma 3.31 we also see that $\mathfrak{h}_{a^{n_k}\cdot x}^U$ converges as $k \to \infty$ to an Ad_a-normalized Lie algebra $\mathfrak{h}$.

Since $\mu_{a^{n_k}\cdot x}^U$ is invariant under $\exp(v_k)$ for all $v_k \in \mathfrak{h}_{a^{n_k}\cdot x}^U$, and since

$$\mu_{a^{n_k}\cdot x}^U \longrightarrow \mu_x^U$$

as $k \to \infty$ it follows that μ_x^U is invariant under $\exp(v)$ for all $v \in \mathfrak{h}$. In other words, $\mathfrak{h} \leqslant \mathfrak{h}_x^U$. Since both algebras have the same dimension it follows that $\mathfrak{h} = \mathfrak{h}_x^U$ is Ad_a-normalized. Since $\exp : \mathfrak{h} \to I_x^U$ is a polynomial isomorphism into the connected component of I_x^U with a polynomial inverse, the lemma will follow once we have shown that I_x^U is connected.

We may split $\mathfrak{g}_a^-$ and $\mathfrak{h}$ into eigenspaces for Ad_a, writing

$$\mathfrak{g}_a^- = (\mathfrak{g}_a^-)_1 + \cdots + (\mathfrak{g}_a^-)_\ell$$

and

$$\mathfrak{h} = \mathfrak{h}_1 + \cdots + \mathfrak{h}_\ell,$$

where $\mathfrak{h}_i \leqslant (\mathfrak{g}_a^-)_i$ are the respective eigenspaces for Ad_a with eigenvalue λ_i. We may suppose without loss of generality that $1 > \lambda_1 > \cdots > \lambda_\ell$.

Now choose a subspace $V_i \leqslant (\mathfrak{g}_a^-)_i$ transverse to $\mathfrak{h}_i$, so that $(\mathfrak{g}_a^-)_i = \mathfrak{h}_i + V_i$, for each $i = 1, \ldots, \ell$. Thus $\mathfrak{g}_a^- = \mathfrak{h} + V$ where $V = V_1 + \cdots + V_\ell$.

We claim that for any $u \in I_x^U$ there exists some $w \in \mathfrak{h}$ such that

$$\log(u \exp(w)) \in V.$$

If G_a^- is abelian, this may be seen by a straightforward linear algebra argument. In general we show this using the following inductive procedure. Suppose that

$$\log u \in V_1 + \cdots + V_{i-1} + (\mathfrak{g}_a^-)_i + \cdots + (\mathfrak{g}_a^-)_\ell$$

for some $i \leqslant \ell$. Notice that for $i = 1$ the claim is trivially satisfied, so assume that $i > 1$. Then

$$\mathfrak{g}_a^{(i)} = \sum_{j>i} (\mathfrak{g}_a^-)_j$$

is a Lie ideal of $\mathfrak{g}_a^-$ modulo which $(\mathfrak{g}_a^-)_i$ is in the center, by Section 2.1.1. Hence we may split the ith component of $\log u$ in the eigenspace decomposition above into a component in V_i and a component $w_i \in \mathfrak{h}_i$. We replace u by $u\exp(-w_i)$ which gives

$$\log(u\exp(-w_i)) \in V_1 + \cdots + V_i + \mathfrak{g}_a^{(i)},$$

since $\log g_1 g_2 = \log g_1 + \log g_2$ if g_1 and g_2 commute and $\exp(w_i)$ belongs to the center of $U^o / \exp(\mathfrak{g}_a^-)$.

By induction we may therefore assume for $u\exp(-w_i)$ that there exists some $w_i' \in \mathfrak{h}$ such that

$$u\exp(-w_i)\exp(w_i') = u\exp(w_i'') \in V$$

for some $w_i'' \in \mathfrak{h}$. For $i = 1$ this gives the claim.

If $I_x^U \smallsetminus (I_x^U)^o$ is non-empty, we may apply the claim and find some

$$u \in I_x^U \smallsetminus (I_x^U)^o$$

with $v = \log u \in V$. In this case we will find a one-parameter subgroup of I_x^U of the form $\{\exp(tw) \mid t \in \mathbb{R}\}$ with $w \in V$. By definition of $\mathfrak{h}$ and of V this gives a contradiction, and thus shows that I_x^U is connected. So suppose that $\exp(v) \in I_x^U \smallsetminus (I_x^U)^o$. Then

$$\exp\left(m \operatorname{Ad}_a^{n_k}(v)\right) = a^{n_k}\exp(v)^m a^{-n_k} \in I_{a^{n_k}\cdot x}^U$$

and $m \operatorname{Ad}_a^n(v) \in V$ for all $m, n \in \mathbb{Z}$. By Lemma 3.31, the limit

$$\lim_{k\to\infty} \frac{1}{\|\operatorname{Ad}_a^{n_k}(v)\|} \operatorname{Ad}_a^{n_k}(v) = w \in V$$

exists.

We claim that $\exp(tw) \in I_x^U$ for all $t \in \mathbb{R}$. Indeed, given $t \in \mathbb{R}$ we may choose an integer sequence (m_k) such that

$$m_k \|\operatorname{Ad}_a^{n_k}(v)\| \longrightarrow t$$

as $k \to \infty$ (since $\|\operatorname{Ad}_a^{n_k}(v)\| \to 0$ as $k \to \infty$) to obtain

$$\exp(tw) = \lim_{k\to\infty} \underbrace{\exp\left(m_k \|\operatorname{Ad}_a^{n_k}(v)\| \frac{1}{\|\operatorname{Ad}_a^{n_k}(v)\|} \operatorname{Ad}_a^{n_k}(v)\right)}_{\in I_{a^{n_k}\cdot x}^U} \in I_x^U$$

as required. □

Problems for Section 11.2.1

Exercise 11.2.1. (a) Show that for $X = \Gamma\backslash \mathrm{SL}_3(\mathbb{R})$ and the action of the full diagonal subgroup A, we can take $c = \frac{1}{2}$ in Theorem 11.6.
(b) Let

$$L = \left\{ \begin{pmatrix} g & \\ & (\det g)^{-1} \end{pmatrix} \mid g \in \mathrm{GL}_2(\mathbb{R}) \right\}$$

and

$$C = \left\{ \begin{pmatrix} \lambda & & \\ & \lambda & \\ & & \lambda^{-2} \end{pmatrix} \mid \lambda > 0 \right\}.$$

Suppose that $\Gamma \leqslant \mathrm{SL}_3(\mathbb{R})$ is a discrete subgroup with the property that the orbits ΓL and ΓC have finite volume. Deduce that the conclusion of Theorem 11.6 cannot hold for any $c < \frac{1}{2}$ (including not the conclusion that μ is algebraic).

Part III
Applications to Number Theory

(discussion and relation to first volume?)

Chapter 12
Mixing and Counting

In this chapter we will briefly mention some classical questions going back to work of Gauss, and then discuss in greater detail more recent developments due primarily to Duke, Rudnick, Sarnak, Eskin and McMullen. We will not aim for maximal generality as we only want to expose the striking connection between equidistribution problems and asymptotic counting problems. Using this connection we will also be able to calculate the co-volume of some natural lattices, including that of $\mathrm{SL}_3(\mathbb{Z})$ in $\mathrm{SL}_3(\mathbb{R})$.

12.1 Equidistribution and the Gauss Circle Problem

In this section we will outline a connection between an equidistribution result and a lattice-point counting problem in the classical setting of the Gauss circle problem. This problem asks for estimates of the number of integral points in the disk of radius R. The basic observation here is that the first estimate is given by the area of the disk, so the emphasis concerns controlling the error term. Write $\|\cdot\|$ for the Euclidean norm on $\mathbb{R}^2$.

Proposition 12.1. *For any $R > 0$ let*

$$N(R) = \left|\{n \in \mathbb{Z}^2 \mid \|n\| \leqslant R\}\right|. \tag{12.1}$$

Then

$$N(R) = \pi R^2 + \mathrm{O}(R).$$

The proof is geometric, and does not require an equidistribution result. Indeed, the main term πR^2 is the area of the 2-dimensional ball of radius R, and the error term is related to the area of an annulus, as indicated in Figure 12.1.

PROOF. Consider the unit square $S = [-\frac{1}{2}, \frac{1}{2}) \times [-\frac{1}{2}, \frac{1}{2})$, which is a fundamental domain for $\mathbb{Z}^2 < \mathbb{R}^2$. Then, as indicated in Figure 12.1, we have

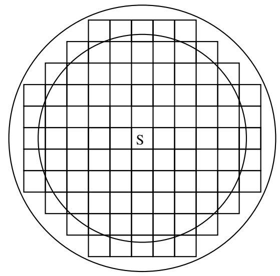

Fig. 12.1 Containing the error term for $N(R)$ inside an annulus.

$$B_{R-\frac{1}{\sqrt{2}}}(0) \subseteq S + \{n \in \mathbb{Z}^2 \mid \|n\| \leqslant R\} \subseteq B_{R+\frac{1}{\sqrt{2}}}(0).$$

By taking areas, we conclude that

$$\left(R - \frac{1}{\sqrt{2}}\right)^2 \pi \leqslant N(R) \leqslant \left(R + \frac{1}{\sqrt{2}}\right)^2 \pi$$

as required. □

It is conjectured that$^{(38)}$

$$N(R) = \pi R^2 + \mathrm{O}_\varepsilon\left(R^{\frac{1}{2}+\varepsilon}\right) \tag{12.2}$$

for all $\varepsilon > 0$. We refer to the paper of Ivić, Krätzel, Kühleitner and Nowak [**?**] for a survey of the many partial results towards this conjecture.

To motivate later arguments, we want to sketch an argument giving a connection between the error term in $N(R)$ and equidistribution properties of large circles in $\mathbb{R}^2$ modulo $\mathbb{Z}^2$. It seems unlikely that this would help in proving the conjecture, but with some effort such an equidistribution result could give

$$N(R) = \pi R^2 + \mathrm{O}\left(R^{1-\delta}\right)$$

for some $\delta > 0$ (see Exercise 12.1.4).

Let

$$\mathrm{T}^1\mathbb{R}^2 = \{(x,v) \mid x \in \mathbb{R}^2, v \in \mathbb{R}^2, \|v\| = 1\} = \mathbb{R}^2 \times \mathbb{S}^1$$

be the unit tangent bundle of $\mathbb{R}^2$, and let

$$\mathrm{T}^1\mathbb{T}^2 = \{(x,v) \mid x \in \mathbb{T}^2, v \in \mathbb{R}^2, \|v\| = 1\} = \mathbb{T}^2 \times \mathbb{S}^1$$

be the unit tangent bundle of $\mathbb{T}^2$. Also write

$$(x,v) \pmod{\mathbb{Z}}^2 = (x \pmod{\mathbb{Z}}^2, v)$$

for the canonical map from $\mathrm{T}^1\mathbb{R}^2$ to $\mathrm{T}^1\mathbb{T}^2$. We also write $\mathrm{d}(x,v)$ for the canonical volume form in $\mathrm{T}^1\mathbb{T}^2$.

Proposition 12.2. *Let*

$$\begin{aligned}\gamma_R : [0,1] &\longrightarrow \mathrm{T}^1\mathbb{R}^2\\ t &\longmapsto (R\mathrm{e}^{2\pi\mathrm{i}t}, \mathrm{e}^{2\pi\mathrm{i}t})\end{aligned}$$

be the constant speed parametrization of the outward tangent vectors on the circle of radius R. Then

$$\int_0^1 f\left(\gamma(t) \pmod{\mathbb{Z})^2}\right)\,\mathrm{d}t \longrightarrow \int_{\mathrm{T}^1\mathbb{T}^2} f(x,v)\,\mathrm{d}(x,v)$$

as $R\to\infty$, for every $f\in C\left(\mathrm{T}^1\mathbb{T}^2\right)$.

The idea of proof for this proposition is that a piece of a circle equidistributes on the torus when expanded (which can be proven using Fourier series). Since this is true for any piece of the circle, this gives the equidistribution result on the unit tangent bundle of the torus. We refer to [**?**] for more details.

We are now ready to present a modest improvement to the error term in Proposition 12.1 using the equidistribution from Proposition 12.2. As before we write $N(R) = \left|\{\mathbf{n}\in\mathbb{Z}^2 \mid \|\mathbf{n}\|\leqslant R\}\right|$.

Theorem 12.3. $N(R) = \pi R^2 + \mathrm{o}(R)$.

Once again we only sketch the argument and refer to [**?**] for more details. In order to take advantage of the equidistribution result, one needs to find a function defined on the unit tangent bundle with the property that the integral along a line segment relates to the difference between an area calculation and a lattice point count. To this end, first define the function h on the fundamental domain

$$\mathrm{T}^1 S = [-\tfrac{1}{2},\tfrac{1}{2}) \times [-\tfrac{1}{2},\tfrac{1}{2}) \times \mathbb{S}^1$$

using Figure 12.2 (recall that S is the unit square $[-\frac{1}{2},\frac{1}{2}) \times [-\frac{1}{2},\frac{1}{2})$).

In other words, $h(x,v)$ is the difference between an area calculation and the simple lattice count of whether or not 0 belongs to the polygon. Using this function we also define $f : S\times\mathbb{S}^1\to\mathbb{R}$ by

$$f(x,v) = \frac{h(x,v)}{\text{length of } L(x,v)},$$

where $L(x,v)$ is the line segment in S going through x and normal to v, as illustrated in Figure 12.2.

One then needs to show that f is Riemann integrable (so that the equidistribution result from Proposition 12.2 can be used) and that the integral along

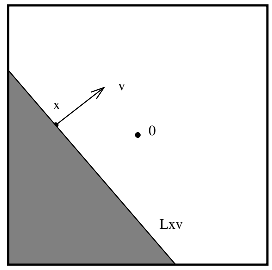

Fig. 12.2 The value $h(x, v)$ is the area of the polygon determined by S and the half-space with x on its boundary and v as outward normal, minus 1 if and only if 0 lies in the polygon.

a piecewise linear version of the circle (defined by the various translates of the fundamental domain) and the circle are very close to each other.

Exercises for Section 12.1

Exercise 12.1.1. Let $d \geqslant 2$. Prove that

$$N^*(R) = |\{n \in \mathbb{Z}^d \mid n \text{ is primitive and } \|n\| \leqslant R\}|$$

satisfies $N^*(R) = (\zeta(d)^{-1} V_d + o(1))R^2$ as $R \to \infty$. Here V_d is the volume of the unit ball in $\mathbb{R}^d$ and $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ denotes the Riemann zeta function.

Exercise 12.1.2. Prove Proposition 12.2 and Theorem 12.3.

Exercise 12.1.3. Generalize Proposition 12.1 to an effective statement in the form

$$\left| \int_0^1 f\left(\gamma_R(t) \pmod{\mathbb{Z}}^2\right) \,\mathrm{d}t - \int_{\mathrm{T}^1\mathbb{T}^2} f(x, v) \,\mathrm{d}(x, v) \right| \ll R^{-\delta_1} (f),$$

for some fixed $\delta_1 > 0$, (f) denotes some Sobolev norm, and $f \in C^\infty\left(\mathrm{T}^1\mathbb{T}^2\right)$ is arbitrary.

Exercise 12.1.4. Using Exercise 12.1.3, sharpen Theorem 12.3 to the form

$$N(R) = \pi R^2 + \mathrm{O}\left(R^{1-\delta_2}\right)$$

for some $\delta_2 > 0$.

12.2 Counting Points in $\Gamma \cdot \mathrm{i} \subseteq \mathbb{H}$

In this section we outline a lattice point count on the hyperbolic plane. For any lattice Γ in $\mathrm{PSL}_2(\mathbb{R})$ we define

$$N(R) = \left|\{\gamma(\mathrm{i}) \mid \mathsf{d}(\gamma(\mathrm{i}),\mathrm{i}) < R, \gamma \in \Gamma\}\right|.$$

Theorem 12.4 (Selberg). $N(R) = \dfrac{\mathrm{volume}\left(B_R^{\mathbb{H}}(\mathrm{i})\right)}{\mathrm{volume}\left(\Gamma\backslash\mathbb{H}\right)|\,\mathrm{Stab}_\Gamma(\mathrm{i})|} + \mathrm{o}\left(\mathrm{volume}\left(B_R^{\mathbb{H}}(\mathrm{i})\right)\right)$ *as* $R \to \infty$.

Selberg [?] uses a completely different spectral method to prove this theorem(39), and obtains additional information about the error term. We will present an approach following Eskin and McMullen [?] that uses mixing and equidistribution following the set-up of Duke, Rudnick and Sarnak [?] (which we will discuss in Section 12.3). As we saw in Section 12.1, the simple argument for counting problems connected to the lattice $\mathbb{Z}^2 \leqslant \mathbb{R}^2$ that simply tiles the ball of radius R using translates of a fundamental domain will give a heuristic rationale for the main term, but the following lemma describes a crucial but well-known difference between the Euclidean and hyperbolic plane (the details may be found in [?] or any hyperbolic geometry textbook).

Lemma 12.5. $\mathrm{volume}\left(B_R^{\mathbb{H}}(\mathrm{i})\right) = 2\pi\left(\cosh(R) - 1\right)$ *for all* $R > 0$.

This complicates the study of the lattice since $\mathrm{volume}\left(B_R^{\mathbb{H}}(\mathrm{i})\right)$ is asymptotic to $\pi\mathrm{e}^R$, and so the volume $\mathrm{volume}\left(B_{R+c}^{\mathbb{H}}(\mathrm{i}) \smallsetminus B_R^{\mathbb{H}}(\mathrm{i})\right)$ of an annulus is comparable in size to the volume $\mathrm{volume}\left(B_R^{\mathbb{H}}(\mathrm{i})\right)$ of the ball. In other words, the error term produced by the annulus has the same order of magnitude as the main term.

Another complication arises in the case $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, where the fundamental domains are unbounded with respect to the metric, so in order to use an annulus to capture all of them we should use $c = \infty$ (or at least some large value to capture most of the translates of the fundamental domain).

Because of this — a manifestation of the hyperbolic geometry at work here — the study of the boundary effects is much more important than it is in the case of $\mathbb{Z}^2 < \mathbb{R}^2$, where the volume of the annulus is asymptotically negligible in comparison with the volume of the ball.

To estimate these boundary effects we will also need the following equidistribution result concerning large circles as illustrated in Figure 12.3 (which will be a consequence of mixing). Below we will work with $\mathrm{PSL}_2(\mathbb{R})$ but will still use the matrix notation for the elements of $\mathrm{PSL}_2(\mathbb{R})$, for example writing

$$k_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \in K = \mathrm{SO}(2)/\{\pm 1\} = \{k_\phi \mid \phi \in [0,1)\}.$$

Theorem 12.6 (Equidistribution of Large Circles). *For any point z in $\mathbb{H}$, the circles obtained by following geodesics from z in all directions for time t equidistribute in $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{T}^1\mathbb{H}$. Indeed, for any finite volume quotient*

$$X = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$$

we have[†]

$$\frac{1}{\pi}\int_0^{\pi} f\left(gk_\phi \bullet x\right)\,\mathrm{d}\phi \longrightarrow \frac{1}{m_X(X)}\int_X f\,\mathrm{d}m_X \tag{12.3}$$

as $g\to\infty$ *in* $\mathrm{PSL}_2(\mathbb{R})$.

We note that the proof of the theorem uses the same argument as in Chapter 5, and leave the details as an exercise.

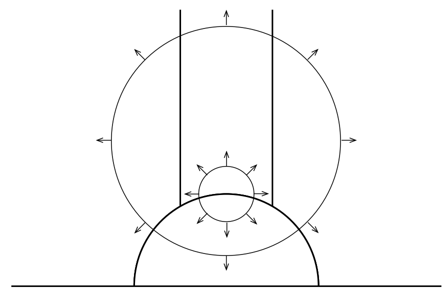

Fig. 12.3 Equidistribution of large circles in the modular surface becomes visible after the circle is moved into the fundamental domain using the isometries in Γ.

Theorem 12.4 now follows from the discussion in the next section by setting $H=K$.

12.3 The Counting Method of Duke–Rudnick–Sarnak and Eskin–McMullen

Eskin and McMullen [**?**] use mixing to establish asymptotic counting results in a more general context. For this (and in preparation for other special cases to be considered later) we describe in this section the general set-up for the work of Duke, Rudnick and Sarnak [**?**] (which is also used in the work of Eskin and McMullen [**?**]) on how to relate a counting problem for points in Γ-orbits on $V=G/H$ to the equidistribution problem for 'translated' H-orbits of the form

$$gH\Gamma\subseteq X={}^{G}\!/_{\Gamma}.$$

In many cases (for example, in the context of affine symmetric spaces), the methods of this chapter can be used to give the asymptotic of the counting for the number of integer points on varieties. In fact, suppose G and H consist of

[†] In this section we will not normalize the Haar measures on the quotient spaces to be probability measures, and instead will assume that they are compatible with a canonically chosen Haar measure on the group.

the $\mathbb{R}$-points of algebraic groups defined over $\mathbb{Q}$, $V = G/H$ can be identified with a variety defined over $\mathbb{Q}$, and $V(\mathbb{Z})$ is non-empty. Then we get that $V(\mathbb{Z})$ is a disjoint union

$$V(\mathbb{Z}) = \bigsqcup_i G(\mathbb{Z})v_i$$

of different $\Gamma = G(\mathbb{Z})$-orbits. Frequently this is a finite union, and then one gets the asymptotic for $|V(\mathbb{Z}) \cap B_t|$ by assembling the results for the individual counts $|G(\mathbb{Z})v_i \cap B_t|$. We will discuss the details of such integer point counting problems in special cases in the remaining sections of this chapter, and we refer to the papers of Duke, Rudnick and Sarnak and of Eskin and McMullen [**?**, **?**] for a detailed discussion of the general problem of counting lattice points in affine symmetric spaces.

12.3.1 Compatibility of all Haar measures involved

In order to state the result, we have to briefly describe the necessary compatibility of all the Haar measures involved. Let m_G be a Haar measure on a unimodular group G, and let $\Gamma < G$ be a lattice, on which we choose counting measure as the Haar measure. As we know m_G induces in a natural way a Haar measure m_X on $X = G/\Gamma$, giving total mass $m_X(X) = m_G(F)$ where $F \subseteq G$ is a Borel fundamental domain for (the right action of) Γ.

Assume that $H < G$ is a closed unimodular subgroup with Haar measure m_H. Then (see Appendix B) we may define a measure $m_{G/H}$ with the following compatibility property which is analogous to Fubini's theorem. If $f \in L^1_{m_G}(G)$ then the function F defined by the relation

$$F(gH) = \int_H f(gh)\,\mathrm{d}m_H(h) \tag{12.4}$$

exists for almost every $g \in G$, and the measure $m_{G/H}$ satisfies

$$\int_{G/H} F(gH)\,\mathrm{d}m_{G/H} = \int_G f\,\mathrm{d}m_G. \tag{12.5}$$

12.3.2 *First step: Dynamics gives an Averaged Counting Result*

12.3.2.1 Dynamical Assumptions on X

Let $\Gamma < G$ be a lattice, and assume that $H < G$ is a closed subgroup with the property that $\Gamma \cap H < H$ is also a lattice. We make the following[†] *equidistribution assumption*

$$\text{the translated } H\text{-orbits } gH\Gamma \text{ equidistribute in } X = G/\Gamma \tag{12.6}$$

as $gH \to \infty$ in G/H.

12.3.2.2 Averaged Counting Result

The assumptions above already imply a weak* version of our desired counting result in the following sense. We let $\{B_t \mid t \in \mathbb{R}\}$ be a collection of subsets of G/H each with finite Haar measure, and define a modified orbit-counting function $F_t : X \to \mathbb{R}_{\geqslant 0}$ by

$$F_t(g\Gamma) = \frac{1}{m_{G/H}(B_t)} \left| g\Gamma H \cap B_t \right|,$$

which counts elements within the Γ-orbit of $H \in G/H$ translated by g belonging to the set B_t. If $m_{G/H}(B_t) \to \infty$ as $t \to \infty$, then we have the *weak*-convergence*

$$F_t \, \mathrm{d}m_X \longrightarrow \frac{m_{H/\Gamma\cap H}(H/\Gamma \cap H)}{m_X(X)} \, \mathrm{d}m_X \tag{12.7}$$

as $t \to \infty$.

