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 [98], Ratner [156], Margulis [127], and Dani [25] 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 Raghunathan's conjecture, to later chapters.

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 [20, Conjecture I], as an analogue of Raghunathan's conjecture, which will be described below) and proved by Ratner [154], [153], [152].

The classification results will generally take the form of asserting that an initially unknown measure has some algebraic structure.

Definition 6.1 (Algebraic measure). A probability measure μ on $\Gamma\backslash G$ is called *algebraic* (or *homogeneous*) if 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 $x_0 \in \operatorname{supp}\mu$).

Theorem 6.2 (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 algebraic.*

In this result (unlike the following ones), it is sufficient to assume that Γ is discrete or even just closed. Theorem 6.2, the theorem of Dani and Smillie [27], (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 [155] generalized all of these results in the following theorem.

Theorem 6.3 (Ratner's equidistribution theorem). *Let $X = \Gamma\backslash G$ where Γ 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*
- $\displaystyle\frac{1}{T}\int_0^T f(u_s \centerdot x_0)\,\mathrm{d}s \longrightarrow \frac{1}{\mathrm{vol}(L\centerdot x_0)}\int_{L\centerdot x_0} f\,\mathrm{d}m_{L\cdot x_0}$ *as $T \to \infty$.*

It is interesting to note that Theorem 6.3 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 6.3 also suggests that the closures of orbits under the action of a unipotent one-parameter subgroup should have some algebraic structure. A more general version of that statement is the famous conjecture of Raghunathan[(23)] that motivated all of the theorems above, and was proved by Ratner [154] using the above results as stepping stones.

Definition 6.4 (Algebraic orbit closure). The orbit closure of a point x_0 in $\Gamma\backslash G$ under the action of a closed subgroup H is called *algebraic* (or *homogeneous*), if 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.

Theorem 6.5 (Raghunathan's conjecture; Ratner's orbit closure theorem). *Suppose that $X = \Gamma\backslash G$, with G a connected Lie group and Γ a lattice. Let $H < G$ be a closed subgroup generated by one-parameter unipotent subgroups. Then the orbit closure of any $x_0 \in X$ is algebraic.*

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 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 [139].

Theorem 6.6 (Mozes–Shah equidistribution theorem). † *Let X be the homogeneous space $\Gamma\backslash G$ with G a connected Lie group and Γ a lattice, and let (H_n) be a sequence of subgroups of G 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) *$\mu = m_{L\cdot y}$ is the L-invariant probability measure on a closed finite volume orbit $L\cdot y$ for the closed connected group $L = \mathrm{Stab}_G(\mu)^\circ \leqslant G$. Moreover, μ is invariant and ergodic for the action of a one-parameter unipotent subgroup. Furthermore, suppose that $x_n = \varepsilon_n\cdot x \in \operatorname{supp}\mu_n$ for $n \geqslant 1$ and some $x \in X$ with $\varepsilon_n \to I$ as $n \to \infty$, and suppose the connected subgroups (L_n) satisfy $\mu_n = m_{L_n\cdot x_n}$ 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 $m_{L\cdot x}$ is almost from within the orbit $L\cdot x$. In fact, after modifying the measures in the sequence only slightly by the elements ε_n we get

$$\operatorname{supp}\left((\varepsilon_n)_*^{-1}\mu_n\right) = \varepsilon_n^{-1}L_n\cdot x_n = \varepsilon_n^{-1}L_n\varepsilon_n\cdot x \subseteq L\cdot x = L\cdot y = \operatorname{supp}\mu$$

for $n \geqslant N$.

We will prove special cases of the theorems above.

† 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 $\mathsf{X}_2 = \mathrm{SL}_2(\mathbb{Z})\backslash \mathrm{SL}_2(\mathbb{R})$, which we already understand in some detail (see Section 1.2, Chapter 5, and [52, 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 $\mathrm{SL}_2(\mathbb{Z})\boldsymbol{\cdot}(z, v)$.

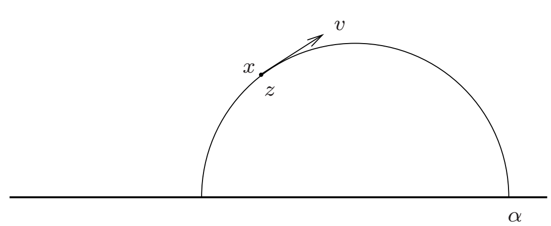

Fig. 6.1: The geodesic determined by x.

In general the answer is given by the following result found by Borel and Prasad [10]. A more general version of this result was obtained more recently by Tomanov [179].

Theorem 6.7. *Let* $\mathsf{X}_d = \mathrm{SL}_d(\mathbb{Z})\backslash \mathrm{SL}_d(\mathbb{R})$*,* $x = \Gamma g \in X$*, 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* 6.2 *and* 6.3 *(respectively Theorem* 6.5*) 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* 6.6 *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* 6.6*.*

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 the 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. We will discuss this further in Section 7.7.

6.3 First Ideas in Unipotent Dynamics

The structure of proof of Theorem 6.2 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 6.2, 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 6.8. *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 $\mathrm{Stab}(\mu)\cdot x_0$ 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 $\mathrm{Stab}(\mu)^\circ\cdot x_0$ also has finite volume. Since H is connected, $H \subseteq \mathrm{Stab}(\mu)^\circ$ and so $\mathrm{Stab}(\mu)^\circ\cdot x_0$ is a H-invariant subset of positive measure. Hence μ is the $\mathrm{Stab}(\mu)^\circ$-invariant Haar measure on $\mathrm{Stab}(\mu)^\circ\cdot x_0 = \mathrm{Stab}(\mu)\cdot x_0$. Finally by Corollary 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.[(24)]

Recall that $x \in X$ is said to be *generic* with respect to μ and a one-parameter flow $\{u_s \mid s \in \mathbb{R}\}$ if

$$\frac{1}{T}\int_0^T f(u_s\cdot x)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $T \to \infty$ for all $f \in C_c(X)$. Using the pointwise ergodic theorem [52, 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 6.12).

Lemma 6.9 (Centralizer Lemma). *If $x, y = h \cdot x \in X$ 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.*

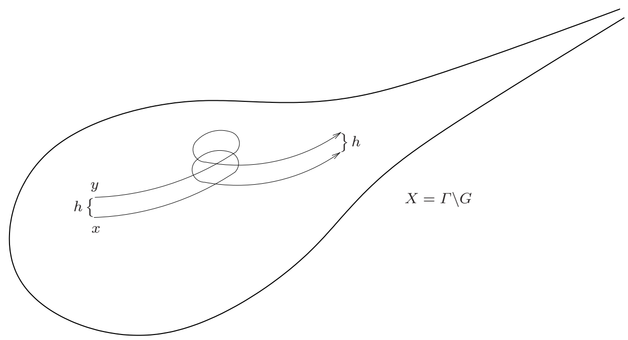

Fig. 6.2: If $y = xh^{-1}$ with $h \in C_G(V)$, then the two orbits are parallel. If in addition both x and y are generic, then the orbits equidistribute (that is, approximate μ), which gives Lemma 6.9.