12.3.3 *Second step: Additional Geometric Assumptions imply the Counting Result*

12.3.3.1 Geometric Assumption on B_t

In order to be able to obtain from the averaged counting result above the desired counting result, we need to assume that the sets B_t are well behaved in a geometric manner. We say that a monotonically increasing family

[†] Alternatively, we may just assume some equidistribution on average — in a sense to be made clear in the proof.

$$\{B_t \mid t \in \mathbb{R}\}$$

of subsets of G/H is *well-rounded* if

$$m_{G/H}(B_t) \longrightarrow \infty$$

as $t \to \infty$, and for every $\delta > 0$ there exists a neighborhood U of $I \in G$ with

$$B_{t-\delta} \subseteq \bigcap_{g \in U} gB_t \subseteq B_t \subseteq \bigcup_{g \in U} gB_t \subseteq B_{t+\delta},$$

and furthermore for every $\varepsilon > 0$ there exists $\delta > 0$ with

$$\frac{m_{G/H}(B_{t+\delta})}{m_{G/H}(B_t)} < 1 + \varepsilon$$

for all $t \geqslant 0$

12.3.3.2 Asymptotic Counting Result

If the translated H-orbits equidistribute as assumed in (12.6), and the family of sets $\{B_t\}$ is well-rounded as above, then we have the asymptotic

$$\lim_{t \to \infty} \frac{1}{m_{G/H}(B_t)} |\Gamma H \cap B_t| = \frac{m_{H/\Gamma \cap H}(H/\Gamma \cap H)}{m_X(X)} \tag{12.8}$$

for the orbit-point counting problem.

12.3.4 Proofs

We now turn to considering the components of the outlined argument in greater detail.

Proof of weak*-convergence in (12.7). For simplicity of notation we set $Y = H/\Gamma \cap H$ and accordingly let m_Y denote the Haar measure induced by m_H on Y. We assume (12.6), or more precisely that the normalized translation

$$\frac{1}{m_Y(Y)} g_* m_{m_Y}$$

of the Haar measure m_Y on

$$Y = {}^{H}/\Gamma \cap H \subseteq X = {}^{G}/\Gamma$$

translated by $gH \in G/H$ converges to the normalized Haar measure

$$\frac{1}{m_X(X)}m_X$$

in the following averaged sense. For $\alpha \in C_c(X)$ we require that

$$\frac{1}{m_Y(Y)m_{G/H}(B_t)}\int_{B_t}\int_Y \alpha\,(gh\mathit{\Gamma})\,\mathrm{d}m_Y(h\mathit{\Gamma})\,\mathrm{d}m_{G/H}(gH) \\ \longrightarrow \frac{1}{m_X(X)}\int \alpha\,\mathrm{d}m_X \quad (12.9)$$

as $t \to \infty$. This is certainly satisfied if both

$$\frac{1}{m_Y(Y)}\int_Y \alpha\,(gh\mathit{\Gamma})\;\mathrm{d}m_Y \longrightarrow \frac{1}{m_X(X)}\int \alpha\,\mathrm{d}m_X$$

as $Hg \to \infty$ in G/H and

$$m_{G/H}(B_t) \to \infty$$

as $t \to \infty$, but (12.9) is a weaker requirement because of the additional averaging.

We wish to deduce from this assumption that

$$\int_X F_t(x)\alpha(x)\,\mathrm{d}m_X \longrightarrow \frac{m_Y(Y)}{m_X(X)}\int_X \alpha\,\mathrm{d}m_X$$

as $t \to \infty$.

The proof is relatively short, and consists of an application of the folding/unfolding trick using the spaces

$$\begin{tikzcd} & G/\mathit{\Gamma}\cap H \arrow[dl] \arrow[dr] & \\ G/\mathit{\Gamma} & & G/H. \end{tikzcd}$$

By definition,

$$A_t^{\alpha} = \int_X F_t(x)\alpha(x)\,\mathrm{d}m_X = \\ \frac{1}{m_{G/H}(B_t)}\int_{G/\mathit{\Gamma}}\sum_{\gamma\in\mathit{\Gamma}/\mathit{\Gamma}\cap H}\mathbb{1}_{B_t}(g\gamma H)\alpha(g\mathit{\Gamma})\,\mathrm{d}m_X(g\mathit{\Gamma}),$$

in which the sum over $\gamma \in \mathit{\Gamma}/\mathit{\Gamma}\cap H$ denotes the sum over a list of representatives of the cosets of $\mathit{\Gamma}\cap H$ in $\mathit{\Gamma}$. Thus by using the compatibility of the Haar measures in (12.4)–(12.5) we get

$$\begin{aligned}
A_t^\alpha &= \frac{1}{m_{G/H}(B_t)} \int_{G/\Gamma\cap H} \mathbb{1}_{B_t}(gH)\alpha(g\Gamma)\,\mathrm{d}m_{G/\Gamma\cap H}(g(\Gamma\cap H))\\
&= \frac{1}{m_{G/H}(B_t)} \int_{G/H} \mathbb{1}_{B_t}(gH) \int_{H/\Gamma\cap H} \alpha(gh\Gamma)\,\mathrm{d}m_Y(h\Gamma)\,\mathrm{d}m_{G/H}(gH)\\
&= \frac{1}{m_{G/H}(B_t)} \int_{B_t}\int_Y \alpha(gh\Gamma)\,\mathrm{d}m_Y(h\Gamma)\,\mathrm{d}m_{G/H}(Hg),
\end{aligned}$$

which converges (by assumption) to

$$\frac{m_Y(Y)}{m_X(X)}\int_X \alpha\,\mathrm{d}m_X$$

as $t\to\infty$. □

Proof of the pointwise count in (12.8). We now suppose that the weak*-convergence discussed above holds, and that the family of sets B_t is well-rounded as defined in Section 12.3.3. From this we wish to derive the asymptotic

$$\frac{1}{m_{G/H}(B_t)}\left|(\Gamma H)\cap B_t\right| \longrightarrow \frac{m_Y(Y)}{m_X(X)},$$

using the same argument as was used for the counting problem on $\mathbb{H}$.

Let $\varepsilon>0$ be arbitrary, and choose $\delta>0$ so that

$$\frac{m_{G/H}(B_{t+\delta})}{m_{G/H}(B_t)} < 1+\varepsilon$$

for all t, and choose a symmetric neighborhood $U=U^{-1}\subseteq G$ of $I\in G$ with

$$UB_t\subseteq B_{t+\delta}$$

for all t. Further let $\alpha\in C_c(X)$ be an approximate identity at the identity coset, in the sense that $\alpha\geqslant 0$

$$\int_X \alpha\,\mathrm{d}m_X = 1,$$

and $\mathrm{Supp}(\alpha)\subseteq U\Gamma$. Then we have for any $g\in B_\delta^G$ that

$$\begin{aligned}
F_{t+\delta}(g) &= \frac{1}{m_{G/H}(B_{t+\delta})}\left|g\Gamma H \cap B_{t+\delta}\right| \\
&= \frac{1}{m_{G/H}(B_{t+\delta})}\left|\Gamma H \cap \underbrace{g^{-1}B_{t+\delta}}_{\supseteq B_t}\right| \\
&\geqslant \frac{m_{G/H}(B_t)}{m_{G/H}(B_{t+\delta})}\frac{1}{m_{G/H}(B_t)}\left|\Gamma H \cap B_t\right| \\
&\geqslant \frac{1}{1+\varepsilon}\frac{1}{m_{G/H}(B_t)}\left|\Gamma H \cap B_t\right|.
\end{aligned}$$

Multiplying by α, integrating with respect to m_X and letting $t \to \infty$ gives

$$\limsup_{t\to\infty} \frac{1}{m_{G/H}(B_t)}\left|\Gamma H \cap B_t\right| \leqslant (1+\varepsilon)\frac{m_Y(Y)}{m_X(X)}.$$

The second inequality is derived in the same way. □

12.4 Counting Integer Points on Quadratic Hypersurfaces

[†]In this section we study our first class of examples of affine symmetric varieties, namely the case of quadratic hypersurfaces. Let Q be any non-degenerate indefinite quadratic form with integer coefficients in $d \geqslant 3$ variables and $a \in \mathbb{Z}\smallsetminus\{0\}$. Then $V(\mathbb{R}) = \{v \in \mathbb{R}^d \mid Q(v) = a\}$ can be identified with $\mathbb{G}(\mathbb{R})/\mathbb{H}(\mathbb{R})$ for $\mathbb{G} = \mathrm{SO}(Q)$ and $\mathbb{H} = \mathrm{Stab}_{\mathbb{G}}(v_0)$ for some $v_0 \in V$ as the $\mathbb{G}(\mathbb{R})$-action is transitive[‡] by Witt's theorem[(40)]. Let us assume that $V(\mathbb{Z})$ is non-empty and that $v_0 \in V(\mathbb{Z})$. If now $\mathbb{H}(\mathbb{Z})$ is a lattice in $\mathbb{H}(\mathbb{R})$ (which is always the case for $d \geqslant 4$ and in many cases also for $d = 3$) then we can derive from the methods of the last section the asymptotics for the counting problem of $V(\mathbb{Z})$.

We wish to discuss this now in greater detail. For the following calculations it is not really necessary but convenient to fix a particular quadratic form. So we set $Q(x_1,\dots x_m,y_1,\dots,y_n) = x_1^2+\dots+x_m^2-y_1^2-\dots-y_n^2$ with $m>1$ and $n \geqslant 1$.

Corollary 12.7 (Counting on quadratic hypersurfaces). *Let $a \in \mathbb{Z}\smallsetminus\{0\}$, let $V = \{v \mid Q(v) = a\}$, and assume that $V(\mathbb{Z})$ is non-empty. Suppose fur-*

[†] The counting problem in this section requires more algebraic background, and in particular more of the language of algebraic groups. We note that the application in the next section is easier in that respect.

[‡] We will see throughout this section enough elements of $\mathbb{G}(\mathbb{R})$ to derive this transitivity directly.

thermore that either $m+n \geqslant 4$ *or that* $m=2$, $n=1$, *and* a *is not a square in* $\mathbb{Z}$. *Define* $B_R^V = \{v \in V(\mathbb{R}) \mid \|v\| \leqslant R\}$. *Then there exists constants*[†] $c>0$ *and* $c'>0$ *such*

$$|V(\mathbb{Z}) \cap B_R^V| \sim c\,\mathrm{volume}_V(B_R^V) \sim c' R^{m+n-2},$$

where volume_V *denotes the* G*-invariant Haar measure on* $V(\mathbb{R})$.

We note that we define (and normalize) the Haar measure volume_V on $V(\mathbb{R})$ by the Lebesgue measure in $\mathbb{R}^{m+n}$ using the formula

$$\mathrm{volume}_V(B) = m_{\mathbb{R}^{m+n}}(\{tv \mid t \in [0,1], v \in B\}) \tag{12.10}$$

for any measurable $B \subseteq V(\mathbb{R})$.

For the following proof we further define $G = \mathrm{SO}(Q)(\mathbb{R})^\circ$ to be the connected component of the associated orthogonal group and $\Gamma = \mathrm{SO}(Q)(\mathbb{Z}) \cap G$.

12.4.1 Reduction to orbit counting problems

Recall from the beginning of the section that $V(\mathbb{R})$ is a single $\mathbb{G}(\mathbb{R})$-orbit and so every connected component V° of V can be identified with a quotient G/H where $H = \mathrm{Stab}_G(v_0)$ for some $v_0 \in V^\circ$. By a theorem of Borel the integer points

$$V(\mathbb{Z}) = \bigsqcup_i \mathbb{G}(\mathbb{Z}) v_i$$

are a finite union of different $\mathbb{G}(\mathbb{Z})$-orbits. As the connected component

$$G = \mathbb{G}(\mathbb{R})^\circ$$

has finite index in $\mathbb{G}(\mathbb{R})$, the same also holds for

$$\Gamma = \mathbb{G}(\mathbb{Z}) \cap G.$$

Hence it suffices to derive the desired result for each individual orbit Γv_0 for some $v_0 \in V(\mathbb{Z})$ (and add the corresponding asymptotic formulas together).

Hence we choose one $v_0 \in V(\mathbb{Z})$ and set $H = \mathrm{Stab}_G(v_0)$.

At some point the chapter on S-algebraic groups may contain a proof of this, in which case this note needs to be replaced by a link. If this doesn't happen we need a reference.

[†] As we will see the constants can be expressed using the volumes of the associated homogeneous spaces that arise in the proof.

12.4.2 Finite volume assumptions

The standing assumptions in Section 12.3 were that $\Gamma < G$ and $\Gamma \cap H < H$ are both lattices. We now check these assumptions in the setting of Corollary 12.7.

Since $\mathbb{G} = \mathrm{SO}(Q)$ is a semi-simple algebraic group defined over $\mathbb{Q}$ it follows by Theorem 7.7 that Γ is a lattice in G.

If $m + n \geqslant 4$ then $\mathbb{H} = \mathrm{Stab}_{\mathbb{G}}(v_0)$ is again a semi-simple algebraic group (since $a \neq 0$ it is simply the orthogonal group of the quadratic form on the orthogonal complement) defined over $\mathbb{Q}$. Hence in that case $\mathbb{H}(\mathbb{Z})$ is a lattice in $\mathbb{H}(\mathbb{R})$ which once more implies that $\Gamma \cap H = \mathbb{H}(\mathbb{Z}) \cap G$ is a lattice in $H = \mathbb{H}(\mathbb{R}) \cap G$.

In the remaining case where $m = 2$, $n = 1$, we see that $\mathbb{H}$ is a one-dimensional torus subgroup. We claim that $\mathbb{H}$ is $\mathbb{Q}$-anisotropic if $a \in \mathbb{Z} \smallsetminus \{0\}$ is a not a square. Assuming this, our assumptions in Corollary 12.7 and Theorem 7.7 imply that $\mathbb{H}(\mathbb{Z})$ is a lattice in $\mathbb{H}(\mathbb{R})$. As above, this gives that $\Gamma \cap H$ is a lattice in H.

So assume indirectly that the one-dimensional $\mathbb{Q}$-torus $\mathbb{H}$ is $\mathbb{Q}$-split. It follows that $\mathbb{H}$ is diagonalizable over $\mathbb{Q}$, meaning that there exists a rational basis w_1, w_2 of the orthogonal complement W of v_0 consisting of eigenvectors of $\mathbb{H}$. This implies that the quadratic form on W with respect to this basis has the form $Q(u_1 w_2 + u_2 w_2) = c u_1 u_2$, which in turn implies that the discriminant† of Q on W equals $-c^2$. Therefore the discriminant of Q on $\mathbb{Q}^3$ with respect to the basis v_1, w_1, w_2 equals $-ac^2$. Since it equals -4 with respect to the standard basis, it follows that a must be a square.

12.4.3 Proving the Equidistribution

The main dynamical assumption in Section 12.3 (and Section 12.3.2 in particular) is the equidistribution of $gH\Gamma$ in $X = G/\Gamma$ as $gH \to \infty$ in G/H. We claim that this follows in the context of this section once again from the same 'mixing argument' that was used in Chapter 5 and also in Section 12.2. We will not repeat this argument here, but will provide the technical input that reduces this repetition of the mixing argument into a straightforward exercise.

What is needed in order to do this is an analog of the local co-ordinate system $P_a G_a^-$ from Section 5.2 and NAK from Section 12.2. For this, we start by defining a one-parameter diagonalizable subgroup

† We recall that the discriminant is defined to be 4 times the determinant of the companion matrix of the quadratic form. Allowing rational coordinate changes the discriminant stays well-defined up to scalar multiples by a square.

$$A = \left\{ a_s = \begin{pmatrix} \cosh s & 0 & \sinh s & 0 \\ 0 & I_{m-1} & 0 & 0 \\ \sinh s & 0 & \cosh s & 0 \\ 0 & 0 & 0 & I_{n-1} \end{pmatrix} \mid s \in \mathbb{R} \right\},$$

and the compact subgroup

$$K = (\mathrm{SO}(m)(\mathbb{R}) \times \mathrm{SO}(n)(\mathbb{R})) \cap G.$$

The next lemma is not yet the analogous decomposition we are seeking, but is needed nonetheless.

Lemma 12.8. *$G = KAH$.*

Proof. Let $g \in G$ be an arbitrary element, and define

$$v = gv_0 = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

for $w_1 \in \mathbb{R}^m$ and $w_2 \in \mathbb{R}^n$. Then

$$Q(v) = Q(v_0) = \|w_1\|^2 - \|w_2\|^2 = 1.$$

Since $m > 1$ and $n \geqslant 1$ there exists some $k \in K$ such that

$$kv = \|w_1\| e_1 \pm \|w_2\| e_{m+1}.$$

Let $s \in \mathbb{R}$ be chosen so that $\cosh s = \|w_1\|$ and $\sinh s = \pm\|w_2\|$. Then

$$kv = kgv_0 = a_s v_0,$$

equivalently $a_{-s}kg = h \in H$, or

$$g = k^{-1} a_s h \in KAH$$

as required. □

As K is compact, the requirement that $g_n H \to \infty$ in G/H is equivalent to $g_n = k_n a_{s_n} h_n$ with $a_{s_n} \to \infty$ as $n \to \infty$. Furthermore, one can show (and we have done this before, in the beginning of the proof of Theorem 12.6) that the sequence (k_n) has no effect on the desired equidistribution claim. Thus we can simply assume that $g_n = a_{s_n}$ with $s_n \to \infty$ or $s_n \to -\infty$ as $n \to \infty$. Below we will assume that $s_n \to \infty$ as $n \to \infty$ (the other case is similar). We now define the local coordinate system that is needed in the proof of the equidistribution statement.

Lemma 12.9. *The stable horospherical subgroup $G^-_{a_s}$ has the property that*

$$G^-_{a_s} AH$$

contains an open neighborhood of the identity.

PROOF. For the proof it is convenient to again assume that $v = e_1$, and to switch to the Lie algebra. The Lie algebra element corresponding to A is

$$h = \begin{pmatrix} 0\,0\,1\,0 \\ 0\,0\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,0 \end{pmatrix},$$

where (for example) the second 0 represents $(m-1)$ zeros in a row. We claim that the Lie algebra of G_a^- contains all vectors of the form

$$W = \begin{pmatrix} 0 & -w_1^{\mathrm{t}} & 0 & w_2^{\mathrm{t}} \\ w_1 & 0 & w_1 & 0 \\ 0 & w_1^{\mathrm{t}} & 0 & -w_2^{\mathrm{t}} \\ w_2 & 0 & w_2 & 0 \end{pmatrix} \tag{12.11}$$

for all $w_1 \in \mathbb{R}^{m-1}$ and $w_2 \in \mathbb{R}^{n-1}$. This requires two calculations, as follows. Since $WJ + JW^{\mathrm{t}} = 0$ for the companion matrix

$$J = \begin{pmatrix} I_m & \\ & -I_n \end{pmatrix},$$

all elements W of the form (12.11) belong to the Lie algebra of G. Moreover, since

$$\begin{aligned} [h, W] &= hW - Wh \\ &= \begin{pmatrix} 0\,0\,1\,0 \\ 0\,0\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,0 \end{pmatrix} \begin{pmatrix} 0 & -w_1^{\mathrm{t}} & 0 & w_2^{\mathrm{t}} \\ w_1 & 0 & w_1 & 0 \\ 0 & w_1^{\mathrm{t}} & 0 & -w_2^{\mathrm{t}} \\ w_2 & 0 & w_2 & 0 \end{pmatrix} - \begin{pmatrix} 0 & -w_1^{\mathrm{t}} & 0 & w_2^{\mathrm{t}} \\ w_1 & 0 & w_1 & 0 \\ 0 & w_1^{\mathrm{t}} & 0 & -w_2^{\mathrm{t}} \\ w_2 & 0 & w_2 & 0 \end{pmatrix} \begin{pmatrix} 0\,0\,1\,0 \\ 0\,0\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & w_1^{\mathrm{t}} & 0 & -w_2^{\mathrm{t}} \\ 0 & 0 & 0 & 0 \\ 0 & -w_1^{\mathrm{t}} & 0 & w_2^{\mathrm{t}} \\ 0 & 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & 0 \\ w_1 & 0 & w_1 & 0 \\ 0 & 0 & 0 & 0 \\ w_2 & 0 & w_2 & 0 \end{pmatrix} \\ &= -W, \end{aligned}$$

it follows that W belongs to the Lie algebra of G_a^-.

If now $P = G_a^- A$ and $\mathfrak{p} = \operatorname{Lie} P$ is its Lie algebra, then we see from the inverse function theorem that Pv_0 must contain a neighborhood of $v_0 \in V$, since $\mathfrak{p}v_0$ contains $\{0\} \times \mathbb{R}^{m+n-1}$ (which coincides with the tangent space at $v_0 \in V$).

It follows that if g is sufficiently close to the identity, then $gv_0 = pv_0 \in Pv_0$ for some $p \in P$, which gives $p^{-1}g = h \in H$ and $g = ph$ as required. □

Theorem 12.10. *$gH\Gamma$ equidistributes in $X = G/\Gamma$ as $gH \to \infty$ in G/H.*

Sketch of Proof. We may assume that $g = a_s$ with $s \to \infty$. We set

$$P = G_a^- A$$

and deduce from Lemma 12.9 and Lemma 1.22 that the Haar measure on m_G restricted to PH equals the direct product of the Haar measures on P and on H respectively.

Assume at first that $H\Gamma \subseteq X$ is compact. Then there exists some uniform injectivity radius $\delta > 0$ for all points in $H\Gamma$. Let $B = B_\delta^P$ be the corresponding neighborhood of the identity in P, and set $T = BH\Gamma$, which we should think of as a *tubular neighborhood* of $H\Gamma \subseteq X$.

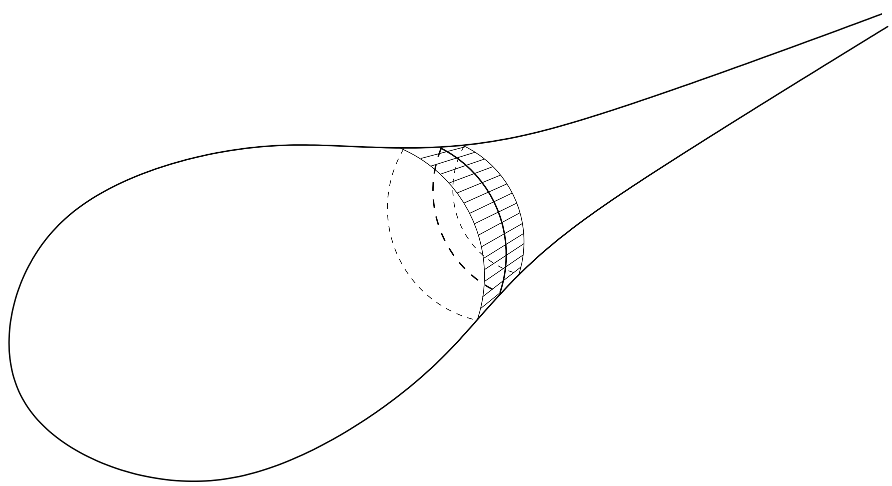

Fig. 12.4 The shaded region depicts the tubular neighborhood T of the orbit $H\Gamma$ in the 'center of T'.

Now pick some $f \in C_c(X)$, some $\varepsilon > 0$, ensuring that $\delta > 0$ is sufficiently small to work for f and ε, and apply mixing (Theorem 2.4) to obtain the desired contradiction for f, up to a precision controlled by ε.

If $H\Gamma$ is not compact, then the outline above needs to be adjusted (for otherwise, the failure of injectivity in a cusp makes the proof break down). Fortunately this case is not difficult either. Let $\kappa > 0$ be arbitrarily small and let $K \subseteq H\Gamma$ be a compact set of measure

$$\frac{1}{m_{H\Gamma}(H\Gamma)} m_{H\Gamma}(K) > 1 - \kappa.$$

Now apply the argument for the compact case above, with $BH\Gamma$ replaced by BK to obtain the desired conclusion up to a precision $\varepsilon + \mathrm{O}_f(\kappa)$. $\square$

12.4.4 *The Asymptotics of* $\mathrm{volume}_V(B_R^V)$ *and Well-Roundedness*

Recall from Section 12.3.1 that we can define the Haar measure on V using the ($\mathbb{G}$-invariant) Lebesgue measure on $\mathbb{R}^{m+n}$. Using this, we get the following result.

Lemma 12.11. $\mathrm{volume}_V\left(B_R^V\right) \sim \dfrac{1}{(m+n)(m+n-2)}|a|^{(m+n)/2}\left(\dfrac{R}{2}\right)^{m+n-2}.$

Proof. We will assume that $a=1$ (which by a scaling argument allows the case $a>0$ to be deduced; the case $a<0$ then follows by swapping m and n) and in any case that $R>\sqrt{|a|}$. Choose $S>0$ with $R^2=\cosh 2S$. Let $U\subseteq\mathbb{R}^{m-1}$ be open and Jordan measurable and let $\phi: U\to\mathbb{S}^{m-1}$ be a smooth parameterization† of $\mathbb{S}^{m-1}$ up to a set of measure 0. Similarly, let $V\subseteq\mathbb{R}^{n-1}$ and $\psi: V\to\mathbb{S}^{n-1}$ be a smooth paramaterization of $\mathbb{S}^{n-1}$. By **??** we have

$$
\begin{aligned}
\mathrm{volume}_V\left(B_V^R\right) &= m_{\mathbb{R}^{m+n}}\left(\left\{t\begin{pmatrix}\cosh(s)\phi(u)\\ \sinh(s)\psi(v)\end{pmatrix} \mid t\in[0,1], s\in[0,S], u\in U, v\in V\right\}\right)\\
&= \int_0^1\int_0^S\int_U\int_V \det\left(\begin{pmatrix}\cosh(s)\phi(u)\\ \sinh(s)\psi(v)\end{pmatrix},\begin{pmatrix}t\sinh(s)\phi(u)\\ t\cosh(s)\psi(v)\end{pmatrix},\right.\\
&\qquad\left.\begin{pmatrix}t\cosh(s)\,\mathrm{D}_u\,\phi(u)\\ 0\end{pmatrix},\begin{pmatrix}0\\ t\sinh(s)\,\mathrm{D}_v\,\psi(v)\end{pmatrix}\right)\,\mathrm{d}v\,\mathrm{d}u\,\mathrm{d}s\,\mathrm{d}t\\
&= \int_0^1\int_0^S\int_U\int_V t^{m+n-1}\cosh(s)^{m-1}\sinh(s)^{n-1}\\
&\qquad \det\left(\begin{pmatrix}\cosh(s)\phi(u)\\ \sinh(s)\psi(v)\end{pmatrix},\begin{pmatrix}t\sinh(s)\phi(u)\\ t\cosh(s)\psi(v)\end{pmatrix},\begin{pmatrix}\mathrm{D}_u\,\phi(u)\\ 0\end{pmatrix},\begin{pmatrix}0\\ \mathrm{D}_v\,\psi(v)\end{pmatrix}\right)\\
&\qquad \mathrm{d}v\,\mathrm{d}u\,\mathrm{d}s\,\mathrm{d}t,
\end{aligned}
$$

where $\mathrm{D}_u\,\phi$ and $\mathrm{D}_v\,\psi$ are the total derivatives of ϕ and of ψ respectively. Now notice that every column of

$$\begin{pmatrix}\mathrm{D}_u\,\phi(u)\\ 0\end{pmatrix}$$

is orthogonal to all the other columns in the matrix above, and similarly for

$$\begin{pmatrix}0\\ \mathrm{D}_v\,\psi(v)\end{pmatrix}.$$

This allows us to split the determinant above into three factors, giving

† For example, using generalized spherical co-ordinates.

$$
\begin{aligned}
\mathrm{volume}_V\left(B_V^R\right) &= \frac{1}{m+n}\int_0^S\int_U\int_V (\cosh s)^{m-1}(\sinh s)^{n-1} \\
&\qquad \det\left(\begin{pmatrix}\phi(u)\\0\end{pmatrix},\begin{pmatrix}0\\\psi(v)\end{pmatrix},\begin{pmatrix}\mathrm{D}_u\,\phi(u)\\0\end{pmatrix},\begin{pmatrix}0\\\mathrm{D}_v\,\psi(v)\end{pmatrix}\right) \\
&\qquad \det\begin{pmatrix}\cosh s & \sinh s & \\ \sinh s & \cosh s & \\ & & I_{m+n-2}\end{pmatrix}\mathrm{d}v\,\mathrm{d}u\,\mathrm{d}s \\
&= \frac{1}{m+n}\int_U \det\left(\phi(u),\mathrm{D}_u\,\phi(u)\right)\,\mathrm{d}u\int_V\det\left(\psi(v),\mathrm{D}_v\,\psi(v)\right)\,\mathrm{d}v \\
&\qquad \int_0^S \cosh(s)^{m-1}\sinh(s)^{n-1}\,\mathrm{d}s \\
&= \frac{1}{m+n}\,\mathrm{volume}\left(\mathbb{S}^{m-1}\right)\mathrm{volume}\left(\mathbb{S}^{n-1}\right) \\
&\qquad \int_0^S\left(\frac{1}{2^{m+n-2}}\mathrm{e}^{s(m+n-2)}+\mathrm{O}(\mathrm{e}^{s(m+n-4)})\right)\mathrm{d}s \\
&= \frac{1}{m+n}\,\mathrm{volume}\left(\mathbb{S}^{m-1}\right)\mathrm{volume}\left(\mathbb{S}^{n-1}\right)\frac{1}{2^{m+n-2}}\frac{1}{m+n-2}\mathrm{e}^{(m+n-2)S} \\
&\qquad +\mathrm{O}\left(\mathrm{e}^{(m+n-4)S}\right).
\end{aligned}
$$

Recall that

$$R^2=\cosh 2S\sim\frac{1}{2}\mathrm{e}^{2S},$$

so that $\mathrm{e}^S\sim\sqrt{2}R$. This gives

$$\mathrm{volume}_V\left(B_R^V\right)\sim\frac{1}{(m+n)(m+n-2)}\,\mathrm{volume}\left(\mathbb{S}^{m-1}\right)\mathrm{volume}\left(\mathbb{S}^{n-1}\right)\left(\sqrt{2}R\right)^{m+n-2}.$$

□

Query about point of explicit constants if we don't calculate the volume initially. On the other hand if keeping the explicit constants is no harder then it makes the calculation more easy to follow in some ways.

Lemma 12.12. *The sets* $B_t=B_{\mathrm{e}^t}^V=\{v\in V\mid\|v\|\leqslant R=\mathrm{e}^t\}$ *are well-rounded in the sense of Section* 12.3*.*

Proof. Let $\delta>0$ and choose a neighborhood U of $I\in G$ such that

$$\max\left(\|g\|,\|g^{-1}\|\right)<\mathrm{e}^\delta$$

for all $g\in G$. Let $v\in B_{t-\delta}$, so that $\|v\|\leqslant\mathrm{e}^{t-\delta}$, $\|g^{-1}v\|<\mathrm{e}^t$, and therefore $g^{-1}v\in B_t$ for $g\in U$. This gives

$$B_{t-\delta}\subseteq\bigcap_{g\in V}gB_t.$$

Similarly we see that

$$\bigcup_{g\in U} gB_t \subseteq B_{t+\delta}.$$

Finally, we have

$$\frac{\mathrm{volume}_V(B_{t+\delta})}{\mathrm{volume}_V(B_t)} \sim \frac{\mathrm{e}^{(m+n-2)(t+\delta)}}{\mathrm{e}^{(m+n-2)t}} = \mathrm{e}^{(m+n-2)\delta} < 1+\varepsilon$$

for small enough δ. Therefore

$$\frac{\mathrm{volume}_V(B_{t+\delta})}{\mathrm{volume}_V(B_t)} < 1+\varepsilon \tag{12.12}$$

for all $t > T$. From the proof of Lemma 12.11, we also see that $\mathrm{volume}_V(B_t)$ depends continuously on t, so we can make δ even smaller if necessary to ensure that 12.12 also holds for all $t \in [0,T]$. □

Exercises for Section 12.4

Exercise 12.4.1. a) Upgrade the sketch of proof of Theorem 12.10 to a real proof.
b) The previous exercise notwithstanding, notice that the sketch proof applies a result without the right hypotheses. If $m = n = 2$, then G is not simple but only semi-simple, and so the Howe–Moore theorem in the form of Theorem 2.4 cannot be applied. However, Theorem 2.7 does apply in this case. Decide whether or not this makes a difference to Theorem 12.10.

12.5 Counting Integer Matrices with Given Determinant

In this section we want to apply the results of Section 12.3 to prove[†] the following corollary.

Theorem 12.13. *Let $d \geqslant 2$ and $a \in \mathbb{Z}\smallsetminus\{0\}$. Then there exists a positive constant c_a such that*

$$|\{M \in \mathrm{Mat}_{dd}(\mathbb{Z}) \mid \det M = a \textit{ and } \|M\| \leqslant R\}| \sim c_a R^{d(d-1)}.$$

is this the correct formula for the exponent in general?

We note that this contains in particular the asymptotic counting result of the lattice elements of $\mathrm{SL}_d(\mathbb{Z})$.

[†] The cases $d = 2, 3$ will be self-contained, but we will not prove here the needed asymptotics of the measure of the ball, nor the well-roundedness for $d \geqslant 4$.

12.5.1 Reduction to orbit counting problems

We note that the asymptotic counting problem in Theorem 12.13 for a and $-a$ are equivalent (by simply changing the sign of one column). So let us assume $a > 0$. We define the set

$$V = \{M \in \mathrm{Mat}_d(\mathbb{R}) \mid \det(M) = a\} \cong \mathrm{SL}_d(\mathbb{R}) \times \mathrm{SL}_d(\mathbb{R})/\Delta_{\mathrm{SL}_d(\mathbb{R})},$$

where $\Delta_{\mathrm{SL}_d(\mathbb{R})} = \{(g,g) \mid g \in \mathrm{SL}_d(\mathbb{R}\}$. In fact $G = \mathrm{SL}_d(\mathbb{R}) \times \mathrm{SL}_d(\mathbb{R})$ acts on matrices $M \in V$ via

$$(g_1, g_2)\raisebox{0.1ex}{\scriptsize$\bullet$}M = g_1 M g_2^{-1}.$$

Then, if $M_0 = \sqrt[d]{a} I$,

$$\mathrm{Stab}_G(M_0) = \Delta_{\mathrm{SL}_d(\mathbb{R})}$$

and transitivity is easy.

Next we outline the proof that

$$V(\mathbb{Z}) = \{M \in \mathrm{Mat}(\mathbb{Z}) \mid \det M = a\}$$

is a finite union of $\Gamma = \mathrm{SL}_d(\mathbb{Z}) \times \mathrm{SL}_d(\mathbb{Z})$-orbits. Let $M \in V(\mathbb{Z})$ be arbitrary. Applying elements of $G(\mathbb{Z})$ to M correspond to certain types of row and column operations on M. Indeed using the elementary unipotent matrices of $\mathrm{SL}_d(\mathbb{Z})$ we can add any multiple of a row (or column) to any other row (or column). Similarly we may permute rows and columns (potentially switching the sign of one of them).

These steps allow a type of Euclidean algorithm: Assuming that the top left corner is already the smallest nonzero entry of absolute value we may either reduce the remaining entries on the first row and column to zero or produce a smaller entry. Hence eventually we create a block matrix with a nonzero entry in the top left corner, zeroes on the remainder of the first row and column, and some matrix in the lower right block. We may repeat this procedure and arrive at a diagonal integer matrix. As there are only finitely many integer diagonal matrices with determinant equal to a the result follows.

As $V(\mathbb{Z})$ is a finite union of Γ-orbits it suffices to establish the counting result for each individual Γ-orbit.

12.5.2 Finite volume assumptions

We now check the standing assumptions of Section 12.3 that both homogeneous spaces appearing there have finite volume.

By Theorem 1.18 Γ is a lattice in G, i.e. $X = G/\Gamma$ has finite volume as required.

By the argument in the previous section it suffices to study the counting problem for ΓM where $M \in V(\mathbb{Z})$ is a diagonal matrix. This defines our second group

$$H = \mathrm{Stab}_G(M) = \{(g,h) \mid g,h \in \mathrm{SL}_d(\mathbb{R}) \text{ and } h = M^{-1}gM\},$$

which clearly is isomorphic to $\mathrm{SL}_d(\mathbb{R})$. In this isomorphism $\Gamma \cap H$ corresponds to $\{g \in \mathrm{SL}_d(\mathbb{Z}) \mid M^{-1}gM \in \mathrm{SL}_d(\mathbb{Z})\}$. The latter is a finite index subgroup of $\mathrm{SL}_d(\mathbb{Z})$ (one way to see this is to note that it contains the congruence subgroup $\{g \in \mathrm{SL}_d(\mathbb{Z}) \mid a \text{ divides } g - I\}$), which implies that $H/(\Gamma \cap H)$ has finite volume as required.

12.5.3 Proving the Equidistribution

The main dynamical assumption in Section 12.3 (see Section 12.3.2) is the equidistribution of $gH\Gamma$ in $X = G/\Gamma$ as gH goes to infinity in G/H. The argument is similar[†] to that used in Section 12.4.3 (and hence also to the arguments in Section 5.2 and in Section 12.2). We will not repeat the 'mixing argument', but will instead discuss the technical requirements that make it work.

For these preparations we set $H_0 = \Delta_{\mathrm{SL}_2(\mathbb{R})}$.

Lemma 12.14. *Let $A \leqslant \mathrm{SL}_d(\mathbb{R})$ denote the full positive diagonal subgroup, and define*

$$\widetilde{A} = \{(a, a^{-1}) \mid a \in A\},$$

$$K = \mathrm{SO}(d)(\mathbb{R}) \times \mathrm{SO}(d)(\mathbb{R}).$$

Then $G = K\widetilde{A}H_0$.

Proof. Multiplying $(g_1, g_2) \in G$ on the right by $(g_2^{-1}, g_2^{-1}) \in H_0$ we see that it is enough to study elements of the form (g, I) for $g = g_1 g_2^{-1} \in \mathrm{SL}_d(\mathbb{R})$.

Let $g = k_1 a k_2$ be a KAK decomposition of g, and let $a_1 \in A$ be a square root of a. Then

$$(g_1, g_2) \in (g, I)H_0 = (k_1 a_1^2 k_2^{-1})H_0 = (k_1, k_2^{-1})(a_1, a_1^{-1})H_0 \subseteq K\widetilde{A}H_0$$

as required. $\square$

If we now consider a sequence $(g_n H_0)$ going to infinity in G/H_0, then it is clear that we may replace g_n by $k_n \widetilde{a_n} \in K\widetilde{A}$. As K is compact, we may suppress the elements $k_n \in K$ (see the beginning of the proof of Theorem 12.6 on p. 369) and consider simply the case

[†] This is not a coincidence, as both are special cases of the class of affine symmetric spaces.

$$\widetilde{a_n} H_0 \longrightarrow \infty$$

in G/H_0, with $\widetilde{a_n} \in \widetilde{A}$.

Moreover, we may assume that $\widetilde{a_n} = (a_n, a_n^{-1})$ is chosen so that the diagonal entries of a_n are monotonically increasing. To see that this is possible, notice that if some $\widetilde{a} = (a, a^{-1})$ does not have this property, then we may find a permutation[†] matrix $g \in \mathrm{SL}_d(\mathbb{R})$ so that gag^{-1} has increasing diagonal entries. With this matrix we may then consider

$$\widetilde{a}H_0 = \widetilde{a}(g^{-1}, g^{-1})H_0 = (g^{-1}, g^{-1})(g, g)\widetilde{a}(g^1, g^{-1})H_0 \in K\widetilde{b}H_0,$$

where $\widetilde{b} \in \widetilde{A}$ now has the required property. As before, we may treat

$$(g^{-1}, g^{-1}) \in K$$

as part of the test function in the desired equidistribution result and simply continue working with $\widetilde{b}H_0$. We will write $\widetilde{A}^{+}$ for the set of pairs (a, a^{-1}) where a has increasing diagonal entries.

Lemma 12.15. *Let $N \leqslant \mathrm{SL}_d(\mathbb{R})$ be the upper-triangular unipotent subgroup, and let*

$$\widetilde{N} = \{(n_1 n_2^{\mathrm{t}}) \mid n_1, n_2 \in N\}.$$

Then $\widetilde{N}\widetilde{A}H_0$ contains the identity in its interior.

Proof. We have

$$\begin{aligned} \operatorname{Lie} H_0 &= \{(v, v) \mid v \in \mathfrak{sl}_d(\mathbb{R})\}, \\ \operatorname{Lie} \widetilde{A} &= \{(h, -h) \mid h \text{ diagonal }, \operatorname{tr}(h) = 0\}, \end{aligned}$$

and $\operatorname{Lie} \widetilde{N}$ is the direct product of the upper and lower nilpotent triangular Lie subalgebras of $\mathfrak{sl}_d(\mathbb{R})$. It is easy to see that these subspaces are transversal, and their dimension sums to the dimension of the Lie algebra of G. The lemma follows by the inverse function theorem. □

Theorem 12.16. *$gH\Gamma$ equidistributes in $X = G/\Gamma$ as $gH \to \infty$ in G/H.*

Sketch of proof. Here $H = \mathrm{Stab}_G(M)$ for a general (or diagonal) matrix $M \in V(\mathbb{Z})$. As we may assume that $a > 0$, we see that $a^{1/d}M \in \mathrm{SL}_d(\mathbb{R})$ which shows that H is conjugate to $H_0 = \Delta_{\mathrm{SL}_d(\mathbb{R})}$ via the element

$$\widetilde{g} = (I, a^{-1/d}M).$$

It follows that

$$H\Gamma = \widetilde{g}H_0\widetilde{g}\Gamma,$$

[†] In which we will allow one entry to be -1 in order to ensure that $\det(g) = 1$.

and it is enough to show that $gH_0\widetilde{g}\Gamma$ equidistributes as $gH_0 \to \infty$ in G/H_0. By Lemma 12.15 we may safely assume that $g = k\widetilde{a}$ with $\widetilde{a} \in \widetilde{A}$, and even that $g = \widetilde{a} \in \widetilde{A}^+$.

Let $\kappa > 0$ be arbitrarily small, and choose $K \subseteq H_0\widetilde{g}\Gamma$ such that

$$m_{H_0\widetilde{g}\Gamma}(K) > (1-\kappa)m_{H_0\widetilde{g}\Gamma}(H_0\widetilde{g}\Gamma).$$

Fix $f \in C_c(X)$ and $\varepsilon > 0$, and choose $\delta > 0$ smaller than the injectivity radius on K and small enough to ensure that

$$\mathsf{d}(x_1, x_2) < \delta \implies |f(x_1) - f(x_2)| < \varepsilon.$$

Set $P = \widetilde{N}\widetilde{A}$ and let $B = B_\delta^P$ be the δ-neighborhood of $I \in P$. Now replace $H_0\widetilde{g}\Gamma$ first by K and then by BK, use the mixing property (Theorem 2.7), use the fact that $g_n = \widetilde{a_n} \in \widetilde{A}^+$ contracts N, and deduce the proof of the theorem. □

12.5.4 The Asymptotics of $\mathrm{volume}_V(B_R^V)$ and Well-Roundedness

Not really an isomorphism as V not a group

Clearly for any $a > 0$ there is a bijection

$$V = \{M \in \mathrm{Mat}_{dd}(\mathbb{R}) \mid \det M = a\} \longleftrightarrow \mathrm{SL}_d(\mathbb{R}),$$

obtained by multiplying by $a^{-1/d}$. Thus it is sufficient to study the volume of 'balls' in $\mathrm{SL}_d(\mathbb{R})$.

Proposition 12.17 (Asymptotics of balls in $\mathrm{SL}_d(\mathbb{R})$). *The asymptotic growth in the volume of balls in $\mathrm{SL}_d(\mathbb{R})$ has the form*

$$m_{\mathrm{SL}_d(\mathbb{R})}\left(\{g \in \mathrm{SL}_d(\mathbb{R}) \mid \|g\| < R\}\right) \sim c_d R^{d(d-1)} \text{ for some } c_d > 0.$$

Corollary 12.18. *The set $B_t = \{g \in \mathrm{SL}_d(\mathbb{R}) \mid \|g\| < \mathrm{e}^t\}$ are well-rounded in the send of Section 12.3.3*

The proof of Corollary 12.18 is very similar to that of Lemma 12.12, and is therefore left to the reader.

We normalize the Haar measure on $\mathrm{SL}_d(\mathbb{R})$ by giving the definition

$$m_{\mathrm{SL}_d(\mathbb{R})}(B) = m_{\mathbb{R}^{d^2}}\left(\{tb \mid t \in [0,1], b \in B\}\right)$$

for any measurable $B \subseteq \mathrm{SL}_d(\mathbb{R})$.

Proof of Proposition 12.17 for $d = 2$. Strictly speaking, we do not need to give a new proof of the result as it is essentially also contained in Lemma 12.5 (and the connection between the Haar measures on $\mathrm{SL}_2(\mathbb{R})$ and $\mathbb{H}$). However, as the proof of Proposition 12.17 for $d = 3$ below uses a more complicated version of the following calculation, we include the $d = 2$ case here.

We define $B_R = \{g \in \mathrm{SL}_2(\mathbb{R}) \mid \|g\| \leqslant R\}$. As mentioned above,

$$\begin{pmatrix} r & \\ & r^{-1} \end{pmatrix} \mathrm{k}_\psi, tk_\phi \begin{pmatrix} r & \\ & r^{-1} \end{pmatrix} \begin{pmatrix} & 1 \\ -1 & \end{pmatrix} k_\psi |$$
$$\mathrm{d}\psi \, \mathrm{d}\phi \, \mathrm{d}r \, \mathrm{d}t,$$

$$m_{\mathrm{SL}_2(\mathbb{R})}(B_R) = m_{\mathbb{R}^4}(\{tg \mid g \in B_R, t \in [0,1]\}) = \int_0^1 \int_1^{R_0} \int_0^{2\pi} \int_0^{\pi} |\det \left(k_\phi \begin{pmatrix} r & \\ & r^{-1} \end{pmatrix} k_\psi, tk_\phi \begin{pmatrix} 1 & \\ & -r^{-2} \end{pmatrix} k_\psi, \ 1 - \right.$$
$$\begin{pmatrix} r & \\ & r^{-1} \end{pmatrix} \mathrm{k}_\psi, tk_\phi \begin{pmatrix} r & \\ & r^{-1} \end{pmatrix} \begin{pmatrix} & 1 \\ -1 & \end{pmatrix} k_\psi |$$
$$\mathrm{d}\psi \, \mathrm{d}\phi \, \mathrm{d}r \, \mathrm{d}t,$$

where we used the KAK decomposition to parameterize

$$k_\phi \begin{pmatrix} r & \\ & r^{-1} \end{pmatrix} k_\psi \in \mathrm{SL}_2(\mathbb{R}),$$

Need to check order of variables in integral; should pair up in layers

the parameter $R_0 > 1$ is chosen so that $\sqrt{R_0^2 + R_0^{-2}} = R$, and the 2×2 matrices in the determinant above are the partial derivatives of this parameterization, and these should be converted into ordinary 4-dimensional vectors before the determinant is taken. This calculation leads to

Surely R_0 can't be allowed to be small?

$$m_{\mathrm{SL}_2(\mathbb{R})}(B_R) = \int_0^1 t^3 \, \mathrm{d}t \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\psi \int_1^{R_0} \det \begin{pmatrix} r & 1 \\ r^{-1} & r^{-2} \end{pmatrix} \det \begin{pmatrix} -r & -r^{-1} \\ r^{-2} & r \end{pmatrix} \mathrm{d}r$$

by taking the factor t out of the determinant, noticing that the matrices k_ϕ, k_ψ on the left (respectively, right) appear in each matrix and hence do not affect the total determinant, and by splitting the resulting determinant into the determinant of the diagonal (respectively, the determinant of the off-diagonal) entries. Thus

$$\begin{aligned} m_{\mathrm{SL}_2(\mathbb{R})}(B_R) &\propto \int_1^{R_0} 2r^{-1}\left(r^2 - r^{-2}\right) \mathrm{d}r \\ &\propto \left(R_0^2 - 1\right) + \left(R_0^{-2} - 1\right) \sim R_0^2 \sim R, \end{aligned}$$

which shows the proposition. □

Proof of Proposition 12.17 for $d = 3$. We set $K = \mathrm{SO}(3)(\mathbb{R})$ and let

$$U \ni \phi \longmapsto k_\phi \in K$$

be a piecewise smooth parameterization, and write

$$a_{r_1,r_2} = \begin{pmatrix} r_1 & & \\ & r_2 & \\ & & 1/r_1 r_2 \end{pmatrix}.$$

As in the case $d = 2$, we have

$$\begin{aligned} m_{\mathrm{SL}_3(\mathbb{R})}\left(B_R\right) &= m_{\mathbb{R}^9}\left(\{tg \mid g \in B_R, t \in [0,1]\}\right) \\ &\propto \int_0^1 \int_{(r_2,r_2)\in S_R} \int_{\phi\in U} \int_{\psi\in U} \\ &\qquad \det\big(k_\phi a_{r_1,r_2} k_\psi^{\mathrm{t}}, t k_\phi \delta_{r_1} a_{r_1,r_2} k_\psi, t k_\phi \delta_{r_2} a_{r_1,r_2} k_\psi, \\ &\qquad\quad t\delta_{\phi_1} k_\phi a_{r_1,r_2} k_\psi^{\mathrm{t}}, t k_\phi a_{r_1,r_2} \delta_\phi k_\psi^{\mathrm{t}}\big)\,\mathrm{d}\psi\,\mathrm{d}\phi\,\mathrm{d}r_1\,\mathrm{d}r_2\,\mathrm{d}t, \end{aligned}$$

where

$$S_R = \left\{(r_1,r_2) \in \mathbb{R}^2 \mid r_1 \geqslant r_2 \geqslant \tfrac{1}{r_1 r_2} > 0 \text{ and } \sqrt{r_1^2 + r_2^2 + \tfrac{1}{r_1^2 r_2^2}} \leqslant R\right\}.$$

The integration with respect to t produces the factor $\frac{1}{9}$.