Proof of Lemma 6.9. We refer to Figure 6.2 for a depiction of the proof. We know that

$$\frac{1}{T}\int_0^T f(u_s \cdot y)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

for any $f \in C_c(X)$. On the other hand

$$\begin{aligned}\frac{1}{T}\int_0^T f(u_s \cdot y)\,\mathrm{d}s &= \frac{1}{T}\int_0^T f(u_s \cdot (h \cdot x))\,\mathrm{d}s \\ &= \frac{1}{T}\int_0^T \underbrace{f(h \cdot (u_s \cdot x))}_{f^h(u_s \cdot x)}\,\mathrm{d}s \qquad (\text{since } h \in C_G(\{u_s\})) \\ &\longrightarrow \int_X f^h\,\mathrm{d}\mu = \int_X f(h \cdot z)\,\mathrm{d}\mu\end{aligned}$$

so μ is h-invariant. □

Lemma 6.9 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 extremely 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 $y = \varepsilon{\scriptstyle\bullet}x$ is close to x, and $\varepsilon \in G$ is the local displacement between x and y, then $u{\scriptstyle\bullet}y = u{\scriptstyle\bullet}\varepsilon{\scriptstyle\bullet}x = u\varepsilon u^{-1}{\scriptstyle\bullet}(u{\scriptstyle\bullet}x)$ and so a displacement between $u{\scriptstyle\bullet}x$ and $u{\scriptstyle\bullet}y$ 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 $u{\scriptstyle\bullet}x$ and $u{\scriptstyle\bullet}y$. 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 $y = \varepsilon{\scriptstyle\bullet}x$, 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 [128].

‡ 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 6.10. 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

$$\left| \frac{1}{T} \int_0^T f(u_s \boldsymbol{\cdot} x) \, \mathrm{d}s - \int_X f \, \mathrm{d}\mu \right| < \varepsilon$$

for all $T \geqslant T_0$ and all $x \in K$.

Proposition 6.11 (Polynomial divergence leads to invariance). *Suppose that $(x_n), (y_n)$ are sequences of uniformly generic points with $y_n = \varepsilon_n \boldsymbol{\cdot} x_n$ for all $n \geqslant 1$ where $\varepsilon_n \to I$ 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

$$p \colon \mathbb{R} \longrightarrow \mathfrak{g}$$

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 6.9. Part of the argument for Proposition 6.11 is illustrated in Figure 6.3.

Proof of Proposition 6.11. 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

$$\mathsf{d}(h_1, h_2) < \delta \implies |f(h_1 \boldsymbol{\cdot} x) - f(h_2 \boldsymbol{\cdot} x)| < \varepsilon$$

for all $x \in X$. Furthermore, choose $\kappa > 0$ so that

$$\mathsf{d}(\exp p(r), \exp p(r_0)) < \delta/2$$

for $r \in [r_0 - \kappa, r_0]$. Then there is an N such that we also have†

† This is the formal version of the statement in Figure 6.3 that the last 1% are parallel.

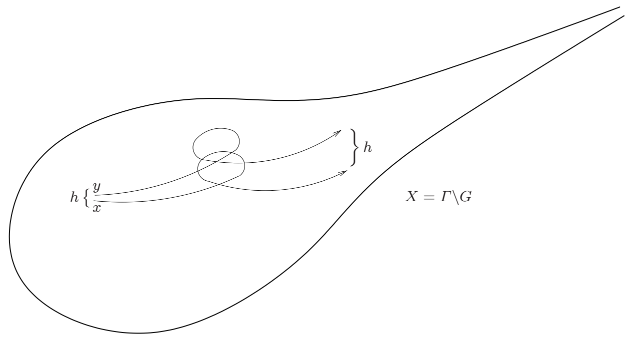

Fig. 6.3: If $y = x\varepsilon^{-1}$ with $\varepsilon \notin C_G(V)$ close to the identity, then the orbits of x and y move away from each other at polynomial speed. If x and y are generic then the last 1% of these pieces of orbits are almost parallel and equidistribute.

$$\mathsf{d}(\exp p_n(r), \exp p(r_0)) < \delta \tag{6.1}$$

for $n \geqslant N$ and $r \in [r_0 - \kappa, r_0]$. We know by the uniform genericity of x_n that

$$\frac{1}{r_0 T_n} \int_0^{r_0 T_n} f(u_s \cdot x_n)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $n \to \infty$, and

$$\frac{1}{(r_0 - \kappa) T_n} \int_0^{(r_0-\kappa) T_n} f(u_s \cdot x_n)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $n \to \infty$. Taking the correct linear combination ($\kappa > 0$ is fixed) and replacing f by $f^{\exp(p(r_0))}$, we get[†]

$$\frac{1}{\kappa T_n} \int_{(r_0-\kappa) T_n}^{r_0 T_n} f^{\exp p(r_0)}(u_s \cdot x_n)\,\mathrm{d}s \longrightarrow \int_X f^{\exp p(r_0)}\,\mathrm{d}\mu$$

as $n \to \infty$ and, by the same argument, we also have

$$\frac{1}{\kappa T_n} \int_{(r_0-\kappa) T_n}^{r_0 T_n} f(u_s \cdot y_n)\,\mathrm{d}s \longrightarrow \int_X f\,\mathrm{d}\mu$$

as $n \to \infty$. However, using the definition of v_n and p_n we have

$$u_s \cdot y_n = u_s \exp(v_n) \cdot x_n = \exp\left(\mathrm{Ad}_{u_s}(v_n)\right) u_s \cdot x_n = \exp\left(p_n(s/T_n)\right) u_s \cdot x_n$$

for all $x \in \mathbb{R}$.

[†] In Figure 6.3 we referred to this as the equidistribution of the last 1% of the orbit.

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

$$\mathsf{d}(u_s \bullet y_n, \exp p(r_0) u_s \bullet x_n) < \delta,$$

and so

$$|f(u_s \bullet y_n) - f(\exp p(r_0) u_s \bullet x_n)| < \varepsilon$$

for every $s \in [(r_0 - \kappa)T_n, r_0 T_n]$. Using this estimate in the integrals above gives

$$\left| \frac{1}{\kappa T_n} \int_{(r_0-\kappa)T_n}^{r_0 T_n} f^{\exp p(r_0)}(u_s \bullet x_n) \, \mathrm{d}s - \frac{1}{\kappa T_n} \int_{(r_0-\kappa)T_n}^{r_0 T_n} f(u_s \bullet y_n) \, \mathrm{d}s \right| < \varepsilon,$$

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 6.12 (Almost full measure 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 $\{u_s \mid s \in \mathbb{R}\}$. 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 [52, Cor. 8.15] for every $f_\ell \in D$ we have

$$\frac{1}{T} \int_0^T f_\ell(u_s \bullet x) \, \mathrm{d}s \longrightarrow \int_X f_\ell \, \mathrm{d}\mu$$

almost everywhere with respect to μ as $T \to \infty$, equivalently for every $\varepsilon > 0$

$$\mu\left(\left\{ x \in X \mid \sup_{T > T_0} \left| \frac{1}{T} \int_0^T f_\ell(u_s \bullet x) \, \mathrm{d}s - \int_X f_\ell \, \mathrm{d}\mu \right| > \varepsilon \right\}\right) \longrightarrow 0$$

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

$$\mu\left(\left\{ x \in X \mid \sup_{T > T_{\ell,n}} \left| \frac{1}{T} \int_0^T f_\ell(u_s \bullet x) \, \mathrm{d}s - \int_X f_\ell \, \mathrm{d}\mu \right| > \frac{1}{n} \right\}\right) < \frac{\rho}{2^{\ell+n}}.$$

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. This will lead to a strengthening of Dani's theorem (Theorem 5.3), due to Ratner, and is the key to the general case.

Exercises for Section 6.3

Exercise 6.3.1. Show that the limit polynomial in Proposition 6.11 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 6.2) 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 6.13. *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 6.11. 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 $y = g\cdot x_0 \in X'$, then $g \in C_G(U) = G$, so $g \in \mathrm{Stab}_G(\mu)$ by Lemma 6.9 and $y \in \mathrm{Stab}_G(\mu)\cdot x_0$ also. It follows that $X' \subseteq \mathrm{Stab}_G(\mu)\cdot x_0$ has full measure, and we deduce that μ must be the Haar measure on $\mathrm{Stab}_G(\mu)\cdot x_0$ 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 $j \in \{0, \ldots, k-1\}$. 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 6.12. Choose some

$$x_0 \in K \cap \mathrm{supp}(\mu|_K). \tag{6.2}$$

We distinguish between two possible scenarios.