Just as in the case $d = 2$, the main interest arises from the integration over $(r_1, r_2) \in S_R$. There are, however, some differences between the two cases. Firstly, the domain S_R is more complicated, and for part of the calculation we will slightly simplify this domain by using the set

$$\widetilde{S_R} = \left\{(r_1,r_2) \in \mathbb{R}^2 \mid r_1 \geqslant r_2 \geqslant \tfrac{1}{r_1 r_2} > 0 \text{ and } \sqrt{r_1^2 + r_2^2} \leqslant R\right\}.$$

Secondly, the paramaterization of K does not have a constant Jacobian

$$J(\phi)\frac{\mathrm{d}m_K}{\mathrm{d}(k\cdot)_*(m_{\mathbb{R}})},$$

and as we do not care about scale factors like $m_K(K)$ we instead concentrate on the distortion of measure in moving from $[0,1] \times K \times S_R \times K$ to $\mathbb{R}^9$. We do this by working with the left-invariant vector fields

$$k_\phi^{u,u} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, k_\phi^{v,v} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, k_\phi^{w,w} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

for $\phi \in U$ instead of the partial derivatives with respect to ϕ_1, ϕ_2, ϕ_3 (and similarly ψ_1, ψ_2, ψ_3). This makes the determinant function invariant under ϕ (respectively, under ψ) since each of the matrices in the determinant again has k_ϕ on the left and k_ψ on the right. This gives

$$m_{\mathrm{SL}_3(\mathbb{R})} = \int_{S_R} |\det(a_{r_1,r_2}, \delta_{r_1} a_{r_1,r_2}, \delta_{r_2} a_{r_1,r_2}, u a_{r_1,r_2}, \\ v a_{r_1,r_2}, w a_{r_1,r_2}, a_{r_1,r_2} u, a_{r_1,r_2} v, a_{r_1,r_2} w)| \,\mathrm{d}r_1\,\mathrm{d}r_2.$$

As before, each matrix a_{r_1,r_2} and so on should be thought of as a 9-dimensional vector, so that we can take the determinant of the resulting 9×9 matrix. The matrix has block form, with

$$a_{r_1,r_2} = \begin{pmatrix} 0 & & \\ & r_2 & \\ & & 1/r_1r_2 \end{pmatrix}, \delta_{r_1} a_{r_1,r_2} = \begin{pmatrix} 1 & & \\ & 0 & \\ & & -1/r_1^2 r_2 \end{pmatrix}, \delta_{r_2} a_{r_1,r_2} = \begin{pmatrix} 0 & & \\ & 1 & \\ & & -1/r_1 r_2^2 \end{pmatrix}$$

forming one 3×3 block, and

$$u a_{r_1,r_2} = \begin{pmatrix} & r_2 & \\ -r_1 & & \\ 0 & 0 & 0 \end{pmatrix}, a_{r_1,r_2} u = \begin{pmatrix} 0 & r_1 & 0 \\ -r_2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

respectively

$$v a_{r_1,r_2} = \begin{pmatrix} 0 & 0 & 1/r_1r_2 \\ 0 & 0 & 0 \\ -r_1 & 0 & 0 \end{pmatrix}, a_{r_1,r_2} v = \begin{pmatrix} 0 & 0 & r_1 \\ 0 & 0 & 0 \\ -1/r_1r_2 & 0 & 0 \end{pmatrix},$$

respectively

$$w a_{r_1,r_2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/r_1r_2 \\ 0 & -r_2 & 0 \end{pmatrix}, a_{r_1,r_2} w = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & r_2 \\ 0 & -1/r_1r_2 & 0 \end{pmatrix}$$

the other three 2×2 blocks. This allows us to calculate the determinant without too much pain, giving

$$m_{\mathrm{SL}_3(\mathbb{R})}(B_R) \propto \int_{S_R} \left(\frac{3}{r_1r_2}\right)(r_1^2 - r_2^2)\Bigl(\underbrace{r_1^2 - \frac{1}{r_1^2 r_2^2}}_{\sim r_1^2}\Bigr)\left(r_2^2 - \frac{1}{r_1^2 r_2^2}\right)\mathrm{d}r_1\,\mathrm{d}r_2.$$

Notice that $\frac{1}{r_1r_2} \to 0$ as $r_1 \to \infty$, so we may replace $r_1^2 - \frac{1}{r_1^2r_2^2}$ by r_1^2 without affecting the asymptotic behavior.

As mentioned earlier, the original domain of integration is difficult to work with, so we instead consider the set $\widetilde{S_R}$ (which is only slightly less annoying). Denote by $\widetilde{B_R} \supseteq B_R$ the set corresponding to $\widetilde{S_R}$, and calculate

$$m_{\mathrm{SL}_3(\mathbb{R})}(\widetilde{B_R}) \sim \int_{\widetilde{S_R}} \frac{r_1}{r_2}(r_1^2 - r_2^2)\left(r_2^2 - \frac{1}{r_1^2r_2^2}\right)\mathrm{d}r_1\,\mathrm{d}r_2$$

by splitting $\widetilde{S_R}$ into the triangle-like region

$$\left\{(r_1, r_2) \mid r_1 \geqslant r_2 \geqslant \frac{1}{r_1 r_2} > 0 \text{ and } r_1 \leqslant \frac{R}{\sqrt{2}}\right\}$$

and the region

$$\left\{(r_1, r_2) \mid r_1 \geqslant r_2 \geqslant \frac{1}{r_1 r_2} > 0, r_1 \geqslant \frac{R}{\sqrt{2}}, \text{ and } \sqrt{r_1^2 + r_2^2} \leqslant R\right\}.$$

This gives

$$m_{\mathrm{SL}_3(\mathbb{R})}(\widetilde{B_R}) \sim (acalculation) R^6.$$

Recall that $B_R \subseteq \widetilde{B_R}$. However, for any $\kappa > 0$ we also have $\widetilde{B}_{R-\kappa} \subseteq B_R$ for all sufficiently large R. Therefore, $m_{\mathrm{SL}_3(\mathbb{R})}(B_R)$ has the same asymptotics, giving the proposition. □

12.6 Computing the Volume of 2 and 3

In this section we will describe a method for calculating volume(d) without actually finding a fundamental domain for $\mathrm{SL}_d(\mathbb{Z}) < \mathrm{SL}_d(\mathbb{R})$. Of course the answer depends on a normalization of the Haar measure on $\mathrm{SL}_d(\mathbb{R})$. Both the normalization and the method to find the volume work inductively.

Theorem 12.19 (Volume of 2). *Normalize the Haar measure on* $\mathrm{SL}_2(\mathbb{R})$ *by setting* $m_{\mathrm{SL}_2(\mathbb{R})} = m_{NA} \times m_K$ *in the* NAK *coordinates, where* $m_K(K)$ *is the Haar measure on* $K = \mathrm{SO}(2)(\mathbb{R})$ *with* $m_K(K) = 2\pi$ *and* m_{NA} *is the left Haar measure on* $NA \cong \mathbb{H}$ *discussed in Section* 1.2. *Then* volume$(2) = \frac{\pi^2}{3}$.

Theorem 12.19 can also be deduced from the Gauss–Bonnet formula in hyperbolic geometry (see [**?**] for the details), but we will give an independent proof based on Section 12.3. Using the same kind of argument in higher dimensions gives the following result.

Need to come back and fix the factor of 2 as this gives volume of X_2 to be $\pi^2/6$ (which seems to be the normal normalization of Haar measure)

Theorem 12.20 (Volume of $d+1$). *For* $d \geqslant 1$ *define the subgroup*

$$H = \left\{\begin{pmatrix} 1 & w \\ 0 & g \end{pmatrix} \mid g \in \mathrm{SL}_d(\mathbb{R}), w \in \mathbb{R}^d\right\}.$$

Assume by induction on d that $m_{\mathrm{SL}_d(\mathbb{R})}$ *has been defined. Using the Lebesgue measure on* $\mathbb{R}^d$ *(and Lemma* 1.22*) this defines a normalization of the Haar measure* $m_H = m_{\mathrm{SL}_d(\mathbb{R})} \times m_{\mathbb{R}^d}$. *Using the identification*

$$V = \mathrm{SL}_d(\mathbb{R})/H \cong \mathbb{R}^{d+1} \smallsetminus \{0\},$$

normalize $m_{\mathrm{SL}_{d+1}(\mathbb{R})}$ *to be compatible with the Lebesgue measure on* $\mathbb{R}^{d+1}$. *Then we have* $\mathrm{volume}(d+1) = \zeta(d+1)\,\mathrm{volume}(d)$.

The standing assumption in Section 12.3 were that

$$\mathrm{volume}(G/\Gamma)$$

and

$$\mathrm{volume}(H/\Gamma \cap H)$$

were both finite. This follows in the case at hand from Theorem 1.18 and the fact that the unipotent radical

$$\left\{ \begin{pmatrix} 1 & w^{\mathrm{t}} \\ & 1 \end{pmatrix} \mid w \in \mathbb{R}^d \right\}$$

intersects $\mathrm{SL}_d(\mathbb{Z})$ in a lattice.

The dynamical assumption of equidistribution of $gH\Gamma$ in

$$X = d + 1 = G/\Gamma$$

for $G = \mathrm{SL}_{d+1}(\mathbb{R})$, $\Gamma = \mathrm{SL}_{d+1}(\mathbb{Z})$ and $gH \longrightarrow \infty$ in G/H is easy to establish — where it is true. In order to do this, we again need to exploit two decompositions of G.

Lemma 12.21. *Write*

$$K = \mathrm{SO}(d+1)(\mathbb{R})$$

and

$$A = \left\{ \begin{pmatrix} a & \\ & a^{-1/d} I \end{pmatrix} \mid a > 0 \right\}.$$

Then $G = KAH$.

Lemma 12.22. *Let*

$$a_1 = \begin{pmatrix} \mathrm{e} & \\ & \mathrm{e}^{-1/d} I \end{pmatrix} \in A$$

and

$$G_{a_1}^{-} = \left\{ \begin{pmatrix} 1 & \\ v & 1 \end{pmatrix} \mid v \in \mathbb{R}^d \right\}.$$

Then $G_{a_1}^{-} AH$ *contains* I *in its interior.*

Theorem 12.23. $gH\Gamma$ *equidistributes on average as* $gH \longrightarrow \infty$ *in* G/H.

Outline of proof. (1) By Lemma 12.21 it is sufficient to consider the case $gH = a_t H$ where

$$a_t = \begin{pmatrix} \mathrm{e}^t & \\ & \mathrm{e}^{-t/d} I \end{pmatrix} \in A$$

with $|t| \to \infty$.
(2) Notice that taking $t \leqslant 0$ in a_t corresponds to non-zero elements in the unit ball of $\mathbb{R}^{d+1}$, and that the unit ball has finite Haar measure on

$$G/H \cong \mathbb{R}^{d+1} \smallsetminus \{0\}$$

(since the Haar measure coincides with the Lebesgue measure). As a result we may ignore the case $t \to -\infty$ in the equidistribution claim sought.
(3) The remaining case $t \to \infty$ may be carried out as in Theorem 12.16. □

The geometric hypothesis that the sets

$$B_t = B_{\mathrm{e}^t}^{\mathbb{R}^d} = \left\{ v \in \mathbb{R}^d \mid \|v\| \leqslant \mathrm{e}^t \right\} \tag{12.13}$$

are well-rounded is easy to check (see Exercise 12.6.4).

Proof of Theorem 12.20. For $d = 2$ we have

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

and $\mathrm{volume}(H/\Gamma \cap H) = 1$. Furthermore, notice that

$$\mathrm{SL}_2(\mathbb{Z}) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left(\mathbb{Z}^2\right)^* = \left\{ n \in \mathbb{Z}^2 \mid \gcd(n) = 1 \right\}.$$

Hence

$$\lim_{R \to \infty} \frac{\left| (\mathbb{Z}^2)^* \cap B_R^{\mathbb{R}^2} \right|}{\pi R^2} = \frac{1}{m_2(2)}$$

by (12.8). By Exercise 12.1.1 we also know that

$$\lim_{R \to \infty} \frac{\left| (\mathbb{Z}^2)^* \cap B_R^{\mathbb{R}^2} \right|}{\pi R^2} = \frac{1}{\zeta(2)},$$

which implies that $m_2(2) = \zeta(2)$.

For $d \geqslant 2$ we have

$$\mathrm{volume}\,(H/\Gamma \cap H) = \mathrm{volume}(d)$$

and

$$\mathrm{SL}_d(\mathbb{Z}) e_1 = \left(\mathbb{Z}^d\right)^* = \left\{ n \in \mathbb{Z}^d \mid \gcd(n) = 1 \right\}. \tag{12.14}$$

Combining 12.8 and Exercise 12.1.1 we get once more

$$\lim_{R \to \infty} \frac{\left| \left(\mathbb{Z}^d\right)^* \cap B_R^{\mathbb{R}^d} \right|}{V_d R_d} = \frac{1}{\zeta(d)} = \frac{\mathrm{volume}(d)}{\mathrm{volume}(d+1)}$$

which gives the theorem. □

We leave the details of the arguments above to the exercises below.

Problems for Section 12.6

Exercise 12.6.1. Prove Lemma 12.21.

Exercise 12.6.2. Prove Lemma 12.22.

Exercise 12.6.3. Prove Theorem 12.23.

Exercise 12.6.4. Prove that the sets

$$B_t = B_{\mathrm{e}^t}^{\mathbb{R}^d} = \left\{ v \in \mathbb{R}^d \mid \|v\| \leqslant \mathrm{e}^t \right\}$$

are well-rounded (see (12.13)).

Exercise 12.6.5. Prove (12.14) for $d \geqslant 2$.

Exercise 12.6.6. Prove that

$$N_R = \left| \left\{ W \leqslant \mathbb{R}^d \mid \dim(W) = m, W \text{ is rational, and } (W \cap \mathbb{Z}^d) \leqslant R \right\} \right|$$

has an asymptotic of the form $N_R \sim cR^d$.

Query: An asymptotic count where mixing is not sufficient, then some other applications: Jens Marklof? McMullen on $\sqrt{n}$ mod 1

Notes to Chapter 12

[38](Page 365) The error term $N(R) - \pi R^2$ was shown to be bounded above by $2\sqrt{2}\pi R$ by Gauss. Hardy [**?**] and Landau [**?**] found a *lower* bound for the error by showing that the error is not $\mathrm{o}(R^{1/2}(\log R)^{1/4})$. It is conjectured that the upper bound is $\mathrm{O}_\varepsilon(R^{\frac{1}{2}+\varepsilon})$. The power of R must be at least $\frac{1}{2}$ by the lower bound of Hardy and Landau, and has been shown to be less than or equal to $\frac{131}{208}$ by Huxley [**?**].
[39](Page 369) For the history and primary references of these developments we refer to the paper of Phillips and Rudnick [**?**].
[40](Page 376) This was shown by Witt [**?**], and a modern treatment may be found in the monograph of Elman, Karpenko and Merkurjev [**?**].

Chapter 13
Diophantine Approximation

In this chapter we want to explore the rich interaction between homogeneous dynamics and Diophantine approximation. We will start by proving some relatively old and quite easy theorems using this connection (which is known as the Dani correspondence), and work our way up to more recent and more difficult theorems. As usual we will not attempt to be exhaustive, but will try to give a flavor of the methods and the type of results.

Maybe we should also mention Artin and Cassels Swinnerton-Dyer who probably used this connection to a lesser extend before Dani

13.1 Dirichlet's Theorem and Dani's Correspondence

We start with a classical result on simultaneous Diophantine approximation. In stating this we will write $\frac{p}{q}$ for the vector $(\frac{p_j}{q})_j$ if $p = (p_j)_j$ is a vector and $q \in \mathbb{N}$.

Theorem 13.1 (Dirichlet's Theorem [?]). *For any $v \in \mathbb{R}^d$ and any integer Q there exist an integer q with $1 \leqslant q \leqslant Q$, and an integer vector $p \in \mathbb{Z}^d$ with*

$$\left\| v - \tfrac{p}{q} \right\|_\infty \leqslant \tfrac{1}{qQ}.$$

The classical proof of Theorem 13.1. Consider the $(Q^d + 1)$ points

$$0, v, \dots, Q^d v \pmod{\mathbb{Z}^d} \tag{13.1}$$

as elements of $\mathbb{T}^d \cong [0,1)^d$. Now partition $[0,1)$ into the Q intervals

$$[0, \tfrac{1}{Q}), [\tfrac{1}{Q}, \tfrac{2}{Q}), \dots, [\tfrac{Q-1}{Q}, 1),$$

and correspondingly divide $[0,1)^d$ into Q^d cubes with sides chosen from the partition of each of the d axes. By the pigeonhole principle (which is for this reason sometimes called Dirichlet's principle$^{(41)}$) there exist two integers k, ℓ

with $0 \leqslant k < \ell \leqslant Q^d$ such that the points kv and ℓv considered modulo $\mathbb{Z}^d$ from (13.1) belong to the same subcube. Letting $q = \ell - k$ gives

$$\|qv - p\|_\infty \leqslant \tfrac{1}{Q}$$

for some $p \in \mathbb{Z}^d$ as required. □

Our first connection between Diophantine analysis and homogeneous dynamics concerns the following notion.

Definition 13.2. Fix $\lambda \in (0,1]$. A vector $v \in \mathbb{R}^d$ is called λ-*Dirichlet improvable* if for every large enough Q there exists an integer q satisfying

$$1 \leqslant q \leqslant \lambda Q^d \text{ and } \left\| v - \tfrac{p}{q} \right\|_\infty \leqslant \lambda \frac{1}{qQ} \tag{13.2}$$

for some $p \in \mathbb{Z}^d$. A vector is simply called *Dirichlet-improvable* if it is λ-improvable for some $\lambda < 1$.

In order to describe the correspondence between this notion and homogeneous dynamics, we write† $\mathit{\Lambda}_v = u_v \mathbb{Z}^{d+1}$ where

$$u_v = \begin{pmatrix} 1 & \\ v & I_d \end{pmatrix},$$

and

$$g_Q = \begin{pmatrix} Q^{-d} & \\ & QI_d \end{pmatrix}.$$

Proposition 13.3 (Dani correspondence). *Let $v \in \mathbb{R}^d$, $Q > 1$, and λ be given with $0 < \lambda \leqslant 1$. Then there exists an integer q satisfying* (13.2) *if and only if the lattice in $\mathbb{R}^{d+1}$ corresponding to $g_Q \mathit{\Lambda}_v$ intersects $[-\lambda, \lambda]^{d+1}$ non-trivially.*

Proof. Suppose that the integer q satisfies (13.2) for some $p \in \mathbb{Z}^d$. Then the vector

$$\begin{pmatrix} q \\ qv - p \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} \begin{pmatrix} q \\ -p \end{pmatrix} \in \mathit{\Lambda}_v$$

belongs to the lattice corresponding to v, and

$$g_Q \begin{pmatrix} q \\ qv - p \end{pmatrix} \in g_Q \mathit{\Lambda}_v$$

satisfies

---

† Most of the research papers concerning the interaction between homogeneous dynamics and Diophantine approximation use the description $X = G/\mathit{\Gamma}$ instead of $X = \mathit{\Gamma} \backslash G$, so we will adhere to this tradition here.

$$\left\| g_Q \begin{pmatrix} q \\ qv - p \end{pmatrix} \right\|_\infty = \max\left(|qQ^{-D}|, Q\|qv-p\|_\infty\right) \leqslant \lambda. \qquad (13.3)$$

Now suppose on the other hand that there is a non-trivial vector

$$\begin{pmatrix} q \\ qv - p \end{pmatrix} \in \Lambda_v$$

satisfying (13.3). We claim that $q \neq 0$. Assuming this for the moment, we may also assume that q is positive†, and then (13.3) is equivalent to (13.2).

To prove the claim, suppose that $q = 0$. However, in this case (13.3) becomes

$$\left\| g_Q \begin{pmatrix} 0 \\ p \end{pmatrix} \right\| = Q\|p\| \leqslant \lambda \leqslant 1,$$

which forces $p = 0$ since $Q > 1$. This contradicts our assumption that $\begin{pmatrix} q \\ qv - p \end{pmatrix}$ is non-trivial, which proves the claim and completes the proof. $\square$

This correspondence allows Dirichlet's theorem to be proved using homogeneous dynamics.

Proof of Theorem 13.1 using dynamics. Set $\lambda = 1$ in Proposition 13.3, and notice that any unimodular lattice in $\mathbb{R}^{d+1}$ has to intersect $[-1,1]^{d+1}$ by Theorem 1.14. $\square$

Using ergodicity of the dynamics of

$$a_t = \begin{pmatrix} \mathrm{e}^{-dt} & \\ & \mathrm{e}^{t} I_d \end{pmatrix}$$

on $d+1 = \mathrm{SL}_{d+1}(\mathbb{R})/\mathrm{SL}_{d+1}(\mathbb{Z})$ we can prove the following, recovering a result of Davenport and Schmidt [**?**].

Corollary 13.4. *Almost no vector $v \in \mathbb{R}^d$ is Dirichlet-improvable.*

Proof. Let $\lambda \in (0,1)$ and define the open neighborhood

$$O_\lambda = \{\Lambda \in d+1 \mid \Lambda \cap [-\lambda,\lambda]^{d+1} = \{0\}\}$$

of $\mathbb{Z}^{d+1}$. Furthermore, let $O'_\lambda \subseteq O_\lambda$ be a non-trivial open subset with $\overline{O'_\lambda} \subseteq O_\lambda$, so that in particular the Hausdorff distance

$$\varepsilon = \mathsf{d}(O'_\lambda, d+1 \smallsetminus O_\lambda)$$

from O'_λ to the complement of O_λ is positive. By Exercise 2.3.7, almost every $v \in \mathbb{R}^d$ has the property that the a_t-orbit of $\Lambda_v = u_v\mathbb{Z}^{d+1}$ is dense.

† For otherwise we may replace $\begin{pmatrix} q \\ qv - p \end{pmatrix}$ by $\begin{pmatrix} -q \\ -qv + p \end{pmatrix}$.

Now let $t \geqslant 0$ be very large and such that $a_t \boldsymbol{\cdot} \mathit{\Lambda}_v \in O'_\lambda$. We set $Q = \lfloor \mathrm{e}^t \rfloor$ and deduce that

$$g_Q a_t^{-1} = \begin{pmatrix} Q^d \mathrm{e}^{-td} & \\ & Q^{-1} \mathrm{e}^t I_d \end{pmatrix}$$

is very close to I_{d+1}, in particular $\mathsf{d}(g_Q a_t^{-1}, I) < \varepsilon$ for sufficiently large t. From this we conclude that

$$g_Q \mathit{\Lambda}_v = g_Q a_t^{-1} a_t \mathit{\Lambda}_v \in O_\lambda.$$

By Proposition 13.3, there is no integer q satisfying (13.2) for Q. It follows that v is not λ-Dirichlet improvable.