It could be that there is some $\delta > 0$ such that $y = \varepsilon\cdot x_0 \in K$ with

$$\mathsf{d}(\varepsilon, I) < \delta$$

implies that $\varepsilon \in C_G(U)$. In this case (6.2) implies that the U-invariant set $C_G(U)\cdot x_0$ has positive measure, and so by ergodicity we have

$$\mu(C_G(U)\cdot x_0) = 1.$$

Now set $G' = C_G(U) \lneqq G$ and $\Gamma' = \mathrm{Stab}_{G'}(x_0)$ so that we may also consider μ as a U-invariant probability measure on

$$\Gamma'\backslash G' \cong G'\cdot x_0$$

(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 $(y_n = \varepsilon_n\cdot x_0)$ 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 6.11 converges to a non-constant polynomial $p\colon \mathbb{R} \to \mathfrak{g}$. By Proposition 6.11 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 6.11)

$$p_n(r) = \mathrm{Ad}_{u(T_n r)}(\log \varepsilon_n) \in \log \varepsilon_n + \mathfrak{g}_{j+1}$$

for all r, where

$$\mathfrak{g}_{j+1} = \operatorname{Lie} G_{j+1} = [\mathfrak{g}, \mathfrak{g}_j].$$

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

$$(\operatorname{Stab}_G(\mu) \cap G_{j+1})^o$$

is a non-trivial subgroup. Clearly U normalizes this subgroup and its Lie algebra, and since $\operatorname{Ad}_{u(t)}$ is unipotent for all $t \in \mathbb{R}$, it follows that there exists a one-parameter unipotent subgroup

$$U' = \{u'_t \mid t \in \mathbb{R}\} \leqslant \operatorname{Stab}_G(\mu) \cap G_{j+1} \cap C_G(U).$$

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 [52, Th. 6.2, 8.20] that the ergodic decomposition allows us to write

$$\mu = \int_X \mu_x^{\mathcal{E}'} \,\mathrm{d}\mu, \tag{6.3}$$

where $\mu_x^{\mathcal{E}'}$ is the conditional measure for the σ-algebra

$$\mathcal{E}' = \{B \in \mathcal{B}_X \mid \mu(u'_t \cdot B \triangle B) = 0 \text{ for all } t\}$$

and that for μ-almost every x the conditional measure $\mu_x^{\mathcal{E}'}$ is a U'-invariant and ergodic probability measure on X with $x \in \operatorname{supp} \mu_x^{\mathcal{E}'}$.

By applying the inductive hypothesis to μ-almost every $\mu_x^{\mathcal{E}'}$ we obtain a function $x \mapsto L_x$ that assigns to x the connected subgroup L_x for which $\mu_x^{\mathcal{E}'}$ is the L_x-invariant probability measure on the closed orbit $L_x \cdot x$. We claim that there is a connected subgroup L such that $L_x = L$ for μ-almost every x. Indeed, since $U = \{u(t) \mid t \in \mathbb{R}\}$ preserves μ and leaves the σ-algebra $\mathcal{E}'$ invariant (since U' and U commute) we get

$$(u_t)_* \mu_x^{\mathcal{E}'} = \mu_{u_t \cdot x}^{\mathcal{E}'} \tag{6.4}$$

for every $t \in \mathbb{R}$ and μ-almost every x by [52, Cor. 5.24]. Since $\mu_x^{\mathcal{E}'}$ is L_x-invariant, it follows from (6.4) that $(u_1)_* \mu_x^{\mathcal{E}'}$ is $u_1 L_x u_1^{-1}$-invariant, which implies that

$$u_1 L_x u_1^{-1} \subseteq L_{u(1) \cdot x}$$

and, by a similar argument for the reverse inclusion,

$$u_1 L_x u_1^{-1} = L_{u_1 \cdot x}.$$

† In fact U' never acts ergodically with respect to μ.

Iterating this relationship shows that

$$u_1^n L_x u_1^{-n} = L_{u(n)\bullet x} \tag{6.5}$$

for μ-almost every 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

$$u_1 L_x u_1^{-1} = L_x$$

for μ-almost every x. Notice that for any such x we also get

$$u(t) L_x u(t)^{-1} = L_x$$

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 6.14 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 $L\bullet x$ is for μ-almost every x closed, with finite L-invariant measure and with the property that $U' \leqslant L$ acts ergodically on $L\bullet x$. 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 $M\bullet x$ is compact for μ-almost every x and postpone the proof to Lemma 6.16.

Next we claim that $N_G^1(M)\bullet x$ is a closed orbit for μ-almost every x, see Lemma 6.17. This implies that μ is supported on a single orbit $x_0 N_G^1(M)$ 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 $\pi_M \colon G \to G/M$ 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

$$\int_X f(x)\,\mathrm{d}\mu(x) = \int_X f(m\bullet x)\,\mathrm{d}\mu(x) \tag{6.6}$$

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 6.14 (Measurability of stabilizer). *Let G be a Lie group, $\Gamma \leqslant G$ a discrete subgroup, and let $X = \Gamma\backslash G$. Then the map*

$$\mathcal{M}(X) \ni \mu \longmapsto \mathrm{Stab}_G(\mu)^o$$

from the space $\mathcal{M}(X)$ 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 $\mathcal{M}(X)$ to

$$\{e\} \sqcup \bigsqcup_{\ell=1}^{\dim G} \mathrm{Grass}_\ell(\mathrm{Lie}\,G),$$

which is a compact metric space and hence has a measurable structure via the Borel σ-algebra.

Proof of Lemma 6.14. Let $d = \dim G$, so that

$$\mathcal{M}_d = \{\mu \in \mathcal{M}(X) \mid \dim \mathrm{Stab}_G(\mu) = d\} = \{m_X\}$$

and $\mathcal{M}_d \ni \mu \mapsto \mathrm{Stab}_G(\mu)^o$ is trivially measurable.

Fix k with $0 \leqslant k \leqslant d$ and suppose that we have already shown that the sets

$$\mathcal{M}_\ell = \{\mu \mid \dim \mathrm{Stab}_G(\mu) = \ell\}$$

for $\ell \geqslant k$ and the map $\mathcal{M}_{k+1} \ni \mu \mapsto \mathrm{Stab}_G(\mu)^o$ are measurable.

Let $\mu_n \in \mathcal{M}_{\geqslant k} = \mathcal{M}_k \cup \cdots \cup \mathcal{M}_d$ 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 $\mathrm{Stab}_G(\mu)^o$. As

$$\{I\} \cup \bigcup_{1\leqslant \ell \leqslant d} \mathrm{Grass}_\ell(\mathrm{Lie}\,G)$$

is compact, we may choose a subsequence and assume also 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 $\mu \in \mathcal{M}_{\geqslant k}$. It follows that $\mathcal{M}_{\geqslant k}$ is closed and hence measurable, which implies that $\mathcal{M}_k = \mathcal{M}_{\geqslant k} \smallsetminus \mathcal{M}_{\geqslant k+1}$ is also measurable.

The argument above also shows that the assumption $\mu_n \in \mathcal{M}_k$ for all $n \geqslant 1$ and $\mu_n \to \mu \in \mathcal{M}_k$ as $n \to \infty$ implies that $\mathfrak{h}_n \to \mathfrak{h}$ as $n \to \infty$, with $\dim \mathfrak{h} = k$. Therefore

$$\mathcal{M}_k \ni \mu \longmapsto \operatorname{Stab}_G(\mu)^o$$

is actually continuous on the measurable set $\mathcal{M}_k$.