Applying this to $\lambda = 1 - \frac{1}{n}$ gives the corollary. □

A general theme in the theory of Diophantine approximation is to try and show inheritance of Diophantine properties on $\mathbb{R}^d$ to submanifolds, or even more generally to fractals(42). We will simply prove a few sample results in this direction. The following is a corollary of the quantitative non-divergence result Theorem 4.9, and is a special case of a more recent result of Weiss.

Corollary 13.5. *For $d \geqslant 2$ there exists some $\lambda_0 \in (0,1)$ such that almost no $t \in \mathbb{R}$ has the property that*

$$v(t) = \begin{pmatrix} t \\ t^2 \\ \vdots \\ t^d \end{pmatrix} \in \mathbb{R}^d$$

is λ_0-Dirichlet improvable.

For the proof we will need the following geometric input regarding the dynamics of g_Q on $\mathbb{R}^{d+1}$.

Lemma 13.6. *Let $W \subseteq \mathbb{R}^{d+1}$ be a k-dimensional subspace, and let*

$$c = \mathsf{d}(e_1, W) = \inf_{w \in W} \|e_1 - w\| \geqslant 0.$$

If $w_1, \dots, w_k \in W$ is an orthonormal basis of W then

$$\|g_Q w_1 \wedge \cdots \wedge g_Q w_k\| \geqslant c Q^k.$$

Proof. Let $w_i = \varepsilon_i e_1 + w'_i$ with $\varepsilon_i \in \mathbb{R}$ for $i = 1, \dots, d$. Clearly

$$w_1 \wedge \cdots \wedge w_k = w'_1 \wedge \cdots \wedge w'_k + \sum_{i=1}^{k} \varepsilon_i \underbrace{w'_1 \wedge \cdots \wedge e_1 \wedge \cdots \wedge w'_k}_{\text{with } e_1 \text{ in place of } w'_i}. \tag{13.4}$$

We calculate the norm using the fact that e_1 is normal to $w'_1, \dots, w'_k$, the identity (13.4), the fact that $\|w_1 \wedge \cdots \wedge w_k\| = 1$ to obtain

$$\begin{aligned}\|w_1' \wedge \cdots \wedge w_k'\| &= \|e_1 \wedge w_1' \wedge \cdots \wedge w_k'\| \\ &= \|e_1 \wedge w_1 \wedge \cdots \wedge w_k\| = c \cdot 1\end{aligned}$$

since the distance from e_1 to W is c. We now apply the map g_Q to get

$$\begin{aligned}\|g_Q w_1 \wedge \cdots \wedge g_Q w_k\|^2 = \|g_Q w_1' \wedge \cdots \wedge g_Q w_k'\|^2 \\ + \Big\| \sum_{i=1}^{k} \varepsilon_i g_Q w_1' \wedge \cdots g_Q e_1 \wedge \cdots g_Q w_k' \Big\|^2 \\ \geqslant Q^k c,\end{aligned}$$

where we have used the fact that

$$g_Q w_1' \wedge \cdots \wedge g_Q w_k' = Q^k w_1' \wedge \cdots \wedge w_k'$$

is orthogonal to the sum. $\square$

Proof of Corollary 13.5. We set $\eta = 1$ and wish to apply Theorem 4.9 on $d+1$. We will define the precise polynomial for the application of Theorem 4.9 below, but for now let us agree that this will be a modified version of the polynomial

$$p_0(t) = \begin{pmatrix} 1 & \\ v(t) & I_d \end{pmatrix} = \begin{pmatrix} 1 & & & \\ t & 1 & & \\ \vdots & & \ddots & \\ t^d & & & 1 \end{pmatrix},$$

so that the parameter D in Theorem 4.9 is already determined. By that theorem, and the equivalence of the norms $\|\cdot\|$ and $\|\cdot\|_\infty$ on $\mathbb{R}^{d+1}$, there exists some $\varepsilon > 0$ so that for any polynomial $p(t)$ (with the same D as p_0) satisfying

$$\sup_{t\in[0,T]} (V, T) \geqslant 1 \tag{13.5}$$

for all rational subspaces $V \subseteq \mathbb{R}^d$ has

$$\frac{1}{T} \left|\left\{t \in [0,T] \mid p(t)\mathbb{Z}^{d+1} \notin O_\varepsilon\right\}\right| \leqslant \frac{1}{2}. \tag{13.6}$$

Now assume that the corollary is false for $\lambda_0 = \varepsilon$. Then

$$DT_\varepsilon = \{t \in \mathbb{R} \mid v(t) \text{ is } \varepsilon\text{-Dirichlet improvable}\}$$

must have a Lebesgue density point. In particular, there exists an interval $[\alpha, \beta] \subseteq \mathbb{R}$ such that

$$\frac{1}{\beta - \alpha} \left|\{t \in [\alpha, \beta] \mid v(t) \text{ is } \varepsilon\text{-Dirichlet improvable}\}\right| \geqslant \frac{9}{10}.$$

Using the definition of ε-Dirichlet improvable (and the basic property of measures), we find some Q_0 such that

$$\frac{1}{\beta-\alpha}\left|\{t\in[\alpha,\beta]\mid v(t) \text{ satisfies (13.2) with } \lambda=\varepsilon \text{ for every } Q\geqslant Q_0\}\right|\geqslant\frac{3}{4}.$$

Using the matrix $p_0(t)=\begin{pmatrix}1 & \\ v(t) & I_d\end{pmatrix}$ and Dani's correspondence (Proposition 13.3), we can also phrase this as

$$\frac{1}{\beta-\alpha}\left|\left\{t\in[\alpha,\beta]\mid g_Qp_0(t)\mathbb{Z}^{d+1}\notin Q_\varepsilon \text{ for every } Q\geqslant Q_0\right\}\right|\geqslant\frac{3}{4}. \tag{13.7}$$

To get a contradiction to (13.6), we have to show the assumption (13.5) for

$$p(t)=g_Qp_0(\alpha+t)$$

and $T=\beta-\alpha$.

Now assume that (13.5) does not hold, meaning that for every $Q\geqslant Q_0$ there is a rational subspace $V_Q\subseteq\mathbb{R}^d$ with

$$\sup_{t\in[\alpha,\beta]}(V_Q,t)<1,$$

where we are using $g_Qp(t)$ for the definition of (V_Q,t). This implies that

$$\sup_{t\in[0,1]}(V_Q,t)\ll_{\alpha,\beta}1.$$

We set $t=\frac{i}{d}$, $W=p\left(\frac{i}{d}\right)V_Q$ and obtain

$$\mathsf{d}(e_1,W)\leqslant\|g_Qw_1\wedge\cdots\wedge g_Qw_k\|=(V_Q,\tfrac{i}{d})\ll_{\alpha,\beta}1$$

from Lemma 13.6. Applying $p_0\left(\frac{i}{d}\right)^{-1}$, this gives

$$\mathsf{d}\left(p_0\left(\tfrac{i}{d}\right)^{-1}e_1,V_Q\right)\ll_{\alpha,\beta}Q^{-1} \tag{13.8}$$

for $i=1,\dots,d$. However, the vectors $p_0\left(\frac{i}{d}\right)^{-1}e_1$ for $i=0,1,\dots,d$ are linearly independent (see Exercise 13.1.2), and for large enough Q the condition (13.8) forces $V_Q=\mathbb{R}^{d+1}$ (see Exercise 13.1.3). Since

$$(\mathbb{R}^{d+1},t)=1$$

this contradicts our choice of V_Q, which proves (13.5) for large enough Q and gives a contradiction between (13.6) and (13.7). □

Using unipotent dynamics (Ratner's measure classification and the full force of the linearization technique) Shah significantly strengthened Corollary 13.5, giving in particular the following result.

Theorem (Shah [?]). Let $d \geqslant 2$. Then almost no $t \in \mathbb{R}$ has the property that

$$v(t) = \begin{pmatrix} t \\ \vdots \\ t^d \end{pmatrix}$$

is Dirichlet improvable.

This result is a consequence of a more general equidistribution theorem, a special case of which is the following.

Theorem (Shah [?]). Let $d \geqslant 2$, let $I \subseteq \mathbb{R}$ be a non-trivial compact interval, and let μ_I be the image of the Lebesgue measure under the map

$$I \ni t \longmapsto \begin{pmatrix} 1 & \\ v(t) & I_d \end{pmatrix} \mathbb{Z}^d.$$

Then

$$(g_Q)_* \mu_I \longrightarrow m_{d+1}$$

in the weak * topology as $Q \to \infty$.

Exercises for Section 13.1

Exercise 13.1.1. Show that there exists some $\lambda_0 \in (0,1)$ such that if $v \in \mathbb{R}$ is λ_0-Dirichlet improvable, then $v \in \mathbb{Q}$.

Exercise 13.1.2. We let $v_i = p(i/d)^{-1} e_1$ for $i = 0, \ldots, d$. Show that these vectors are linearly independent in $\mathbb{R}^{d+1}$.

Exercise 13.1.3. Let $W \subseteq \mathbb{R}^{d+1}$ be a subspace and let v_i for $i = 0, \ldots, d$ be a basis of $\mathbb{R}^{d+1}$. Show there there exists some ε such that if $d(v_i, W) \leqslant \varepsilon$ for $i = 0, \ldots, d$ then $W = \mathbb{R}^{d+1}$.

Exercise 13.1.4. A vector $v \in \mathbb{R}^d$ is called *singular* if v is λ-Dirichlet improvable for all $\lambda \in (0,1)$. Prove that v is singular if and only if

$$a_t \begin{pmatrix} 1 & \\ v & I_d \end{pmatrix} \mathbb{Z}^{d+1}$$

diverges to ∞ as $t \to \infty$.

Exercise 13.1.5. Show that Shah's equidistribution result implies the result on Dirichlet improvability, where in fact it would suffice to know that

$$\frac{1}{T} \int_0^T (a_t)_* \mu_I \,\mathrm{d}t \longrightarrow m_{d+1}.$$

Exercise 13.1.6. In this exercise we discuss one idea in the work of Shah, revealing the connection to unipotent dynamics which is not immediately apparent. We assume that the derivative of

$$v(t) = \begin{pmatrix} t \\ \vdots \\ t^d \end{pmatrix}$$

is non-zero on I, and choose a continuous map

$$I \ni t \longmapsto k(t) \in \mathrm{SL}_d(\mathbb{R})$$

such that $e_1 k(t) = v'(t)$. We now modify the measure μ_I and define $\widetilde{\mu_I}$ to be the normalized image of the Lebesgue measure under the map

$$I \ni t \longmapsto \begin{pmatrix} 1 & \\ & k(t) \end{pmatrix} \begin{pmatrix} 1 & \\ v(t) & I_d \end{pmatrix} \mathbb{Z}^{d+1}.$$

(a) Prove that any weak*-limit of $(a_t)_* \widetilde{\mu_I}$ as $t \to \infty$ is invariant under the action of

$$\begin{pmatrix} 1 & \\ e_1 & I_d \end{pmatrix}.$$

(b) Show that if $(a_t)_* \widetilde{\mu_I}$ converges along a subsequence to m_{d+1}, then $(a_t)_* \mu_I$ also converges to m_{d+1}.

Exercise 13.1.7. Show that m_3 is the only probability measure that is simultaneously invariant under the action of $\{a_t \mid t \in \mathbb{R}\}$ and the action of

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

Combine this with the result in Exercises 13.1.5–13.1.6 to obtain the special case $d = 3$ in the Theorem of Shah on Dirichlet improvability on page 403.

13.2 Well and Badly Approximable Vectors

Let us note an immediate corollary of Dirichlet's theorem (Theorem 13.1).

Corollary 13.7. *For any $v \in \mathbb{R}^d$ there are infinitely many $p \in \mathbb{Z}^d$ and integers $q \geqslant 1$ with*

$$\left\| v - \frac{p}{q} \right\|_\infty \leqslant \frac{1}{q^{1+\frac{1}{d}}}.$$

Definition 13.8. A vector $v \in \mathbb{R}^d$ is called *well-approximable* if, for every $\varepsilon > 0$, the inequality

$$\left\| v - \frac{p}{q} \right\|_\infty \leqslant \frac{\varepsilon}{q^{1+\frac{1}{d}}} \tag{13.9}$$

has infinitely many solutions $p \in \mathbb{Z}^d$, $q \geqslant 1$. A vector $v \in \mathbb{R}^d$ is called *badly approximable* or BA if there exists some $\varepsilon > 0$ such that the inequality (13.9) has no solution.

Recall from Section 13.1 the notation

$$a_t = \begin{pmatrix} \mathrm{e}^{-dt} & \\ & \mathrm{e}^t I_d \end{pmatrix}$$

and

$$\varLambda_v = \begin{pmatrix} 1 & \\ v & I_d \end{pmatrix} \mathbb{Z}^{d+1}.$$

Proposition 13.9 (Dani's correspondence). *A vector $v \in \mathbb{R}^d$ is* well-approximable *if and only if the forward orbit*

$$\left\{ a_t \begin{pmatrix} 1 & \\ v & I_d \end{pmatrix} \mathbb{Z}^{d+1} \mid t \geqslant 0 \right\}$$

is unbounded (that is, has non-compact closure in $d+1$), and is badly approximable *if and only if the forward orbit is bounded (that is, has compact closure).*

Proof. Suppose that v is well-approximable, $\varepsilon > 0$, and $p \in \mathbb{Z}^d$, $q \geqslant 1$ have the property (13.9). Then

$$\begin{pmatrix} q \\ qv - p \end{pmatrix} \in \varLambda_v = \begin{pmatrix} 1 & \\ v & I_d \end{pmatrix} \mathbb{Z}^{d+1}$$

is a non-trivial vector, and we may choose $t \geqslant 0$ with

$$\mathrm{e}^{-td} q = \varepsilon^{1/2}$$

and so we also have

$$\|qv - p\|_\infty \leqslant \frac{\varepsilon}{q^{1/d}} = \varepsilon^{1-\frac{1}{2d}} \mathrm{e}^{-t}.$$

Thus $a_t \varLambda_v$ contains the vector

$$\begin{pmatrix} \mathrm{e}^{-td} q \\ \mathrm{e}^t (qv - p) \end{pmatrix}$$

of length $\ll \varepsilon^{1/2}$. As $\varepsilon > 0$ was arbitrary, Mahler's compactness criterion (Theorem 1.17) shows the orbit is unbounded.

Now suppose that the orbit is unbounded, let $\varepsilon \in (0,1)$ and suppose that $a_t \varLambda_v$ for $t \geqslant 0$ contains a non-trivial ε-short vector, say

$$\begin{pmatrix} \mathrm{e}^{-td} q \\ \mathrm{e}^t (qv - p) \end{pmatrix}$$

with

$$\left\| \begin{pmatrix} \mathrm{e}^{-td} q \\ \mathrm{e}^{t}(qv - p) \end{pmatrix} \right\|_{\infty} \leqslant \varepsilon.$$

Notice that if $q = 0$ then we would also have $p = 0$. Hence $q \neq 0$ and we may therefore suppose that $q \geqslant 1$. We have $\mathrm{e}^{-t} q^{1/d} \leqslant \varepsilon^{1/d}$ and $\mathrm{e}^{t}\|qv - p\| \leqslant \varepsilon$, so by taking the product we show (13.9) (with ε replaced by $\varepsilon^{1+\frac{1}{d}}$).

The argument above also implies the Dani correspondence for badly approximable vectors. $\square$

The following is immediate from Dani's correspondence and Exercise 2.3.7 (but see also Exercise 13.2.1 for a deeper result in this direction).

Corollary 13.10. *Almost every $v \in \mathbb{R}^d$ is well-approximable.*

W. Schmidt proved in [**?**] proved that the set of badly approximable vectors has full Hausdorff dimension in $\mathbb{R}^d$. For this purpose he invented a type of game, now known as a *Schmidt game.* We refer to the more recent papers of Kleinbock and Weiss [**?**] and McMullen [**?**] on the details for this game, more recent modifications and connections to dynamics, and Diophantine approximation.

Exercises for Section 13.2

Exercise 13.2.1. Show that the equidistribution theorem of Shah on page 403 implies that the vector

$$\begin{pmatrix} t \\ \vdots \\ t^d \end{pmatrix}$$

is well-approximable for almost every $t \in \mathbb{R}$.

Notes to Chapter 13

(41)(Page 397) While this principle — in this finite form particularly — must date from antiquity, Dirichlet [**?**] seems to have been one of the first to use it in a formal way with this kind of application in mind, calling it the *Schubfachprinzip* (drawer or shelf principle).
(42)(Page 400) We will not discuss this more general framework concerning the inheritance of Diophantine properties to 'sufficiently curved smooth manifolds' and simply mention here some of the key developments. Davenport and Schmidt [**?**] showed that almost every point of $\mathbb{R}^d$ is not Dirichlet-improvable and later showed in [**?**] that almost every point on the curve (t, t^2) is not $(1/4)$-improvable. Baker [**?**] extended this to the same statement for almost every point on a sufficiently smooth curve in $\mathbb{R}^2$, and to almost every point on a sufficiently smooth curved manifold by Dodson, Rynne, and Vickers [**?**]. Bugeaud [**?**] extended the result to the specific curve $(t, t^2, \dots, t^d)$. Kleinbock and Weiss [**?**] used the correspondence introduced by Dani [**?**], [**?**] and the machinery of Kleinbock and Margulis [**?**]

to formulate some of these questions in homogeneous dynamics, and the argument used for the proof of Corollary 13.5 is the argument used in [?]. We refer to a paper of Shah [?] for more details on the background and for another direction of similar results for curves that do not lie in translates of proper subspaces.

Started a new page to push the empty sections onto a new page, in case we want to post this chapter before the remaining sections get written ....

13.3 Khintchin's Theorem and Homogeneous Dynamics

definition and discussion of Diophantine exponent

13.4 Higher-rank Phenomena

Littlewood's conjecture, almost every point on the middle third Cantor set is well-approximable but not Dirichlet improvable

13.5 Diophantine Approximation on Quadratic Surfaces

new work of Kleinbock et al.

Chapter 14
Oppenheim's Conjecture

Conjecture 14.1. Let Q be a non-degenerate† indefinite quadratic from in $d \geqslant 3$ variables. Then either

- $\lambda Q \in \mathbb{Z}[x_1, \dots, x_d]$ for some $\lambda \in \mathbb{R}$ — in which case Q is said to be *rational* — and $Q(\mathbb{Z}^d)$ is discrete in $\mathbb{R}$; or
- $\lambda Q \notin \mathbb{Z}[x_1, \dots, x_d]$ for all $\lambda \in \mathbb{R}$ — in which case Q is said to be *irrational* — and $Q(\mathbb{Z}^d)$ is dense in $\mathbb{R}$.

Remark 14.2. (a) Conjecture 14.1 should be compared to the case of a linear form $L : \mathbb{R}^d \to \mathbb{R}$, where a similar dichotomy holds (and is easy to show; see Exercise 14.0.1).
(b) It is clearly necessary that Q be indefinite, since otherwise $Q(\mathbb{Z}^d)$ is always discrete‡.
(c) The assumptions that $d \geqslant 3$ and Q is non-degenerate are also necessary because of badly approximable numbers. For example,

$$Q(x, y) = (x + \sqrt{2}y)(x + y)$$

is irrational and satisfies

$$Q(\mathbb{Z}^2) \cap (0, \varepsilon) = \varnothing$$

for some $\varepsilon > 0$. To see this, notice that

$$|(x + \sqrt{2}y)(x - \sqrt{2}y)| = |x^2 - 2y^2| \geqslant 1$$

unless $x = y = 0$. If $Q(x, y)$ is small but non-zero for integral x, y, then $x+y \neq 0$ and so again $|x + y| \geqslant 1$, and hence $x + \sqrt{2}y$ must be small. In that

† This will be defined precisely later, but for now take it to mean that Q really involves all d variables in an essential way.

‡ Though in the positive-definite case it is conjectured that $Q(\mathbb{Z}^d)$ should become more and more dense as the values it takes on become large, but this is unknown.

case $|x-\sqrt{2}y|$ and $|x+y|$ are both approximately $\|\begin{pmatrix}x\\y\end{pmatrix}\|$, which proves the claim.

(d) As the rational case is easy, we will assume that Q is irrational.

Raghunathan's conjecture (Theorem 6.3) was motivated by Oppenheim's conjecture and the following connection between the two. Suppose that the $\mathrm{SO}(Q)(\mathbb{R})^o$-orbit of the point $\mathrm{SL}_d(\mathbb{Z})$ (corresponding to the lattice $\mathbb{Z}^d$) in $X=\mathrm{SL}_d(\mathbb{Z})\backslash\mathrm{SL}_d(\mathbb{R})$ is dense in X, then $Q(\mathbb{Z}^d)\subseteq\mathbb{R}$ is dense also by the following argument. Notice that

$$\overline{Q(\mathbb{Z}^d)}=\overline{Q(h\mathbb{Z}^d)}$$

for any $h\in H=\mathrm{SO}(Q)(\mathbb{R})^o$ and so this closure must contain $Q(v)$ for any $v\in\mathbb{R}^d$, because any $v\in\mathbb{R}^d$ can be extended to some unimodular lattice Γ, and then there is a sequence $(h_k)_{k\geqslant1}$ in H with $h_k\mathbb{Z}^d\to\Gamma$ as $k\to\infty$. This shows that there is an integral sequence $(n_k)_{k\geqslant1}$ in $\mathbb{Z}^d$ with $h_kn_k\to v$ as $k\to\infty$ and hence with $Q(n_k)\to Q(v)$. In particular, this argument shows that

$$\overline{Q(\{n=(n_1,n_2,n_3)\in\mathbb{Z}^3\mid\gcd(n_1,n_2,n_3)=1\})}=\mathbb{R}\tag{14.1}$$

(see Exercise 14.0.2).

Another reduction will be useful for us.

Lemma 14.3. *If Q is a non-degenerate, indefinite, irrational quadratic form in $d>3$ variables, then there exists a rational 3-dimensional subspace $V\subseteq\mathbb{R}^d$ such that $Q|_V$, once expressed in a rational basis of V, is non-degenerate, indefinite, and irrational.*

Outline Proof. Since Q is non-degenerate and indefinite, we can find 3 vectors v_1,v_2,v_3 with $|Q(v_i)|=1$ for $i=1,2,3$ but with $Q(v_1),Q(v_2),Q(v_3)$ not all of the same sign. Moreover, we can also require them to be orthogonal in the sense that

$$\langle v_i,v_j\rangle=\tfrac{1}{2}\left(Q(v_i+v_j)-Q(v_i)-Q(v_j)\right)=0$$

if $i\neq j$. This implies that $Q|_{\langle v_1,v_2,v_3\rangle}$ is non-degenerate and indefinite. However, as both non-degenerate and indefinite are open conditions (that is, expressible using strict inequalities on values of continuous functions) the same must hold for any 3-dimensional subspace with a basis $\{w_1,w_2,w_3\}$ that can be chosen close to $\{v_1,v_2,v_3\}$. We indicate now how to find such a basis $w_1,w_2,w_3\in\mathbb{Q}^3$ so as to ensure that the restricted quadratic form is irrational when expressed in the basis w_1,w_2,w_3.

Choose initially $w_1,w_2,w_3'\in\mathbb{Q}^3$ close to v_1,v_2,v_3, and let $a\in\mathbb{Q}(w_1)$. If $\mathbb{Q}(w_3')\notin\mathbb{Q}a$ let $w_3=w_3'$ and we are done. Otherwise, notice that by the assumed irrationality there exists some $x\in\mathbb{Q}^3$ with $\mathbb{Q}(x)\notin\mathbb{Q}a$. Now define $w_3=w_3'+\alpha x$ with $\alpha\in\mathbb{Q}$ small enough to ensure that w_3 is still close to v_3 and chosen to ensure that

$$\mathbb{Q}(w_3) = \mathbb{Q}(w_3') + \alpha^2\mathbb{Q}(x) + 2\alpha\,\langle w_3', x\rangle \notin \mathbb{Q}a.$$

To see that this is always possible, assume that

$$\alpha^2 Q(x) + 2\alpha\,\langle w_3', x\rangle \in \mathbb{Q}a$$

for all small choices of $\alpha \in \mathbb{Q}$. However this assumption implies that $Q(x) \in \mathbb{Q}a$ (which contradicts the choice of x) because

$$\det\begin{pmatrix}\alpha_1^2 & 2\alpha_1\\ \alpha_2^2 & 2\alpha_2\end{pmatrix} = 2\alpha_1\alpha_2(\alpha_1-\alpha_2) \neq 0$$

for distinct non-zero α_1, α_2. □

Thus we may (and will) assume from now on that Q is a non-degenerate, indefinite, irrational quadratic form in $d=3$ variables. By Sylvester's inertia theorem [**?**] (see Lang [**?**, XV, Sec. 4] for a modern treatment) there is some $g \in \mathrm{SL}_3(\mathbb{R})$ for which

$$Q = \lambda Q_0 \circ g$$

with $\lambda \in \mathbb{R}\smallsetminus\{0\}$ (a negative choice of λ may be used to switch from signature $(2,1)$ to signature $(1,2)$ if needed) and with

$$Q_0(x) = Q_0\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = 2x_1x_3 - x_2^2.$$

Now recall that $\mathrm{SO}(Q_0)(\mathbb{R})$ is locally isomorphic† to $\mathrm{SL}_2(\mathbb{R})$. For this, notice that the adjoint action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathfrak{sl}_2(\mathbb{R})$ preserves the quadratic form det on $\mathfrak{sl}_2(\mathbb{R})$. Now choose the basis of $\mathfrak{sl}_2(\mathbb{R})$ so that the coordinates x_1, x_2, x_3 correspond to the matrix

$$\begin{pmatrix}x_2 & -2x_1\\ x_3 & -x_2\end{pmatrix}$$

with determinant $2x_1x_3 - x_2^2$. Using this basis we calculate that

$$\begin{aligned}\mathrm{Ad}_{\begin{pmatrix}1 & s\\ & 1\end{pmatrix}}\begin{pmatrix}x_2 & -2x_1\\ x_3 & -x_2\end{pmatrix} &= \begin{pmatrix}1 & s\\ & 1\end{pmatrix}\begin{pmatrix}x_2 & -2x_1\\ x_3 & -x_2\end{pmatrix}\begin{pmatrix}1 & -s\\ & 1\end{pmatrix}\\ &= \begin{pmatrix}x_2+sx_3 & -2(x_1+x_2s+\frac{s^2}{2}x_3)\\ x_3 & -(x_2+sx_3)\end{pmatrix}\end{aligned}$$

corresponds to the 3×3 matrix

† Notice that there are two different finite-index phenomena at work here: $\mathrm{SO}(Q_0)(\mathbb{R})^o$ has index two in $\mathrm{SO}(Q_0)(\mathbb{R})$ and $\mathrm{SO}(Q_0)(\mathbb{R})^o \cong \mathrm{PSL}_2(\mathbb{R})$, as will be shown later.

$$\begin{pmatrix} 1 & s & \frac{s^2}{2} \\ & 1 & s \\ & & 1 \end{pmatrix}$$

preserving $Q_0(x) = 2x_1x_3 - x_2^2$ on $\mathbb{R}^3$. Similarly we calculate that

$$\mathrm{Ad}_{\begin{pmatrix} 1 & \\ \frac{s}{2} & 1 \end{pmatrix}} \begin{pmatrix} x_2 & -2x_1 \\ x_3 & -x_2 \end{pmatrix} = \begin{pmatrix} x_2 + x_1 s & -2x_1 \\ x_3 + x_2 s + x_1 \frac{s^2}{2} & -(x_2 + x_1 s) \end{pmatrix},$$

which corresponds to

$$\begin{pmatrix} 1 & & \\ s & 1 & \\ \frac{s^2}{2} & s & 1 \end{pmatrix}$$

preserving Q_0 once again. Moreover,

$$\begin{pmatrix} \mathrm{e}^{2t} & & \\ & 1 & \\ & & \mathrm{e}^{-2t} \end{pmatrix}$$

also preserves Q_0, and this matrix corresponds in the same way to

$$\begin{pmatrix} \mathrm{e}^{t} & \\ & \mathrm{e}^{-t} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

Let us now prove the earlier claim.

Lemma 14.4. $\mathrm{SO}(2,1)(\mathbb{R})^o$ *is isomorphic to* $\mathrm{PSL}_2(\mathbb{R})$ *and is a maximal connected subgroup of* $\mathrm{SL}_3(\mathbb{R})$.

Proof. Sending $g \in \mathrm{SL}_2(\mathbb{R})$ to the matrix representation $\phi(g)$ of the map Ad_g (described above) defines a map from $\mathrm{SL}_2(\mathbb{R})$ into $\mathrm{SO}(2,1)(\mathbb{R})^o$. For $g = -e$, we have $\mathrm{Ad}_g = I$, so we a homomorphism

$$\phi : \mathrm{PSL}_2(\mathbb{R}) \longrightarrow \mathrm{SO}(2,1)(\mathbb{R})^o.$$

It is easy to check that the map ϕ is proper. Since $\mathrm{PSL}_2(\mathbb{R})$ has no non-trival normal subgroup (see Exercise **??**), and the homomorphism is non-trivial, we deduce that the homomorphism is injective. To show that the image is indeed $\mathrm{SO}(2,1)(\mathbb{R})^o$, it is enough to check that the image is a maximal connected subgroup.

To achieve this, we analyze the finite-dimensional representation

$$\mathrm{Ad}_{\phi(g)} : \mathfrak{sl}_3(\mathbb{R}) \to \mathfrak{sl}_3(\mathbb{R})$$

of $\mathrm{SL}_2(\mathbb{R})$. The algebra $\mathfrak{sl}_3(\mathbb{R})$ contains the Lie algebra $\mathfrak{h}$ of the image H of ϕ, which is generated by the matrices

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

These have weight 2, -2 and 0 respectively for the Cartan subgroup of $\mathrm{SL}_2(\mathbb{R})$. The Lie algebra $\mathfrak{sl}_3(\mathbb{R})$ also contains

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which has weight 4. However, by the finite-dimensional representation theory of $\mathrm{SL}_2(\mathbb{R})$, this forces there to be a decomposition

$$\mathfrak{sl}_3(\mathbb{R}) = \mathfrak{h} \oplus V_5,$$

where V_5 is an *irreducible* 5-dimensional representation of $\mathrm{SL}_2(\mathbb{R})$. It follows that every closed connected subgroup containing H must either be equal to H or to $\mathrm{SL}_3(\mathbb{R})$. In particular, we deduce that $H = \mathrm{SO}(2,1)(\mathbb{R})^o$. □

Since we have already classified all measures that are invariant under the action of subgroups locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$, one might now hope that we are not far from a proof of Oppenheim's conjecture (Conjecture 14.1). This hope is misplaced. For example, it is not clear how to find even one H-invariant measure on the closure of an H-orbit

$$\overline{H{\scriptstyle\bullet}x} \leqslant \mathrm{SL}_3(\mathbb{Z})\backslash\mathrm{SL}_3(\mathbb{R})$$

since H is not amenable (see [**?**, Sec. 8.4] for background on amenability).

Is there a good source for statements like this — that this or that group is not amenable. Zimmer? Day?

Instead...

one should work with a unipotent subgroup of H using rational maps and entropy arguments along the same lines as used in Section 6.6.

(Need to return to this)

Query: to do this?

Alternatively, one can also prove Oppenheim's conjecture by arguments in topological dynamics, and this was done by Margulis [**?**].

Theorem 14.5 (Margulis). *Oppenheim's conjecture holds in $d \geqslant 3$ variables.*

Dani and Margulis [**?**] then improved the argument to obtain the special case of Raghunathan's conjecture.

Theorem 14.6 (Dani and Margulis). *With the notation from p. 410, for any $x \in X$, either $\overline{Hx} = Hx$ or $\overline{H{\scriptstyle\bullet}x} = X$, and in particular we have* (14.1).

There is one more connection to explain, namely why does the non-dense possibility $\overline{H{\scriptstyle\bullet}x} = H{\scriptstyle\bullet}x$ really correspond to Q being a rational form? This will be discussed in Section **??**.

Exercises for Section 14

Exercise 14.0.1. Show that every orbit of the map

$$\begin{aligned} R_\alpha : \mathbb{T} &\longrightarrow \mathbb{T} \\ x &\longmapsto x+\alpha \end{aligned}$$

with $\alpha \notin \mathbb{Q}$ is dense. Deduce that if L is an irrational linear form, then $L(\mathbb{Z}^2)$ is dense in $\mathbb{R}$.

Exercise 14.0.2. Give the details of the argument sketched on p. 410, and deduce (14.1).

14.1 Temporary: $\mathbf{SL_2}(\mathbb{Z})$ is a maximal lattice

Proposition 14.7. $\mathrm{SL}_2(\mathbb{Z})$ *is a maximal lattice in* $\mathrm{SL}_2(\mathbb{R})$*. Moreover, any subgroup* $H < \mathrm{SL}_2(\mathbb{R})$ *that contains* $\mathrm{SL}_2(\mathbb{Z})$ *properly is dense.*

Proof of Proposition 14.7. Let

$$g_0 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}),$$

and write

$$\Gamma = \langle \mathrm{SL}_2(\mathbb{Z}), g \rangle$$

for the subgroup generated by $g_0 \in \mathrm{SL}_2(\mathbb{R})$ and $\mathrm{SL}_2(\mathbb{Z})$. In the following we will multiply $g = g_0$ several times on the left or on the right by elements of $\mathrm{SL}_2(\mathbb{Z})$. Modifying g in this way does not change Γ, and we will abuse notation slightly by using the letter $g = (g_{ij})$ for the matrix obtained by these successive modifications.

Below we will frequently apply 'division with remainder' to g_{21} and g_{22}. Indeed, if $n \in \mathbb{Z}$ then the calculation

$$g \begin{pmatrix} 1 & \\ n & 1 \end{pmatrix} = \begin{pmatrix} g_{11} + n g_{12} & g_{12} \\ g_{21} + n g_{22} & g_{22} \end{pmatrix}$$

shows that we may apply division with remainder to replace g with

$$g' = \begin{pmatrix} g'_{11} & g_{12} \\ g'_{21} & g_{22} \end{pmatrix} = g \begin{pmatrix} 1 & \\ n & 1 \end{pmatrix}$$

with the property that $|g'_{21}| \leqslant \frac{1}{2}|g_{22}|$. If in this new matrix we have $g_{21} \neq 0$ then we may use a similar calculation for

$$g \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix}$$

to similarly reduce g_{22} in size. There are two possibilities to consider.
CASE 1: If the initial entries c and d in g are *incommensurable*† then the division with remainder argument above may be applied indefinitely, with the size of the g_{21} and g_{22} entries being rendered smaller than half of the g_{22} and g_{21} entries at each stage respectively. In particular, after finitely many steps we may assume that $\|(g_{21}, g_{22})\| < 1$.
CASE 2: If c and d are commensurable, then the procedure stops when one of the entries vanishes. We may assume without loss of generality that g_{21} vanishes. If $|g_{22}| < 1$ we continue to the next step of the proof. Suppose therefore that $g_{21} = 0$ and $|g_{22}| \geqslant 1$, which forces $|g_{11}| = |g_{22}^{-1} \leqslant 1$. Now notice that

$$\begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix} = \begin{pmatrix} 0 & -g_{22} \\ g_{11} & g_{12} \end{pmatrix}.$$

If $g_{22} = g_{11} = \pm 1$ and $g_{12} \in \mathbb{Z}$ then $g \in \mathrm{SL}_2(\mathbb{Z})$ and hence also $g_0 \in \mathrm{SL}_2(\mathbb{Z})$ (since all the steps taken involve multiplication on left or right by elements of $\mathrm{SL}_2(\mathbb{Z})$ and so are invertible). If $|g_{11}| < 1$ or $|g_{11}| = 1$ but $g_{12} \notin \mathbb{Z}$, then the division by remainder argument again leads (after renaming the resulting matrix) to the situation with $\|(g_{21}, g_{22})\| < 1$.

So now suppose that

$$g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \smallsetminus \mathrm{SL}_2(\mathbb{Z})$$

with $\|(g_{21}, g_{22}\| < 1$. Applying the Möbius transformation corresponding to g to i gives

$$\Im(g\mathrm{i}) = \frac{1}{|g_{21}\mathrm{i} + g_{22}|^2} > 1.$$

Let $z = \mathrm{i}y$ with $y > 1$ chosen close to 1. Then $\Im(gz)$ is close to $\Im(g\mathrm{i})$ and we may assume that

$$\Im(gz) > \Im(g\mathrm{i}). \tag{14.2}$$

Now recall that for any discrete subgroup $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ and point z in the upper-half plane which is not fixed by any nontrivial element $\gamma \in \Gamma$ we can define the Dirichlet domain (or Dirichlet region)

$$D_z = \{z' \in \mathbb{H} \mid \mathsf{d}(z', z) = \min\{\mathsf{d}(z', \gamma(z)) \mid \gamma \in \Gamma\}\}$$

which (apart from the boundary) is then a fundamental domain for Γ (see [**?**, Sec. 11.1]).

Assume now that $\Gamma = \langle g_0, \mathrm{SL}_2(\mathbb{Z})\rangle = \langle g, \mathrm{SL}_2(\mathbb{Z})$ is still discrete even though $g \notin \mathrm{SL}_2(\mathbb{Z})$. Then $\Gamma_g \supseteq \mathrm{SL}_2(\mathbb{Z})$ is a lattice (since its covolume is bounded above by the co-volume of $\mathrm{SL}_2(\mathbb{Z})$, which we know to be finite), and we may apply the Dirichlet domain construction to the point z as above. As

† That is, linearly independent over $\mathbb{Q}$.

illustrated in Figure 14.1, the inequality (14.2) shows that D_z has compact closure.

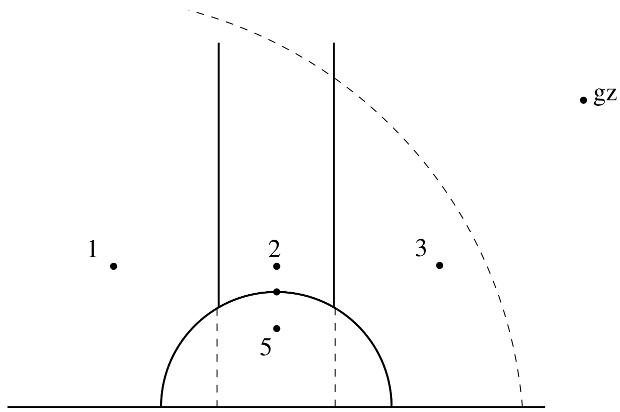

Fig. 14.1 The lines drawn are the boundaries of the usual fundamental domain, and are also the geodesics consisting of the set of points equidistant from z and $\gamma(z)$ for various $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. The dotted line is the set of points that are equidistant from z and $g(z)$. The Dirichlet domain D_z is contained in the shaded region.

Since Γ is assumed to be discrete, it is therefore a lattice, and so the index $[\Gamma_g : \mathrm{SL}_2(\mathbb{Z})]$ must be finite (since this quantity is also the ratio of the volume of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ and of $\Gamma_g\backslash\mathbb{H}$. However, if

$$\Gamma = \bigcup_{i=1}^{n} \mathrm{SL}_2(\mathbb{Z}) h_i$$

then

$$\bigcup_{i=1}^{n} h_i(D_z)$$

would be a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ in $\mathbb{H}$ with compact closure. As $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ is not compact, this is a contradiction.

It follows that $L = \overline{\Gamma_g}$ is closed and not discrete, and so has a non-trivial Lie algebra $\mathfrak{l}$. Since $\mathrm{SL}_2(\mathbb{Z}) \subseteq \Gamma_g$, the Lie algebra $\mathfrak{l}$ is preserved by Ad_γ for γ in $\mathrm{SL}_2(\mathbb{Z})$. By Borel density (Theorem 3.30) this implies that $\mathfrak{l} \lhd \mathfrak{sl}_2(\mathbb{R})$ and hence $\mathfrak{l} = \mathfrak{sl}_2(\mathbb{R})$. This implies that $L = \mathrm{SL}_2(\mathbb{R})$ since $\mathrm{SL}_2(\mathbb{R})$ is connected. $\square$

maybe we include some applications found by Jens: free path length in Boltzmann-grad limit and Frobenius numbers?

Part IV
Odds and ends without permanent residence yet

Lemma 14.8 (Tangent space for non-Archimedean fields). *Let $\Bbbk = \mathbb{F}_q((s))$ be a completion of a global field $\mathbb{K}|\mathbb{F}_p(t)$ with positive characteristic at a place σ, and let $V \subseteq \Bbbk^n$ be the set of $\Bbbk$-points of a variety defined over $\mathbb{K}$, and suppose that $e \in V$ is a smooth point of V. Then every $v_0 \in V$ sufficiently close to e is of the form*

$$v_0 = e + w + \mathrm{O}\left(|v_0 - e|_\sigma\right) \tag{14.3}$$

for some w in the $\Bbbk$-tangent space of V at e. Similarly, for any sufficiently small w_0 in the tangent space there exists some $v \in V$ with

$$v = e + w_0 + \mathrm{O}\left(|w_0|_\sigma\right). \tag{14.4}$$

Proof. We may assume (by translation) that $e = 0$ and (by applying a linear map) that the tangent space is given by $\Bbbk^d \times \{0\}^{n-d}$. By definition of the tangent space[(43)] of a variety, this implies that the ideal $\mathscr{J}$ defining the variety contains polynomials of the form

$$T_j + f_j(T_1, \dots, T_n) \tag{14.5}$$

for $j = d+1, \dots, n$ where f_j is a polynomial with no constant or linear terms. This already implies that any vector

$$v_0 = \begin{pmatrix} x_1 \\ \vdots \\ x_d \\ x_{d+1} \\ \vdots \\ x_n \end{pmatrix}$$

close to 0 must satisfy $|x_j|_\sigma = \mathrm{O}\left(|v_0|_\sigma^2\right)$ for $j = d+1, \dots, n$, which is precisely (14.3).

Localizing the ring $R = \Bbbk[x_1, \dots, x_n]$ at the ideal (0), we obtain the local ring

$$R_{(0)} = \left\{ \tfrac{f}{g} \mid f, g \in R, g(0) \neq 0 \right\}.$$

Then the ideal $\mathscr{J} R_{(0)} \subseteq R_{(0)}$ is generated by the polynomials in (14.5). Therefore, for any vector

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_d \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

belonging to the tangent space of V at 0 and sufficiently close to 0, we can use the relations in (14.5) to find a vector

$$v = \begin{pmatrix} x_1 \\ \vdots \\ x_d \\ x_{d+1} \\ \vdots \\ x_n \end{pmatrix} \in V$$

as needed for (14.4). $\square$

Chapter 15
Some Homeless Fragments

15.1 Každan's Property (T)

Property (T), introduced(44) by Každan in 1967, is a versatile notion from the representation theory of locally compact groups that has found applications in many branches of mathematics. It has a formulation(45) as a fixed point property, but we describe only a characterization in terms of unitary representations. It turns out — though we do not show this — that many Lie groups with rank at least two (and lattice inside them) have Každan's property (T). Thus, for example, $\mathrm{SL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{Z})$ have Každan's property (T) for $n \geqslant 3$.

In ergodic theory, property (T) is related to the notion of *strong ergodicity*(46), where it seems complementary to amenability. Despite this, amenability and property (T) are not complementary; there are important groups that fall between the two definitions, an example being $\mathrm{SL}_2(\mathbb{Z})$; on the other hand compact groups have both properties.

Definition 15.1. Let G be a locally compact group with a countable basis for its topology, and let $\pi : G \to U(\mathscr{H})$ be a continuous homomorphism from G to the group of unitary operators on some Hilbert space $\mathscr{H}$ (such a homomorphism is called a *unitary representation*). Given $\varepsilon > 0$ and $K \subseteq G$ a compact set, a vector $v \in \mathscr{H}$ with $\|v\| = 1$ is called (ε, K)*-invariant* if

$$\|\pi(g)v - v\| < \varepsilon \text{ for all } g \in K.$$

The unitary representation π *almost has invariant vectors* if there is an (ε, K)-invariant vector for each $\varepsilon > 0$ and K compact.

An extreme way in which G could almost have invariant vectors is to have an invariant vector: if there is some non-trivial vector v with the property that $\pi(g)v = v$ for all $g \in G$, then clearly v is (ε, K)-invariant for all ε and K.

Definition 15.2. Let G be a locally compact group with a countable basis for its topology. Then G has *property (T)* or the *Každan property* if any unitary representation of G which almost has invariant vectors has non-trivial invariant vectors.

Example 15.3. The regular representation $\pi : \mathbb{Z} \to U\left(L^2(\mathbb{Z})\right)$ is defined by $(\pi(n)f)_k = f_{k+n}$ for $f = (f_k) \in L^2(\mathbb{Z})$. Given any $\varepsilon > 0$ and finite set $F \subseteq \mathbb{Z}$, let E be a finite set with $|E| > \frac{2}{\varepsilon}|F|$. Define f by

$$f_k = \begin{cases} 1/|E| & \text{if } k \in E, \\ 0 & \text{if not.} \end{cases}$$

Then for any $n \in F$, $\|\pi(n)f - f\| \leqslant 2|F|/|E| < \varepsilon$, showing that π almost has invariant vectors. However, it is clear that π has no invariant vectors: if f is invariant, then the sequence f_k is constant. Thus $\mathbb{Z}$ does not have Každan's property (T).

Každan [**?**] explored some algebraic consequences of property (T), the simplest of which is the following.

Proposition 15.4. *If G is a discrete group with Property (T), then G is finitely generated.*

Proof. Enumerate the elements of G as $\{g_1, g_2, \dots\}$, and write

$$G_n = \langle\!\langle g_1, g_2, \dots, g_n \rangle\!\rangle$$

for the subgroup generated by the first n elements. Let

$$\pi_n : G \to U\left(L^2(G/G_n)\right)$$

be the translation representation

$$\left(\pi_n(h)f\right)(g + G_n) = f(h + g + G_n);$$

this is the representation induced by the identity representation on the subgroup G_n. Finally, let $\pi = \bigoplus \pi_n$ be the direct sum of the representations $\{\pi_n\}$. For each $n \geqslant 1$, the function $\mathbb{1}_{1_G + G_n}$ is an invariant vector for π_n, so π almost has invariant vectors — and therefore has non-trivial invariant vectors. Let $f \in \bigoplus L^2(G/G_n)$ be such a vector. Each of the projections f_n of f onto $L^2(G/G_n)$ will also be invariant, and for some n the projection f_n must be non-trivial. It follows that some π_n has non-trivial invariant vectors, which can only happen if G/G_n is a finite group. Thus G is a finite extension of a group generated by n elements, so is finitely generated. $\square$

Problems for Section 15.1

Exercise 15.1.1. Extend Example 15.3 to show that $\mathbb{R}^d$ and $\mathbb{Z}^d$ do not have Každan's property (T) for any $d \geqslant 1$.

15.2 Ideas and Questions; Things to Include

Completely positive entropy implies mixing of all orders for a single transformation? For a $\mathbb{Z}^d$-action (Kaminski/Conze)? For amenable group actions (Rudolph–Weiss; very difficult result)?

15.2.1 Some comments from Anish

Since Minkowski's theorems are followed by adelic Mahler's compactness, it may be helpful to mention in passing that the adelic analogue of Minkowski's theorems have been proved by Bombieri-Vaaler (http://www.ams.org/mathscinet-getitem?mr=707346) and have found use in number theoretic contexts. Actually, they were proved earlier by Mcfeat (http://www.ams.org/mathscinet-getitem?mr=318104) unbeknownst to B-V, but Mcfeat's paper is hard to find and his result has weaker constants.

I look forward to the spectral gap section in particular. Some audience requests

1) How about a section/chapter on SL(2, R)? I mean a quick discussion of principal/complementary/discrete followed by the explicit calculation of decay, for example Howe-Tan style.

2) How about a discussion of (a baby case of) Einsiedler-Margulis-Venkatesh? It is very closely related to spectral gap and property tau after all?

Notes to Chapter 15

(43)(Page 419) For background in algebraic geometry used here, we refer to Reid [**?**] for an introduction and to Hartshorne [**?**] for a sophisticated account. Some of the language needed is reviewed in Chapter 3.

(44)(Page 421) This property was introduced by Každan [**?**], who showed that many semisimple Lie groups have the property. The theory was developed significantly by Delaroche and Kirillov [**?**]; see also Zimmer [**?**, Chap. 7].