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 6.15. *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 6.16. *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. 124) there exist elements $\lambda_1, \dots, \lambda_k \in \Lambda$ with

$$C_L(L) = C_L(\lambda_1, \dots, \lambda_k).$$

Thus Lemma 6.15 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 6.17. *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 6.15, 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

$$\Gamma g_n M \cong \big((g_n^{-1} \Gamma g_n) \cap M\big) \backslash M,$$

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 6.13 we can establish the equidistribution theorem (Theorem 6.3) and the orbit closure theorem (Theorem 6.5) 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 6.18. *Let G be a connected nilpotent Lie group, let $\Gamma < G$ be a lattice in 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.*

PROOF. (to come) □

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 [154], the observation that this particular case has a short and relatively easy proof was made in [47].

† 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 $\overline{H\bullet x} \subseteq X$.

Theorem 6.19 (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 M that is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$ and that projects non-trivially to each simple almost direct factor. Furthermore, we let $U \leqslant M$ be the subgroup corresponding to the upper unipotent subgroup in $\mathrm{SL}_2(\mathbb{R})$. By Proposition 2.11 the subgroup U, and hence also M, satisfies the Mautner phenomenon for H. Since H acts ergodically, so does the subgroup. So we may assume that H is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$.

By the structure theory of finite-dimensional representations of $\mathrm{SL}_2(\mathbb{R})$ (see [46], Fulton and Harris [64], or Knapp [103, 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 $V < \mathfrak{g}$. We note that we have no reason to expect that V is a Lie algebra, and that this step uses crucially the fact that H is semi-simple.

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

$$y_n = g_n \cdot x_n,$$

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 6.11. 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 $p\colon \mathbb{R} \to \mathfrak{g}$ 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

$$Y = \left\{ x \in X \mid \int_{B_\delta^L} \mathbb{1}_K(\ell \cdot x) \, \mathrm{d}m_L(\ell) > 0.9 m_L(B_\delta^L) \right\}.$$

We claim first that $\mu(Y) > 0.9$, which may be seen by looking at the complement as follows:

$$\begin{aligned}
\mu(X \smallsetminus Y) &= \mu\left(\left\{ x \in X \mid \int_{B_\delta^L} \mathbb{1}_{X \smallsetminus K}(\ell \cdot x) \, \mathrm{d}m_L(\ell) \geqslant 0.1 m_L(B_\delta^L) \right\} \right) \\
&\leqslant \frac{1}{0.1 m_L(B_\delta^L)} \int_X \int_{B_\delta^L} \mathbb{1}_{X \smallsetminus K}(\ell \cdot x) \, \mathrm{d}m_L(\ell) \, \mathrm{d}\mu \\
&= \frac{1}{0.1 m_L(B_\delta^L)} \int_{B_\delta^L} \underbrace{\int_X \mathbb{1}_{X \smallsetminus K}(\ell \cdot x) \, \mathrm{d}\mu}_{=\mu(X \smallsetminus K)} \, \mathrm{d}m_L \qquad \text{(by Fubini)} \\
&= \frac{\mu(X \smallsetminus K)}{0.1} < \frac{0.01}{0.1} = 0.1,
\end{aligned}$$

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

$$x' = \ell_x \cdot x \in K, \tag{6.10}$$

$$y' = \ell_y \cdot y \in K, \tag{6.11}$$

and

$$y' = \exp(v) \cdot x' \tag{6.12}$$

with $v \in V$. To see this, notice that if δ is sufficiently small, then (by the inverse mapping theorem) the map

$$\begin{aligned}
\psi \colon B_{2\delta}^L \times B_{2\delta}^V(0) &\longrightarrow G \\
(\ell, v) &\longmapsto \ell \exp(v)
\end{aligned}$$

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 $y = g \cdot x$. 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 $x' = \ell_x \cdot x \in K$, and at least 90% of all $\ell_y \in B_\delta^L$ satisfy $y' = \ell_y \cdot y \in K$. However, we need to do this while ensuring that (6.13) (or equivalently, (6.12)) holds. So define the map

$$\begin{aligned}\phi\colon B_\delta^L &\longrightarrow B_{2\delta}^L\\ \ell_x &\longmapsto \ell_y\end{aligned}$$

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

$$\begin{aligned}m_L\left(\phi\left(\left\{\ell_x \in B_\delta^L \mid \ell_x \cdot x \in K\right\}\right) \cap B_\delta^L\right) &> 0.9 m_L\left(\phi\left(\left\{\ell_x \in B_\delta^L \mid \ell_x \cdot x \in K\right\}\right)\right)\\ &> 0.8 m_L\left(\left\{\ell_x \in B_\delta^L \mid \ell_x \cdot x \in K\right\}\right)\\ &> (0.8)(0.9) m_L\left(B_\delta^L\right) > 0.7 m_L\left(B_\delta^L\right).\end{aligned}$$

Together with

$$m_L\left(\{\ell_y \in B_\delta^L \mid \ell_y \cdot y \in K\}\right) > 0.9 m_L\left(B_\delta^L\right),$$

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

$$z \in \mathrm{supp}\left(\mu|_Y\right).$$

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 6.9 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 6.11 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. □

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 for the horocycle flow under a weaker assumption. Note that the assumption made in Theorem 6.20 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 general measure classification results, and indeed is a result of Ratner (see Theorem 6.2 and the survey [157]). The use of entropy below goes back to work of Margulis and Tomanov [128].

Theorem 6.20 (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.*

† 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 $\Gamma\backslash \mathrm{SL}_3(\mathbb{R})$.

‡ The entropy discussion has not been finished; for now we refer to Pisa notes [43] (Einsiedler and Lindenstrauss, 'Diagonal actions on locally homogeneous spaces', Clay Math. Proc. 2010) where the same material is presented.

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 $\{u_s\}$ (see Exercise 6.3.1). We start by generalizing Section 6.3.1.

Lemma 6.21 (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 $x, y = g\boldsymbol{.}x \in X$ 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.

Proof of Lemma 6.21. 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

$$g u_s g^{-1} = u_{\lambda s}$$

for some fixed $\lambda \in \mathbb{R}^\times$. Hence the piece of the orbit

$$u_{[-T,T]}\boldsymbol{.}x = \{u_s\boldsymbol{.}x \mid s \in [-T,T]\}$$

is mapped under g to

$$g u_{[-T,T]}\boldsymbol{.}x = \{u_{\lambda s} g\boldsymbol{.}x \mid s \in [-T,T]\} = u_{[-|\lambda|T, |\lambda|T]}\boldsymbol{.}y.$$

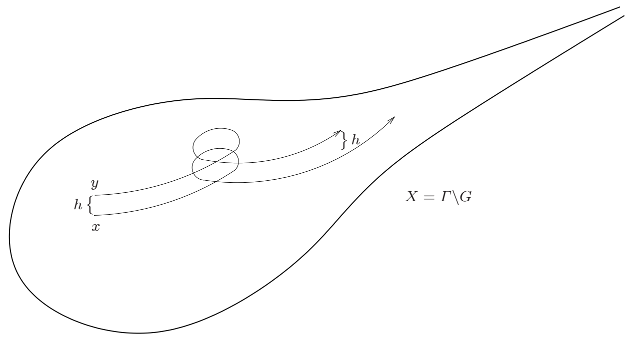

Fig. 6.4: If $y = xh^{-1}$ with $h \in N_G(V)$, then the two orbits are again parallel as in Figure 6.2, but xu_s^{-1} may not be close to yu_s^{-1} but instead be close to yu_r^{-1} for some r.

As before, the normalized Lebesgue measure on $u_{[-T,T]}\bullet x$ and $u_{[-|\lambda|T,|\lambda|T]}\bullet y$ both approximate μ as $T \to \infty$, and we deduce that g preserves μ. □

Just as in the discussion in Section 6.3.1, for the proof of Theorem 6.20 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 6.20. 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

$$B_\delta(z) \cap K \subseteq U\bullet z,$$

then $U\bullet z$ 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

$$y_n = g_n\bullet x_n$$

and $g_n \notin U$ for all $n \geqslant 1$ and $g_n \to I$ 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 6.21 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 6.11, but we already learned from the proof of Lemma 6.21 that we might have to use different clocks for the parametrization of the orbits of x and y. We have seen before (see (2.5) in the proof of Proposition 2.5 on p. 55) 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

$$\mathsf{d}(h_1, h_2) < \delta \implies |f(h_1 \cdot x) - f(h_2 \cdot x)| < \varepsilon \tag{6.16}$$

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

$$\mathsf{A}_n = \frac{1}{\kappa T_n} \int_{(\frac{1}{2}-\kappa)T_n}^{\frac{1}{2}T_n} f^{\phi(1/2)}(u_s \cdot x_n) \, \mathrm{d}s \longrightarrow \int_X f^{\phi(1/2)} \, \mathrm{d}\mu.$$

Now

$$f^{\phi(1/2)}(u_s \cdot x_n) = f\left(\phi(1/2) u_s \cdot x_n\right)$$

is within ε of

$$f\left(\phi_n(c_n s) u_s \cdot x_n\right),$$

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

$$\left| \mathsf{A}_n - \frac{1}{\kappa T_n} \int_{(\frac{1}{2}-\kappa)T_n}^{\frac{1}{2}T_n} f\Big(u_{a_n s/(d_n - c_n s)} \underbrace{g_n \cdot x_n}_{y_n}\Big) \, \mathrm{d}s \right| < \varepsilon.$$

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

$$\left|\mathsf{A}_n - \frac{1}{4\kappa T_n}\int_{(\frac{1}{2}-\kappa)T_n}^{\frac{1}{2}T_n} f\Bigl(\,u_{\underbrace{a_n s/(d_n - c_n s)}_{t}}\Bigr) y_n \cdot \frac{\mathrm{d}t}{\mathrm{d}s}\,\mathrm{d}s\right| < \varepsilon\left(1 + \|f\|_\infty\right),$$

and equivalently

$$\left|\mathsf{A}_n - \frac{1}{4\kappa T_n}\int_{(\frac{1}{2}-\kappa)\frac{a_n T_n}{d_n - \frac{1}{2}+\kappa}}^{\frac{1}{2}\frac{a_n T_n}{d_n - \frac{1}{2}}} f\left(u_t \cdot y_n\right)\,\mathrm{d}t\right| < \varepsilon\left(1 + \|f\|_\infty\right).$$

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

$$\frac{1}{\text{total length}}\int_{(\frac{1}{2}-\kappa)\frac{a_n T_n}{d_n - \frac{1}{2}+\kappa}}^{\frac{1}{2}\frac{a_n T_n}{d_n - \frac{1}{2}}} f\left(u_t \cdot y_n\right)\,\mathrm{d}t \longrightarrow \int_X f\,\mathrm{d}\mu.$$

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

$$h_\mu(a) = 2\log|\lambda|,$$

[†] 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 [45] and the more refined arguments of Part II.

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

$$G = \mathrm{ASL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2 = \left\{\begin{pmatrix}A & v\\ 0 & 1\end{pmatrix} \mid A \in \mathrm{SL}_2(\mathbb{R}), v \in \mathbb{R}^2\right\},$$

$\Gamma = \mathrm{ASL}_2(\mathbb{Z}) = G \cap \mathrm{SL}_3(\mathbb{R})$, 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 $u_s = \begin{pmatrix}1 & 0 & s\\ & 1 & 0\\ & & 1\end{pmatrix}$.

(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

[†] The reader may check whether these are all of them, up to conjugation.

- $\left\{u_s = \left(\begin{pmatrix}1 & s\\ & 1\end{pmatrix}, \begin{pmatrix}1 & s\\ & 1\end{pmatrix}\right) \middle| s \in \mathbb{R}\right\}$.

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 6.2).

Theorem 6.22 (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 [150, 149, 151] and will not use entropy theory but only works in a special case (more precisely, only for product lattices).

Start of proof of Theorem 6.22 using transverse divergence. Let $K \subseteq X$ be a set of uniformly generic points of measure $\mu(K) > 0.99$ and suppose that $y = g{\boldsymbol\cdot}x$ 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

$$u_{s_y}{\boldsymbol\cdot}y = u_{s_y} g u_{-s_x}{\boldsymbol\cdot}(u_{s_x}{\boldsymbol\cdot}x)$$

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)

$$u_{s_y} g u_{-s_x} = \left(\begin{pmatrix}a + cs_y & b - as_x + s_y(d - cs_x)\\ c & d - cs_x\end{pmatrix},\right.$$
$$\left.\begin{pmatrix}a' + c's_y & b' - a's_x + s_y(d' - c's_x)\\ c' & d' - c's_x\end{pmatrix}\right).$$

We again set $s_y = \frac{as_x}{d - cs_x}$, so that the above simplifies to

† In some portions of the proof it will be convenient to refer to Proposition 5.7 where we assumed that $\Gamma\backslash G$ is isomorphic to an orbit in X_d for some $d \geqslant 2$. So strictly speaking we should assume this. Alternatively see Exercise 6.7.2.

$$\phi(s_x) = \left(\begin{pmatrix} a + cs_y & b \\ c & d - cs_x \end{pmatrix}, \begin{pmatrix} a' + c's_y & b' - a's_x + s_y(d' - c's_x) \\ c' & d' - c's_x \end{pmatrix} \right), \tag{6.19}$$

and this already ensures that any non-constant limit of functions like ϕ (with b approaching 0) does not take values in U.

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 6.23. *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‡ $u\colon \mathbb{R}^k \to U$ 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 that $s_y(0) = 0$ and

$$\phi(s_x) = u(s_y) g u(-s_x)$$

is constant where defined. Then $g \in N_G(U)$.

Proof. Let $s_x \in \mathbb{R}^k$ be close to 0. Then

$$\phi(s_x) = u(s_y) g u(-s_x) = \phi(0) = g$$

is equivalent to

$$g u(s_x) g^{-1} = u(s_y) \in U.$$

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 6.21 we also have $g \in \mathrm{Stab}_G(\mu)$. Suppose that $x \in \operatorname{supp} \mu|_K$, and that we are in this case for all $y = g \cdot x \in K$ sufficiently close to x. Then $\mathrm{Stab}_G(\mu) \cdot x$ 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

$$y_n = g_n \cdot x$$

† 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.

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

$$\Phi_n(s_x) = (d_n - c_n s_x, (d_n - c_n s_x)\phi_n(s_x))$$

is a tuple of polynomials with not both being constant. We define the speeding-up parameters $T_n > 0$ such that

$$\sup_{r\in[0,1]} \|\Phi_n(T_n r) - \Phi_n(0)\|_\infty = 1.$$

We may choose a subsequence† such that

$$\Phi_n(T_n r) \longrightarrow \Psi(r)$$

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,

$$\frac{1}{1-\alpha r}\psi_0(r) = \psi(r)$$

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 behaviour of the function

$$r \mapsto \psi(r) \in G$$

for $r \neq \frac{1}{\alpha}$, without calculating it explicitly.