(45)(Page 421) A locally compact group G with a countable basis for its topology has Každan's property (T) if every continuous action of G by isometries of a Hilbert space has a fixed point.

(46)(Page 421) Let T be an action of a locally compact second countable group G on the Lebesgue space $(X, \mathscr{B}, \mu)$. A sequence of measurable sets (A_n) is called *asumptotically invariant* if for every $g \in G$, $\lim_{n\to\infty} \mu\left(A_n \triangle T_g(A_n)\right) = 0$, and is called *trivial* if $\lim_{n\to\infty} \mu(A_n)\left(1 - \mu(A_n)\right) = 0$. The action is called *strongly ergodic* if the only asymptotically invariant sequences are the trivial ones. This notion gives a characterization of Každan's property (T) as follows. K. Schmidt [**?**] showed that if every ergodic action of G is strongly ergodic, then G has property (T); Connes and Weiss [**?**] showed that the converse holds. The proof of the converse requires methods from probability; a convenient place for both results is the book by Glasner [**?**, Chap. 13]. A striking result of Glasner and Weiss [**?**] gives another characterization of property (T). For a countable group G their results are as follows. If G has property (T), then for any action of G by homeomorphisms of a compact metric space (X, d), if the space $\mathscr{M}^G(X)$ of G-invariant probability measures on X is non-empty, then the extreme points (the ergodic measures) are a closed subset of $\mathscr{M}^G(X)$; such a simplex is called a *Bauer simplex*. In the reverse direction, they show that G does not have property (T) if and only if $\mathscr{M}^G(X)$ is the *Poulsen simplex*, characterized by the property that the extreme points (the ergodic measures) are dense.

Appendix A: Topological Groups

Definition A.1. A collection $\mathscr{U}$ of neighborhoods of a point x in a topological space is a *neighborhood basis* if for any open neighborhood $V \ni x$ there is some neighborhood $U \in \mathscr{U}$ with $U \subseteq V$.

A topological space is *first countable* if every point has a countable neighborhood basis.

A metric space (X, d) is automatically first countable, since the open neighborhoods $B_{1/n}(x)$ for $n \in \mathbb{N}$ form a countable neighborhood basis at x(47).

Recall that a topological group is a group G together with a Hausdorff topology $\mathscr{T}$ with respect to which the maps $g \mapsto g^{-1}$ and $(g, h) \mapsto gh$ are continuous. This means that

- if U is a neighborhood of a product $gh \in G$, then there are neighborhoods $U_1 \ni g$ and $U_2 \ni h$ with $U_1U_2 \subseteq U$;
- if U is an open neighborhood of $g \in G$, then there is an open neighborhood V of g^{-1} with $V^{-1} \subseteq U$.

Lemma A.2 (Birkhoff–Kakutani [?], [?]). *The following properties of a topological group G are equivalent.*(48)

(1) *G has a left-invariant metric, that is a metric d giving the topology which additionally has $\mathsf{d}(gh_1, gh_2) = \mathsf{d}(h_1, h_2)$ for all $g, h_1, h_2 \in G$.*
(2) *Each $g \in G$ has a countable basis of open neighborhoods.*
(3) *The identity $1 \in G$ has a countable basis of open neighborhoods.*

Proof. It is clear that (1) $\Rightarrow$ (2) $\Leftrightarrow$ (3) since the rotation $g \mapsto gh$ is a homeomorphism of G for any $h \in G$. So we will assume (2). Let $\mathscr{U} = \{V_1, V_2, \dots\}$ be a countable neighborhood basis at the identity I consisting of open sets. Without loss of generality we may assume that

$$V_1 \supseteq V_2 \supseteq \cdots \supseteq \{I\}, \tag{A.1}$$

and since G is Hausdorff we have

$$\bigcap_{n \geqslant 1} V_n = \{I\}.$$

We wish to construct sets that mimic the behavior of nested metric open sets. To that end we use the continuity properties of the two group operations to construct from the sequence of sets (V_n) another nested sequence of open neighborhoods of the identity

$$U_1 = G \supseteq U_{1/2} \supseteq U_{1/2^2} \supseteq \cdots \supseteq \{I\} \tag{A.2}$$

with the property that $U_{1/2^n}^{-1} = U_{1/2^n}$ (each set is symmetric), $U_{1/2^n} \subseteq V_n$, and $U_{1/2^{n+1}} U_{1/2^{n+1}} \subseteq U_{1/2^n}$ for each $n \geqslant 1$. It follows that

$$\bigcap_{n \geqslant 0} U_{1/2^n} \subseteq \bigcap_{n \geqslant 1} V_n = \{I\}. \tag{A.3}$$

For any rational of the form $\frac{a}{2^n}$ with $a \in \{1, \dots, 2^n - 1\}$ and $n \in \mathbb{N}$ we define

$$U_{a/2^n} = U_{1/2^{n_1}} \cdots U_{1/2^{n_r}}$$

where

$$\frac{a}{2^n} = 2^{-n_1} + \cdots + 2^{-n_r}$$

is the binary expansion of $\frac{a}{2^n}$ arranged in the natural order with

$$1 \leqslant n_1 < \cdots < n_r.$$

By construction[†]

$$U_{a/2^n} U_{1/2^n} \subseteq U_{(a+1)/2^n} \tag{A.4}$$

for $n \geqslant 1$ and $1 \leqslant a \leqslant 2^n - 1$. Hence the sets $U_{a/2^n}$ are nested in the sense that[‡]

$$0 < a < b \leqslant 2^n \Rightarrow U_{a/2^n} \subseteq U_{b/2^n}.$$

Using this neighborhood basis we can define a function f on G by

$$f(x) = \inf\{\tfrac{a}{2^n} \mid x \in U_{a/2^n}\}.$$

We claim that f has the following properties[§]:

[†] If there is no carry in the binary addition of $a/2^n$ and $1/2^n$ this is just the definition, if there is a carry one uses the defining properties of $U_{1/2^n}$.

[‡] Suppose $a = 2^{-n_1} + \cdots + 2^{-n_r}$. If $b = a + \frac{c}{2^{n_r+1}}$ for an integer $c \geqslant 1$ this follows simply from the definition of U_b. If $b = a + \frac{1}{2^{n_r}}$ this follows from (A.4). If $b > a + \frac{1}{2^{n_r}}$ the conclusion follows from the latter case and induction on n_r.

[§] Notice that the existence of such a function would follow easily from the conclusion we seek. If G has a left-invariant metric d defining its topology, then the function f defined by $f(g) = \mathsf{d}(e, g)$ has the three properties claimed.

(a) $f(g) > 0$ for $g \in G \smallsetminus \{I\}$ and $f(I) = 0$;
(b) the collection $\bigl\{\{g \in G \mid f(g) < \frac{1}{n}\} \mid n \in \mathbb{N}\bigr\}$ form a neighborhood base at the identity $I \in G$; and
(c) for any $\varepsilon > 0$ there is an open neighborhood $U \ni e$ such that

$$|f(hg) - f(h)| \leqslant \varepsilon$$

for all $g \in U$ and $h \in G$.

Property (b) holds since the collection of sets $\{U_{a/2^n}\}$ form a neighborhood basis at the identity, and (a) follows from (A.3). The uniform continuity property (c) is a consequence of (A.4). Indeed if $\varepsilon > 0$ is arbitrary we may choose some n with $\frac{1}{2^n} < \varepsilon$ and set $U = U_{1/2^{n+1}}$. We let $h \in G$ and $g \in U$ be arbitrary. If $f(h) \in [\frac{a}{2^{n+1}}, \frac{a+1}{2^{n+1}})$ then $h \in U_{(a+1)/2^{n+1}}$, and so $hg \in U_{(a+2)/2^{n+1}}$ which implies $f(hg) < \frac{a+2}{2^{n+1}} \leqslant f(h) + \frac{2}{2^{n+1}} < f(h) + \varepsilon$. Using $U^{-1} = U$ the inequality $f(h) < f(hg) + \varepsilon$ follows from the former.

We now define a metric-like function $\mathsf{d}^f : G \times G \to [0, \infty)$ by

$$\mathsf{d}^f(g_1, g_2) = \sup_{h \in G} |f(hg_1) - f(hg_2)|. \tag{A.5}$$

Clearly

$$\mathsf{d}^f(g_1, g_2) = \mathsf{d}^f(g_2, g_1)$$

and

$$\mathsf{d}^f(hg_1, hg_2) = \mathsf{d}^f(g_1, g_2)$$

for all $g_1, g_2, h \in G$, $\mathsf{d}^f(g, g) = 0$, and d^f obeys the triangle inequality. That is, d^f is a left-invariant *pseudometric* on G. Now assume that $\mathsf{d}^f(g_1, g_2) = 0$. Then $f(g_1^{-1}g_2) \leqslant \mathsf{d}^f(I, g_1^{-1}g_2) = 0$ implies that $g_1 = g_2$ by (a). Hence we see that d^f is a metric on G.

It remains to show that the metric topology induced from d^f is the original group topology. As both topologies make G into a topological group it is sufficient to study the neighborhoods of I with respect to both topologies. Any $h \in G$ with $\mathsf{d}^f(h, I) < \frac{1}{2^n}$ satisfies $h \in U_{1/2^n}$, which shows that a neighborhood in the original topology is also a neighborhood in the the metric topology. Now let $\varepsilon > 0$ and let U be a neighborhood as in the uniform continuity property (c). Then $\mathsf{d}^f(h, I) \leqslant \varepsilon$ for all $h \in U$, which shows that a metric neighborhood is also a neighborhood in the original topology. □

Notice in particular that this means the groups encountered in this volume, like $\mathrm{GL}_d(\mathbb{R})$ and $\mathrm{SL}_d(\mathbb{R})$, have left-invariant metrics that give the group topology.

Exercises for Appendix A

Let $d \geqslant 2$. Show that $G = \mathrm{SL}_d(\mathbb{R})$ cannot be equipped with a metric that gives the standard topology and is bi-invariant, that is, satisfies $\mathsf{d}(gh_1, gh_2) = \mathsf{d}(h_1, h_2) = \mathsf{d}(h_1 g, h_2 g)$ for all $g, h_1, h_2 \in G$.

Notes to Appendix A

(47)(Page 425) First countable topological spaces are not automatically metrizable. An example to see this is the *Sorgenfrey line* [**?**], the space $\mathbb{R}$ with the topology formed by using the half-open intervals $[a, b)$ with $a < b$ as basis. It is clear that this is first countable, since the sets $[a, a + \frac{1}{n})$ for $n \in \mathbb{N}$ form a countable neighborhood basis at a. Much less clear is the fact that it is not metrizable, and we refer to Kelley [**?**] for the details.
(48)(Page 425) If d_ℓ is a left-invariant metric then $\mathsf{d}_r(x, y) = \mathsf{d}_\ell(x^{-1}, y^{-1})$ is a right-invariant metric defining the same topology. A bi-invariant metric with

$$\mathsf{d}(xgy, xhy) = \mathsf{d}(g, h)$$

for all $x, y, g, h \in G$ only exists in special cases: we refer to [**?**, Lem. C.2] for the simple case that a compact metrizable group has a bi-invariant metric, and [**?**, Ex. C.3] for an explanation of why $\mathrm{GL}_2(\mathbb{C})$ has no bi-invariant metric. A striking result of Milnor [**?**] is that a connected Lie group admits a bi-invariant metric if and only if it is isomorphic to $K \times \mathbb{R}^n$ for some compact Lie group K. The proof of Lemma A.2 given here is taken from the monograph of Montgomery and Zippin [**?**, Sec. 1.22] and Tao's blog [**?**].

Appendix B: Haar Measure on Quotients of Groups

[†]Let G be a unimodular group with Haar measure m_G, and let $\Gamma < G$ be a discrete subgroup (which is also unimodular, and has the counting measure as its Haar measure). Then, as we have discussed in Section 1.1, the measure m_G induces a G-invariant measure $m_{\Gamma\backslash G}$ on $\Gamma\backslash G$ which we also call Haar measure. Moreover,

$$m_{\Gamma\backslash G}\,(\Gamma\backslash G) = m_G(F)$$

for a Borel fundamental domain $F \subseteq G$ for Γ. We will generalize this construction in this section to allow quotients $H\backslash G$ by certain closed unimodular subgroups $H < G$. We refer to the work of Knapp [?] or Raghunathan [?] for a more general treatment of the existence of Haar measure on quotients. Notice that we may also define the quotient topology on $H\backslash G$ by constructing a quotient metric $\mathsf{d}_{H\backslash G}$ along the lines of (1.1) on p. 8 (see also Exercise 1.1.2).

B.1 Compact Subgroups

As a first case, assume that G is σ-compact, locally compact and unimodular, equipped with a left-invariant metric, and assume that $H < G$ is a compact subgroup with Haar measure m_H. Then m_G and m_H together induce a G-invariant measure $m_{H\backslash G}$, which we will again call Haar measure on $H\backslash G$. In fact, if $B \subseteq H\backslash G$ is a Borel subset then we can define[‡]

[†] The material in this chapter will only be used in Chapter 12. It extends the results of Section 1.1.

[‡] The normalization is compatible with the more general one presented in Section B.1.1. Also notice that the construction from Section 1.1 will not work here if H is not discrete. A fundamental domain for H, called a cross-section in this context, has Haar measure zero, as seen for example in the case

$$m_{H\backslash G}(B) = \frac{m_G(\pi^{-1}(B))}{m_H(H)} \tag{B.6}$$

where $\pi : G \to H\backslash G$ is the canonical right G-equivariant† projection map

$$\pi : g \longmapsto Hg.$$

B.1.1 Subgroups with Uniform Lattices

Once again let G be a σ-compact and locally compact unimodular group equipped with a left-invariant metric. The case we will be interested in later is the case where the closed unimodular subgroup $H < G$ contains a uniform lattice $\Gamma_H < H$. We can use this lattice as a tool for generalizing (B.6) as follows. For a Borel subset $B \subseteq H\backslash G$, define its measure by

$$m_{H\backslash G}(B) = \frac{m_{\Gamma_H\backslash G}(\pi^{-1}(B))}{m_{\Gamma_H\backslash H}(\Gamma_H\backslash H)} \tag{B.7}$$

where $\pi : \Gamma_H\backslash G \to H\backslash G$ is the canonical right G-equivariant projection map

$$\pi : \Gamma_H g \longmapsto Hg.$$

First notice that (B.7) generalizes (B.6), since one can take $\Gamma_H = \{I\}$ if H is compact. Next notice that the right-hand side of (B.7) seemingly depends on the choice of a lattice $\Gamma_H < H$, but we will see below that this is not the case. As a first indication of this phenomenon, one may quickly check that $m_{H\backslash G}$ defined by (B.7) does not change if Γ_H is replaced by a finite-index subgroup $\Gamma_H' < \Gamma_H$ (by unfolding the definitions on the right-hand side of (B.7) using (1.5)). This would not be the case without the normalization from the denominator.

The induced Haar measure $m_{H\backslash G}$ is right G-invariant since

$$m_{H\backslash G}(Bg) = \frac{m_{\Gamma_H\backslash G}(\pi^{-1}(Bg))}{m_{\Gamma_H\backslash H}(\Gamma_H\backslash H)} = \frac{m_{\Gamma_H\backslash G}(\pi^{-1}(B)g)}{m_{\Gamma_H\backslash H}(\Gamma_H\backslash H)} = m_{H\backslash G}(B)$$

for a Borel subset $B \subseteq H\backslash G$ and any $g \in G$.

We summarize and extend the ideas above in the following lemmas.

Lemma B.1 (Haar measure on $H\backslash G$). *Suppose that G is a σ-compact, locally compact, unimodular group equipped with a left-invariant metric. Let H*

$$H = \mathrm{SO}(2,\mathbb{R}) < G = \mathrm{SL}_2(\mathbb{R}),$$

and the definition in (B.6) also does not work if $m_H(H) = \infty$.

† That is, intertwining the natural G-actions.

be a closed subgroup of G, and let $\Gamma_H < H$ be a uniform lattice in H. Then $m_{H\backslash G}$ as defined in (B.7) *is a locally finite G-invariant Borel measure on $H\backslash G$, also called a Haar measure.*

Lemma B.2 (Continuous folding and unfolding). *Let $H < G$ be as above. The Haar measures m_G, m_H, and $m_{H\backslash G}$ have the following compatibility relation, which may be viewed as an analog of the Fubini theorem for product measures. If $f \in L^1_{m_G}(G)$, then the function F on $H\backslash G$ defined by the relation*

$$F(Hg) = \int_H f(hg)\,\mathrm{d}m_H(h) \tag{B.8}$$

exists for m_G-almost every $g \in G$ (or equivalently for $m_{H\backslash G}$-almost every Hg), and

$$\int_{H\backslash G} F(Hg)\,\mathrm{d}m_{H\backslash G} = \int_G f\,\mathrm{d}m_G. \tag{B.9}$$

Moreover, (B.9) *holds (up to a multiplicative scalar constant) for any nonzero locally finite G-invariant Borel measure on $H\backslash G$.*

Proof of Lemma B.1 and Lemma B.2. Let $m_{H\backslash G}$ be as in (B.7). Using the measure we define a measure ν on G by

$$\nu(B) = \int_{H\backslash G} m_H\left(\{h \in H \mid hg \in B\}\right)\mathrm{d}m_{H\backslash G}(Hg)$$

for any Borel set $B \subseteq G$ (so that $\nu(B)$ is the left-hand side of (B.9) applied to $f = \mathbb{1}_B$). We will show that ν is locally finite, positive on open sets, and right G-equivariant (which will imply that $\nu = cm_G$ for some $c \in (0,\infty)$ by the uniqueness of Haar measure). To see that μ is locally finite, it is enough to show that π is proper.

Since $\Gamma_H < H$ is assumed to be a uniform lattice, we have

$$\pi^{-1}(Hg) = \{\Gamma_H hg \mid h \in H\} = \Gamma_H K_H g$$

for some compact set $K_H \subseteq H$. Now any compact set in $H\backslash G$ is of the form HK for some compact set $K \subseteq G$. In fact, if $L \subseteq H\backslash G$ is compact, and $U \subseteq G$ is an open neighborhood of $I \in G$ with compact closure, then

$$L \subseteq \bigcup_{Hg \in L} HgU$$

is an open cover, so there is a finite subcover

$$L \subseteq \bigcup_{i=1}^{n} Hg_iU$$

and hence

$$K = \left(\overline{\bigcup_{i=1}^{n} g_i U}\right) \cap \{g \in G \mid Hg \in L\}$$

satisfies the claim. Furthermore, we have

$$\pi^{-1}(HK) = \Gamma_H HK = \Gamma_H K_H K,$$

which means that the pre-image of any compact subset of $H\backslash G$ is compact in $\Gamma_H\backslash G$. This implies that $m_{H\backslash G}$ is finite on compact subsets, and that ν is locally finite.

If $B \subseteq G$ is measurable and $g_1 \in G$, then

$$\begin{aligned}\nu(Bg_1) &= \int_{H\backslash G} m_H\left(\{h \in H \mid hg \in Bg_1\}\right) \mathrm{d}m_{H\backslash G}(Hg)\\ &= \int_{H\backslash G} m_H\left(\{h \in H \mid hgg_1^{-1} \in B\}\right) \mathrm{d}m_{H\backslash G}(Hg)\\ &= \nu(B)\end{aligned}$$

by the right-invariance of $m_{H\backslash G}$. Now let $O \subseteq G$ be open and non-empty. If $\nu(O) = 0$, then we can cover any compact $K \leqslant G$ by finitely many translates

$$K = \bigcup_{i=1}^{n} Og_i$$

of O, which implies that $\nu(K) = 0$. Since G is σ-compact, this forces $\nu = 0$ which contradicts the definition. It follows that ν is a Haar measure on G, and so $\nu = cm_G$ for some $c \in (0, \infty)$. Notice that the argument thus far would have worked equally well for any non-zero locally finite Haar measure on $H\backslash G$.

Finally, let $F \subseteq G$ be a fundamental domain for Γ_H, where Γ_H is viewed as a discrete subgroup of G, let $K \subseteq G$ be compact with non-empty interior, and define

$$B = HK \cap F.$$

The set HK is closed (because K is compact and H is closed), so B is measurable. By definition of ν, $m_{H\backslash G}$ and $m_{\Gamma_H\backslash G}$, we have

$$\begin{aligned}\nu(B) &= \int_{H\backslash G} \underbrace{m_H\left(\{h \in H \mid hg \in HK \cap F\}\right)}_{=m_{\Gamma_H\backslash H}(\Gamma_H\backslash H)\mathbb{1}_{HK}(Hg)} \mathrm{d}m_{H\backslash G}(Hg)\\ &= m_{\Gamma_H\backslash G}\left(\pi^{-1}(HK)\right) = m_G(B).\end{aligned}$$

It follows that $c = 1$, and so (B.8) and (B.9) hold for characteristic functions $f = \mathbb{1}_B$ of Borel subsets $B \subseteq G$. Using linearity and monotone convergence, this extends first to simple functions, then to all non-negative measurable functions, and finally to all integrable functions. □

B.1.2 Other Lattices of H

We wish to extend the discussion above in two ways. Firstly, the measure $m_{H\backslash G}$ defined in (B.7) using a uniform lattice $\Gamma_H < H$ is independent of the chosen lattice, in the sense that it only depends on the choice of Haar measures m_H and m_G. Secondly, the formula B.7 also holds for non-uniform lattices in H.

Lemma B.3. *Let G, H, Γ_H, and $m_{H\backslash G}$ be as in Lemma* B.1*, and assume that $\Lambda_H < H$ is another lattice. Then*

$$m_{H\backslash G}(B) = \frac{m_{\Lambda_H\backslash G}(\pi^{-1}(B))}{m_{\Lambda_H\backslash H}(\Lambda_H\backslash H)}$$

Proof. Let $B \subseteq G$ be a measurable set. Then $f = \mathbb{1}_B$ is a measurable function on G, and by applying (B.8)–(B.9) to the function f we get

$$m_G(B) = \int_{H\backslash G}\int_H \mathbb{1}_B(hg)\,\mathrm{d}m_H(h)\,\mathrm{d}_{H\backslash G}(Hg).$$

Let us write

$$\pi : \Lambda_H\backslash G \longrightarrow H\backslash G$$

and

$$\pi_G : G \longrightarrow H\backslash G$$

for the canonical projections. Now let $F \subseteq G$ be a fundamental domain for Λ_H, let $B \subseteq H\backslash G$ be a measurable set, and apply the identity above to $\pi^{-1}(B) \cap F$ to get

$$\begin{aligned} m_{\Lambda_H\backslash G}\left(\pi^{-1}(B)\right) &= m_G\left(\pi_G^{-1}(B)\cap F\right)\\ &= \int_{H\backslash G}\int_H \mathbb{1}_{\pi_G^{-1}(B)\cap F}(hg)\,\mathrm{d}m_H(h)\,\mathrm{d}m_{H\backslash G}(Hg). \end{aligned} \tag{B.10}$$

We claim that

$$\int_H \mathbb{1}_{\pi_G^{-1}(B)\cap F}(hg)\,\mathrm{d}m_H(h) = \mathbb{1}_B(Hg)m_{\Lambda_H\backslash H}\left(\Lambda_H\backslash H\right).$$

Indeed, if $Hg \notin B$, then $hg \notin \pi_G^{-1}(B)$ for all $h \in H$. On the other hand, if $Hg \in B$, then $Hg \subseteq \pi_G^{-1}(B)$ and

$$\begin{aligned} \int \mathbb{1}_{\pi_G^{-1}(B)\cap F}(hg)\,\mathrm{d}m_H(h) &= \int \mathbb{1}_{Fg^{-1}}(h)\,\mathrm{d}m_H(h)\\ &= m_H(Fg^{-1}\cap H)\\ &= m_{\Lambda_H\backslash H}\left(\Lambda_H\backslash H\right) \end{aligned}$$

since $Fg^{-1} \cap H$ is a fundamental domain for Λ_H in H. Together with (B.10) this proves the lemma. □

B.1.3 Sketch of the General Case

As not all unimodular groups (indeed, not even all unimodular Lie groups) admit compact quotients by lattices, the discussion above does not handle all possible cases. A more general definition of $m_{H\backslash G}$ with similar properties can be obtained by using (B.8) and (B.9) more directly. For any $f \in C_c(G)$ the function F defined as in (B.8) has $F \in C_c(H\backslash G)$. One can now show that for any $F \in C_c(H\backslash G)$ there exists some $f \in C_c(G)$ satisfying (B.8), and that (B.9) can be used as a definition of $\int F \,\mathrm{d}m_{H\backslash G}$ (but for this one has to show that the definition is independent of the function $f \in C_c(G)$ that gives rise to F).