Lemma 6.24. *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 $u\colon \mathbb{R}^k \to U$ such that $u(0) = I$. Let $\{M_t \mid U \to U \mid t \in \mathbb{R}\}$ 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 $\left(s_y^{(n)}\colon \mathbb{R}^k \to \mathbb{R}^k\right)$, well-defined except possibly on proper subvarieties, such that‡

$$\psi_n(r) = M_{t_n}\big(u(s_y^{(n)}(r))\big) g_n M_{t_n}(u(-r))$$

† As usual we simplify notation by not introducing a further subscript to denote the subsequence.

‡ As the reader may notice, the automorphism M_{t_n} does the speeding-up in this more general setting.

converges uniformly on some open subset $O \subseteq \mathbb{R}^k$ to some function

$$\psi\colon O \to G.$$

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 $M_t\colon U \to U$ 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

$$\psi(r) = \lim_{n\to\infty} u_n'\psi_n(r)u$$

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 6.20 and Lemma 6.24 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 6.24 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 2.9 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 6.22 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 almost every 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 6.19 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 for every measurable set X' of full measure there exist $g\bullet x, x \in X'$ with $g \in G_a^+ \smallsetminus U^+$. 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 6.25. *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 $y = g\bullet x$ and x with $g \in G_a^+ \smallsetminus U^+$, and infinitely many $n \geqslant 1$ with*

$$a^{-n}\bullet y = a^{-n} g a^n \bullet (a^{-n}\bullet x) \in K \text{ and } a^{-n}\bullet x \in K.$$

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 2.9 μ is also ergodic with respect to the action of a. Let

$$X' = \left\{ x \in X \mid \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_K(a^{-n} \bullet x) > \tfrac{9}{10} \right\}, \tag{6.21}$$

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 $y = g \bullet x$ and $x \in X'$ with $g \in G_a^+ \smallsetminus U^+$, and the lemma follows. □

Applying Lemma 6.25 in our case, we find two points $y = g \bullet x$ 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 $y_n = a^{-n} \bullet y$, $x_n = a^{-n} \bullet x$ in K. Notice that

$$g_n = a^{-n} g a^n = \left(\begin{pmatrix} 1 & \\ \lambda^{2n} c & 1 \end{pmatrix}, \begin{pmatrix} 1 & \\ \lambda^{2n} c' & 1 \end{pmatrix} \right).$$

We now start the arguments from the very beginning of the proof again, using the points $y_n = g_n \bullet x_n, x_n \in K$. The calculation there now simplifies, since[†]

$$u\left(\tfrac{s_x}{1-cs_x}\right) g u(-s_x) = \left(\begin{pmatrix} * & 0 \\ c & 1 - cs_x \end{pmatrix}, \begin{pmatrix} * & -s_x + s_y(1 - c's_x) \\ c' & 1 - c's_x \end{pmatrix} \right),$$

which after conjugation by a^{-n} gives

$$u\left(\lambda^{-2n} \tfrac{s_x}{1-cs_x}\right) g_n u\left(-\lambda^{-2n} s_x\right) = \left(\begin{pmatrix} * & 0 \\ \lambda^{2n} c & 1 - cs_x \end{pmatrix}, \begin{pmatrix} * & \lambda^{-2n} s_x^2 \frac{c-c'}{1-cs_x} \\ \lambda^{2n} c' & 1 - c's_x \end{pmatrix} \right).$$

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

$$\frac{T_n}{\lambda^{-2n}} \longrightarrow 0$$

and, in particular, $c\lambda^{2n} T_n \to 0$ and $c'\lambda^{2n} T_n \to 0$, as $n \to \infty$. This shows (after choosing a converging subsequence and taking the limit) that

[†] We write $*$ for an entry of the matrices that we do not care about.

$$\psi(r) = \left(\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \begin{pmatrix} 1 & \kappa r^2 \\ & 1 \end{pmatrix} \right)$$

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 6.22 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 $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.

(ii) One of the two measures is a periodic orbit measure, but the other is Haar measure.

(iii) 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 $\mu = \int \mu^{\mathcal{A}}_{(x,x')} \, \mathrm{d}\mu(x, x')$ into conditional measures for the σ-algebra

$$\mathcal{A} = \mathcal{B}_{X_1} \times \{\varnothing, X_2\},$$

we notice that u_s preserves every element of $\mathcal{A}$ modulo μ (since μ-almost everywhere u_s does not change the first component). By [52, Cor. 5.24] this implies a.s. that

$$(u_s)_* \mu^{\mathcal{A}}_{(x,x')} = \mu^{\mathcal{A}}_{u_s \cdot (x,x')} = \mu^{\mathcal{A}}_{(x,x')}.$$

To summarize we have that $\mu^{\mathcal{A}}_{(x,x')}$ 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

$$m_{X_2} = (\pi_2)_* \mu = (\pi_2)_* \int \mu^{\mathcal{A}}_{(x,x')} \, \mathrm{d}\mu(x, x') = \int (\pi_2)_* \mu^{\mathcal{A}}_{(x,x')} \, \mathrm{d}\mu_{(x,x')}$$

expresses m_{X_2} as an integral convex combination of other horocycle-invariant probability measures, it follows by ergodicity that

$$(\pi_2)_*\mu^{\mathcal{A}}_{(x,x')} = m_{X_2},$$

or equivalently

$$\mu^{\mathcal{A}}_{(x,x')} = \delta_x \times m_{X_2}$$

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 212, 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 $\mathcal{A} = \mathcal{B}_{X_1} \times \{\varnothing, X_2\}$ and consider the conditional measures $\mu^{\mathcal{A}}_{(x,x')}$ which describes μ on the atom $[(x,x')]_{\mathcal{A}} = \{x\} \times X_2$ for μ-almost every (x,x').

Claim. If $\mu^{\mathcal{A}}_{(x,x')}$ is not atomic almost everywhere, then $\mu = m_{X_1} \times m_{X_2}$ is the trivial joining.

Proof of Claim. 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 $\mu^{\mathcal{A}}_{(x,x')}$ is not atomic almost everywhere, we see that $\mu^{\mathcal{A}}_{(x,x')}|_{X'}$ is not atomic μ-almost everywhere. Therefore there exists some set K_m and some $(x,x') \in X$ such that $\mu^{\mathcal{A}}_{(x,x')}|_{K_m}$ is non-atomic. As $\operatorname{supp}\mu^{\mathcal{A}}_{(x,x')} \subseteq \{x\} \times X_2$, we see that there exists a sequence of points

$$x_n, y_n = g_n \cdot x_n \in K_m$$

with $x_n \neq y_n$, $\pi_1(x_n) = x = \pi_1(y_n)$ where the displacement satisfies

$$g_n = \left(I, \begin{pmatrix} a'_n & b'_n \\ c'_n & d'_n \end{pmatrix}\right) \neq I$$

[†] This element a will be used later in the proof, but its entropy properties will not be used.

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 212 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 6.9), or because we are back in Case I of the transverse divergence argument. □

Thus we may suppose that $\mu^{\mathcal{A}}_{(x,x')}$ is atomic almost everywhere, in which case we make the following claim.

Claim. There exists some $m \geqslant 1$ and functions[†] $f_1, \dots, f_m \colon X_1 \to X_2$ such that the measure $\mu^{\mathcal{A}}_{(x,x')}$ may be expressed in the form[‡]

$$\mu^{\mathcal{A}}_{(x,\cdot)} = \frac{1}{m} \sum_{i=1}^{m} \delta_{(x, f_i(x))}$$

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 $\{f_1, \dots, f_m\}$

Proof of Claim. We define a function

$$f((x,x')) = \mu^{\mathcal{A}}_{(x,\cdot)}\left(\{(x,x')\}\right)$$

which by the previous claim is positive almost surely. Notice that $u_s^{-1}\mathcal{A} = \mathcal{A}$, so that

$$\begin{aligned} f\left(u_s \centerdot (x,x')\right) &= \mu^{\mathcal{A}}_{u_s \centerdot (x,\cdot)}\left(\{u_s \centerdot (x,x')\}\right) \\ &= (u_s)_* \mu^{u_s^{-1}\mathcal{A}}_{(x,\cdot)}\left(\{u_s \centerdot (x,x')\}\right) \\ &= \mu^{\mathcal{A}}_{(x,\cdot)}\left(\{(x,x')\}\right) = f((x,x')) \end{aligned}$$

[†] 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 $\mu^{\mathcal{A}}_{(x,\cdot)} = \mu^{\mathcal{A}}_{(x,x')}$ as the conditional measure does not depend on the second coordinate x'.

by [52, 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

$$\mu^{\mathcal{A}}_{(x,\cdot)}\left(\{(x,x')\}\right) = f\left((x,x')\right) = f\left((x,y')\right) = \mu^{\mathcal{A}}_{(x,\cdot)}\left(\{(x,y')\}\right)$$

if both (x,x') and (x,y') belong to this full-measure set and share the same first coordinate. As $\mu^{\mathcal{A}}_{(x,\cdot)}$ 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

$$\mu^{\mathcal{A}}_{(x,\cdot)} = \frac{1}{m}\sum_{i=1}^{m}\delta_{(x,f_i(x))},$$

for μ-almost every (x,x') (or equivalently for m_{X_1}-almost every x). □

Claim. We may choose the functions $f_1, f_2, \dots, f_m \colon X_1' \to X_2$ to be measurable on a subset $X_1' \subseteq X_1$ of full measure.

Proof of Claim. We let X_1' be the set on which $\mu^{\mathcal{A}}_{(x,\cdot)}$ 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 $(\mathcal{P}_n)$ such that

$$\mathcal{P}_n \leqslant \sigma\left(\mathcal{P}_{n+1}\right)$$

and

$$\mathcal{B}_{X_2} = \bigvee_{n=1}^{\infty}\sigma\left(\mathcal{P}_n\right).$$

We also order the elements of

$$\mathcal{P}_n = \{P_{n,1},\dots\}$$

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

$$f_1^{(1)}(x) = y_{1,1} \text{ on } B_{1,1} = \{x \in X_1 \mid \mu^{\mathcal{A}}_{(x,\cdot)}(\{x\}\times P_{n,1}) > 0\}$$
$$f_1^{(1)}(x) = y_{1,2} \text{ on } B_{1,2} = \{x \in X_1 \mid \mu^{\mathcal{A}}_{(x,\cdot)}(\{x\}\times P_{n,2}) > 0\}\smallsetminus B_{1,1},$$

[†] This function is also measurable, which the reader may check by exhibiting f as a pointwise limit of a sequence of measurable functions using $\mu^{\mathcal{A}}_{(x,x')}(B)$ 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 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 $\mathcal{P}_1$. We can ensure this by requiring that each $P_{1,i}$ is split into finitely many partition elements of $\mathcal{P}_2$, and the subsets of $P_{1,i}$ appear before the subsets of $P_{1,j}$ in the enumeration of the elements of $\mathcal{P}_2$ 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 $f_1\colon X_1' \to X_2$. By construction there exists for every n some $Q_n \in \mathcal{P}_n$ with$f_1(x) \in \overline{Q_n}$,

$$\mu^{\mathcal{A}}_{(x,\cdot)}(\{x\} \times Q_n) > 0,$$

and so also $\mu^{\mathcal{A}}_{(x,\cdot)}(Q_n) \geqslant 1/m$. Since Q_n has diameter no larger than $1/n$ we see that $\bigcap_{n=1}^{\infty} \overline{Q_n} = \{(x, f_1(x))\}$ which gives $\mu^{\mathcal{A}}_{(x,\cdot)}(\{(x, f_1(x))\}) = 1/m$ for all $x \in X_1'$. If $m > 1$ then we remove $(x, f_1(x))$ from $\mu^{\mathcal{A}}_{(x,\cdot)}$ by replacing the measure with

$$\mu^{\mathcal{A}}_{(x,\cdot)} - \tfrac{1}{m}\delta_{(x,f_1(x))}$$

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 $f_1, \ldots, f_m$. 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 $f_1, \ldots, f_m$), we get the 'equivalence properties' for the functions $f_1, \ldots, f_m$. In fact

$$\{f_1(u\bullet x), \ldots, f_m(u\bullet x)\} = u\bullet\{f_1(x), \ldots, f_m(x)\}$$

almost surely, and similarly with u replaced by a. Indeed, by [52, Cor. 5.24] we have

$$\begin{aligned}(u,u)\bullet(\{x\}\times\{f_1(x),\dots,f_m(x)\}) &= (u,u)\bullet\operatorname{supp}\mu^{\mathcal{A}}_{(x,\cdot)}\\ &= \operatorname{supp}(u,u)_*\mu^{(u,u)^{-1}\mathcal{A}}_{(x,\cdot)}\\ &= \operatorname{supp}\mu^{\mathcal{A}}_{(u\bullet x,\cdot)},\end{aligned}$$

which in turn may be written as

$$\operatorname{supp}\mu^{\mathcal{A}}_{(u\bullet x,\cdot)} = \{u\bullet x\}\times\{f_1(u\bullet x),\dots,f_m(u\bullet x)\},$$

almost everywhere with respect to m_{X_1}. This is the claimed equivariance property of the set of functions for u, and 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

$$\{f_1(v_t\bullet x),\dots,f_m(v_t\bullet x)\} = v_t\bullet\{f_1(x),\dots,f_m(x)\}, \tag{6.22}$$

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 [52, Sec. 9.5]). Consider the points

$$(x,f_1(x)) \text{ and } (v_t\bullet x, f_1(v_t\bullet x)) = (v_t\bullet x, g\bullet f_1(x))$$

(with $g=v_t$ being our goal). Applying the equivariance property for a to f_1 we obtain

$$\begin{aligned}f_1(a^{-n}v_t\bullet x) = a^{-n}gf_1(x) &= a^{-n}ga^n\bullet f_1(a^{-n}\bullet x)\\ &= f_1(v_{\lambda^{2n}t}a^{-n}\bullet x).\end{aligned} \tag{6.23}$$

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 $f_1(a^{-n}\bullet x)$ 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[†] $a^{-n}ga^n\to\infty$ as $n\to\infty$.

[†] 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.

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 $i = 1, \dots, m$. We define

$$Y_1 = \left\{ x \in X_1' \mid \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{1}_K \left(u^\ell \cdot x\right) \geqslant \tfrac{9}{10} \text{ for all } L \geqslant 1 \right\}$$

and

$$Y_2 = \left\{ x \in X_1' \mid \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{Y_1 \cap K} \left(a^{-n} \cdot x\right) > \tfrac{1}{2} \right\}.$$

By the maximal ergodic theorem applied to the action of u we have

[†] 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.

$$m_{X_1}(Y_1) \geqslant \tfrac{2}{3},$$

hence

$$m_{X_1}(Y_1 \cap K) > \tfrac{1}{2},$$

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.