For a general closed subgroup H of a locally compact σ-compact group G the existence of a Haar measure $m_{H\backslash G}$ on the quotient is equivalent to a compatibility condition of the Haar measures m_H and m_G with respect to the respective modular characters. We refer to Raghunathan [**?**] for a complete treatment.

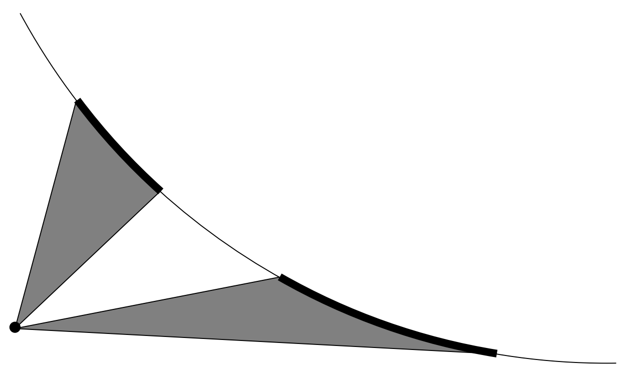

Fig. B.1 For quotients $H\backslash G$ that can be realized in terms of hyperplanes in $\mathbb{R}^d$, the Lebesgue measure of cones can often be used to describe the Haar measure on $H\backslash G$.

Exercises for Appendix B

Show that the Haar measure $m_{H\backslash G}$ (that is, the G-invariant locally finite, non-zero measure on $H\backslash G$) is unique up to a scalar multiple.

a) Let $G = \mathrm{SL}_2(\mathbb{R})$ and $H = U$. Find an explicit description of the elements of $U\backslash G$, and in terms of this description describe $m_{U\backslash G}$,
(b) Let $G = \mathrm{SL}_2(\mathbb{R})$ and

$$H = B = \left\{ \begin{pmatrix} a & s \\ & a^{-1} \end{pmatrix} \mid a \neq 0, s \in \mathbb{R} \right\}.$$

Show that there is no G-invariant locally finite measure on $B\backslash G$.

Let $Q(x, y) = x_1^2 + \cdots + x_p^2 - y_1^2 - \cdots - y_q^2$ be a quadratic form with signature (p, q), and let

$$\begin{aligned} G = \mathrm{SO}(p,q)(\mathbb{R}) &= \{g \in \mathrm{SL}_{p+q}(\mathbb{R}) \mid g \text{ preserves } Q\} \\ &= \left\{g \in \mathrm{SL}_{p+q}(\mathbb{R}) \mid g I_{p,q} g^{\mathrm{t}} = I_{p,q}\right\} \end{aligned}$$

where $I_{p,q}$ is the matrix $\begin{pmatrix} I_p & \\ & -I_q \end{pmatrix}$.

Let $v_0 = (1, 0, \ldots, 0) \in \mathbb{R}^{p+q}$ and

$$H = \mathrm{Stab}_G(v_0) = \{g \in G \mid v_0 g = v_0\}.$$

Show that

$$G/H \cong v_0 G = \{w \in \mathbb{R}^{p+q} \mid w_1^2 + \cdots + w_p^2 - w_{p+1}^2 - \cdots - w_{p+q}^2 = 1\}$$

and that

$$m_{G/H}(B) = m_{\mathbb{R}^{p+q}}\left(\{\lambda v_0 g \mid \lambda \in [0,1], g \in B\}\right)$$

defines a Haar measure on G/H.

Appendix C: Something Algebraic

Appendix D: Adeles and Local Fields?

Appendix E: Modular Characters on Lie Groups?

cf. page **??**.

Appendix F: General Case of Proposition 9.13

If we include this, to contain proof from Pisa notes of the general case to make this volume more self-contained.

Hints for Selected Exercises

Exercise 1.1.5 (p. 20): Treat the case of a unipotent matrix separate from the case of a scalar matrix and show that these cases suffice.

Exercise 1.1.7 (p. 20): Apply the standard version of Poincaré recurrence to every element of a countable basis of the topology of X.

Exercise 1.1.10 (p. 21): Assuming the opposite consider the characteristic polynomial of the matrix in $G \cap \mathrm{SL}_d(\mathbb{Z})$ which is responsible for a small injectivity radius.

Exercise 1.1.11 (p. 21): Let $F \subseteq G$ be a fundamental domain for Γ in G such that $\overline{F}$ is compact (and hence bounded). Let $d = \mathrm{diam}(\overline{F})$, let $\delta > 0$, and show that

$$S = \{\gamma \in \Gamma \mid \mathsf{d}_G(F, \gamma F) < \delta\}$$

generates Γ because $\langle S \rangle F \subseteq G$ is ("uniformly") open and (therefore) closed. (See also Exercise 1.2.1 for the same argument in a more concrete case.)

Exercise 1.1.12 (p. 21): Let $y \in \overline{x(H_1 \cap H_2)} \subseteq (xH_1) \cap (xH_2)$. Applying Proposition 1.13 we find some $\delta > 0$ smaller than the injectivity radius with $yB_\delta^G \cap xH_i = yB_\delta^{H_i}$ (if we use the same metric on G and H_i) for $i = 1, 2$.

Exercise 1.2.1 (p. 28): For (2) prove and use a formula relating $\Im(\gamma(z))$ and $\Im(z)$ for $z \in \mathbb{H}$ and $\gamma \in \mathrm{SL}_2(\mathbb{R})$.

Exercise 1.2.4 (p. 28): Show that the characteristic polynomial of the lattice element corresponding to the period is irreducible over $\mathbb{Q}$. See also Section 3.3 where this correspondence is considered in greater generality.

Exercise 1.3.1 (p. 42): Applying Exercise 1.1.11 (which in this case is much easier to understand geometrically) one sees that $\Lambda = \mathbb{Z}v_1 + \ldots + \mathbb{Z}v_k$ for finitely many vectors $v_1, \ldots, v_k \in \mathbb{R}^d$. By the structure theory of finitely generated abelian torsion-free groups (that is, torsion-free modules over the principal ideal domain $\mathbb{Z}$) we may assume that $v_1, \ldots, v_k$ are linearly independent over $\mathbb{Q}$. Show that if $v_1, \ldots, v_k$ do not span $\mathbb{R}^d$ over $\mathbb{R}$, then $\mathbb{R}^d/\Lambda$

cannot have finite volume. Finally show that if $v_1, \dots, v_k$ are not linearly independent over $\mathbb{R}$, then Λ cannot be discrete.

Alternatively, go through the proof of Theorem 1.15 to extract a proof of this fact.

Exercise 1.3.2 (p. 42): Apply Theorem 1.15 to each $\Lambda \cap W < W$ for each subspace W with $(\Lambda \cap W)$ close to $\inf\{(\Lambda \cap V) \mid V \subseteq \mathbb{R}^d$ is a subspace of rank $k\}$.

Alternatively consider the vectors in the lattice in $\bigwedge^k \mathbb{R}^d$ obtained by taking exterior products of elements of Λ.

Exercise 1.3.6 (p. 42): Notice that the map

$$\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_d \end{pmatrix} \longmapsto \left(\log \tfrac{a_1}{a_2}, \log \tfrac{a_2}{a_3}, \dots, \log \tfrac{a_{d-1}}{a_d}\right)$$

is an isomorphism $A \to \mathbb{R}^{d-1}$. For

$$U = \left\{ \begin{pmatrix} 1 & u_{12} & \dots & u_{1d} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \mid u_{ij} \in \mathbb{R} \right\},$$

show that the Haar measure is given by integration with respect to

$$\mathrm{d}u_{12} \cdots \mathrm{d}u_{1d}\, \mathrm{d}u_{23} \cdots \mathrm{d}u_{(d-1)d}.$$

Exercise 1.3.8 (p. 42): Show that the first step of the algorithm does not change the function θ but the third step of the algorithm decreases its value by a fixed amount (depending on $t < \frac{\sqrt{3}}{2}$). Conclude by showing that there is a lower bound on the values of θ that can be attained for a given lattice.

Exercise 1.3.9 (p. 43): For (1) and (2) generalize the calculations in the proof of Lemma 1.25. For (3) use the fact that Haar measures on any quotient $H\backslash G$ are unique up to scalars (if one exists).

Exercise 2.1.1 (p. 52): Study the abelian subalgebra

$$\mathfrak{h} = \left\{ \begin{pmatrix} a_{11} & & \\ & \ddots & \\ & & a_{dd} \end{pmatrix} \in \mathfrak{sl}_d(\mathbb{R}) \right\}$$

and its eigenspaces (corresponding to $\mathfrak{h}$ and its roots). Then study a Lie ideal $\mathfrak{f} \lhd \mathfrak{sl}_d(\mathbb{R})$, which automatically is a sum of eigenspaces.

Exercise 2.1.2 (p. 52): Show that the complexification is isomorphic to

$$\mathfrak{sl}_d(\mathbb{C}) \times \mathfrak{sl}_d(\mathbb{C}),$$

and that $\mathfrak{g}$ corresponds to the real subalgebra $\{(u, \overline{u}) \mid u \in \mathfrak{sl}(\mathbb{R})\}$.

Exercise 2.1.3 (p. 53): Start by showing that $\mathrm{Ad}_g(\mathfrak{f}) \subseteq \mathfrak{f}$ for $g \in G$. Then set $g(t) = \exp(tu)$ and compute the derivative at $t = 0$.

Exercise 2.1.4 (p. 53): Let $F \lhd \mathrm{SL}_d(\mathbb{R})$ be a non-central normal subgroup, and let $g \in F$ be any non-central element. Consider some $v \in \mathfrak{sl}_d(\mathbb{R})$ such that $\exp(tv)$ and g do not commute for some $t \in \mathbb{R}$. Then $\gamma(t) = \exp(tv) g \exp(-tv)$ defines a smooth curve taking values in F. For a suitable t_0 the curve $\gamma_0(t) = \gamma(t + t_0)\gamma(t_0)^{-1}$ takes values in F and has a non-trivial derivative at $t = 0$. Now consider conjugates of γ_0 to obtain many such curves whose derivatives form a basis of the Lie algebra of G and apply the inverse function theorem.

Exercise 2.1.5 (p. 53): Notice that the left and the right action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathrm{Mat}_{22}(\mathbb{R})$ preserves the quadratic form det on $\mathrm{Mat}_{22}(\mathbb{R})$.

Exercise 2.2.1 (p. 56): For (a) consider the space

$$X = \prod_{\substack{\Gamma < \mathbb{Z}^d; \\ |\mathbb{Z}^d/\Gamma| < \infty}} \mathbb{Z}^d/\Gamma.$$

For (b) use the spectral theory of $\mathbb{R}^d$ (see Lemma 5.8).

Exercise 2.2.2 (p. 56): Show that $g^{\mathrm{t}}g$ is symmetric and positive, and so there is a positive diagonal matrix D with $g^{\mathrm{t}}g = k_2{}^{\mathrm{t}}Dk_2$; take the positive diagonal square root a of D and show that $gk_2{}^{\mathrm{t}}a^{-1} = k_1 \in K$ is orthogonal.

Exercise 2.3.4 (p. 62): Consider the left regular unitary action on $\ell^2(\mathrm{SL}_d(\mathbb{Q}_p)/H)$ and conclude that H has finite index. Now study $H \cap U$ for the unipotent one-parameter subgroups and apply Lemma 1.24.

Exercise 2.3.7 (p. 65): Notice first that if x has an equidistributed orbit, then so does $u^- \cdot x$ for any $u^- \in G_a^-$ since these two points have asymptotical orbits. In the case of x and $g \cdot x$ with $g \in C_G(a)$ the orbits are not asymptotic but parallel, but this leads to the same conclusion. Combining these two cases the statement follows.

Exercise 3.1.1 (p. 83): Write $u_1^2 - u_1u_2 - u_2^2 = (u_1 - \rho u_2)(u_1 + \rho^{-1}u_2)$ with $\rho = \frac{1+\sqrt{5}}{2}$.

Exercise 3.1.4 (p. 84): (1) If $w_1, \dots, w_k$ is a basis of V, define

$$v = w_1 \wedge \cdots \wedge w_k \in \bigwedge^k \mathbb{R}^d,$$

so that $L_V = \mathrm{Stab}_{\mathrm{SL}_d}(v)$. (2) Apply (1) to the rational subspace $V = Vg_0^{-1}$. (3) Either check that the proof of Proposition 3.8 extends to that statement or apply Exercise 1.1.12.

Exercise 3.5.1 (p. 120): Show first that the quadratic form Q (or equivalently the symmetric matrix A_Q) is up to a scalar multiple uniquely determined by the orthogonal group. Then use the Borel density theorem to see that rational equations define the one-dimensional line $\mathbb{R}A_Q$ in the space of symmetric matrices.

Exercise 3.6.1 (p. 124): For (b) generalize the reduction theory presented in Section 1.3. Here the A part in the NAK (or Iwasawa) decomposition for $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ comprises the product of the diagonal subgroups of the two factors. Use the non-trivial unit in $\mathbb{Z}[\sqrt{D}]$ to find a subset of A that is infinite in only one of the two directions of A and suffices in the definition of a Siegel domain for Γ.

Exercise 4.1.1 (p. 131): See Corollary 4.10.

Exercise 4.4.1 (p. 154): Combine Theorem 4.11, and Exercise 3.5.1.

Exercise 4.4.2 (p. 154): Combine Proposition 3.8 with Theorem 4.11.

Exercise 5.1.1 (p. 157): For (a), let

$$U^+ = \left\{ u_s^+ = \begin{pmatrix} 1 & \\ s & 1 \end{pmatrix} \mid s \in \mathbb{R} \right\}$$

be the expanding horocycle. Set $z = u_s^+ \cdot x$ and

$$a_T \cdot z = u_{\mathrm{e}^T s} \cdot (a_T \cdot x) = u_{\mathrm{e}^T s}^+ g \cdot y$$

where $g \in \mathrm{SL}_2(\mathbb{R})$ is the displacement of $a_T \cdot x = g \cdot y$ and y with $\mathsf{d}(g, I) < \varepsilon$. Show that $g = u_r^+ a_\delta u$ for some r, δ with $|r| \ll \varepsilon$, $|\delta| \ll \varepsilon$, $u \in U$ with $\mathsf{d}(u, I) \ll \varepsilon$.

For (b), start for example with two nearby points x, y that are periodic with different periods and that satisfy the requirements in Figure 5.1. Now use part (a) infinitely often to find an orbit that (approximately) alternates between the two periodic orbits in a non-periodic fashion.

Exercise 5.1.2 (p. 158): Write $a_T \cdot x = g \cdot x$ for some $g \in G$ with $\mathsf{d}(g, I) \leqslant \varepsilon$. Try to replace x by $u_s^+ \cdot x$ and calculate

$$a_T \cdot (u_s^+ \cdot x) = u_{\mathrm{e}^T s}^+ g u_s^+ \cdot (u_s^+ \cdot x) = g' \cdot (u_s^+ \cdot x).$$

Now choose s close to zero such that

$$g' = \begin{pmatrix} a' & b' \\ & d' \end{pmatrix}.$$

Repeat with

$$u_r = \begin{pmatrix} 1 & r \\ & 1 \end{pmatrix}.$$

Exercise 5.1.3 (p. 158): See the hints for Exercise 5.1.1.

Exercise 5.1.4 (p. 158): For (a) notice that G has near the identity a local coordinate system of the form $g = u^+u^-c$ with $u^+ \in G_a^+$, $u^- \in G_a^-$, $c \in C_G(a)$. If $a^N\boldsymbol{\cdot}x = g\boldsymbol{\cdot}x$ we may write $g = u_g^+u_g^-c_g$ and replace x by $x_1 = a^{-N}(u^+)^{-1}a^N\boldsymbol{\cdot}x$. Then

$$a^N\boldsymbol{\cdot}x_1 = (u_g^+)^{-1}g\boldsymbol{\cdot}x = \underbrace{u_g^-c_ga^{-N}u_g^+a^N}_{u_1}\boldsymbol{\cdot}x_1.$$

Now write $g_1 = u_1^+u_1^-c_1$ and prove the estimate $\mathsf{d}(u_1^+, I) \ll \varepsilon\lambda^N$ for some $\lambda < 1$ that only depends on a. Iterate the construction and show that the sequence (x_n) found converges to y with $a^N\boldsymbol{\cdot}y = h\boldsymbol{\cdot}y$ with $\mathsf{d}(h, I) \ll \varepsilon$ and $h = u_h^-c_h$. Now replace y by $z = u^-\boldsymbol{\cdot}y$ for a suitable $u^- \in U^-$ to conclude the proof.

For (b), show that for sufficiently large N the point $a' \in A$ is generic. Therefore $z = \Gamma g$, $ga' = \gamma g$ for some generic $\gamma \in \Gamma$. Show that $\Gamma C(\gamma)$ is closed (and compact), therefore $\Gamma C(\gamma)g = \Gamma gA$ is compact too.

For (c), show that the characteristic polynomial of γ is irreducible over $\mathbb{Q}$ (see also the paper of Shapira and Weiss [**?**]).

For (d) use Poincaré recurrence to produce points as in (b) or (c).

For (e) additional ideas are needed, particularly if X is not compact.

Exercise 5.2.1 (p. 161): Conjugate again and study $m_{G_a^-}(B_0\triangle((a^nKa^{-n})B_0))$ instead.

Exercise 5.2.2 (p. 161): Show that any limit point of the left hand side defines a G_a^--invariant probability measure.

Exercise 5.4.1 (p. 173): For (a) simply iterate Theorem 4.9 for $U_1, U_2, \ldots$ (using $\delta, \delta^2, \delta^3, \ldots$, for example). In this case $o(1) = 0$.

For (b) choose a one-parameter subgroup $U' < U$ that does not fix any Λ_x-rational subspaces V of covolume $< \eta^{\dim V}$. Apply Theorem 4.9 to find some $u' \in U'$ such that $u'\boldsymbol{\cdot}x$ belongs to a compact subset that only depends on η. Now use the assumption that F_n is a Følner sequence to show that (a) implies (b).

Exercise 5.4.2 (p. 174): Let μ_n be the normalized measure on $F_n\boldsymbol{\cdot}x$ induced by the Haar measure on $U = G_a^-$. Suppose μ is a weak* limit of a subsequence of μ_n. By Exercise 5.4.1(a) μ is a U-invariant probability measure on X. Now combine Exercise 5.4.1(b) and the argument in the proof of Theorem 5.7.

Exercise 6.6.1 (p. 205): For (1) notice that the orbit xU for $U = \left\{\begin{pmatrix} I & * \\ 0 & 1 \end{pmatrix}\right\}$ is compact for any $x \in X$. For (2) repeat the matrix calculations and the appropriate argument giving Theorem 6.18. For (3) use Proposition 6.9 if the measure is not supported on the orbit of the centralizer.

Exercise 6.7.1 (p. 225): Start as in Exercise 6.6.1 (2) with arguments similar to the proof of Theorem 6.18 leading in the interesting case to invariance under the geodesic flow (possibly conjugated). Consider the factor map $\pi: X \to 2$ and the conditional measures for the σ-algebra $\pi^{-1}\mathscr{B}_2$ in X. Now treat the case where these conditional measures are atomic similar to the arguments in this section.

Exercise 9.1.2 (p. 275): Suppose first that μ is finite and show for the conditional measures that

$$\mathrm{d}\nu_x^{\mathscr{A}} = \frac{f\,\mathrm{d}\mu_x^{\mathscr{A}}}{\int f\,\mathrm{d}\mu_x^{\mathscr{A}}}$$

almost surely.

Exercise 9.2.3 (p. 297): Let $B_G = \pi^{-1}(B) \subseteq G$, and apply Lemma 9.1 to a compact set $Y \subseteq G$ and $Y' = Y \cap B_G$.

Exercise 9.4.1 (p. 312): It may be helpful to consult a simpler case in [**?**, Exercise 2.1.9] or the material in Appendix D.

Exercise 12.6.6 (p. 395): Here is a guide to the problem in a series of steps.

(a) Identify the space W as in the counting problem (up to sign) with

$$w_1 \wedge w_2 \wedge \cdots \wedge w_m,$$

where $w_1, w_2, \ldots, w_m$ form a $\mathbb{Z}$-basis of $W \cap Z^d$.

(b) Show that

$$V = \{(w_1 \wedge w_2 \wedge \cdots \wedge w_m, w_1, w_2, w_1 \wedge w_2 \wedge \cdots \wedge w_m) \in \mathbb{R}^4\}$$

is a single $\mathrm{SL}_d(\mathbb{R})$-orbit (if $m < d$) in the space $\bigwedge^m \mathbb{R}^d \cong \mathbb{R}^{\binom{d}{m}}$. In the case $d = 4$, $m = 2$, and

$$w_1 \wedge w_2 = (a_{ij} = \det(\pi_{ij} w_1, \pi_{ij} w_2 \mid i < j)\,,$$

the space

$$V = \{(a_{ij} \mid i < j) \mid a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23} = 0\}$$

is a hypersurface.

(d) Let $H = \mathrm{Stab}_{\mathrm{SL}_4(\mathbb{R})}(e_1 \wedge w_2 \wedge \cdots \wedge w_m)$, and prove that $gH\,\mathrm{SL}_d(\mathbb{Z})$ equidistributes as $gH \to \infty$ in $\mathrm{SL}_d(\mathbb{R})/H$.

(e) Show that balls in $\mathrm{SL}_d(\mathbb{R})/H \cong V(\mathbb{R})$ are well-rounded. To do this, describe the Haar measure on

$$L = \left\{ \begin{pmatrix} \mathrm{e}^{t/m} I_m & \\ & \mathrm{e}^{-t/n} I_n \end{pmatrix} \begin{pmatrix} I_m \\ Y I_n \end{pmatrix} \mid t > 0, Y \in \mathrm{Mat}_{nm}(\mathbb{R}) \right\},$$

notice that $L \cdot e_1 \wedge e_2 \cdots \wedge e_m$ is an open subset of V, and so the Haar measure on L needs to coincide with the restriction of the Haar measure on V to the L-orbit. As the Haar measure on V is smooth and $\mathrm{SO}(d)(\mathbb{R})$-invariant we conclude that

$$\mathrm{d}m_V \propto f(r)\,\mathrm{d}r\,\mathrm{d}\sigma,$$

where σ denotes the natural surface area measure on

$$\mathbb{S}_V = \{v \in V \mid \|v\| = 1\}$$

and we describe V in generalized polar coordinates rv_1 with $r > 0$ and $v_1 \in \mathbb{S}_V$. Comparing m_V and m_L, calculate $f(r)$ and then integrate.

(f) Combine parts (a)–(e) with Section 12.3 to give the result.

Exercise 11.2.1 (p. 359): For (b) recall the fact that the geodesic flow on a finite volume quotient of $\mathrm{SL}_2(\mathbb{R})$ has many invariant measures of positive entropy.

Exercise 13.1.2 (p. 403): Show that $v_i^{(1)} = v_{i+1} - v_i$ has (as a polynomial in i) degree one less in each entry, $v_i^{(2)} = v_{i+1}^{(1)} - v_i^{(1)}$ also. Hence $v_0^{(d)}$ is a linear combination of $v_0, \ldots, v_d$ and is a nonzero multiple of e_{d+1}. This shows that e_{d+1} is in the linear hull.

Exercise 13.1.3 (p. 403): If ε is sufficiently small, then the determinant of the matrix obtained from the best approximation $w_i \in W$ of the basis vectors v_i is close to the determinant of the matrix obtained from v_i.

Exercise 13.1.5 (p. 403): Modify the proof of Corollary 13.5 accordingly.

Exercise 13.1.7 (p. 404): Apply Ratner's measure classification theorem to the U-ergodic component of such a measure. Show that the resulting subgroup is normalized by a_t, and study the various cases.

Exercise 13.2.1 (p. 406): Assume the opposite and use a Lebesgue density point similar to the proof of Corollary 13.5.

Exercise 14.0.1 (p. 414): Reduce to the case $L(x_1, x_2) = \alpha x_1 + x_2$ with α irrational.

Author Index

Index of Notation

General Index