Lemma 6.26 (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 $f_1, \ldots, f_m$ 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 6.27 (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*

$$\frac{1}{L} \left| \left\{ \ell \in \{0, \ldots, L-1\} \mid u^\ell \boldsymbol{\cdot} x \in K \text{ and } \mathsf{d}(u^\ell \boldsymbol{\cdot} x, u^\ell \boldsymbol{\cdot} y) < \kappa \right\} \right| \geqslant p.$$

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 X_2 in Section 4.1. We let $\rho \in (0,1])$ be chosen so that 2ρ is an injectivity radius on K, and let

$$S = \{\ell \in \{0, \ldots, L\} \mid u^\ell \boldsymbol{\cdot} x \in K \text{ and } \mathsf{d}(u^\ell \boldsymbol{\cdot} y, u^\ell \boldsymbol{\cdot} x) < \rho\}.$$

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 $u^\ell \bullet y, u^\ell \bullet x$ respectively $u^m \bullet y, u^m \bullet x$ 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

$$u^\ell \begin{pmatrix} a & b \\ c & d \end{pmatrix} u^{-\ell} = \begin{pmatrix} a + c\ell & b + (d-a)\ell - c\ell^2 \\ c & d - c\ell \end{pmatrix}$$

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

$$ad = ad - bc + O(L^{-2}) = 1 + O(L^{-2}).$$

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

$$(a-1)(d+1) = ad - 1 + a - d = O(\tfrac{\rho}{L} + \tfrac{1}{L^2})$$

implies $|a-1| \ll_p \frac{1}{L}$. The estimate $|d-1| \ll_p \frac{1}{L}$ follows from $|(d-a)L| \ll_p 1$.

[†] Note that possibly not all of these integers k belong to S due to the additional requirement $u^k \bullet x \in K$ 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.

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

$$\frac{1}{T}\left|\{t \in \{0, \ldots, T-1\} \mid |f(t)| \leqslant \frac{\kappa}{\rho}\|f\|_{\infty,T}\}\right| < \frac{p}{3}$$

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

$$\frac{1}{T}\left|\{t \in \{0, \ldots, T-1\} \mid \mathsf{d}(u^t h u^{-t}, I) \leqslant \kappa\}\right| < \frac{p}{3}$$

if $h \in B_G^\rho$ is such that $\mathsf{d}(u^{-1} h u, I) \geqslant \rho$ 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 $\{0, \ldots, L\}$ 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

$$P_{[\ell]} = \{0, \ldots, L-1\},$$

and so the lemma by the first part of the proof. Hence, by our choice of κ,

$$\frac{1}{|P_{[\ell]}|}|\underbrace{\{k \in P_{[\ell]} \mid \mathsf{d}(u^{k-\ell} g_\ell u^{-(k-\ell)}, I) \leqslant \kappa\}}_{\mathrm{Bad}_{[\ell]}}| < \frac{p}{3}.$$

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

$$\mathrm{Bad} = \{\ell \in \{0,\dots,L-1\} \mid \mathsf{d}(u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} y, u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} x) < \kappa \text{ and } u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} x \in K\} \subseteq \bigcup_{[\ell]} \mathrm{Bad}_{[\ell]}$$

and obtain

$$|\,\mathrm{Bad}\,| \leqslant \sum_{[\ell]} |\,\mathrm{Bad}_{[\ell]}\,| \leqslant \sum_{[\ell]} \frac{p}{3}|P_{[\ell]}| \leqslant \frac{2}{3}pL.$$

However, this contradicts our assumptions. Hence there can only be one equivalence class and the lemma follows. □

We return to the setting of Theorem 6.22 and apply the lemma above.

Proof of Lemma 6.26. 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 6.27 to

$$X = X_2 = \varGamma_2\backslash \mathrm{SL}_2(\mathbb{R})$$

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

$$x, y = g\raisebox{0.1ex}{\scriptsize$\bullet$} x \in K, g \in \mathrm{SL}_2(\mathbb{R}) \text{ with } \mathsf{d}(g, I) < \delta \Longrightarrow \mathsf{d}(f_i(y), f_i(x)) < \kappa$$

for $i = 1,\dots,m$.

Suppose now that $t \in (-\delta, \delta)$ and $x, v_t\raisebox{0.1ex}{\scriptsize$\bullet$} x \in K \cap Y_1$. 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 $u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} x, u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} y \in K$ and so

$$\mathsf{d}\left(f_i\left(u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} y\right), f_i\left(u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} x\right)\right) < \kappa$$

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

$$\mathsf{d}\left(u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} f_i(y), u^\ell\raisebox{0.1ex}{\scriptsize$\bullet$} f_j(x)\right) < \kappa. \tag{6.24}$$

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 6.27 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 $|b| \ll 1$. □

We continue with the proof of Theorem 6.22. 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 6.26 to y_n and x_n. It follows that for every i there exists some j such that†

$$f_i(y_n) = g_n \cdot f_j(x_n) = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \cdot f_j(x_n)$$

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 $x = a^n \cdot x_n$ and $y = a^n \cdot y$ 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

$$|\lambda^{2n} b_n| \ll \lambda^{2n}.$$

† As was mentioned before, we do not know at this stage any relationship between these displacement g_n for different n's.

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 217 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

$$(x,x') = (x, f_j(x)), (y,y') = (v_t \cdot x, f_i(v_t \cdot x)) = (v_t, v_c) \cdot (x,x')$$

with $t \neq c$. Applying this to the set in (6.21) and continuing the argument as just after the proof of Lemma 6.25 we conclude that μ is the trivial joining (which actually contradicts our description of the conditional measures $\mu^{\mathcal{A}}_{(x,\cdot)}$).

Since we now may assume $c = t$ for almost every $x \in Y_1$ and since both sets $\{f_1(v_t \cdot x), \ldots, f_m(v_t \cdot x)\}$ and $\{v_t \cdot f_1(x), \ldots, v_t \cdot f_m(x)\}$ 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

$$\begin{aligned}
&\int_{X_1 \times X_2} f((v_t, v_t) \cdot (x,x')) \,\mathrm{d}\mu = \\
&\quad \int_{X_1} \int_{\{x\} \times X_2} f((v_t, v_t) \cdot (x,x')) \,\mathrm{d}\mu_{(x,\cdot)}(x,x') \,\mathrm{d}m_{X_1}(x) = \\
&\quad \int_{X_1} \frac{1}{m} \sum_{i=1}^{m} f((v_t, v_t) \cdot (x, f_i(x))) \,\mathrm{d}m_{X_1}(x) = \\
&\quad \int_{X_1} \frac{1}{m} \sum_{i=1}^{m} f(v_t \cdot x, f_i(v_t \cdot x)) \,\mathrm{d}m_{X_1}(x) = \\
&\quad \int_{X_1} \frac{1}{m} \sum_{i=1}^{m} f(x, f_i(x)) \,\mathrm{d}m_{X_1}(x) = \int_{X_1 \times X_2} f \,\mathrm{d}\mu,
\end{aligned}$$

where we used in order the definition of the conditional measures, our description of them, (6.22) for m_{X_1}-almost every x, and the fact that v_t preserves m_{X_1}.

Now note that U as in Theorem 6.22 together with $\{(v_t, v_t) \mid t \in \mathbb{R}\}$ 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 6.19 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.

Theorem 6.28. *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*[‡].

Proof. (to come) □

[†] 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.

Exercises for Section 6.8

Exercise 6.8.1. Find the complete list of all possible connected algebraic subgroups L of $\mathrm{SL}_3(\mathbb{R})$ 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

(23) (Page 182) This appeared in print in the work of Dani [20, Conjecture II].
(24) (Page 186) This is an instance of a more general result due to Weyl [185] giving equidistribution modulo one for the values on the natural numbers of any polynomial with an irrational coefficient. Furstenberg [65] showed that this followed from a general result extending unique ergodicity from irrational circle rotations to certain maps on tori. We refer to [52, Sec. 4.4.3] for a detailed discussion.